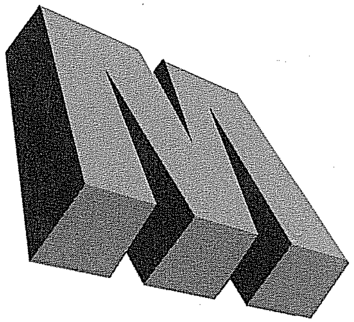
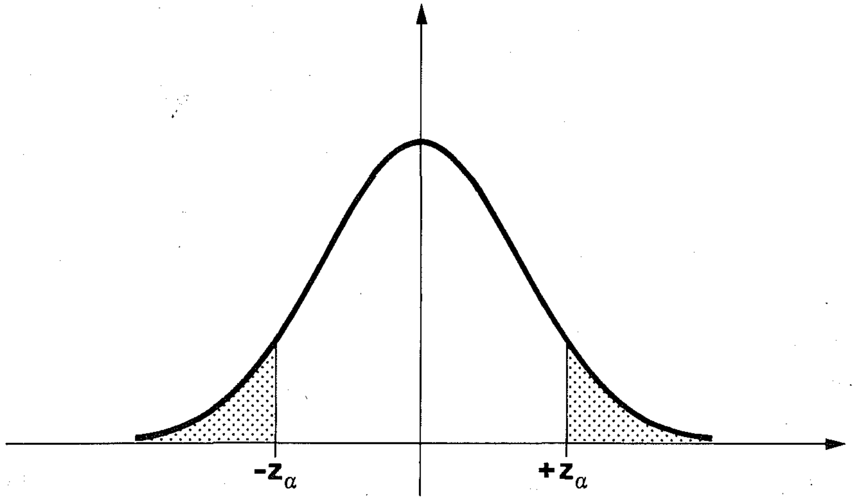
MATEMATIKA





Kritično območje dvostranskega preizkusa

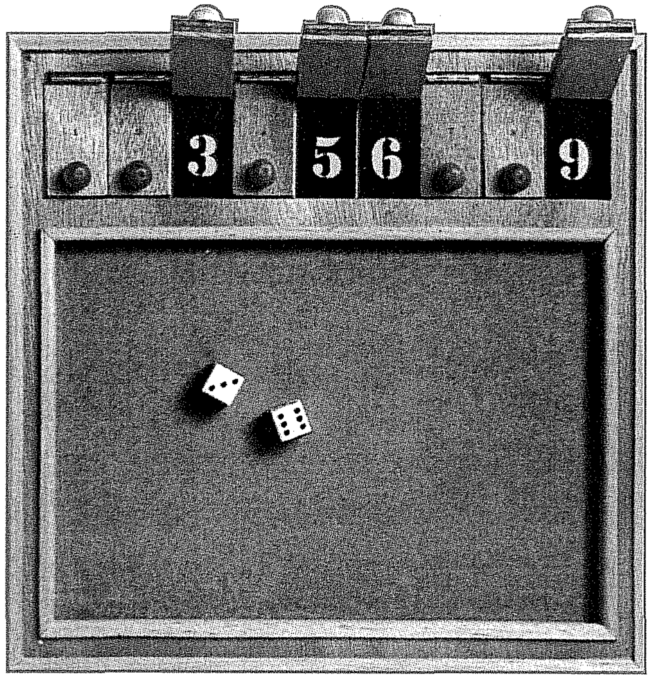
ISBN 86-341-1440-6

9 788634 114409

KOMBINATORIKA • VERJETNOSTNI RAČUN • STATISTIKA

MATEMATIKA

KOMBINATORIKA • VERJETNOSTNI RAČUN • STATISTIKA



DZS

MATEMATIKA

KOMBINATORIKA • VERJETNOSTNI RAČUN • STATISTIKA

JOŽE ANDREJ ČIBEJ

DZS

Učbenik je napisal mag. Jože Andrej Čibej

Rokopis sta strokovno pregledala:
prof. dr. Mihael Perman in
Branko Roblek, prof.

Konzulentka Zavoda Republike Slovenije za šolstvo in šport je bila
Nada Marčič, prof.

Rokopis je jezikovno pregledala Danijela Čibej, prof.

Strokovni svet RS za vzgojo in izobraževanje je na seji dne 9. 3. 2000 sprejel sklep št. 612-58/00 o potrditvi učbenika MATEMATIKA — Kombinatorika. Verjetnostni račun. Statistika Jožeta Andreja Čibeja za uporabo pri pouku matematike v 3. in 4. letniku srednjih šol za dobo petih let.

Ponovna recenzija ob potrditvi:
prof. dr. Mihael Perman in mag. Bogdan Kejžar

znanje uresničuje sanje
DZS, d.d., DIVIZIJA ZALOŽNIŠTEV
IZOBRAŽEVALNO ZALOŽNIŠTVO
http://www.dzs.si
e-pošta: info.narocila@dzs.si
tel. št.: 01/ 241 04 26

CIP - Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

519.1(075.3)
519.2(075.3)

ČIBEJ, Jože Andrej
Matematika / Jože Andrej Čibej ; [tehnične risbe narisal Darko Simeršek]. - 4. prenovljena izd., 4. natis. - Ljubljana : DZS, 2004

Vsebina na nasl. str.: Kombinatorika ; Verjetnostni račun ; Statistika

ISBN 86-341-1440-6

128373760

KAZALO

1. UVOD

Verjetnostni račun in statistika sodita med tiste matematične discipline, ki so potrebovale nekoliko dlje, da so si pridobile takšno spoštovanje med uporabniki, kot ga po vsej pravici zaslužijo. Razlogov za to je več in vsaj nekateri vzroki ležijo pregloboko, da bi lahko o njih razpravljali v srednješolskem učbeniku. Zato samo preprosto, v treh zaporednih stopnjah, ugotovimo sedanje stanje, ki pa je - enim na srečo, drugim na žalost - bistveno drugačno:

a) Brez osnovnih znanj s področja **statistike** in statistične analize podatkov ne moremo razumeti niti člankov v dnevnem časopisju, kaj šele strokovne literature, pa naj gre za tehnično, družboslovno ali kako drugo področje dela. Še hujša nevarnost za okolje kot nerazumevanje tujega pisanja pa je kakopak nekvalificirana aktivna uporaba statističnih orodij, ki jih (tudi nepoučenemu) uporabniku prijazno ponujajo različni računalniški programski paketi, denimo programi za delo s preglednicami. (Napačnih napovedi volilnih rezultatov in drugih nerodnosti, ki so posledica neznanja, raje ne omenjamo.)

b) Kdor misli, da lahko zares razume statistiko, ne da bi vsaj v glavnih črtah obvladal **verjetnostni račun**, živi v globoki zmoti in zanj dejansko lahko obveljajo znani poenostavljeni reki o tem, kaj je statistika, še več je res, zanj *statistika lahko postane največja laž* celo v primerih, ko so bili podatki sicer korektno predstavljeni.

c) Nazadnje je treba reči, da se po tehnični plati marsikatera naloga iz verjetnostnega računa prevede na ugotavljanje moči neke končne množice, ali bolj konkretno, na preštevanje izidov določenega poskusa; podobno velja za vzorčenje in druge tehnike, ki jih potrebujemo pri statističnem sklepanju. Za udobnejše delo pri zahtevnejših preštevanjih je najpomembnejše orodje **kombinatorika**.

Povedano po eni strani pojasni, zakaj se ustrezne vsebine vedno manj sramežljivo pojavljajo v srednješolskih učnih načrtih matematike, pa ne samo tam: katalogi znanj za maturo in zaključni izpit jasno kažejo, da bo tudi v *vélikem finalu* tekla beseda o dogodkih, verjetnosti, frekvenčnih poligonih in podobnem.

S tem je na drugi strani vsaj v grobem določena tudi vsebina tega učbenika. Velik del knjige pomeni (tako vsaj upamo) *očiščenje in pomlajenje*, v dobršni meri pa tudi razširitev avtorjevih dosedanjih srednješolskih učbenikov *Kombinatorika* in *Verjetnostni račun in statistika*. Temu je dodano poglavje o statističnem sklepanju, ki formalno sodi med izbirne vsebine (podobno vlogo je imelo to poglavje v učnem načrtu naravoslovnomatematičnih gimnazij), po vsebini pa bi morda moralo biti vsaj informativno branje za vsakega srednješolca...

Razširitve obvezne snovi (oziroma tisto, kar po maturitetnem katalogu sodi pretežno na višjo raven) so tradiciji na ljubo povečini označene s pravokotnikom, na zahtevnejša razmišljanja ali vaje opozarjajo trikotniki. Slike so oštevilčene po poglavjih, zgledi (ki jih je bistveno več kot slik) pa po razdelkih. Tako na primer "zgled 2.4.6" pomeni šesti zgled v četrtem razdelku drugega poglavja; dokler smo znotraj tega razdelka, ga dostikrat omenjamo kar kot "zgled 6". Od tabel oziroma preglednic so oštevilčene samo tiste, na katere se večkrat sklicujemo.

Vsaj za razumevanje obveznega dela bi moralo zadoščati znanje iz prvega letnika srednje šole v obsegu učbenika Legiša, *Matematika za prvi letnik. - Realna števila, linearna funkcija.*, nekatere najpomembnejše resnice za vsak slučaj ponovimo, ko je potrebno. Kako se bo prebijal skozi učbenik, pa je precej odvisno tudi od bralca in njegovih siceršnjih navad. Če je domač z osebnim računalnikom (kar bi moralo biti prej pravilo kot izjema), bo lahko dobršen del snovi praktično doživel na zaslonu; nekaj idej za tovrstne naloge smo navrgli med besedilom, ostale prepuščamo njegovi (pa njegovih učiteljev in sošolcev) domišljiji.

Na koncu, pa ne najmanj - veliko veselja!

Avtor

Trbovlje – Ljubljana, poleti 1994

2. KOMBINATORIKA

V tem poglavju se naučimo uporabljati najpomembnejše kombinatorične prijeme in preštevati različne - urejene in neurejene - razporedbe elementov dane množice. Spoznamo permutacije, variacije in kombinacije, računamo z binomskimi simboli in jih uporabimo v binomskem izreku, bolj radovedni se srečajo z načelom vključitev in izključitev. Vse, česar se naučimo, so predvsem orodja za kasnejše delo.

2.1 Osnovni kombinatorični prijemi

Tipične kombinatorične naloge sprašujejo po številu različnih razporeditev, izborov ali sestavljenih odločitev, na primer:

- Koliko različnih besed s po petimi različnimi črkami lahko sestavimo iz 25 črk slovenske abecede?
- Na koliko načinov lahko posadimo v vrsto pet otrok, če ne postavljamo nobenih dodatnih pogojev, in na koliko načinov, če Miha in Peter ne smeta sedeti skupaj?
- Koliko različnih tričlanskih delegacij lahko sestavimo v kolektivu, v katerem je 200 delavcev?

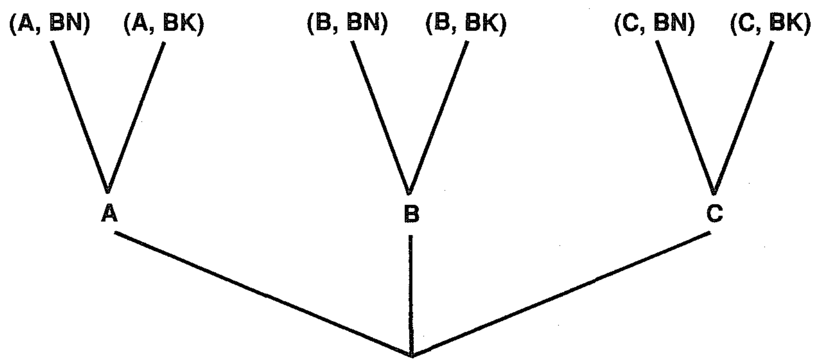
Vse take naloge lahko obravnavamo kot probleme odločanja v več stopnjah ali fazah. Za nazornejši prikaz korakov v odločanju dostikrat uporabljamo t.i. **kombinatorično drevo** (slika 2.1 na naslednji strani).

ZGLED 2.1.1:

Proizvajamo tri različne izdelke (A, B in C), vsakega od njih lahko glede na naročnikovo željo obarvamo bodisi z običajno ali s kovinsko barvo. Koliko različnih izdelkov pravzaprav ponujamo?

Nalogo lahko interpretiramo takole: najprej gre za izbiro izdelka, pri čemer imamo tri različne možnosti, zaradi česar narišemo iz skupnega izhodišča tri veje. Po opravljeni odločitvi v prvi fazi imamo ne glede na to, kaj smo v tej fazi izbrali, dve možnosti za izbiro barve; iz vsake od treh vej rasteta po dve veji, ki ponazarjata odločitvi v drugi fazi (navadna ali kovinska barva). Skupno število sestavljenih odločitev je enako številu načinov, na katere lahko pridemo od "korenin" drevesa do skrajnih vej, teh pa je $3 \cdot 2 = 6$.

Značilnost dvostopenjskega odločanja v opisanem zgledu je dejstvo, da je - po predpostavkah naloge - število odločitev v drugi fazi (izbira barve)



Slika 2.1: Kombinatorično drevo - dve stopnji odločanja

neodvisno od tega, kako smo se odločili na prvi stopnji (izbira izdelka). Zato - in samo zato - je bilo število sestavljenih izborov produkt števila možnih odločitev v prvi fazi in ustreznega števila možnosti, ki jih ponuja druga faza. Ugotovitev lahko posplošimo tudi na primere, ko poteka izbiranje v več kot dveh fazah:

Naj bo proces odločanja takšen, da poteka v k zaporednih fazah, pri čemer je v prvi fazi n_1, v drugi n_2, ... in v k-ti fazi n_k možnih odločitev, število izborov v posamezni fazi pa je neodvisno od tega, katere možnosti smo izbrali v predhodnih fazah. Potem je mogoče sestavljeni izbor opraviti na natanko

$$n = n_1 \cdot n_2 \cdot \ldots \cdot n_k$$

načinov.

To ugotovitev običajno imenujemo **osnovni izrek kombinatorike**, dostikrat tudi **pravilo produkta**.

ZGLED 2.1.2:

Koliko različnih "besed" (v smislu poljubne skupine črk) s po petimi črkami slovenske abecede lahko sestavimo, če dovoljujemo ponavljanje posameznih črk, in koliko, če ponavljanja ne dovolimo?

V prvem primeru lahko na prvo mesto postavimo katerokoli od 25 črk, neodvisno od tega, katero smo izbrali, na drugo mesto spet katerokoli od petindvajsetih ... do vključno petega mesta, torej $25 \cdot 25 \cdot 25 \cdot 25 \cdot 25 = 9\,765\,625$ besed.

Pri drugi različici imamo za prvo mesto 25 kandidatk, za drugo - ker ne dovolimo ponavljanja - samo še 24, za tretje 23,..., za peto mesto 21 možnosti. Katero črko imamo še na voljo v posamezni fazi, je seveda odvisno od tega, katere smo že porabili v prejšnjih, koliko jih imamo na izbiro, pa je povsem neodvisno od izbire v prejšnjih fazah, zato osnovni izrek kombinatorike še zmeraj velja: vseh takih besed je $25 \cdot 24 \cdot 23 \cdot 22 \cdot 21 = 6\,375\,600$.

OPOMBA: S podobnimi nalogami se bomo še večkrat srečali, zato opozorimo, da izraz **abeceda** v splošnem uporabljamo za vsako končno množico simbolov, v kateri je dan t.i. *naravni* ali *slovarski (tudi: leksikografski) vrstni red.* Z urejenostjo abecede je potem predpisana tudi urejenost v množici vseh možnih **besed** kot končnih nizov, sestavljenih iz simbolov dane abecede. V smislu te definicije je za kombinatoriko niz CŽFF ravno tako beseda dolžine 4 kot ENKA, čeprav v vsakdanjem jeziku nima nobenega pomena.

ZGLED 2.1.3:

A in B naj bosta končni množici, prva z močjo $m(A) = M$ in druga z močjo $m(B) = N$. Koliko elementov ima kartezični produkt $A \times B$?

Moč kartezičnega produkta $A \times B$ je enaka številu urejenih parov s prvo komponento iz množice A in z drugo komponento iz množice B; prvo komponento lahko izberemo na M načinov in drugo - neodvisno od tega, kateri element smo izbrali v množici A - na N načinov, toliko je elementov v množici B. Po osnovnem izreku kombinatorike je torej

$$m(A \times B) = m(A) \cdot m(B) = M \cdot N$$

Tako smo ponovno dognali resnico, ki jo poznamo iz prvega letnika: *Moč kartezičnega produkta je produkt moči obeh množic, ki v njem nastopata.*

Osnovni izrek kombinatorike velja samo v primeru, ko je število izborov v posamezni fazi neodvisno od tega, kaj se je zgodilo v prejšnjih. V praksi ta pogoj dostikrat ni izpolnjen, na ustreznem kombinatoričnem drevesu se to pokaže z nesimetrično rastjo.

ZGLED 2.1.4:

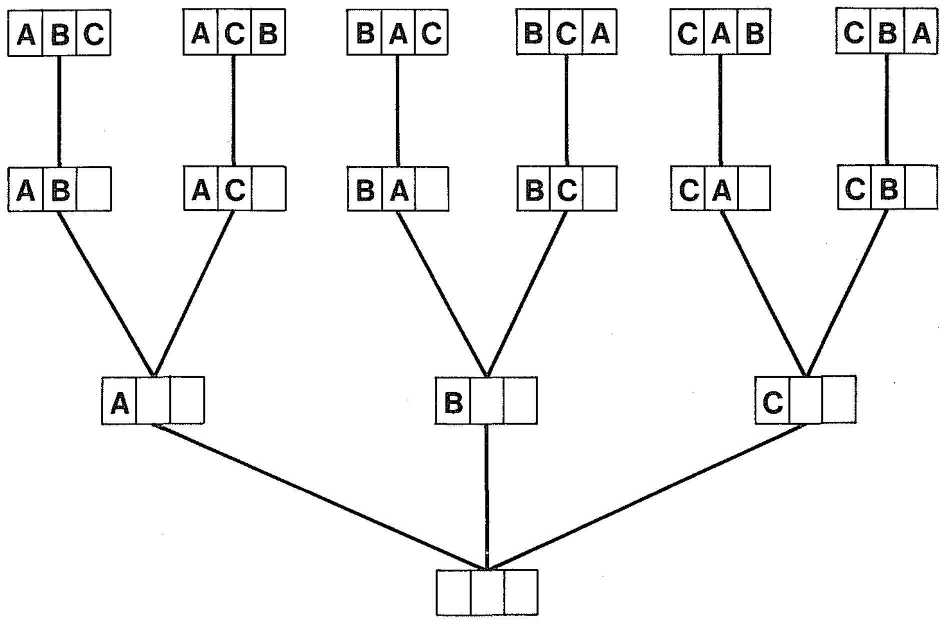
Na kongresu politične stranke so se delegati odločili, da bodo osebe A, B in C zasedle predsedniški, podpredsedniški in tajniški položaj. S kombinatoričnim drevesom prikažimo, kako se lahko razporedijo na teh treh položajih,

a) če dopuščamo poljubno zasedbo funkcij;

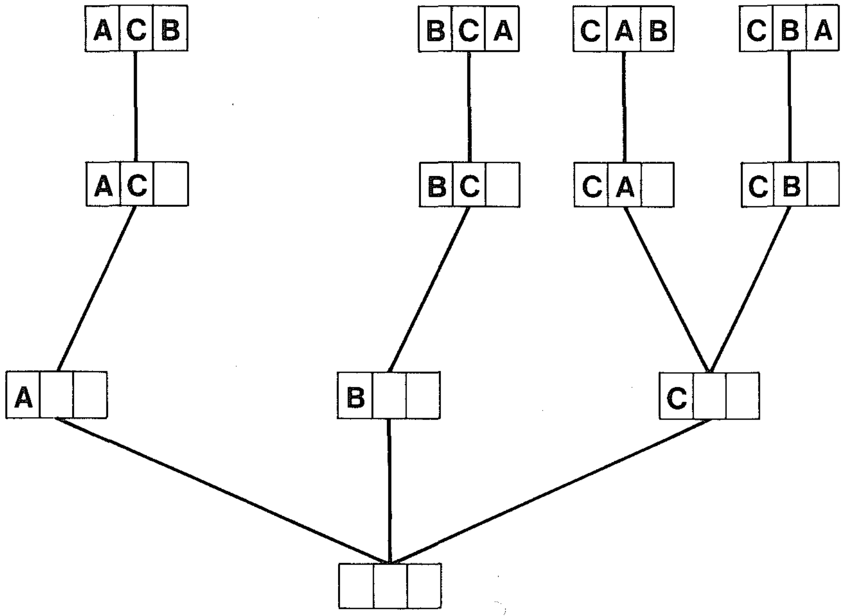
b) če se zavedamo, da zaradi notranjih odnosov v stranki ni sprejemljivo, da A in B zasedeta oba najvišja položaja.

V prvem primeru lahko predsedniško mesto zasedemo na tri načine, podpredsedniško neodvisno od tega, kdo je predsednik, na dva načina, na tajniško mesto posadimo preostalega člana izvoljene trojice (slika 2.2).

V drugem primeru odpadejo vse tiste veje, kjer se na najvišjih položajih srečata A in B in drevo je nesimetrično (slika 2.3).



Slika 2.2: Izbor brez omejitev - simetrično drevo



Slika 2.3: Izbor z omejitvami - nesimetrično drevo

Kljub "okleščenosti" pa seveda tudi v drugem primeru še vedno velja, da je število sestavljenih izborov enako številu načinov, na katere pripotujemo od korenin do najvišjih vej.

Pravilo produkta, ki ga upravičeno imenujemo osnovni izrek kombinatorike, pri navedenih pogojih uporabljamo v primerih, ko je treba opraviti dvoje ali več zaporednih izborov. Obstaja pa tudi t.i. **pravilo vsote**, ki govori o izbiranju iz ene od dveh (ali več) med seboj tujih (disjunktnih) množic. Takole pravi:

Če se lahko pri izbiranju odločimo bodisi za eno od n_1 možnosti iz prve množice izborov ali pa za eno od n_2 možnosti iz druge množice izborov, ki so nezdružljivi z izbori iz prve množice, imamo v celoti

$$n = n_1 + n_2$$

možnih izborov.

Obrazec lahko posplošimo na poljubno končno število tujih množic.

ZGLED 2.1.5:

Micka ima 6 kap in 8 klobukov. Na koliko načinov lahko izbere pokrivalo?

Če predpostavimo, da sta ustrezni množici tuji (da Micka nobenega pokrivala ni štela dvakrat, kot kapo in kot klobuk), je načinov $6 + 8 = 14$.

Praktične izkušnje pa kažejo, da je treba v splošnem pazljivo presojati, ali so pogoji za uporabo pravila vsote izpolnjeni. Nekdo, ki je slišal, da bo na izletu 25 gledališčnikov in 40 članov pevskega zbora, je nabavil 65 klobas in se čudil, ko mu jih je nekaj ostalo, čeprav je vsak izletnik pospravil po eno... Zakaj?

ZGLED 2.1.6:

Iz kraja A v kraj C lahko pridemo skozi kraj B ali kraj D ali pa po avtocesti, ki vodi neposredno iz A v C. Med krajema A in B so kar štiri možne poti, pa tudi od B do C so na voljo tri ceste. Od kraja A do D vodi ena sama pot, zato pa imamo med C in D spet dve možnosti. Na koliko načinov lahko tedaj pridemo iz A v C?

Na videz zapletena geografija postane povsem preprosta, če si - kot je v geografiji pač navada - narišemo zemljevid (kar bo bralec najbrž z zadovoljstvom postoril sam). Potem imamo tri med seboj tuje načine potovanja, A $\to$ B $\to$ C, A $\to$ C in A $\to$ D $\to$ C. Na prvi poti je po pravilu produkta $4 \cdot 3 = 12$ različnih možnosti in na tretji sta možnosti 2. Skupaj imamo po pravilu vsote tako $12 + 1 + 2 = 15$ možnosti za potovanje iz A v C.

V A J E

1. Nekdo ima 4 obleke, 8 srajc in 6 kravat. Na koliko načinov se lahko obleče?

2. Na jedilniku sta dve predjedi, dve vrsti juhe, štiri glavne jedi, tri prikuhe, tri vrste solate in štirje različni poobedki. Koliko menujev lahko sestavimo,
a) če vključimo po eno jed iz vsake od naštetih skupin;
b) brez predjedi;
c) brez poobedka?

3. Koliko besed s po tremi simboli (koliko besed dolžine 3) lahko sestavimo s simboli abecede $\{A, B, C, D\}$?

4. Koliko besed dolžine 4 lahko sestavimo s simboli iz abecede $\{1, 2, 3, 4, 5\}$?

5. Koliko je vseh štirimestnih dvojiških števil, če mednje štejemo tudi tiste, ki se začenjajo z eno ali več ničlami?

6. $\triangle$ Dokaži osnovni izrek kombinatorike s pomočjo načela popolne indukcije.

7. Prva delegacija ima sedem, druga pa šest članov. Na koliko načinov je mogoče izbrati vodjo posveta in njegovega namestnika, če naj bo vodja iz prve, namestnik pa iz druge delegacije?

8. Na vrh gore vodi 5 poti. Na koliko načinov se lahko povzpnemo nanjo in se spustimo z nje,
a) če se lahko vrnemo po isti poti;
b) če moramo obvezno izbrati drugo pot?

9. Ob neki železniški progi je 25 postaj. Koliko kart (za vožnjo med katerimakoli postajama) morajo natisniti? (Ne pozabi, da vlaki vozijo v obe smeri.)

10. Koliko je pravih petmestnih števil (takih, ki se ne začnejo z 0), ki so enake, če jih preberemo z leve ali z desne (npr. 12321)?

11. Na koliko načinov si lahko 26 tekačev na 100 metrov razdeli tri medalje na državnem prvenstvu? (Možnost, da bi si dva tekača delila isto mesto, zanemarimo.)

12. Šahista igrata dvoboj, ki se konča z zmago tistega, ki prvi doseže dve zmagi, ali pa se konča neodločeno, če igralca dvakrat remizirata. Predstavi vse možne načine tega dvoboja s kombinatoričnim drevesom in jih preštej.

13. Andrej in Miha igrata tenis. Dvoboj je končan,
- če eden od njiju dobi tri nize ali
- če eden od njiju dobi dva niza zapored.

S pomočjo kombinatoričnega drevesa preštej vse načine, na katere se lahko odvija in konča ta dvoboj.

14. Z enim kovancem stavimo na igralnem avtomatu. V vsaki igri se lahko zgodi dvoje:
- če stavo izgubimo, izgubimo tudi vplačani kovanec;
- če stavo dobimo, avtomat vrne vplačani kovanec in izplača še enega povrhu.

Stavili bomo največ štirikrat, že prej pa končamo igro, če k začetnemu zaslužimo še dva kovanca, seveda pa tudi v primeru, ko nam zmanjka kovancev. Prikaži vse možne poteke igre s kombinatoričnim drevesom.

15. Morsejeve znake sestavljamo iz pik in črtic, pri čemer lahko oba znaka nastopata bodisi posamič ali v skupinah po 2, 3 in 4. Koliko je vseh možnih znakov? Nastanek različnih Morsejevih znakov, ki so sestavljeni iz treh osnovnih znakov (pik ali črtic), prikaži tudi s kombinatoričnim drevesom.

16. Stojimo v točki 0 na številski premici. Premaknemo se lahko vsakič samo za en korak (eno enoto) v levo ali en korak (eno enoto) v desno. Prikaži takšno gibanje s kombinatoričnim drevesom, če se ustavimo,
a) kadar dosežemo točko +3 ali točko -3;
b) kadar že drugič pridemo v katerokoli točko, ki ni izhodišče.
Koliko je vseh takšnih poti in koliko se jih konča v +3 ali -3?

17. Precej ljudi ima težave z iskanjem po slovarjih in enciklopedijah, ker so pri obvladanju abecede prišli samo do urejanja po prvi črki. Kakšen je leksikografski vrstni red naslednjih besed: AMANIJA, ANEMONA, AMAI, AMENONA, AMAN, ANAM, ANONIMEN, AMONIT, ANA, ANIMA, AMA, AMIN?

2.2 Permutacije

O **permutacijah** govorimo, kadar urejamo vse elemente neke dane (končne) množice v vrsto, denimo, da od leve proti desni. Ker so elementi v množici različni, lahko takim razporedbam rečemo

Permutacije brez ponavljanja

ker v posamezni permutaciji (razporedbi) vsak element iz dane množice nastopa samo enkrat.

ZGLED 2.2.1:

Družina ima tri člane. Na koliko različnih načinov se lahko razvrstijo pri jutranjem umivanju?

Če družinske člane simbolično označimo z A, B in C ter predpostavimo, da želi biti v kopalnici vsak sam, imamo položaj, ki ga poznamo iz zgleda 2.1.4 in slike 2.2 iz prejšnjega razdelka:
- za odločitev o tem, kdo gre prvi, imamo tri možnosti,
- za drugo mesto konkurirata preostala dva člana,
- tretji se umiva žalujoči ostali...
skupno je tako po osnovnem izreku kombinatorike $3 \cdot 2 \cdot 1 = 6$ možnih vrstnih redov; po leksikografskem redu urejene so to permutacije ABC, ACB, BAC, BCA, CAB in CBA. Povedano seveda velja samo ob predpostavki, da so možni vsi vrstni redi (kar v družinski praksi ni prav pogosto).

ZGLED 2.2.2:

Dana je množica $\{a_1, a_2, a_3, \ldots, a_n\}$. Na koliko načinov lahko njene elemente razporedimo v vrsto?

Naloga je posplošitev prejšnjega zgleda. Na prvo mesto lahko postavimo kateregakoli od n elementov, na drugo neodvisno od tega, kateri element je na prvem mestu, kateregakoli od preostalih $n-1$ elementov; prvi dve mesti lahko tako zasedemo na $n \cdot (n-1)$ načinov. Nadaljevanje premisleka nas pripelje do ugotovitve, da je takih razporedb ravno

$$n \cdot (n-1) \cdot (n-2) \ldots 3 \cdot 2 \cdot 1$$

Produkt naravnih števil od vključno 1 do vključno n se v matematiki pojavlja tako pogosto, da je dobil poseben simbol in ime, pišemo namreč

$$n! = 1 \cdot 2 \cdot 3 \ldots (n-1) \cdot n$$

in oznako $n!$ preberemo **"n fakulteta"**, pa tudi "n faktorsko" ali "n faktorialno".

Dodatno se je treba dogovoriti, da je $0! = 1$, kar je morda na prvi pogled čudno, na drugega pa ne več: dejansko je smiselno reči, da lahko prazno množico linearno uredimo natanko na en način.

Če opazujemo funkcijo $f : \mathbb{N} \cup \{0\} \to \mathbb{N}$, ki preslika n v $n!$, lahko vidimo, da izredno hitro narašča.

Naloga: Na večini žepnih računalnikov najdemo tudi tipko, s katero izračunamo $n!$ za izbrano naravno število n. Izdelaj tabelo vrednosti $n!$ za $1 \leq n \leq 20$ in poskusi razmišljati o praktičnih posledicah tako hitrega naraščanja vrednosti $n!$.

Ena znamenitih nalog te vrste govori o človeku, ki je imel devet prijateljev, s katerimi je vsak večer sedel pri večerji v drugačnem vrstnem redu. Obljubil jim je brezplačno večerjo, ko bodo prvič spet sedeli v začetnem vrstnem redu...

ZGLED 2.2.3:

Izračunaj število permutacij množice $A = \{a, e, i, o, u\}$ in zapiši prvih pet in zadnjih pet permutacij!

Ker je $m(A) = 5$, imamo

$$n! = 5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

permutacij. Izraza "prvih" in "zadnjih" se nanašata na leksikografsko ureditev vseh razporedb, torej glede na naravni vrstni red, ki je pač pri črkah, če ni povedano drugače, določen z abecedo. Tako imamo v našem primeru

PERMUTACIJE MNOŽICE $\{a, e, i, o, u\}$		
$a\,e\,i\,o\,u$	.....	.....
$a\,e\,i\,u\,o$	.....	.....
$a\,e\,o\,i\,u$	.....	.....
$a\,e\,o\,u\,i$	.....	.....
$a\,e\,u\,i\,o$	.....	.....
.....	.....	.....
.....	.....	$u\,o\,a\,i\,e$
.....	.....	$u\,o\,e\,a\,i$
.....	.....	$u\,o\,e\,i\,a$
.....	.....	$u\,o\,i\,a\,e$
.....	.....	$u\,o\,i\,e\,a$

Iz zgleda je očitno, da se je treba pri pregledovanju permutacij natančno držati vrstnega reda, ker se sicer hitro zmotimo.

ZGLED 2.2.4:

Telovadci so oblečeni v bela, modra in rdeča oblačila. Na koliko načinov lahko postavimo v vrsto 4 telovadce v belih, 2 v modrih in 5 v rdečih oblačilih, če morajo telovadci enakih barv stati skupaj?

Razporejanje opravimo v dveh fazah: najprej razporedimo tri "komplete", kar lahko naredimo na $3! = 6$ načinov, nato pa v vsakem "kompletu" razporejamo telovadce na toliko načinov, kolikor je vseh permutacij elementov v ustrezni podskupini. Tako dobimo za skupno število razporedb rezultat

$$3! \cdot 4! \cdot 2! \cdot 5! = 34\,560$$

Razumljivo je, da imamo ob dodatnih omejitvah znatno manj razporedb, kot pa je vseh permutacij enajstih telovadcev; teh bi bilo $11! = 39\,916\,800$.

ZGLED 2.2.5:

Koliko je permutacij množice $\{1, 2, ..., 2n\}$, pri katerih so liha števila na lihih mestih v razporedbi, soda pa na sodih?

Liha števila lahko na $n!$ načinov razporedimo po lihih mestih, soda pa neodvisno od tega, kakšna je razporedba lihih, na $n!$ načinov po položajih, ki so jim namenjeni. Po osnovnem izreku kombinatorike je vseh razporedb predpisane vrste $n! \cdot n! = (n!)^2$.

Permutacije kot preslikave

Razvrščanje elementov dane končne množice, kakršno smo imeli v dosedanjih zgledih, pravzaprav ni drugega kot neka bijektivna preslikava te množice nase: po razporejanju se v nizu pojavi vsak element natanko enkrat, v splošnem na drugem mestu, kot je bil spočetka, ko je bila množica urejena po naravnem vrstnem redu.

$$\begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_k & \dots & a_n \\ a_{i_1} & a_{i_2} & a_{i_3} & \dots & a_{i_k} & \dots & a_{i_n} \end{pmatrix}$$

Zato lahko upravičeno imenujemo vsako bijektivno preslikavo končne množice nase kar permutacija.

Če je naravni vrstni red izbran enkrat za vselej, je ta preslikava natanko določena in popisana že z navedbo indeksov,

$$\begin{pmatrix} 1 & 2 & 3 & \dots & k & \dots & n \\ i_1 & i_2 & i_3 & \dots & i_k & \dots & i_n \end{pmatrix}$$

Še več, ker imamo v prvi vrstici vedno naravni vrstni red, je dovolj, če zapišemo vrstni red indeksov v sliki: permutacija je tako določena z nizom števil $i_1\, i_2\, i_3 \dots i_k \dots i_n$.

Na začetku tega razdelka smo pravzaprav izračunali, da lahko poljubno končno množico z močjo n na natančno $n!$ načinov bijektivno preslikamo nase. Analogno ni težko premisliti, da je toliko bijektivnih preslikav med poljubnima množicama z močjo n. Če ne gre drugače, si elemente ene množice predstavljamo kot predalčke, v katere polagamo elemente druge množice, v vsakega natanko po enega...

ZGLED 2.2.6:

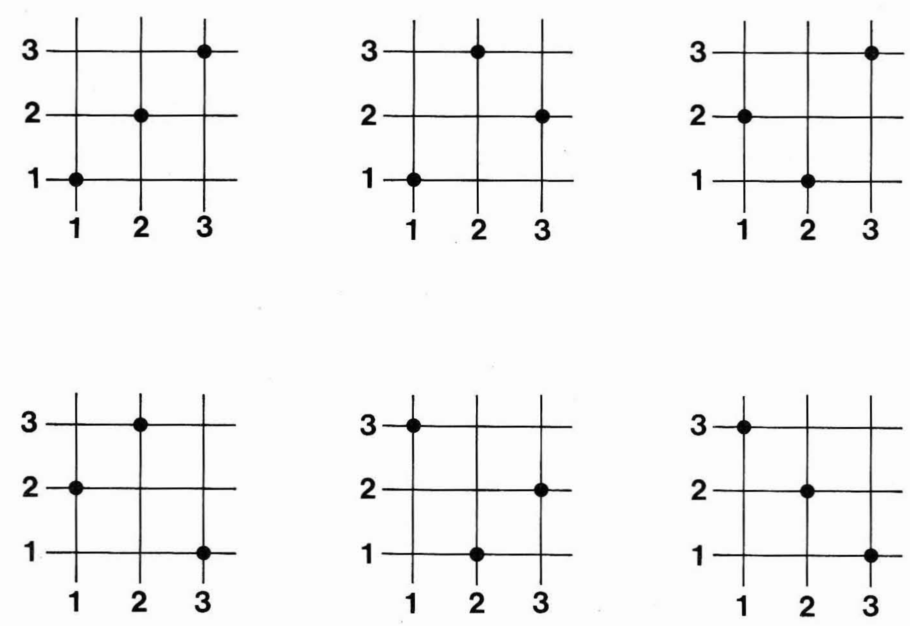
Na zabavi je 8 fantov in 8 deklet. Na koliko načinov si lahko dekleta izberejo plesalce?

Če predpostavimo, da gre za klasičen ples po načelih "natanko en plesalec na plesalko" in "brez lenuhov", ki zagotavljata bijektivnost preslikave, je takih razporedb natanko 8! oziroma 40 320.

ZGLED 2.2.7:

Permutacije elementov množice $\{1, 2, 3\}$ ponazorimo z grafi in s puščičnimi diagrami!

Drugi del naloge puščamo bralcu za vajo, grafi vseh permutacij pa so na sliki 2.4.



Slika 2.4: Grafi vseh permutacij množice števil 1, 2, 3

△ Permutacije dane končne množice $\{a_1, a_2, a_3, \ldots, a_n\}$ so preslikave te množice nase, zato jih lahko poljubno komponiramo. Če denimo prva permutacija (P_1) preslika element a_i v element $a_{i_1} = P_1(a_i)$ in druga permutacija (P_2) preslika a_{i_1} v $a'_{i_1} = P_2(a_{i_1})$, je kompozitum permutacij tista preslikava, ki prenese a_i neposredno v a'_{i_1}.

Pri permutacijah pišemo posamezne preslikave v kompozitumu kar od leve proti desni; tako $P_1 P_2$ pomeni, da najprej opravimo permutacijo P_1 in nato permutacijo P_2.

ZGLED 2.2.8:

Za permutaciji

$$P_1 = \begin{pmatrix} A\,B\,C\,D\,E \\ D\,C\,A\,E\,B \end{pmatrix}, \qquad P_2 = \begin{pmatrix} A\,B\,C\,D\,E \\ C\,A\,D\,B\,E \end{pmatrix}$$

je kompozitum $P_1 P_2$ permutacija

$$\begin{pmatrix} A\,B\,C\,D\,E \\ D\,C\,A\,E\,B \end{pmatrix} \begin{pmatrix} A\,B\,C\,D\,E \\ C\,A\,D\,B\,E \end{pmatrix} = \begin{pmatrix} A\,B\,C\,D\,E \\ B\,D\,C\,E\,A \end{pmatrix}$$

Če se držimo predpostavke o nespremenljivosti naravnega vrstnega reda, lahko kompozitum iz tega zgleda predstavimo tudi samo s slikami, $(DCAEB)\,(CADBE) = (BDCEA)$.

Ker je vsaka permutacija P bijektivna preslikava dane množice nase, obstaja seveda tudi inverzna preslikava oziroma **inverzna permutacija** P^{-1}. Kako deluje, z lahkoto preberemo iz tabele, s katero predstavimo permutacijo P. Po definiciji inverzne preslikave dobimo s kompozitumom permutacije P in k njej inverzne permutacije tako permutacijo P_E, ki ohranja naravni vrstni red.

ZGLED 2.2.9:

$$P = \begin{pmatrix} A\,B\,C\,D\,E \\ D\,C\,A\,E\,B \end{pmatrix} \Longrightarrow P^{-1} = \begin{pmatrix} A\,B\,C\,D\,E \\ C\,E\,B\,A\,D \end{pmatrix}$$

$$P\,P^{-1} = \begin{pmatrix} A\,B\,C\,D\,E \\ D\,C\,A\,E\,B \end{pmatrix} \begin{pmatrix} A\,B\,C\,D\,E \\ C\,E\,B\,A\,D \end{pmatrix} = \begin{pmatrix} A\,B\,C\,D\,E \\ A\,B\,C\,D\,E \end{pmatrix}$$

Bralec se lahko sam prepriča, da je tudi $P^{-1} P = P_E$.

Označimo množico permutacij dane končne množice z močjo n s simbolom $\mathcal{S}_n$. Rečemo lahko, da zanjo veljajo naslednje tri lastnosti:

a) V $\mathcal{S}_n$ je definirana notranja operacija, ki dvema permutacijama priredi natanko določen kompozitum - neko permutacijo dane množice.

b) V množici $\mathcal{S}_n$ obstaja nevtralni element za opisano operacijo, namreč permutacija P_E, ki ohranja naravni vrstni red.

c) Vsaka permutacija P iz $\mathcal{S}_n$ ima v tej množici inverzen element P^{-1}, tako da je $P\,P^{-1} = P^{-1} P = P_E$.

Na kratko povedano, množica $\mathcal{S}_n$ s tako definirano operacijo med permutacijami je **grupa**, ki ji iz razumljivih razlogov rečemo *permutacijska* ali *simetrična grupa*. △

Permutacije s ponavljanjem

Vsi dosedanji zgledi so se nanašali na permutacije brez ponavljanja, saj smo razporejali n različnih elementov. V praksi srečamo tudi primere, ko imamo dve ali več skupin objektov, znotraj katerih objekti glede na karakteristično značilnost niso razločljivi. V tem primeru govorimo o **permutacijah s ponavljanjem**. Razumljivo je, da bo pri nespremenjeni moči množice takih permutacij manj kot onih brez ponavljanja.

ZGLED 2.2.10:

Iz izbrane razporedbe $ABBBB$ bi v primeru, ko B-ji ne bi bili enaki, pri danem položaju A dobili $4! = 24$ permutacij. Če gledamo v nasprotni smeri, ugotovimo, da je zaradi dejstva, da je v celi množici skupina k med seboj nerazločljivih elementov, $k\,!$-krat manj permutacij, kot bi jih bilo sicer.

Lahko torej rečemo takole: *razporedb dolžine n, v katerih na k mestih nastopa enak objekt (ostali pa so od njega in med seboj različni), je $k!$ - krat manj, kot bi bilo permutacij n različnih objektov.*

Tako premislimo tudi splošno pravilo za število permutacij s ponavljanjem, za primer, ko imamo m različnih objektov, iz katerih sestavljamo razporedbe dolžine n, pri čemer $i-$ ti tip objekta nastopa natanko $k_i -$ krat:

$$P_n^{k_1,k_2,\ldots,k_r} = \frac{n\,!}{k_1\,! \cdot k_2\,! \cdot \ldots \cdot k_r\,!}$$

Kadar ni ponavljanj (beri: vsak element se pojavi natanko enkrat), je $P_n = n\,!$ spet število permutacij samih različnih elementov.

ZGLED 2.2.11:

Koliko je razporeditev 4 telovadcev v belih oblačilih, 2 v modrih in 5 v rdečih oblačilih, če ne postavljamo nobenih dodatnih pogojev, vendar pa telovadcev kot oseb ne razlikujemo med seboj (ločimo jih samo po oblačilih)?

$$P_{11}^{4,2,5} = \frac{11\,!}{4\,! \cdot 2\,! \cdot 5\,!} = 6\,930$$

Tokrat je število dovoljenih razporedb povsem neprimerljivo s številom vseh permutacij enajstih elementov, teh bi bilo natanko 11! oziroma 39 916 800.

VAJE

1. Dana je množica $\{a, b, c\}$. Predstavi vse možne permutacije s kombinatoričnim drevesom.

2. Za množico $\{a, b, c\}$ predstavi vse možne permutacije s puščičnimi diagrami in funkcijsko tabelo.

3. Zapiši vse tiste razporedbe množice $\{1, 2, 3, 4\}$, pri katerih se element 4 preslika sam vase.

4. Šola ima na voljo 10 različnih knjig. Na koliko načinov jih lahko razdeli desetim najboljšim učencem, če naj vsak od njih dobi po eno knjigo?

5. Koliko besed dolžine devet je mogoče sestaviti iz simbolov dane abecede { A, E, I, J, L, N, O, S, V }, če mora vsak simbol nastopiti samo po enkrat? Katera od teh razporedb je zadnja, če je leksikografski vrstni red določen s slovensko abecedo?

6. Koliko različnih nizov (razporedb) lahko sestavimo iz šestih raznobarvnih kroglic, če za vsak niz porabimo vse kroglice?

7. Koliko petmestnih števil lahko sestavimo s števkami 1, 2, 3, 4 in 5, če morajo v vsakem številu nastopiti vse naštete števke?

8. $\triangle$ V ravnini je danih osem točk, od teh nobena trojica ne leži na isti premici. Koliko različnih lomljenih črt, ki se "lomijo" natanko v danih točkah, je mogoče narisati?

9. Učenec ve, katerih šest predmetov ima na urniku v ponedeljek, ne more pa se spomniti, v kakšnem vrstnem redu. Spomin si skuša osvežiti tako, da zapiše vse možne vrstne rede, in upa, da bo med njimi prepoznal pravega. Koliko časa bo potreboval v najslabšem primeru za to delo, če za zapis enega vrstnega reda potrebuje 15 sekund?

10. $\triangle$ Na koliko načinov lahko razporedimo v vrsto štiri fante in eno dekle, če naj dekle ne stoji na začetku ali koncu vrste?

11. $\triangle$ Na koliko načinov lahko sestavimo vrstni red govornikov A, B, C in D, če govornik B ne sme nastopiti pred govornikom A?

12. Koliko permutacij brez ponavljanja iz črk A, B, C, Č in D
a) se začne z D;
b) se ne začne niti z A niti z B;
c) se ne konča s skupino DA?

13. Koliko petmestnih števil lahko sestavimo iz cifer 0, 2, 4, 6, 8, če moramo vsakič uporabiti vse cifre in
a) imamo tudi števila, ki se začenjajo z 0, za petmestna;
b) ne dovolimo ničle na prvem mestu?
c) Koliko števil se konča z ničlo?

14. $\triangle$ Na koliko načinov lahko razporedimo v vrsto 5 begonij, 4 primule, 3 pelargonije in 3 fuksije,
a) če naj cvetice iste vrste stoje skupaj;
b) če so lahko poljubno pomešane;
c) če morajo begonije stati skupaj, ostale cvetice pa so lahko pomešane?
Pozor: pravi ljubitelj cvetja razlikuje posamezne primerke iste vrste cvetic.

15. Poenostavi:
a) $10!\,/\,5!$

b) $\dfrac{15! - 14!}{15! + 14!}$

c) $\dfrac{(n-1)! - n!}{(n+1)! + n!}$

16. Zapiši v drugačni obliki s pomočjo simbolov oblike $n!$:

a) $27 \cdot 26$

b) $\dfrac{1}{10 \cdot 9 \cdot 8}$

c) 56

č) $n\,(n^2-1)\,(n^2-4)\,(n^2-9)$

17. Koliko različnih besed dolžine 10 lahko sestavimo iz črk besede matematika?

18. Koliko različnih nizov lahko sestavimo iz (vseh) sedmih kroglic, če so med njimi tri črne, dve beli in dve rdeči kroglici? (Kroglic iste barve med seboj ne razlikujemo.)

19. Pri začetniškem pletenju imamo na voljo samo "leve" in "desne" zanke. Koliko različnih vzorčkov po 10 zank lahko sestavimo iz šestih "levih" in štirih "desnih" zank?

20. $\triangle$ Iz k elementov prve vrste in $n-k$ elementov druge vrste delamo nize po n elementov. Analiziraj, kako je pri dani vrednosti n število permutacij s ponavljanjem odvisno od vrednosti k.

21. S pomočjo obrazca za število permutacij s ponavljanjem in pravila vsote ugotovi, koliko besed dolžine 5 lahko sestaviš bodisi iz 2 črk A in 3 črk B ali obratno, iz 3 črk A in 2 črk B.

22. △ Koliko besed s petimi črkami lahko sestavimo iz črk A, B in C, če lahko A nastopi kvečjemu dvakrat, B kvečjemu enkrat in C kvečjemu trikrat?

23. △ Na koliko načinov lahko petnajst košarkarjev dodelimo trem trenerjem, če naj vsak trener dobi natanko pet igralcev? (Nasvet: obrni nalogo, postavi košarkarje v vrsto in dodeljuj tri trenerje.)

24. △ Posplošimo: na koliko načinov lahko $3n$ različnih predmetov razdelimo med 3 ljudi tako, da vsak od njih prejme natanko n predmetov?

25. Na koliko načinov lahko 10 predmetov razdelimo v prvo, drugo in tretjo skupino tako, da bo v prvi pet predmetov, v drugi trije in v tretji natanko dva?

26. Posplošimo: Na koliko načinov lahko $k+m+n$ predmetov razdelimo v prvo, drugo in tretjo skupino tako, da bo v prvi k predmetov, v drugi m in v tretji natanko n predmetov?

27. Koliko štirimestnih števil lahko sestavimo kot permutacijo cifer števila 1 233? Koliko tako dobljenih števil je večjih od 3 000? Koliko tako dobljenih števil je manjših od 3 000?

28. Na koliko načinov lahko $x^3\, y^3\, z^4$ zapišemo kot produkt ustreznega števila faktorjev x, y in z?

29. △ Kolikšen odstotek vseh permutacij črk besede ANANAS se začenja s črko "S"?

2.3 Variacije

Pri permutacijah smo se ukvarjali z *linearnim urejanjem* vseh elementov dane končne množice (urejanjem v vrsto), oziroma, kot smo ugotovili kasneje, z vprašanjem, koliko je bijektivnih preslikav take množice nase.

Tudi pri variacijah gre za linearno urejanje elementov neke končne množice, vendar tokrat ne jemljemo vseh elementov, ampak samo neko podmnožico, denimo r od skupaj n elementov. Rečemo, da delamo **variacije reda r**. Te so lahko **brez ponavljanja**, kadar sme posamezen element samo enkrat nastopiti v naboru, ali **s ponavljanjem**, če ne postavimo takih omejitev.

Variacije brez ponavljanja

V tem primeru lahko variacije opišemo kot injektivne preslikave množice z močjo r (množice "predalov") v množico z močjo n (torej množico, iz katere izbiramo elemente); seveda mora zaradi zahteve po injektivnosti nujno veljati $r \leq n$.

Po stari navadi tudi zdaj večkrat govorimo o sliki kot o preslikavi in rečemo, da so **variacije reda r brez ponavljanja iz n elementov** dane množice $A = \{a_1, a_2, \ldots a_n\}$ taki nizi dolžine r, v katerih se vsak element a_i pojavi kvečjemu enkrat. Število takih variacij bomo označili z V_n^r.

ZGLED 2.3.1:

Zapišimo prvih nekaj variacij reda 5 brez ponavljanja, ki jih sestavimo iz elementov množice $\{0, 1, 2, \ldots, 9\}$!

Pri variacijah brez ponavljanja imamo po vrsti

01234 01235 01236 01237 01238
01239 01243 01245 01246 01247 ...

Število variacij brez ponavljanja najlaže dobimo z enakim premislekom, kot nas je pripeljal do števila permutacij; razlika je v tem, da moramo sedaj zasesti samo r pozicij. Po osnovnem izreku kombinatorike je torej število

n	$n-1$	$n-2$	$\cdots$	$n-r+1$
1	2	3	$\cdots$	r

variacij reda r iz elementov množice z n elementi (brez ponavljanja) enako

$$V_n^r = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-r+2) \cdot (n-r+1)$$

Če produkt podaljšamo še za $(n-r)\cdot(n-r-1)\cdot\ldots\cdot 2\cdot 1$ in seveda z enakim produktom tudi delimo, lahko obrazec prepišemo v obliko

$$V_n^r = \frac{n!}{(n-r)!}$$

ZGLED 2.3.2:

Vseh variacij reda 5 iz prejšnjega zgleda je

$$V_{10}^5 = \frac{10!}{5!} = 30\,240$$

Pri tem pripominjamo, da si pravzaprav laže zapomnimo, kako to število izračunamo, po osnovnem izreku kombinatorike:
$10\cdot 9\cdot 8\cdot 7\cdot 6 = 30\,240$.

Če pri variacijah brez ponavljanja dovolimo, da r doseže svojo mejno vrednost $r=n$, je vsaka injektivna preslikava med množicama avtomatično tudi surjektivna, s tem pa bijektivna, ali drugače, v takem primeru je število variacij brez ponavljanja enako številu permutacij. To se navsezadnje vidi tudi iz obrazcev za izračun V_n^r, saj imamo

$$r = n \Longrightarrow V_n^r = V_n^n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!$$

S tem smo mimogrede še enkrat videli, zakaj naj bo $0! = 1$.

Variacije s ponavljanjem

ZGLED 2.3.3:

Zapišimo prvih nekaj variacij reda 5 s ponavljanjem, ki jih sestavimo iz elementov množice $\{0, 1, 2, \ldots, 9\}$!

Zdaj gre za začetek naštevanja vseh formalno petmestnih števil ("formalno" zato, ker dopuščamo eno ali več ničel na vodilnih mestih):

00000	00001	00002	00003	00004
00005	00006	00007	00008	00009
00010	00011	00012	00013	00014
00015	00016	00017	00018	00019 ...

Pri nespremenjenih n in r je seveda variacij s ponavljanjem znatno več kot onih brez ponavljanja, saj imamo zdaj za vsakega od "predalčkov" na voljo vseh n različnih elementov (na sosednji strani). Zato je

$$^{(p)}V_n^r = \underbrace{n\cdot n\cdot\ldots\cdot n}_{r} = n^r$$

n	n	n	$\cdots$	n
1	2	3	$\cdots$	r

ZGLED 2.3.4:

Na koliko načinov lahko izpolnimo listek športne napovedi, ki ima trinajst parov, pri vsakem paru pa lahko zapišemo enega od treh simbolov (0, 1 ali 2)?

Pri variacijah s ponavljanjem je treba paziti, kaj je moč množice n in kaj število razredov r. Spet si najbolje pomagamo z r predali in n objekti, ki pa se lahko pojavijo tudi v več predalih. Tako imamo v konkretnem primeru $n=3$ elemente in $r=13$ prostorov, tako da je

$$^{(p)}V_3^{13} = 3^{13} = 1\,594\,323$$

Ker lahko pri variacijah s ponavljanjem vsakemu od r predalov priredimo kateregakoli od n elementov, je število takih variacij enako številu vseh preslikav množice z močjo r v množico z močjo n, ali drugače,

Naj bo $m(A) = r$ in $m(B) = n$ ter $n \geq r$. Potem obstaja natanko n^r preslikav množice A v množico B.

V A J E

1. Zapiši vse variacije reda 2 iz elementov množice $\{A, B, C, D\}$
   a) brez ponavljanja;
   b) s ponavljanjem.

2. Izračunaj V_7^3, V_7^7 in V_6^4.

3. Izračunaj $^{(p)}V_7^3$ in $^{(p)}V_3^7$.

4. Določi število n iz naslednjih pogojev:
   a) $V_n^2 = 20$, *b)* $V_n^2 = 240$, *c)* $V_n^3 = 6\,V_n^2$, *d)* $V_{n+1}^3 = 3n$.

5. Koliko trimestnih števil lahko sestavimo s ciframi 1, 3, 5, 7 in 9, če se cifre ne smejo ponavljati, in koliko, če se lahko ponavljajo?

6. △ Koliko števil iz prejšnje naloge je (v obeh primerih) manjših od 600?

7. Koliko je vseh "pravih" (brez ničle na prvem mestu) petmestnih števil?

8. △ Število variacij s ponavljanjem reda r izmed elementov množice $\{A, B, C, D\}$ je 4 096. Kolikšen je red variacije?

9. Na festivalu sodeluje 13 filmov. Žirija podeljuje nagrade za scenarij, režijo, glavno vlogo in filmsko glasbo. Na koliko načinov je mogoče rezdeliti te nagrade med 13 filmov?

10. V škatli imamo 10 raznobarvnih kroglic. Iz škatle trikrat po vrsti izvlečemo po eno kroglico. Koliko barvnih vzorcev lahko nastane,
    a) če izvlečeno kroglico vsakič vrnemo v posodo;
    b) če izvlečenih kroglic ne vračamo v posodo?

11. Iz kupa 32 kart štirikrat izvlečemo po eno karto, jo pogledamo in vrnemo v kup. Koliko različnih četveric je mogočih? Koliko jih je, če kart ne vračamo?

12. V podjetju je n zaposlenih. Na koliko načinov lahko kolektiv določi svojega predstavnika na treh sejah, če
    a) vse tri seje potekajo sočasno (na različnih mestih);
    b) se lahko isti predstavnik udeleži tudi več kot ene seje?

13. Koliko števil, ki so večja od 2 000 in manjša od 5 000, lahko sestavimo iz cifer 0, 2, 4, 6 in 8, če se cifre ne smejo ponavljati, in koliko, če se lahko poljubnokrat ponovijo?

14. Na zidu so štirje oštevilčeni obešalniki. Na koliko načinov lahko nanje obesimo tri plašče, če naj bo na vsakem kvečjemu en plašč? Na koliko načinov lahko razporedimo plašče, če ne postavljamo te omejitve?

15. Koliko različnih urnikov lahko sestavimo za ponedeljkovih šest ur, če lahko izbiramo med desetimi predmeti in se noben predmet ne sme ponoviti?

16. Koliko stolpcev športne napovedi je treba izpolniti, da imamo zagotovo vsaj 5 pravilno napovedanih rezultatov? (Glej zgled 2.3.4!)

17. S kombinatoričnim drevesom, grafom in puščičnim diagramom predstavi variacije reda 2 iz elementov množice $\{1, 2, 3\}$, tako za variacije brez ponavljanja kot za variacije s ponavljanjem.

2.4 Kombinacije

Razloga za razlikovanje variacij (pri dani množici in danem redu variacije) sta lahko dva:
a) različnost elementov, ki jih izberemo v podmnožico in
b) različen vrstni red v nizu pri sicer enakem izboru.

Če zanemarimo drugi kriterij in dva izbora razlikujemo samo po tem, katere elemente vsebujeta, govorimo o **kombinacijah**. Najprej bomo obravnavali

Kombinacije brez ponavljanja

Kombinacije reda r elementov dane končne množice brez ponavljanja niso nič drugega kot njene ("neurejene") podmnožice z močjo r. Število teh kombinacij bomo označili s C_n^r, določili pa ga bomo s primerjavo števila variacij in kombinacij.

ZGLED 2.4.1:

Zapišimo in preštejmo vse kombinacije reda 3 iz elementov množice $\{a, b, c, d\}$!

VARIACIJE						KOMBINACIJE
abc	acb	bac	bca	cab	cba	abc
abd	adb	bad	bda	dab	dba	abd
acd	adc	cad	cda	dac	dca	acd
bcd	bdc	cbd	cdb	dbc	dcb	bcd

Ko se nehamo zanimati za vrstni red znotraj izbrane trojke, je naborov šestkrat manj, kot jih je bilo prej, ali v obratni smeri – iz vsake kombinacije (= podmnožice) reda $r = 3$ lahko naredimo po $r! = 3! = 6$ različnih variacij.

V splošnem je tedaj pri danih n in r kombinacij za faktor $r!$ manj kot variacij,

$$C_n^r = \frac{V_n^r}{r!} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-r+1)}{r!}$$

Če uporabimo drugo obliko obrazca za število variacij, lahko to pravilo predelamo v

$$C_n^r = \frac{n!}{r!\,(n-r)!} = \binom{n}{r}$$

Simbolu na desni strani rečemo **binomski simbol**, pove nam, koliko različnih podmnožic po r elementov premore množica z n elementi. Kakorkoli bi se to utegnilo bralcu zdeti smešno, bomo ta stavek zapisali še enkrat v nasprotni smeri, v takšnem vrstnem redu, ki ga potrebujemo pri verjetnostnem računu in statistiki:

Na koliko načinov lahko iz množice z n elementi izberemo podmnožico z r elementi, nam pove binomski simbol $\binom{n}{r}$.

ZGLED 2.4.2:

Koliko različnih vzorcev po 3 elemente lahko naberemo iz množice z 10 elementi?

Tu so običajno mišljeni *vzorci brez ponavljanja,* ki jih dobimo pri sočasnem izbiranju predpisanega števila elementov iz množice, ali, kar je ekvivalentno, pri r zaporednih izborih po enega elementa, pri čemer izbranih elementov ne vračamo v množico, tako da je vsak element kvečjemu enkrat izbran v vzorec.

$$\binom{10}{3} = \frac{10!}{3!\,(10-3)!} = \frac{10\cdot 9\cdot 8\cdot 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{3\cdot 2\cdot 1\cdot 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}$$

Po krajšanju v števcu in imenovalcu dobimo

$$\binom{10}{3} = \frac{10\cdot 9\cdot 8}{3\cdot 2\cdot 1} = 120$$

Iz računa se vidi, kako računanje binomskih simbolov skrajšamo, če se držimo prvotne oblike obrazca za število variacij: produkt števil, ki sestavljajo $(n-r)!$ v imenovalcu (v našem primeru $7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1$) in enak "rep" v števcu se vedno krajšata, zato lahko ravnamo takole: v imenovalcu napišemo vse faktorje, ki pripadajo $r!$, v števcu pa produkt enakega števila faktorjev od vključno n navzdol. Pri tem je varno, če v imenovalcu pišemo tudi enojko, da se ne bi zmotili glede števila faktorjev v števcu.

Za računanje z binomskimi simboli veljata dve koristni pravili, ki lahko precej olajšata računanje. Prvo je takšno:

$$\binom{n}{r} = \binom{n}{n-r} \qquad C_n^r = C_n^{n-r}$$

Interpretacija tega pravila je pravzaprav že njegov dokaz: r odlikovanih elementov lahko izberemo na isto število načinov kot tistih $n-r$ elementov, ki jih ne vključimo v izbrano podmnožico, kajti z izbiro podmnožice je natanko določen tudi njen komplement glede na celo množico. (Ta ugotovitev pa naj bralca ne moti pri formalnem dokazu enakosti obeh binomskih simbolov.)

ZGLED 2.4.3:

Nogometno moštvo ima v ekipi enega samega vratarja in 15 kandidatov za igralska mesta. Na koliko načinov lahko sestavimo enajsterico za začetek tekme?

Ker moramo k vratarju izbrati še 10 igralcev, je enajsteric toliko, na kolikor načinov lahko iz množice s petnajstimi elementi izberemo podmnožico z desetimi; po gornjem pravilu pa lahko naredimo račun po krajši poti, če upoštevamo, da je natanko toliko tudi podmnožic po $15-10=5$ elementov. Tako je iskano število

$$\binom{15}{10} = \binom{15}{5} = \frac{15\cdot 14\cdot 13\cdot 12\cdot 11}{5\cdot 4\cdot 3\cdot 2\cdot 1} = 3\,003$$

Ljubitelji nogometa bi se sicer pritožili, da vsak nogometaš res ne sodi na vsako mesto v ekipi, ampak to kombinatorike ne moti...

Drugo pravilo je neke vrste adicijski izrek za binomske simbole:

$$\binom{n}{r} + \binom{n}{r+1} = \binom{n+1}{r+1} \qquad C_n^r + C_n^{r+1} = C_{n+1}^{r+1}$$

Dokaz te relacije je neposreden: zapišemo oba binomska simbola v razviti obliki, ju spravimo na skupni imenovalec in predelamo v obliko, iz katere prepoznamo binomski simbol na desni strani enačaja:

$$\binom{n}{r} + \binom{n}{r+1} = \frac{n!}{r!(n-r)!} + \frac{n!}{(r+1)!(n-(r+1))!} =$$
$$\frac{n!}{r!(n-r-1)!}\left\{\frac{1}{n-r} + \frac{1}{r+1}\right\} =$$
$$\frac{n!}{r!(n-r-1)!}\cdot\frac{r+1+n-r}{(n-r)(r+1)} = \frac{(n+1)!}{(r+1)!(n-r)!} =$$
$$= \frac{(n+1)!}{(r+1)!((n+1)-(r+1))!} = \binom{n+1}{r+1}$$

Ob tem opozorimo še na tri očitna dejstva, ki tudi ne potrebujejo dosti pojasnil:

$$\binom{n}{0} = 1 \qquad \binom{n}{n} = 1 \qquad \binom{n}{1} = n$$

V prvem primeru se spet srečamo s prazno množico in dejstvom, da ima poljubna množica natanko eno prazno podmnožico; pri formalnem računanju tega binomskega simbola (vaja!) se moramo spomniti dogovora, da je $0! = 1$.

Drugi primer govori o tem, koliko svobode imamo, ko je treba iz množice n ljudi izbrati n prostovoljcev, tretje pravilo pa samo prešteje posamezne elemente v množici.

Kombinacije s ponavljanjem

Včasih pri vzorčenju ravnamo tudi takole:
- iz dane množice (na slepo) izberemo en element,
- si ga ogledamo (izmerimo, opišemo,...)
- in ga vrnemo v množico,

nato pa postopek ponovimo tolikokrat, kolikor je zahtevana velikost vzorca.

Na opisani način nastanejo *vzorci s ponavljanjem*, saj posamezen element lahko nastopi v vzorcu tudi več kot enkrat.

ZGLED 2.4.4:

Učitelj na podružnični šoli ima v razredu samo štiri učence. Zastavi jim tri vprašanja. Na koliko načinov lahko razporedi vprašanja med učence,

a) če je pomembno samo to, kolikokrat je odgovarjal posamezen učenec;

b) če je važno tudi to, na katero vprašanje je kdo odgovarjal?

a) Če označimo učence z A, B, C in D, imamo v prvem primeru naslednje možnosti: AAA, AAB, AAC, AAD, ABB, ABC, ABD, ACC, ACD, ADD, BBB, BBC, BBD, BCC, BCD, BDD, CCC, CCD, CDD in DDD. Tako imamo dvajset kombinacij s ponavljanjem. Pri predpostavkah, ki smo jih navedli, niza AAB ne razlikujemo od ABA ali BAA.

b) V drugem primeru imamo opraviti z variacijami s ponavljanjem, ki jih bralec že obvlada; ker imamo tri predalčke (vprašanja) in štiri elemente, ki jih porazdeljujemo (učence), je takih variacij $4^3 = 64$. Spet vabimo bralca, da jih izpiše.

Na tako vzorčenje sicer v praksi redkeje naletimo (in so pogostejši vzorci brez ponavljanja, glej 4. poglavje), vendar ga zaradi izrednega teoretičnega pomena ne moremo zanemariti. Zato si poglejmo, koliko je kombinacij reda r s ponavljanjem iz množice z n elementi. Zaradi enostavnosti vzemimo za premislek kar množico $\mathbb{N}_n = \{1, 2, ... n\}$. Naj bo $i_1 i_2 ... i_r$ poljubna kombinacija s ponavljanjem reda r iz te množice, pri čemer se držimo dogovora, da so v nizu števila vedno zapisana v nepadajočem vrstnem redu (podobno smo se v prejšnjem zgledu držali slovenske abecede). Vsaki kombinaciji $i_1 i_2 ... i_r$ priredimo neko drugo kombinacijo $j_1 j_2 ... j_r$, ki jo dobimo tako, da število na prvem mestu pustimo nespremenjeno, številu na drugem mestu prvotne kombinacije prištejemo 1, onemu na tretjem mestu prištejemo 2, ..., zadnjemu (r- temu) prištejemo $r-1$, torej

$$j_1 = i_1 + 0,\ j_2 = i_2 + 1,\ j_3 = i_3 + 2,\ \dots,\ j_r = i_r + r - 1$$

S tem smo dosegli, da so v nizu $j_1 j_2 ... j_r$ vsa števila različna, ne glede na to, iz katere kombinacije $i-$ jev smo ta niz dobili. To prirejanje je obratno enolično, z nizom $i-$ jev je natančno določen niz $j_1 j_2 ... j_r$, velja pa tudi obratno. Za primer, ko imamo $n = 3$ in $r = 2$, dobimo na opisani način iz kombinacij s ponavljanjem reda 2 iz treh elementov (11, 12, 13, 22, 23 in 33) kombinacije brez ponavljanja reda 2 iz štirih elementov (12, 13, 14, 23, 24, 34). Ni težko videti, da s takšno obratno enolično prireditvijo čisto splošno iz kombinacij s ponavljanjem reda r izmed n elementov dobimo vse kombinacije brez ponavljanja reda r izmed $n+r-1$ elementov. Ker imata končni množici, med katerima obstaja bijektivna preslikava, isto moč, je

$$^{(p)}C_n^r = C_{n+r-1}^r = \binom{n+r-1}{r}$$

ZGLED 2.4.5:

Cvetličarka ima na voljo 4 barve vrtnic, iz katerih sestavlja šopke po 7 cvetic. Koliko različnih šopkov lahko naredi, če jih razlikujemo samo po tem, koliko vrtnic posamezne barve je v šopku?

Ker je barv, ki so na voljo (n), manj kot mora biti vrtnic v šopku (r), se morajo barve nujno ponavljati; iskano število različnih šopkov je

$$^{(p)}C_4^7 = C_{10}^7 = \binom{10}{7} = \binom{10}{3} = 120$$

Vezane kombinacije

Pogosto pa naletimo na primere, ko izbiramo vzorce iz populacije, razdeljene na več med seboj tujih podmnožic (razredov, stratumov), pri čemer zahtevamo predpisano strukturo vzorca glede na to, iz katerega dela množice prihajajo elementi. Ker predpostavimo, da so izbiranja predpisanega števila elementov iz posamezne podmnožice neodvisna od tega, kaj smo izbrali v ostalih podmnožicah, lahko tudi tu uporabimo osnovni izrek kombinatorike:

Naj bo končna množica A z močjo n razdeljena na m med seboj paroma tujih razredov A_i,

$$A_1 \cup A_2 \cup ... \cup A_m = A \qquad A_i \cap A_j = \emptyset \qquad (i \neq j)$$

z močmi

$$n_1,\ n_2,\ ... n_m \qquad (n_1 + n_2 + ... + n_m = n)$$

Potem je število načinov, na katere je mogoče iz množic A_i izbrati natanko po r_i elementov ($r_1 + r_2 + \ldots + r_m = r$), določeno z obrazcem

$$C_{n_1,n_2,\ldots,n_m}^{r_1,r_2,\ldots,r_m} = \binom{n_1}{r_1} \cdot \binom{n_2}{r_2} \cdot \ldots \cdot \binom{n_m}{r_m}$$

Včasih uporabljamo za take izbore izraz **vezane kombinacije**, ker smo pri izbiranju vezani na predpise o strukturi vzorca glede na strukturo celotne populacije.

ZGLED 2.4.6:

V seriji 100 izdelkov je 8 izdelkov pokvarjenih. Iz serije sočasno izberemo v vzorec 5 izdelkov. Koliko je med vsemi vzorci te velikosti takšnih, da je v njih natanko en pokvarjen izdelek?

Če primerjamo podatke s splošnim navodilom, vidimo, da je

$$n = 100,\ n_1 = 92,\ n_2 = 8,\ r = 5,\ r_1 = 4,\ r_2 = 1$$

in uporaba obrazca je neposredna:

$$C_{92,8}^{4,1} = \binom{92}{4} \cdot \binom{8}{1} = \frac{92 \cdot 91 \cdot 90 \cdot 89}{4 \cdot 3 \cdot 2 \cdot 1} \cdot 8 = 22\,353\,240$$

Število takih vzorcev je ogromno, pa še zdaleč ne takšno kot število vseh vzorcev velikosti 5 iz populacije s 100 elementi (vaja!).

Porazdelitve

Pri izpeljavi obrazca za število vezanih kombinacij smo izhajali iz predpostavke, da je množica A razdeljena na m medsebojno tujih množic A_i,

$$A_1 \cup A_2 \cup \ldots \cup A_m = A, \quad A_i \cap A_j = \emptyset \quad (i, j = 1, 2, \ldots, m;\ i \neq j)$$

Zastavimo si vprašanje, koliko je sploh takšnih razdelitev dane množice A na podmnožice, ali drugače, na koliko načinov lahko n različnih elementov razporedimo v m urejenih (razločljivih) predalov tako, da bo v predalih po vrsti $n_1, n_2, \ldots, n_m$ elementov ($n_1 + n_2 + \ldots + n_m = n$), pri čemer vrstni red elementov znotraj posameznega predala ni pomemben.

Porazdelitve, pri katerih razlikujemo predale, (recimo, da so oštevilčeni od 1 do m), bomo imenovali **urejene porazdelitve**. Pri takih porazdelitvah na primer razlikujemo porazdelitev "elementa 1 in 2 v prvi predal, element 3 v drugega" od porazdelitve "element 3 v prvi predal, elementa 1 in 2 v drugega". Število urejenih porazdelitev ugotovimo takole: prvi predal, v katerega moramo dati n_1 elementov, lahko napolnimo na $\binom{n}{n_1}$ načinov; za drugi predal izberemo n_2 elementov izmed preostalih $n - n_1$, kar lahko storimo na $\binom{n-n_1}{n_2}$ načinov; elemente za tretji predal lahko naberemo na $\binom{n-n_1-n_2}{n_3}$ načinov,..., za napolnitev zadnjega predala je $\binom{n-n_1-\ldots-n_{m-1}}{n_m} = \binom{n_m}{n_m} = 1$ možnost, torej je v celoti vseh izborov toliko, kolikor znaša produkt

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n_m}{n_m}$$

Če ta produkt razvijemo in opravimo vsa možna krajšanja, se izkaže (preveri!), da je natanko enak

$$\frac{n!}{n_1!\,n_2!\,n_3!\,\ldots\,n_m!}$$

to pa je (kot se morda še spomnimo) število permutacij n elementov, od katerih se prvi ponavlja n_1 – krat, drugi n_2 – krat, ..., m – ti n_m – krat.

Naloga: Ugotovi, kako lahko pridemo do tega obrazca z neposrednim premislekom! S tem si lahko pomagaš tudi pri naslednjih dveh zgledih.

ZGLED 2.4.7:

Na koliko načinov lahko 9 različnih igrač razdelimo med 4 otroke tako, da najmlajši dobi 3, ostali pa po dve igrači?

Formalno lahko rečemo, da moramo množico z $n = 9$ elementi razdeliti na štiri podmnožice, ki imajo po $n_1 = 3$, $n_2 = 2$, $n_3 = 2$, in $n_4 = 2$ elementa. Porazdelitve so urejene, zato imamo po obrazcu

$$\frac{n!}{n_1!\,n_2!\,n_3!\,n_4!} = \frac{9!}{3!\,2!\,2!\,2!} = 7\,560$$

različnih načinov za razdelitev igrač po opisanih načelih.

ZGLED 2.4.8:

V razredu je 12 učencev, ki jih je treba razdeliti v 3 skupine po štiri učence, ki bodo pisali test iz po enega od treh različnih predmetov. Na koliko načinov lahko opravimo razdelitev?

"Predali" so razločljivi (različni predmeti), zato je teh načinov

$$\frac{12!}{4!\,4!\,4!} = 34\,650$$

Prav pri tem zgledu je lahko zelo koristna uporaba neposrednega premisleka, v katerem zamenjamo vloge: predstavljajmo si, da imamo po štiri bele, rdeče in modre kroglice in 12 učencev; vsakemu predalu (pardon, učencu) dodelimo po eno kroglico. Število porazdelitev je enako številu permutacij 12 objektov, ki so treh različnih tipov in se vsak ponovi po štirikrat.

Za konec poglavja o kombinacijah dodajmo še nekaj koristnih premislekov in zgledov.

ZGLED 2.4.9:

Na koliko načinov lahko množico n učencev razdelimo v I. in II. skupino, če ne postavljamo nobenih dodatnih omejitev glede števila učencev v posamezni skupini?

Pri izbiranju učencev za I. skupino imamo pri vsakem natanko dve možnosti, lahko ga namreč damo vanjo ali pa ga pustimo za drugo skupino. Ker je število možnosti v vsaki fazi izbiranja neodvisno od tega, kako smo se odločili v prejšnjih fazah (pri ostalih učencih), je takšnih porazdelitev po osnovnem izreku kombinatorike

$$\underbrace{2 \cdot 2 \cdot \ldots \cdot 2}_{n \text{ faktorjev}} = 2^n$$

Drugače rečeno, na toliko različnih načinov lahko napolnimo prvo skupino z $0, 1, 2, \ldots, n$ učenci, s tem pa je tudi II. skupina drugega predala enolično določena.

Ob tem zgledu ne smemo pozabiti, da sta upoštevani tudi obe mejni možnosti, ko se vsi učenci znajdejo bodisi v I. ali v II. skupini, preostala skupina pa ostane prazna. Taki primeri niso tako redki: če I. skupina pomeni učence, ki so razred izdelali, II. skupina pa "žalujoče ostale", bi v razredu s 100% uspehom II. skupina pač pomenila prazno množico.

Število načinov, na katere lahko izberemo tiste elemente dane množice z močjo n, ki sodijo med izbrane (I. skupina iz prejšnjega zgleda), je seveda ravno enako številu vseh podmnožic dane množice, zato velja:

Za končno množico z močjo n ima potenčna množica moč 2^n.

Ta pomembni rezultat bomo še enkrat preverili v naslednjem razdelku, tega pa končajmo še z enim zgledom.

ZGLED 2.4.10:

Na koliko načinov lahko množico z 10 elementi razdelimo v dva dela?

Opozorimo, da je problem nekoliko drugačen kot v prejšnjem zgledu. Gre namreč za takšno delitev, pri kateri ne razlikujemo boljšega in slabšega kosa množice. Kaj to pomeni, lahko premislimo ob naslednjem primeru: Vzemimo množico števil $\{1, 2, 3, 4, 5\}$ in eno možnih razdelitev $\{1, 2, 3\}$ in $\{4, 5\}$. Ko preštevamo podmnožice, štejemo enkrat mednje $\{1, 2, 3\}$, drugič pa še $\{4, 5\}$, čeprav obe v bistvu določata eno razdelitev množice v dva dela.

Tako kot besedilo običajno razumemo, bo torej delitev v dva dela natanko dvakrat manj, kot ima dana množica vseh podmnožic, namreč $2^{10}/2 = 512$.

Pri tem pa spet ne pozabimo, da je v številu 512 zajeta tudi tista razdelitev, ko je en del celotna množica, drugi del pa njena prazna

podmnožica. Marsikdo bi rekel, da to sploh ni razdelitev množice na dva dela, vendar kruto življenje ("Delila bova bratsko, meni vse, tebi ostanek!") uči, da nima prav...

V A J E

1. Zapiši vse kombinacije reda 2 (brez ponavljanja in s ponavljanjem) izmed črk A, B, C, Č, D. Pazi na leksikografski vrstni red.

2. Zapiši vse kombinacije brez ponavljanja reda 3, ki jih lahko sestavimo iz črk A, B, C, Č, D.

3. Izračunaj vrednosti naslednjih binomskih simbolov:
a) $\binom{5}{2}$ b) $\binom{7}{3}$ c) $\binom{7}{4}$ č) $\binom{100}{98}$ d) $\binom{99}{5} + \binom{99}{6}$

4. Poenostavi:
a) $\binom{n}{n-2}$ b) $\binom{n+5}{n+3}$ c) $\binom{7}{4}$ č) $\binom{n}{k} : \binom{n-1}{k-1}$ d) $\frac{n}{k} \cdot \binom{n-1}{k-1}$

5. Za katere vrednosti n so izpolnjene naslednje enačbe:
a) $C^n_{n+2} = 10$ b) $C^2_n = C^3_{n+1}$ c) $C^2_n = C^3_{n+2} - 3$

6. Na koliko načinov lahko 59-članska skupščina izbere petčlansko predsedstvo (izmed svojih članov)?

7. Koliko različnih peterk lahko košarkaški trener sestavi iz desetčlanske ekipe?

8. Učenec mora izbrati sedem od desetih vprašanj, ki so zapisana na listku. Na koliko načinov lahko to stori? Koliko je načinov, če mora obvezno odgovoriti na prva tri vprašanja (in na katerakoli štiri od preostalih sedmih)?

9. Učenca zanima pet knjig, vendar si lahko po knjižničnih pravilih izposodi samo tri knjige sočasno. Koliko izborov ima na voljo? Koliko je načinov, če se eni od petih knjig nikakor noče odpovedati?

10. Koliko je pravih štirimestnih števil (ne dovolimo ničle na prvem mestu), pri katerih je vsaka naslednja cifra večja od prejšnje?

11. Koliko je štirimestnih števil, pri katerih je vsaka naslednja cifra manjša od prejšnje?

12. V ravnini je n točk, od katerih nobena trojica ne leži na isti premici. Koliko premic je tedaj določeno s temi n točkami?

13. $\triangle$ Koliko različnih premic določa n točk v ravnini, od katerih jih m $(m < n)$ leži na isti premici, od ostalih pa nobena trojica ne leži na isti premici?

14. V ravnini so dane točke A, B, C, D, E in F, od katerih nobena trojica ne leži na isti premici.
a) Koliko različnih premic določajo te točke?
b) Koliko različnih trikotnikov določajo te točke?
c) Koliko je med temi trikotniki takih, ki imajo točko C za eno od oglišč?
d) Koliko je med temi trikotniki takih, ki imajo daljico AF za eno od stranic?

15. V ravnini je danih n točk, od katerih nobena trojica ne leži na isti premici, določajo pa natanko n različnih trikotnikov. Koliko je teh točk?

16. Na koliko načinov lahko med desetimi člani komisije izberemo štiričlansko predsedstvo, če najstarejša člana komisije ne smeta biti sočasno v predsedstvu?

17. Zapiši vse kombinacije s ponavljanjem reda 4 izmed elementov množice $Z = \{0, 1\}$. Zapisani vrstni red naj bo "naravni vrstni red". Kakšen je (lahko) praktičen pomen teh nizov?

18. Izračunaj števila kombinacij s ponavljanjem:
a) ${}^{(p)}C_6^4$ b) ${}^{(p)}C_4^6$ c) ${}^{(p)}C_{12}^5$ č) ${}^{(p)}C_9^3$

19. Iz elementov neke množice smo dobili 276 kombinacij s ponavljanjem drugega reda. Kolikšna je bila moč množice?

20. Koliko je vseh štirimestnih števil, v katerih je vsaka naslednja cifra manjša ali kvečjemu enaka prejšnji?

21. Tovarna proizvaja frnikole osmih barv in jih prodaja po dvanajst skupaj. Koliko različnih kompletov lahko sestavi?

22. Desetčlanska delegacija mora določiti delegate za tri seje. Na koliko načinov lahko to stori, če je posamezen delegat lahko določen tudi za dve ali celo za vse tri seje?

23. V slaščičarni prodajajo 12 vrst tort. Na koliko načinov lahko izberemo
a) štiri (kakršnekoli) torte;
b) največ štiri (enake ali različne) torte?

24. Izračunaj:
a) $C_{12,7}^{4,3}$ b) $C_{8,8}^{2,2}$ c) $C_{6,5,4}^{3,2,2}$

25. V razredu je 15 fantov in 12 deklet. Na koliko načinov lahko sestavimo delegacijo, v kateri bosta dva fanta in eno dekle?

26. Na matematičnem tekmovanju je moral tekmovalec izbrati dve od petih nalog iz prve skupine in tri od sedmih nalog iz druge skupine. Koliko variant ima v celoti na voljo?

27. Izmed 5 fizikov, 4 kemikov in 3 matematikov je treba izbrati šestčlansko komisijo, tako da bodo v njej trije fiziki, dva kemika in 1 matematik. Koliko različnih komisij je mogoče sestaviti pri teh pogojih?

28. V dvorani je 18 plesalcev in 14 plesalk. Koliko različnih plesnih parov in koliko različnih četveric (po dva plesalca in dve plesalki) je mogoče sestaviti?

29. Koliko besed s petimi črkami lahko sestavimo iz petih samoglasnikov in dvajsetih soglasnikov slovenske abecede, če naj ima vsaka od teh besed tri soglasnike in dva samoglasnika, črke pa se ne smejo ponavljati?

30. Koliko besed iz prejšnje naloge vsebuje črko "ž"?

31. Koliko različnih petčlanskih predsedstev z vnaprej določenim predsednikom lahko izmed svojih članov izbere dvajsetčlanski odbor? (Tudi bodoči predsednik je član odbora.)

32. V škatli imamo 5 belih in 7 črnih kroglic. Iz škatle sočasno izvlečemo 3 kroglice. Koliko različnih vzorcev lahko pri tem dobimo,
a) če kroglice iste barve med seboj razlikujemo;
b) če kroglic iste barve med seboj ne razlikujemo?

33. V razredu je 25 učencev, od tega 5 fantov. Na koliko načinov lahko sestavijo tričlansko delegacijo, v kateri mora biti vsaj en fant?

34. Izmed 4 natakarjev in 5 kuharjev je treba sestaviti strokovno komisijo tako, da bosta v njej dva natakarja in trije kuharji. Na koliko načinov je mogoče to storiti,
a) če ni nobenih dodatnih omejitev;
b) če mora določen natakar zagotovo biti v komisiji;
c) če dva kuharja nočeta biti oba hkrati v komisiji?

35. Na koliko načinov lahko izberemo 3 ljudi izmed 15,
a) če je eden od njih vključen v vsak izbor;
b) če dva od njih nista vključena v noben izbor;
c) če je eden vedno vključen, druga dva pa nikoli sočasno?

36. 60 učencev opravlja sprejemni izpit. Na koliko načinov se lahko izteče ta izpit, če sta možni samo oceni "pozitivno" in "negativno"? ("Pozitivne" učence damo v prvi "predal", "negativne" v drugega... Zapiši število možnih načinov in ga poskusi oceniti, na primer z logaritmi.)

2.5 Binomski izrek

Kvadrat in kub dvočlenika (binoma) nam ne povzročata težav, pri višjih potencah si pomagamo z zaporednimi množenji, upoštevaje, da je

$$(a+b)^n = \underbrace{(a+b)(a+b)\dots(a+b)}_{n \text{ faktorjev}}$$

Pri takem množenju nastane vsota produktov oblike $a^{n-r}\, b^r$. Vsota eksponentov je v vsakem sumandu n. Takšen produkt nastane, kadar iz katerihkoli r faktorjev vzamemo v produkt člen b, iz preostalih $n-r$ faktorjev pa člen a. Faktorje, iz katerih pride v produkt b, lahko izberemo natanko na $\binom{n}{r}$ načinov, zato je vrednost tega binomskega simbola tudi koeficient, ki v razvoju gornjega binoma pripada členu $a^{n-r}\, b^r$. Zato lahko zapišemo n-to potenco binoma takole:

$$(a+b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \dots + \binom{n}{n-1} a\, b^{n-1} + \binom{n}{n} b^n$$

Tu smo že upoštevali, kaj pomeni a^0, a^1, b^0, b^1. S simbolom za vsoto je zapis še bolj eleganten,

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^{n-r}\, b^r$$

Za zapis poljubne potence binoma potrebujemo torej vrednosti ustreznih binomskih simbolov. Če se spomnimo na njihovo simetričnost, jih je dovolj že polovica, če upoštevamo še "adicijski izrek", pa dobimo zelo preprosto shemo za računanje binomskih simbolov, ki jo imenujemo **Pascalov trikotnik** (po francoskem matematiku B. Pascalu, 1623 - 1662):

$$\begin{array}{ccccccccccccccc|c}
 & & & & & & & 1 & & & & & & & & n=0 \\
 & & & & & & 1 & & 1 & & & & & & & n=1 \\
 & & & & & 1 & & 2 & & 1 & & & & & & n=2 \\
 & & & & 1 & & 3 & & 3 & & 1 & & & & & n=3 \\
 & & & 1 & & 4 & & 6 & & 4 & & 1 & & & & n=4 \\
 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & & n=5 \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & & n=6 \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1 & n=7 \\
.. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & \dots
\end{array}$$

Elementi v trikotniku so vrednosti binomskih simbolov, vsak od njih je v skladu z omenjeno aditivno relacijo vsota obeh simbolov, ki stojita nad njim, seveda z izjemo robov, kjer je vrednost vedno 1. (Zakaj?)

Če ima kateri od členov v binomu negativen predznak, se seveda spremenijo predznaki tistih sumandov, kjer ta člen nastopa z lihim eksponentom.

ZGLED 2.5.1:

Po binomskem izreku razvijmo $(x - \sqrt{2}\,)^5$!

Koeficiente v razvoju preberemo kar iz Pascalovega trikotnika pri $n = 5$, po vrsti so enaki 1, 5, 10, 10, 5, 1, zato je

$$\begin{aligned}(x-\sqrt{2})^5 &= 1\cdot x^5 + 5\cdot x^4\cdot(-\sqrt{2}\,) + 10\cdot x^3\cdot(-\sqrt{2}\,)^2 + \\ &+ 10\cdot x^2\cdot(-\sqrt{2}\,)^3 + 5\cdot x\cdot(-\sqrt{2}\,)^4 + 1\cdot(-\sqrt{2}\,)^5 = \\ &= x^5 - 5\sqrt{2}x^4 + 20x^3 - 20\sqrt{2}x^2 + 20x - 4\sqrt{2}\end{aligned}$$

Še posebej je treba paziti na predznake, če sta negativno predznačena oba člena v binomu, v takem primeru bi bilo morda najbolj varno najprej izpostaviti ustrezno potenco števila -1 in nato računati s pozitivnimi členi.

ZGLED 2.5.2:

Zapišimo peti člen v razvoju $(\sqrt{3} - \sqrt[4]{5}\,)^8$!

Očitno je $n = 8$, $a = \sqrt{3}$ in $b = -\sqrt[4]{5}$. Previdno pa je treba opaziti, da je treba zaradi začetka štetja pri $r = 0$ za peti člen vzeti $r = 4$. Tako je iskani člen v razvoju (vaja!) enak 3 150.

ZGLED 2.5.3:

Z binomskim izrekom ocenimo vrednost $1\text{'}004^7$!

$$1\text{'}004^7 = (1 + 0\text{'}004)^7 =$$

$$\binom{7}{0} 1^7 + \binom{7}{1} 1^6\, 0\text{'}004 + \binom{7}{2} 1^5\, 0\text{'}004^2 + \binom{7}{3} 1^4\, 0\text{'}004^3 + \dots$$

Vseh členov ni treba pisati, ker vidimo, da vrednosti zelo hitro padajo, ko imamo opraviti z vedno višjimi potencami majhnega števila $0\text{'}004$. Iz prvega dela razvoja pa dobimo iskano oceno

$$1\text{'}004^7 = \dots \approx 1 + 7\cdot 0\text{'}004 + 21\cdot 0\text{'}000016 = 1\text{'}028336 \approx 1\text{'}0283$$

Pri takih računih je treba presoditi, koliko členov vključimo v obravnavo.

S pomočjo razvoja binoma lahko pokažemo, da ima vsaka množica z močjo n ravno 2^n vseh podmnožic (vključno s samo sabo in s prazno podmnožico). Ker je po binomskem izreku

$$(1+1)^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n-1} + \binom{n}{n}$$

po drugi strani pa je to ravno 2^n, smo trditev že dokazali, če le uvidimo, da vsota binomskih simbolov pomeni seštevek števil vseh podmnožic z močmi $0, 1, 2, \dots, n$.

VAJE

1. Po binomskem izreku izračunaj naslednje potence:

   a) $(x+1)^5$ b) $(y-2)^6$ c) $(x+3y)^4$ č) $(2x+y^2)^5$

   d) $(2x-y)^4$ e) $(a-\sqrt[3]{b})^8$ f) $(3-i)^5$ g) $(-1+2i)^7$

2. Napiši

   a) deseti člen v razvoju $(x^2-2y^3)^{15}$;

   b) šesti člen v razvoju $(\sqrt{x}-2x)^8$.

3. Izračunaj tisti člen v razvoju $(2x^2-y^3)^8$, ki vsebuje potenco x^{10}.

4. Poišči člen v razvoju $(3xy^2+z^2)^7$, ki vsebuje potenco y^6.

5. Poišči prve tri člene v razvoju $(x^2-\frac{2}{x})^{10}$.

6. Poišči prve tri člene v razvoju $(x+\frac{2}{x^2})^{12}$.

7. Izračunaj po binomskem izreku približka za $0{\cdot}994^4$ in $2{\cdot}003^6$ tako, da obakrat upoštevaš samo prve tri člene v razvoju, rezultat pa vsakič primerjaj s pravo vrednostjo, ki jo da kalkulator.

8. Pri razvoju potence $(1+x)^n$ po binomskem izreku smo ugotovili, da sta koeficienta pri x^5 in x^{12} enaka. Kolikšen je bil eksponent n?

9. Koliko nepraznih podmnožic ima množica samoglasnikov slovenske abecede?

10. Na koliko načinov lahko desetčlanski odbor izbere tri ali več delegatov na neko sejo?

△ 11. Pokaži, da velja pri poljubnem naravnem številu n enačba:

$$\binom{n}{0}-\binom{n}{1}+\binom{n}{2}-\binom{n}{3}+\cdots\pm\binom{n}{n}=0$$

2.6 Načelo vključitev in izključitev

Uporaba tega načela nam v različnih primerih omogoča, da ugotovimo, koliko je v dani množici elementov, ki se odlikujejo z eno ali več predpisanimi lastnostmi. Vsebino načela bomo spoznali ob naslednjem zgledu.

ZGLED 2.6.1:

Koliko je naravnih števil, ki so manjša ali enaka 99 in so pri tem deljiva s 3 ali s 5?

Ker je $99:3=33$, je v množici $\mathbb{N}_{99}=1,2,3,\ldots,98,99$ natanko 33 števil, ki so deljiva s 3; podobno bi ugotovili, da je med elementi množice N_{99} 19 elementov, ki so deljivi s 5. Če bi preprosto sešteli $33+19=52$ in trdili, da je to število odgovor na zastavljeno vprašanje, seveda ne bi imeli prav. Število 15 smo na primer šteli dvakrat, enkrat kot mnogokratnik 3, drugič še kot mnogokratnik 5. Isto velja za vsa števila, ki so hkrati deljiva s 3 in 5; teh je šest (15, 30, 45, 60, 75 in 90), zato je pravilen odgovor na gornje vprašanje $33+19-6=46$.

Poskusimo nalogo prevesti v jezik teorije množic! V univerzalni množici (naravnih števil, ki so manjša od 100) smo z lastnostma A_1: "element je deljiv s 3" oziroma A_2: "element je deljiv s 5" opredelili dve podmnožici in vprašali po moči njune unije oziroma po številu elementov, ki se ponašajo z (vsaj) eno od obeh lastnosti. Rešena naloga nas je naučila, da to število izračunamo po obrazcu

$$m(A_1\cup A_2)=m(A_1)+m(A_2)-m(A_1\cap A_2) \qquad (1)$$

V tem obrazcu se lepo vidi, kako deluje načelo vključitev in izključitev: najprej **vključimo** vse elemente s prvo lastnostjo in vse elemente z drugo lastnostjo, nato pa **izključimo** tisto, kar smo šteli dvakrat (število elementov prve množice + število elementov druge množice − število elementov preseka, ker smo le-te šteli dvakrat). Pri tem je vseeno, ali govorimo o moči podmnožic ali pa kar o številu elementov z lastnostjo, ki takšno podmnožico določa. Tako lahko tudi oznako A' pojasnimo na dva načina: kot komplementarno množico množice A (v izbrani univerzalni množici) ali kot lastnost "*ne* A", kakor nam pač v posameznem praktičnem primeru bolj ustreza. Za posamezen element dane množice velja bodisi lastnost A ali pa A', nikakor pa ne obe hkrati, zato je $m(A\cap A')=0$, $m(A\cup A')$ pa po obrazcu (1) kar $m(A)+m(A')$. Ker ima vsak element eno od obeh lastnosti, je $m(A\cup A')$ enako moči množice, recimo n. Tako smo ugotovili:

Naj bo dana poljubna končna množica z močjo n. Potem vsaka lastnost A razdeli to množico v dve med seboj tuji množici, tako da velja

$$m(A)+m(A')=n \qquad (2)$$

Ker gre pri tem v bistvu za računanje s podmnožicami in z njihovimi močmi, po izkušnjah iz teorije množic brez težav premislimo naslednje enačbe, ki jih tu ne bomo posebej dokazovali.

$$(A')' = A \tag{3}$$

$$(A_1 \cup A_2 \cup \ldots \cup A_m)' = A_1' \cap A_2' \cap \ldots \cap A_m' \tag{4}$$

$$(A_1 \cap A_2 \cap \ldots \cap A_m)' = A_1' \cup A_2' \cup \ldots \cup A_m' \tag{5}$$

ZGLED 2.6.2:

Izračunaj $m\,(A_1' \cap A_2')$!

Najprej poglejmo obrazec (4) z desne v levo, to nam da

$$m\,(A_1' \cap A_2') = m\,((A_1 \cup A_2)')$$

to pa je zaradi (2) enako

$$n - m\,(A_1 \cup A_2)$$

Če za $m\,(A_1 \cup A_2)$ vstavimo izraz iz (1), dobimo končno obliko

$$m\,(A_1' \cap A_2') = n - m\,(A_1) - m\,(A_2) + m\,(A_1 \cap A_2) \tag{6}$$

ZGLED 2.6.3:

Na preizkusu gibalnih sposobnosti je od 100 kandidatov opravilo prvo nalogo 65, drugo pa 41 kandidatov; pri tem je samo 14 kandidatov opravilo obe nalogi. Pogoj za oceno "uspešno" je bila vsaj ena od obeh nalog. Koliko kandidatov ni uspešno opravilo preizkusa?

Naj bo lastnost A_1 "je uspešno opravil prvo nalogo" in A_2 "je uspešno opravil drugo nalogo". Računamo torej število tistih, ki so bili obakrat neuspešni; po obrazcu (6) je

$$m(A_1' \cap A_2') = n - m(A_1) - m(A_2) + m(A_1 \cap A_2) =$$

$$100 - 65 - 41 + 14 = 8$$

Da je bilo neuspešnih kandidatov 8, bi lahko izračunali tudi v dveh korakih, najprej število uspešnih po obrazcu (1), nato pa manjkajoče število dobili z odštevanjem. (Tako smo pravzaprav prišli do obrazca (6)!).

Če v množico vpeljemo več lastnosti hkrati in računamo število elementov, ki imajo vsaj eno od teh lastnosti, se obrazec, ki je soroden (1), hitro podaljšuje, vendar ohranja podobno obliko: seštejemo najprej moči posameznih podmnožic, ki ustrezajo $A_1, \ldots, A_r$, od te vsote odštejemo moči

vseh možnih presekov po dveh podmnožic; prištejemo moči vseh možnih presekov po treh podmnožic,..., v znakih

$$\begin{aligned} m\,(A_1 \cup A_2 \cup \ldots \cup A_r) &= m\,(A_1) + m\,(A_2) + \ldots + m\,(A_r) - \\ &- \{m(A_1 \cap A_2) + m(A_1 \cap A_3) + \ldots + m(A_1 \cap A_r) + \ldots + m(A_{r-1} \cap A_r)\} + \\ &+ \{m(A_1 \cap A_2 \cap A_3) + m(A_1 \cap A_2 \cap A_4) + \ldots + m(A_{r-2} \cap A_{r-1} \cap A_r)\} - \\ &- \ldots + (-1)^{r-1}\, m\,(A_1 \cap A_2 \cap \ldots \cap A_{r-1} \cap A_r) \end{aligned} \tag{7}$$

Če hočemo dokazati, da je enakost (7) izpolnjena pri poljubnem r, moramo pokazati, da je vsak element iz $A_1 \cup A_2 \cup \ldots \cup A_r$ na desni strani štet natanko enkrat. Naj bo x poljuben element iz $A_1 \cup A_2 \cup \ldots \cup A_r$ in naj bo vsebovan v k različnih množicah iz te unije, pri čemer seveda $k \leq r$. (To pomeni, da ima element x natanko k lastnosti izmed $A_1, \ldots, A_r$.) Vsekakor lahko vrstni red množic (lastnosti) preuredimo tako, da je x vsebovan v $A_1, A_2, \ldots, A_k$, ne spada pa v $A_{k+1}, \ldots, A_r$. Potem je na desni strani obrazca (7) element x najprej štet k-krat kot element posamezne množice A_i za $i = 1, 2, \ldots, k$, vsakič s pozitivnim predznakom. Nato je štet z negativnim predznakom tolikokrat, na kolikor načinov lahko izberemo po dve množici izmed k množic, torej $\binom{k}{2}$-krat. Med vsemi preseki, ki jih sestavljajo po tri množice, je tudi $\binom{k}{3}$ takšnih, v katerih leži x – tolikokrat ga štejemo s pozitivnim predznakom. Podoben premislek nam pomaga pregledati tudi ostale člene v obrazcu (7) in pove, da je število pojavljanj proučevanega elementa enako

$$\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \ldots + (-1)^{k-1}\binom{k}{k} \tag{8}$$

Pri tem smo že upoštevali, da lahko namesto k pišemo $\binom{k}{1}$. Naša naloga bo končana, če uvidimo, da je vsota v (8) natanko enaka 1. Pri tem si pomagamo z naslednjim razvojem, ki je posledica binomskega izreka:

$$0 = (1-1)^k = \binom{k}{0} - \binom{k}{1} + \binom{k}{2} - \binom{k}{3} + \ldots + (-1)^k\binom{k}{k}$$

in zato

$$\binom{k}{0} - \left[\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \ldots + (-1)^{k-1}\binom{k}{k}\right] = 0$$

Ker je $\binom{k}{0} = 1$, je enaka 1 tudi vsota v oglatem oklepaju, kakor smo želeli pokazati. Z obrazcem (7) pa lahko uženemo naslednji

ZGLED 2.6.4:

Od 100 učencev se jih 28 ukvarja s košarko, 30 z rokometom in 42 z nogometom, 10 s košarko in z nogometom, 5 z rokometom in z nogometom, 3 učenci pa z vsemi tremi športi hkrati. Koliko učencev se ne ukvarja z nobenim od naštetih športov?

Nalogo bomo rešili tako, da bomo izračunali število tistih, ki se ukvarjajo vsaj z enim, nato pa njihovo število odšteli od 100. Če z A, B in C po vrsti označimo lastnosti "je košarkar", "je rokometaš" in "je nogometaš" (oziroma ustrezne množice), imamo te podatke: $m(A) = 28$, $m(B) = 30$ in $m(C) = 42$; $m(A \cap B) = 8$, $m(A \cap C) = 10$, $m(B \cap C) = 5$ in $m(A \cap B \cap C) = 3$. Število tistih, ki se ukvarjajo vsaj z enim športom, je po obrazcu (7) enako

$$\begin{aligned} m(A \cup B \cup C) &= m(A) + m(B) + m(C) - \\ -m(A \cap B) - m(A \cap C) - m(B \cap C) + m(A \cap B \cap C) &= \\ &= 28 + 30 + 42 - 6 - 10 - 5 + 3 = 80 \end{aligned}$$

Potemtakem je število tistih, ki se ne ukvarjajo z nobenim od teh športov, enako $100 - 80 = 20$.

Za konec še en zgled, če z njim radovednost še ne bo potešena, bo treba pogledati v zahtevnejšo literaturo s področja kombinatorike.

△ ZGLED 2.6.5:

Dana je množica $\mathbb{N}_n = 1, 2, 3, \ldots, n$. Koliko je med vsemi $n!$ permutacijami te množice takšnih, da je vsaj eno od števil $1, 2, 3, \ldots, n$ na tistem mestu, ki mu pripada po naravnem vrstnem redu?

Naj za množico vseh permutacij pomeni A_k lastnost "število k je v permutaciji natanko na k-tem mestu". Izračunati moramo moč unije množic $m(A_1 \cup A_2 \cup \ldots \cup A_n)$. Pri tem bomo potrebovali števila $m(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k})$, torej moči množic, ki vsebujejo takšne permutacije, da so na mestih $i_1, i_2, \ldots, i_k$ števila $i_1, i_2, \ldots, i_k$, na preostalih $n - k$ mestih pa so ostala števila poljubno razporejena. Zato je moč takšne množice enaka $(n-k)!$. Ne smemo pozabiti, da je permutacij, ki imajo natanko k elementov na "pravilnih" mestih, $\binom{n}{k}$ – na toliko načinov lahko izmed n mest izberemo k mest, ki jih zasedemo z enakim številom, kot je zaporedna številka mesta. To potem pomeni, da je vsota števil $m(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k})$ - indeksi i_j pri tem pretečejo vse možne vrednosti – enaka

$$\binom{n}{k}(n-k)! = \frac{n!}{k!\,(n-k)!} \cdot (n-k)! = \frac{n!}{k!}$$

Po obrazcu (7) je zato iskano število permutacij z vsaj enim elementom na pravem mestu enako

$$\frac{n!}{1!} - \frac{n!}{2!} + \frac{n!}{3!} - \ldots + (-1)^{n-1}\frac{n!}{n!} =$$

$$= n!\left(1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + \frac{(-1)^{n-1}}{n!}\right)$$

Za $n = 4$ dobimo tako na primer

$$\frac{4!}{1!} - \frac{4!}{2!} + \frac{4!}{3!} - \frac{4!}{4!} = 24 - 12 + 4 - 1 = 15$$

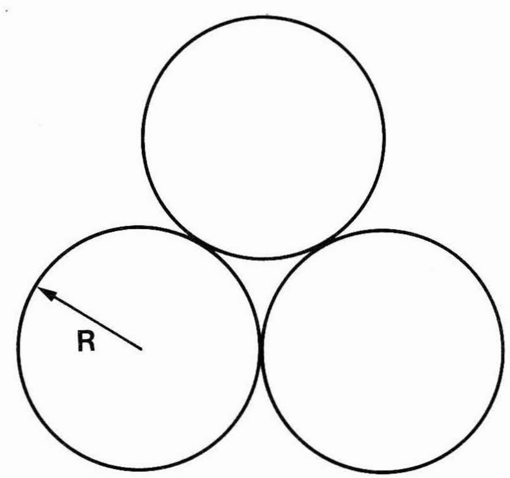
V A J E

1. V mestu so trije kinematografi, A, B in C. 20% prebivalcev obiskuje kinematograf A, 16% B, 14% C, 8% A in B, 5% A in C, 4% B in C ter 2% vse tri kinematografe. Kolikšen odstotek prebivalcev tega mesta zahaja v kino?

2. Koliko naravnih števil, manjših od 1 000, ni deljivih niti z 2 niti s 3 niti s 5?

3. △ Šahovsko ploščo prepotujemo od spodnjega k zgornjemu robu tako, da v vsaki vrstici stopimo na eno polje, pri tem pa lahko tudi posamezen stolpec uporabimo samo enkrat. Koliko je vseh možnih poti? Koliko je med njimi takšnih, pri katerih ne stopimo na nobenega od diagonalnih elementov (diagonala: levo spodaj - desno zgoraj)?

4. △ Trije krogi z enakimi polmeri (denimo R enot) se dotikajo, kot kaže spodnja skica. Kolikšna je ploščina lika med krogi?



Slika 2.5: Geometrijska uporaba načela vključitev in izključitev

PREGLED POMEMBNEJŠIH OBRAZCEV

Permutacije brez ponavljanja

$$P_n = n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n$$

Permutacije s ponavljanjem

$$P_n^{k_1,k_2,\ldots,k_r} = \frac{n!}{k_1! \cdot k_2! \cdot \ldots \cdot k_r!}$$

Variacije brez ponavljanja

$$V_n^r = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-r+1) = \frac{n!}{(n-r)!}$$

Variacije s ponavljanjem

$$^{(p)}V_n^r = \underbrace{n \cdot n \cdot \ldots \cdot n}_{r} = n^r$$

Kombinacije brez ponavljanja

$$C_n^r = \frac{V_n^r}{r!} = \frac{n!}{r!\,(n-r)!} = \binom{n}{r}$$

Lastnosti binomskih simbolov

$$\binom{n}{r} = \binom{n}{n-r} \qquad C_n^r = C_n^{n-r}$$

$$\binom{n}{r} + \binom{n}{r+1} = \binom{n+1}{r+1} \qquad C_n^r + C_n^{r+1} = C_{n+1}^{r+1}$$

$$\binom{n}{0} = 1 \qquad \binom{n}{n} = 1 \qquad \binom{n}{1} = n$$

Kombinacije s ponavljanjem

$$^{(p)}C_n^r = C_{n+r-1}^r = \binom{n+r-1}{r}$$

Vezane kombinacije

$$C_{n_1,n_2,\ldots,n_m}^{r_1,r_2,\ldots,r_m} = \binom{n_1}{r_1} \cdot \binom{n_2}{r_2} \cdot \ldots \cdot \binom{n_m}{r_m}$$

Binomski izrek

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^{n-r}\, b^r$$

3. OSNOVE OPISNE STATISTIKE

V tem poglavju se seznanimo s statistiko in z osnovnimi statističnimi pojmi, kot so populacija, enota, vzorec, statistična spremenljivka. Naučimo se urejati podatke in jih grupirati v frekvenčne razrede. Iz izhodiščne frekvenčne porazdelitve izračunamo relativne frekvence in različne kumulative, za lažjo predstavljivost sežemo po grafičnih orodjih. Spoznamo, kako z nekaj parametri (različnimi srednjimi vrednostmi, podatki o variabilnosti itd.) vsaj grobo opišemo porazdelitev. Za likof na hitro pregledamo še indeksna števila, ta strah in trepet slovenskih novinarjev...

3.1 Osnovni statistični pojmi

Ker stopamo v nove prostore, najprej definicija:

Statistika je veda, ki proučuje množične pojave.

Z zbiranjem, urejanjem, grupiranjem, povzemanjem, prikazovanjem in analiziranjem številskih podatkov o teh pojavih skuša odkriti njihove splošne zakonitosti in nato pridobljena spoznanja izkoristiti za oblikovanje ustreznih napovedi oziroma odločitev.

Resnici na ljubo je treba dodati, da ima izraz "statistika" še kopico drugih pomenov, tako z njim na primer označujemo
a) same (običajno sistematično zbrane) številčne podatke;
b) delo, ki nas pripelje do takih zbirk;
c) organe (službe, institucije...), ki to delo opravljajo
in še kaj. Po tej plati nas statistika v tem učbeniku ne bo zanimala.

Poglejmo si takle primer. Smučanje je v Sloveniji množičen pojav. Prodajalci smučarske opreme zbirajo in proučujejo podatke o smučanju in o tistih pojavih, ki so s smučanjem povezani, da bi odkrili najpomembnejše zakonitosti (npr. delež Slovencev, ki se ukvarja s smučanjem; zvezo med starostjo in nagnjenostjo k temu športu; odvisnost med osebnimi dohodki in sredstvi, ki so jih ljudje pripravljeni (sposobni) porabiti za smučanje, itd.). Te odnose morajo poznati, če hočejo pravilno načrtovati obseg poslovanja.

Statistika se je včasih ukvarjala skoraj izključno z gospodarstvom, naseljenostjo,... in drugimi "državnimi zadevami"; odtod tudi njeno ime, ki prihaja iz latinske besede "status", kar pomeni "stanje", pa tudi "država". Kasneje je svoje delovno področje razširila na najrazličnejše znanosti, ki imajo opravka z množičnimi pojavi, npr. agronomijo, biologijo, ekonomijo, elektroniko, fiziko, kemijo, medicino, psihologijo, sociologijo, itd. Zato lahko mirno rečemo, da dandanašnji ni področja, na katerem bi bili lahko uspešni, ne da bi obvladali vsaj osnovne statistične tehnike.

Skupna značilnost večine statističnih proučevanj je v tem, da

- pojavov ne moremo (ker so preobsežni) ali

- ne želimo (ker bi bilo predrago) v celoti zajeti,

v splošnem smo zato prisiljeni sklepati na osnovi nepopolnih informacij. Zato tudi o dobljenih ugotovitvah ne moremo trditi, da so zagotovo pravilne, ampak je njihova pravilnost samo bolj ali manj **verjetna**.

Za pravilno razumevanje dejanske vrednosti oziroma zanesljivosti rezultatov statističnih proučevanj moramo obvladati vsaj **osnove verjetnostnega računa**. S statistiko, statističnimi podatki in rezultati statističnih raziskovanj se srečujemo vsak dan. Žal so enako pogoste tudi napačne razlage teh rezultatov in zmote o moči statistike nasploh, zato imamo od statistike veliko manjšo korist, kot bi jo sicer lahko imeli.

Poudarimo še enkrat: statistiko zanimajo množični pojavi, torej takšni, ki se v času in prostoru ne pojavljajo posamično, ampak v velikem številu. Tako opazovanje posameznikovega odnosa do smučanja prodajalcem tovrstne opreme kaj malo koristi; če opazujemo nakupe posameznikov, se zdi njihovo pojavljanje povsem neurejeno, določene zakonitosti, o katerih smo govorili (npr. odvisnost med splošno življenjsko ravnjo in izdatki za šport in rekreacijo), se pokažejo šele pri proučevanju nakupov večje množice potrošnikov kot celote, npr. kar vseh Slovencev.

Končno ali neskončno množico, ki jo statistično proučujemo, imenujemo **populacija** ali **statistična množica**. Lahko jo sestavljajo živa bitja (odtod ime, "populus" pomeni v latinščini ljudstvo), lahko pa tudi predmeti, dogodki ali sploh karkoli drugega. Populacijo opišemo tako, da navedemo t.i. **opredeljujoče pogoje**, ki jim morajo elementi zadoščati, da spadajo v populacijo. Posamezen element populacije imenujemo (statistična) **enota**. Tistim značilnostim populacije, ki nas v konkretnem proučevanju zanimajo, so nekoč rekli (statistični) **znaki**, danes jih imenujemo **statistične spremenljivke**. Značilnostim populacije kot celote rečemo (statistični) **parametri**.

V gornjem primeru lahko populacijo opredelimo na primer kot množico vseh prebivalcev Slovenije; vsak od njih predstavlja za to proučevanje statistično enoto, spremenljivka recimo delež dohodka, ki ga ta Slovenec namenja za šport in rekreacijo, parameter pa povprečen delež dohodka, ki ga v tem primeru daje za šport in rekreacijo celotna proučevana populacija.

Spremenljivka je pravzaprav tista značilnost statističnih enot, ki je predmet proučevanja. Za posamezne enote iz populacije ima različne vrednosti, je spremenljiva ali variabilna, tako je dobila ime. Vsebinsko lahko spremenljivke delimo na krajevne (kraj rojstva, bivališče, sedež podjetja...), časovne (npr. datum rojstva) in stvarne. Pri slednjih razlikujemo atributivne spremenljivke, za katere vrednosti izražamo opisno, samo z besedami (spol, poklic,...), od številskih ali numeričnih, katerih vrednosti izražamo s števili (velikost, teža, število družinskih članov, ocena, število točk pri testu). V tem učbeniku se bomo omejili predvsem na numerične spremenljivke.

Proučevana populacija naj bo neka množica $\mathcal{G}$, e pa njen poljubni element (statistična enota). Če elementom priredimo vrednosti $X(e)$ izbrane numerične spremenljivke X, smo s predpisom $e \mapsto X(e)$ definirali preslikavo množice $\mathcal{G}$ v množico realnih števil, ali drugače, vsako številsko spremenljivko lahko obravnavamo kot realno funkcijo, definirano na izbrani populaciji. Če lahko ta funkcija zavzame katerokoli vrednost z nekega intervala, gre za zvezno spremenljivko, medtem ko lahko nezvezne zavzamejo samo nekatere (najpogosteje celoštevilske) vrednosti. Razlika med zveznimi in nezveznimi spremenljivkami je v praksi včasih nekoliko zabrisana, ker imamo zaradi zaokroževanja tudi pri zveznih spremenljivkah v bistvu samo končno mnogo različnih vrednosti, vendar vsebinska razlika vedno ostane: pri zveznih spremenljivkam (starost, teža, višina) lahko ta zavzame vrednosti, ki ležijo med "okroglimi" (ko imamo na primer v tabeli razvrščene prebivalce po starosti, zaokroženi na eno leto natančno, je jasno, da so možne tudi vse vmesne vrednosti), medtem ko je pri nezveznih spremenljivkah med zaporednima vrednostima "bela lisa" (družina npr. ne more imeti 1·5 otroka). Vrednosti nezveznih spremenljivk v praksi najpogosteje ugotavljamo s štetjem, vrednosti zveznih pa z merjenjem. (Bralcu, ki mu je razlika med zveznimi in diskretnimi množicami dovolj jasna, se opravičujemo za morda dolgovezen komentar, ki pa se nam je zdel potreben zaradi kasnejšega lažjega dela.)

Množične pojave analiziramo tako, da proučujemo vrednosti primerno izbranih statističnih spremenljivk na elementih ustrezno opredeljene populacije. Če je populacija neskončna, seveda ni mogoče pregledati vrednosti izbranih spremenljivk na vseh elementih, pa tudi pri končnih populacijah bi to iz različnih razlogov (običajno ekonomskih) ne bilo vedno smiselno.

Poglejmo primer iz kmetijstva! Kvaliteta kokoši nesnic je določena predvsem s številom jajc, ki jih znesejo v letu dni. Kmetovalec, ki vzgaja takšne kokoši za prodajo, seveda ne bo na ta način kontroliral vsake izmed njih, ker bi moral potem tako "iztrošene" kokoši prodajati po bistveno nižji ceni kot mlade nesnice.

Podobno lahko premislimo, zakaj tovarna žarnic, katerih kvaliteto presojamo predvsem po življenjski dobi, ne preizkusi vsake žarnice, preden gre v prodajo. (Najbrž ni treba posebej omenjati, da tudi v tovarni vžigalic ne preizkusijo vsake vžigalice, ali se prižge.)

Zato v praksi običajno proučujemo ustrezno statistično spremenljivko samo na primerno izbrani končni podmnožici celotne populacije. To "odlikovano" podmnožico imenujemo **vzorec**.

Rejec kokoši iz prejšnjega primera od vsake generacije obdrži samo nekaj izbranih kokoši, da lahko iz podatkov o njihovi nesnosti sklepa na kvaliteto celotne generacije in preverja, ali se je od prejšnje generacije kvaliteta bistveno spremenila. V tovarni od cele serije žarnic preizkusijo samo nekatere, recimo vsako stoto.

Ugotovljene vrednosti spremenljivke X lahko bolje ali slabše predstavljajo vrednosti te spremenljivke na celotni populaciji. Osnovno vprašanje, na katero mora odgovoriti statistika, je zato naslednje:

Kaj je mogoče reči o statistični spremenljivki X za celo populacijo na osnovi podatkov, ki smo jih o njej zbrali na vzorcu?

Odgovor na to vprašanje je v pretežni meri odvisen od načina, kako iz cele populacije izberemo vzorec. Najlažji je takrat, kadar imajo vsi elementi enako možnost, da bodo izbrani. Tako dobljen vzorec bomo imenovali **slučajni vzorec**.

Preden lahko karkoli rečemo o sklepanju iz podatkov o izbranem vzorcu, se moramo najprej naučiti, kako obdelamo podatke o vrednostih statistične spremenljivke na enotah iz vzorca, ne glede na to, kako je ta sestavljen. S to nalogo se bomo ukvarjali v naslednjih dveh razdelkih, k odnosu med vzorcem in celotno populacijo pa se bomo vrnili ob koncu učbenika.

V A J E

1. Tabela (Veliki družinski atlas, DZS, Ljubljana 1992, str. 260)

Št.	**Vrh**	**Skupina**	**Pogorje**	**m**
1.	Triglav	Triglav	Julijske Alpe	2 864
2.	Škrlatica	Škrlatica	Julijske Alpe	2 740
3.	Veliki Mangart	Mangart - Ponce	Julijske Alpe	2 679
4.	Jalovec	Jalovec-B. Grint.	Julijske Alpe	2 645
5.	Visoki Rokav	Škrlatica	Julijske Alpe	2 644
6.	Oltar	Škrlatica	Julijske Alpe	2 629
...	...	...	...	...
13.	Grintovec	Grintovec	Kamniške Alpe	2 558
14.	Prisojnik	Razor - Prisojnik	Julijske Alpe	2 547
...	...	...	...	...

prikazuje del podatkov o najvišjih vrhovih v Republiki Sloveniji.

a) Kaj je v tem primeru statistična populacija?
b) Naštej nekaj statističnih enot!
c) Katere statistične spremenljivke so predstavljene v tabeli?
č) Kaj posamezna spremenljivka pomeni po vsebini?
d) Katere vrednosti lahko zavzame numerična spremenljivka?

2. Tabela (Vir: Zavod za statistiko RS, Statistične informacije, Popis 1991)

Zap. št.	**Občina**	**Povr.** (km^2)	**Prebiv. (1991)**	**% kmeč. prebiv.**
1.	Ajdovščina	352·41	22 632	9·7
2.	Brežice	268·45	24 724	15·6
3.	Celje	229·74	64 736	3·2
4.	Cerknica	482·45	15 020	9·1
5.	Črnomelj	486·41	18 374	11·5
6.	Domžale	239·95	44 185	4·1
...	...	...	...	...

pomeni začetek pregleda površin, števila prebivalcev in deležev kmečkega prebivalstva po slovenskih občinah.

a) Kaj je v tem primeru populacija?
b) S katerimi pogoji je opredeljena?
c) Katere statistične spremenljivke so predstavljene v tabeli?
č) Za posamezne spremenljivke opredeli, ali so zvezne ali nezvezne!

3. Kakšna spremenljivka je "zakonsko stanje"? Katere vrednosti lahko zavzame? (Ljudski opis, da je nekdo "na pol oženjen", zanemarimo.)

4. Izvedeti želimo, kaj menijo državljani Slovenije o svojih politikih. V ta namen oblikujemo vzorec tako, da neko jutro na slepo izberemo iz telefonskega imenika 1 000 številk in pokličemo njihove lastnike ter jim zastavimo (denimo) ustrezna vprašanja. Kaj meniš na splošno o takem vzorcu? Kaj bi mu lahko očital?

5. Izberi primer iz lastne šolske prakse, na katerem boš lahko pojasnil osnovne statistične pojme: populacija, enota, spremenljivka, vzorec,... in njihove posamezne značilnosti!

3.2 Urejanje in prikazovanje podatkov

Dolgi nizi vrednosti, ki jih proučevana statistična spremenljivka zavzame na posameznih enotah (cele populacije ali vzorca), so v večini praktičnih primerov nepregledni, zato te vrednosti običajno grupiramo.

Denimo, da smo pri metanju običajne igralne kocke v 80 metih zabeležili po vrsti naslednje izide:

5, 2, 1, 6, 2, 3, 1, 2, 5, 4, 6, 3, 5, 6, 1, 2, 4, 4, 3, 6, 4, 3, 2, 4, 3, 3, 5, 1, 3, 4, 6, 1, 2, 6, 3, 1, 2, 4, 5, 3, 5, 6, 6, 3, 1, 4, 4, 5, 6, 2, 4, 6, 3, 2, 5, 3, 2, 1, 6, 5, 4, 1, 3, 2, 4, 6, 3, 2, 6, 1, 4, 5, 5, 6, 1, 3, 2, 5, 1, 4.

Izid pri metu igralne kocke je število pik, ki se pokaže na gornji ploskvi kocke, ko se ta po metu umiri.

k	f_k
1	12
2	13
3	15
4	14
5	12
6	14
N	80

Namesto s tem nepreglednim nizom izide lepše prikažemo, če preštejemo, kolikokrat je nastopila posamezna od šestih možnih vrednosti.

S tem je postal zapis mnogo bolj pregleden, naše poznavanje populacije pa ni bilo okrnjeno. Populacija je tu množica poskusov, enota posamezen poskus, spremenljivka je število pik; $X(1)$ je na primer število pik pri prvem metu, $X(2)$ število pik pri drugem itd. Simbol N - od *"numerus"* - bomo dosledno uporabljali za označevanje števila enot v populaciji.

Tabela 3.1: Porazdelitev izidov pri 80 metih igralne kocke

Tako smo spoznali najpreprostejši način, ki nas pripelje do **frekvenčne porazdelitve** ali **frekvenčne distribucije**. Spremenljivka, ki smo jo proučevali, je bila nezvezna in je lahko zavzela le šest različnih vrednosti, zato je bilo treba samo prešteti, kolikokrat je dejansko zavzela vsako od teh vrednosti. Drugače rečeno, ugotoviti je bilo treba **(absolutne) frekvence** teh vrednosti.

Iz osnovne frekvenčne distribucije lahko izpeljemo nekatere druge prikaze. O tem, kako so vrednosti proučevane spremenljivke porazdeljene po posameznih enotah, nam dosti povedo **relativne frekvence**.

Porazdelitev relativnih frekvenc f_k^o dobimo tako, da frekvence f_k posameznih vrednosti delimo s številom vseh enot v populaciji (N),

$$f_k^o = \frac{f_k}{N} \tag{1}$$

Relativna frekvenca *posamezne možne vrednosti statistične spremenljivke pove delež tistih enot iz populacije, na katerih je spremenljivka zavzela to vrednost.*

V praksi relativno frekvenco dostikrat izrazimo v odstotkih.

Iz frekvenc oziroma iz relativnih frekvenc lahko izračunamo **kumulative**. Kumulativa absolutnih frekvenc nam pri izbrani vrednosti spremenljivke pove, na koliko enotah populacije je ta zavzela manjšo vrednost; pri kumulativi relativnih frekvenc pa s tem podatkom povemo delež enot v populaciji z manjšo vrednostjo spremenljivke. Za najmanjšo vrednost spremenljivke sta tako obe kumulativi enaki 0,

$$F_1 = 0\,, \qquad F_1^o = 0 \tag{2}$$

za naslednje pa dobimo kumulativi (kot pove že ime) s postopnim prištevanjem frekvenc oziroma relativnih frekvenc:

$$F_k = F_{k-1} + f_{k-1}\,, \quad F_k^o = F_{k-1}^o + f_{k-1}^o \quad (k = 2, 3, ..., r) \tag{3}$$

Pri tem r označuje število različnih vrednosti, ki jih spremenljivka lahko zavzame. Naslednja tabela vsebuje vse štiri prikaze za prej opisani poskus s kocko. Kadar je možnih vrednosti preveč – kar se, če ne prej, zagotovo zgodi

Izid k	Frekvenca f_k	Relativna frekvenca f_k^o	Kumulativa frekvenc F_k	Kumulativa rel. frek. F_k^o
1	12	0·1500	0	0·0000
2	13	0·1625	12	0·1500
3	15	0·1875	25	0·3125
4	14	0·1750	40	0·5000
5	12	0·1500	54	0·6750
6	14	0·1750	66	0·8250
SKUPAJ	80	1·0000	80	1·0000

Tabela 3.2: Tabela frekvenc, relativnih frekvenc in obeh kumulativ

pri zveznih spremenljivkah – te vrednosti razdelimo v **razrede**. Razredi so določeni s primerno delitvijo intervala med največjo in najmanjšo možno vrednostjo dane spremenljivke na določeno število delov, v praksi običajno 5 do 20.

Z grupiranjem podatkov v razrede izgubljamo del informacije o individualnih vrednostih spremenljivke na posameznih enotah, zato je nevarno, če naredimo premajhno število razredov, preveliko število frekvenčnih razredov pa ne izboljša preglednosti. Optimalno število razredov je odvisno od pojava, ki ga preučujemo, in od števila enot v populaciji oziroma vzorcu: za izbiro pravilnega števila razredov je treba poznati vsebino problema in znati nekaj (več) statistike.

Frekvenca posameznega razreda je enaka številu enot proučevane populacije, za katere vrednost številske spremenljivke spada v tisti razred. Sestavljanje frekvenčne porazdelitve si oglejmo na naslednjem praktičnem primeru.

V medicinski raziskavi so ugotavljali prehranjenost učencev neke šole. Pri tehtanju učencev so bile ugotovljene naslednje mase (v kilogramih), ki so že urejene po velikosti:

59·7	60·0	60·6	61·8	62·3	62·9	63·0	63·0	63·1	63·2
63·5	63·5	63·6	63·8	64·0	64·0	64·3	64·4	64·5	64·5
64·8	65·0	65·3	65·5	65·6	65·6	65·7	65·7	65·8	65·9
66·0	66·0	66·0	66·2	66·4	66·5	66·5	66·5	66·7	66·8
66·9	67·0	67·0	67·0	67·0	67·1	67·2	67·2	67·4	67·5
67·5	67·6	67·6	67·7	67·7	67·7	67·8	67·8	67·9	68·0
68·0	68·0	68·2	68·2	68·4	68·6	68·6	68·7	68·8	68·9
68·9	69·0	69·0	69·1	69·1	69·2	69·2	69·2	69·3	69·4
69·5	69·5	69·5	69·5	69·5	69·7	69·9	70·0	70·0	70·1
70·3	71·4	71·5	71·8	72·2	73·0	73·8	74·0	74·0	74·4

Tabela 3.3: Negrupirani podatki o masah učencev

Razred	Masa (kg)	f_k
1	59·5 – 62·5	5
2	62·5 – 65·5	18
3	65·5 – 68·5	42
4	68·5 – 71·5	27
5	71·5 – 74·5	8
	SKUPAJ	100

Interval med najmanjšo maso 59·7 kg in največjo maso 74·4 kg ima dolžino 14·7 (kg). Zaradi enostavnosti smo vzeli nekoliko večji interval od 59·5 kg do 74·5 kg in ga razdelili na pet enako širokih razredov po 3 kg.

Tabela 3.4: Podatki o masah učencev, grupirani po razredih

Pri mejnih točkah, ki ločujejo razrede, se – običajno – odločimo tako, da je interval na zgornji meji odprt; prvi razred bo zato opisan "od (vključno) 59·5 kg do (pod) 62·5 kg", drugi "od (vključno) 62·5 kg do (pod) 65·5 kg" in tako dalje.

V prvem stolpcu smo razrede oštevilčili po vrstnem redu. Ker smo imeli podatke urejene po velikosti, s sestavljanjem frekvenčne porazdelitve ni bilo težav. Če zaporedje podatkov ni urejeno po velikosti, si lahko pri razvrščanju v razrede pomagamo s črtkanjem.

Kako praktično izvedemo črtkanje? Na listu si za vsak razred pripravimo primeren prostor, v katerega narišemo črtico, brž ko v zaporedju vrednosti naletimo na takšno,

ki spada v ta razred. Ko pregledamo vse vrednosti, preštejemo črtice po posameznih razredih in zapišemo frekvence v tabelo.

Bistveno za grupiranje je, da mora vsaka vrednost, ki jo zavzame spremenljivka, pasti v natanko določen frekvenčni razred - nobena ne ostane brez svojega razreda in nobena vrednost ni hkrati v dveh različnih razredih. Zato je treba posebej paziti pri mejah med razredi, ali spadajo v spodnji ali v zgornji razred. Pri nas smo meje 62·5 kg, 65·5 kg, 68·5 kg in 71·5 kg uvrstili v zgornje razrede.

Širina razreda je enaka razliki med zgornjo in spodnjo mejo razreda. Če z_k pomeni zgornjo in je s_k spodnja meja k-tega frekvenčnega razreda, je njegova širina enaka

$$i_k = z_k - s_k$$

Pri ugotavljanju širine je treba paziti: v prvem frekvenčnem razredu so npr. vse vrednosti od 59·5 kg do pod 62·5 kg, zato je širina tega razreda $i_1 = 62{\cdot}5 - 59{\cdot}5 = 3$ (kg). Razredi v naši porazdelitvi so vsi enako široki, kar pa ni nujno, čeprav je za nadaljnje delo najbolj ugodno.

Drug podatek o frekvenčnem razredu je njegova **sredina** y_k, ki jo izračunamo na običajen način kot sredino ustreznega intervala,

$$y_k = \frac{z_k + s_k}{2}$$

Za prvi razred iz našega primera je sredina $(59{\cdot}5 + 62{\cdot}5)/2 = 61{\cdot}0$ (kg).

> Bralec je najbrž opazil, da smo z združitvijo vrednosti v razrede posredno tudi enote iz proučevane množice (celotne populacije ali nekega vzorca) porazdelili v paroma tuje razrede. Logika združevanja številskih podatkov v razrede ima v tej luči naslednji "podaljšek":
>
> Od trenutka, ko smo podatke grupirali v razrede in s tem oblikovali frekvenčno porazdelitev, se pretvarjamo, da ima spremenljivka na vseh enotah, ki sodijo v neki razred, vrednost, ki je enaka sredini razreda, individualne vrednosti so postale nepomembne, če že ne povsem pozabljene.

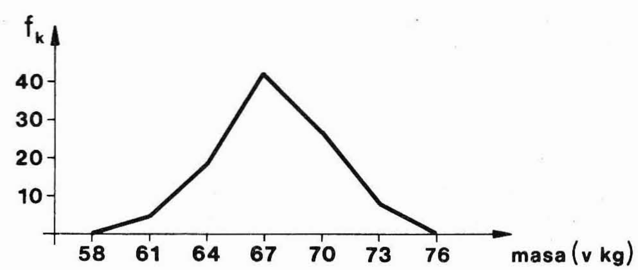
Po tem dodatnem koraku, ko nekako "skoncentriramo" vse vrednosti iz določenega razreda v njegovo sredino (fiziki bi morda govorili o težišču), lahko rečemo, da se nadaljevanje zgodbe ne razlikuje od tistega, kar že poznamo. Na popolnoma enak način (po obrazcih (1), (2) in (3)) izračunamo relativne frekvence in obe kumulativi. Rezultat računanja je naslednja tabela:

Razred	Masa (v kg)	Sredina	f_k	F_k	f_k^o	F_k^o
1	59·5 – 62·5	61·0	5	0	0·05	0·00
2	62·5 – 65·5	64·0	18	5	0·18	0·05
3	65·5 – 68·5	67·0	42	23	0·42	0·23
4	68·5 – 71·5	70·0	27	65	0·27	0·65
5	71·5 – 74·5	73·0	8	92	0·08	0·92
	SKUPAJ		100	100	1·00	1·00

Tabela 3.5: Frekvenčna porazdelitev mas učencev

Frekvenčne porazdelitve in izpeljanke zaradi nazornosti radi prikazujemo tudi grafično. V ta namen uporabljamo predvsem **frekvenčni poligon** in **histogram**.

Pri risanju frekvenčnega poligona na abscisni osi označimo frekvenčne razrede, na ordinatni osi pa frekvence (ali relativne frekvence, če želimo prikazati porazdelitev le-teh). V tako dobljen koordinatni sistem narišemo točke, ki imajo za abscise sredine posameznih razredov, za ordinate pa pripadajoče frekvence. Običajno gremo še korak v levo od prvega oziroma korak v desno od zadnjega frekvenčnega razreda; tema sredinama dodanih razredov pripišemo ordinati 0. Če tako dobljene točke po vrsti povežemo z lomljeno črto, dobimo frekvenčni poligon.

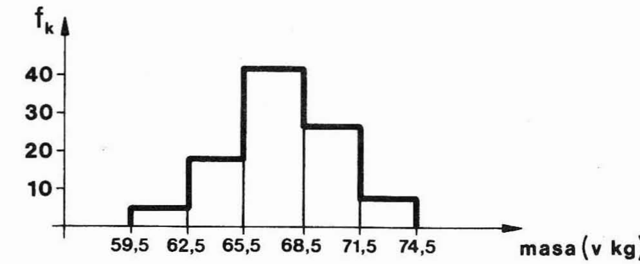


Slika 3.1: Frekvenčni poligon za porazdelitev mas

Sredine razredov so po vrsti enake 61, 64, 67, 70 in 73 (kg), pripadajoče frekvence pa 5, 18, 42, 27 in 8; tem parom smo po dogovoru dodali še točki s koordinatama (58,0) in (76,0); saj je $61 - 3 = 58$, $73 + 3 = 76$. Dobljene točke smo vrisali v koordinatni sistem in jih povezali v frekvenčni poligon.

Histogram sestavljajo pravokotniki – stolpci, katerih osnovnica je določena z daljico na abscisni osi med spodnjo in zgornjo mejo frekvenčnega razreda, ploščina pravokotnika pa je premo sorazmerna frekvenci tega razreda. To v primeru enako širokih razredov seveda pomeni, da so tudi višine

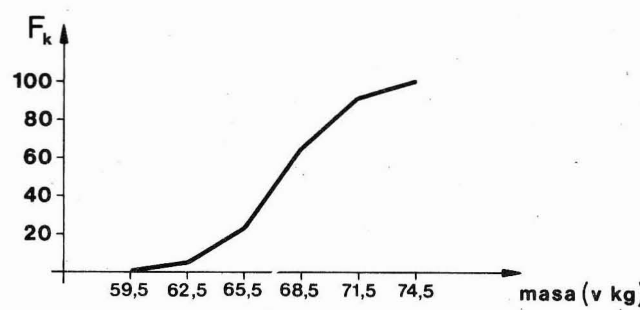
stolpcev premo sorazmerne frekvencam. Če razredi nimajo enake širine, je treba višine stolpcev ustrezno prilagoditi.



Slika 3.2: Histogram za porazdelitev mas

Pri risanju histogramov včasih ne izvlečemo celih navpičnic, ampak narišemo samo obrise ustreznega lika (na sliki 3.2 označeno z debelejšo črto). Dostikrat narišemo histogram in frekvenčni poligon v istem koordinatnem sistemu. (Naredi tako za vajo – združi sliki 3.1 in 3.2!) Točke, ki določajo poligon, so v tem primeru razpolovišča zgornjih stranic pravokotnikov, ki sestavljajo histogram.

V splošnem je poligon primernejši za nezvezne, histogram pa za zvezne spremenljivke. Poligon se loči od histograma še po tem, da njegova oblika nakazuje takšno porazdelitev vrednosti znotraj posameznega frekvenčnega razreda, ki je odvisna tudi od frekvenc obeh sosednjih razredov in ni enakomerna, kot je to pri histogramu.

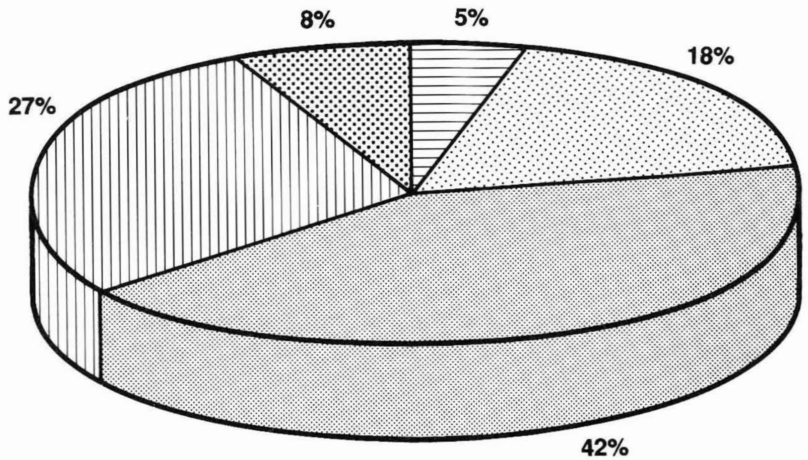


Slika 3.3: Kumulativna frekvenčna porazdelitev mas

Podobno kot frekvenčni poligon lahko narišemo tudi lomljeno črto, ki prikazuje kumulativno frekvenčno porazdelitev. Abscise točk so v tem primeru spodnje meje frekvenčnih razredov, ordinate pa pripadajoče vrednosti F_k. Dobljene točke po vrsti povežemo z lomljeno črto, ki ima v praktičnih primerih najpogosteje obliko stilizirane, razpotegnjene črke S.

Podatke za risanje najdemo v tabeli na str. 56. Pazimo na spodnje meje razredov; zadnja točka je zgornja meja zadnjega razreda, oziroma (74·5,100).

Strukturo po posameznih razredih frekvenčne porazdelitve lepo pokaže tudi t.i. **frekvenčni kolač** oziroma **strukturni krog**, kjer deleže enot, ki sodijo v posamezne razrede, prikazujemo s krogovimi izseki, katerih središčni koti so sorazmerni z relativnimi frekvencami teh razredov. Za porazdelitev mas, ki jo ves čas uporabljamo kot zgled, je frekvenčni kolač prikazan na naslednji sliki.



Slika 3.4: Frekvenčni kolač za porazdelitev mas

Mimogrede še tole. Uporaba strukturnega kroga ni vezana na frekvenčne distribucije (še več, pri teh imata prej obravnavani orodji prednost), princip, ki ga uporabljamo pri prikazovanju, pa je splošen: kolikor večji je določen delež, toliko večji "kos torte" (od tod angleški izraz *pie chart*) mu pripada...

Preden končamo ta razdelek, se vendarle spomnimo, da živimo v računalniških časih. Tudi sicer preproste aritmetične operacije, ki smo jih potrebovali pri dosedanjih izračunih, postanejo dolgočasne, če jih je treba nekaj tisočkrat ponoviti. Zato vsi, ki se resneje ukvarjajo s statistiko, z veseljem posegajo po pripravljenih računalniških orodjih za statistično obdelavo podatkov, najbolj vneti pa uživajo v tem, da jih tudi sami izdelujejo. Če zanemarimo računanje na prste, možgane in kar je drugih takih organov, imamo nekaj razvojnih stopenj v uporabi strojnih in programskih pripomočkov:

a) žepni računalniki splošne vrste;

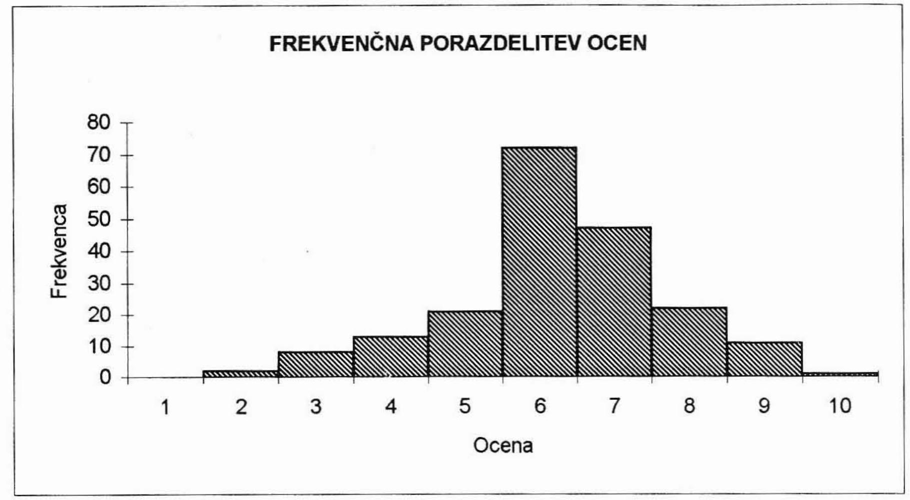
b) specializirani žepni računalniki z dodatnimi vgrajenimi statističnimi funkcijami in (ali) elektronskimi dodatki (na primer razširitvenimi karticami);

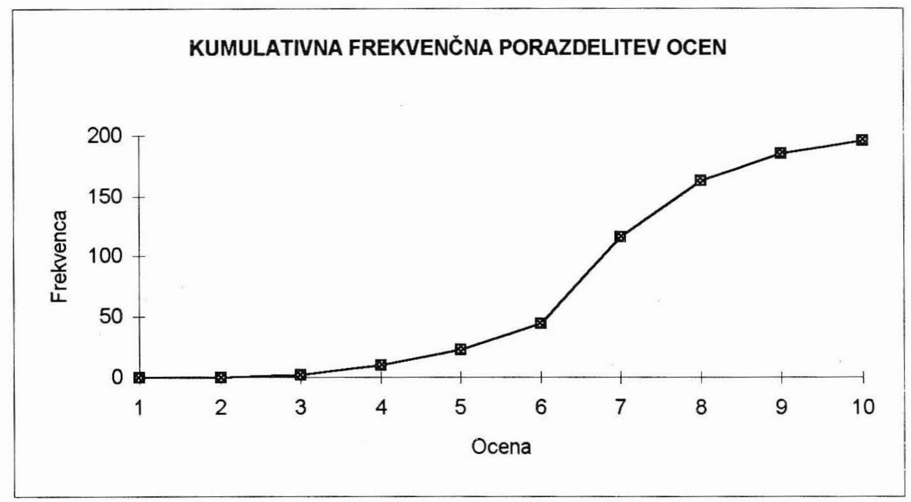
c) večnamenska programska oprema na osebnih in drugih računalnikih, ki - med drugim - omogoča tudi uporabo statističnih funkcij; tipičen primer so elektronske preglednice (Excel, Lotus, Quattro Pro,...);

č) specialna programska oprema za statistično obdelavo podatkov (od programov, ki jih uporabljamo na osebnih računalnikih, sta pri nas najbolj znana SPSS in Statgraphics);

d) za specialne statistične raziskave posebej izdelana programska oprema.

Bralec zagotovo pozna vsaj prvo in tretjo stopnjo, zato priporočamo, da pri vseh nalogah, ki jih bomo še srečali, razmišlja tudi o tem, kako te pripomočke kar najbolj izkoristiti. Morda ga bo presenetilo, ko bo ugotovil, da ima na žepnem računalniku tudi neke tipke, ki jih še nikoli ni uporabil... Za vzorec in spodbudo pa še z enim od programov za delo s preglednicami narejena prikaza neke porazdelitve (visokošolskih) ocen iz matematike.





Slika 3.5: Frekvenčna porazdelitev izpitnih ocen (računalniški izpis)

Ob tem pa moramo opozoriti, da nam noben program ne more zares pomagati pri delu, če ne obvladamo statističnih metod, pri katerih si želimo pomagati z računalnikom. Še bolj pa je res, da nam statistične metode ne morejo pomagati, če ne razumemo problema, s katerim se ukvarjamo, a to je že druga zgodba...

V A J E

1. Od 47 učencev je 7 učencev doseglo oceno odlično, 16 oceno prav dobro, 13 dobro, 5 zadostno in 6 nezadostno. Prikaži frekvenčno porazdelitev ocen (izraženih numerično na običajni način kot 5, 4, 3, 2 oziroma 1) s tabelo, histogramom in frekvenčnim poligonom.

2. V besedilu te naloge ugotovi frekvence samoglasnikov a, e, i, o, u. Sestavi frekvenčno tabelo in izračunaj pripadajoče relativne frekvence.

3. Policija je ugotavljala povprečno maso policistov na vzorcu, ki je zajel 80 policistov. Pri tem so dobili naslednje vrednosti (v kilogamih):

68	73	61	66	96	79	65	86	84	79
65	78	78	62	80	67	75	88	75	82
89	67	73	73	82	73	87	65	61	78
57	81	68	60	74	94	75	85	88	72
90	93	62	77	95	76	78	63	62	71
95	69	60	65	62	76	88	59	78	74
79	71	76	75	76	75	63	68	83	97
53	85	93	75	72	60	71	75	74	77

Sestavi iz teh podatkov frekvenčno porazdelitev z razredi (v kilogramih) 50 – 54, 55 – 59, 60 – 64, ..., 95 – 99. Izračunaj porazdelitev relativnih frekvenc, kumulativno frekvenčno porazdelitev in kumulativno porazdelitev relativnih frekvenc. Nariši vse možne grafične prikaze. Kaj bi lahko očital tako definiranim razredom?

4. Na žagi so imeli pri žaganju desk na predpisane dolžine precej odpadkov, zato so se odločili, da bodo podrobneje proučili dolžine ostankov. Na vzorcu, sestavljenem iz 40 enot, so izmerili dolžine (v cm):

148	155	185	164	177	165	160	174	172	169
176	173	139	168	145	170	155	183	156	164
162	167	196	167	152	155	162	158	160	170
165	193	146	178	184	181	166	188	166	158

Sestavi frekvenčno porazdelitev z razredi
138 – 146, 147 – 155, ..., 192 – 200.
Izračunaj vse ostale porazdelitve (relativne frekvence, kumulativni porazdelitvi frekvenc in relativnih frekvenc). Kolikšen odstotek ostankov je daljših od 155 cm? Nariši histogram in frekvenčni poligon.

5. Nariši frekvenčni poligon za porazdelitev števila otrok v vzorcu 6 570 družin, če so možne vrednosti proučevane spremenljivke

0, 1, 2, 3, 4, 5, 6, 7, 8, 9 in 10,

pripadajoče frekvence pa

1 230, 1 520, 1 545, 962, 537, 301, 174, 108, 69, 51 in 73.

Izračunaj relativne frekvence in kumulativno frekvenčno porazdelitev.

Ugotovi delež družin, ki imajo
a) vsaj enega otroka;
b) več kot dva otroka;
c) kvečjemu tri otroke.

6. V športnem društvu so izmerili višino vseh svojih 120 mlajših dečkov, mladincev in članov in vrednosti razvrstili v razrede (v cm) 110 – 119, 120 – 129, ..., 200 – 209, 210 – 219; merjenje so izvedli na cm natančno. Frekvence posameznih razredov so bile po vrsti enake 1, 4, 17, 28, 25, 18, 13, 6, 5, 2 in 1. Predstavi to frekvenčno porazdelitev s histogramom in frekvenčnim poligonom, izračunaj relativne frekvence in kumulativno frekvenčno porazdelitev ter ugotovi, kolikšen odstotek članov tega kluba meri vsaj 180 cm.

7. Največja cifra števila 15 je 5, največja cifra števila 66 je 6. Za posamezne cifre ugotovi, kolikokrat nastopijo kot največje cifre v naravnih številih od 1 do 99 in sestavi ustrezno frekvenčno porazdelitev. Nariši frekvenčni kolač, s katerim boš predstavil to porazdelitev.

8. Premisli, kako bi predstavil frekvenčno porazdelitev v primeru, ko razredi niso enako široki, da bi bil prikaz čimbolj nazoren.

3.3 Srednje vrednosti

Bralec gotovo pozna izraz **aritmetična sredina**. V najpreprostejšem primeru gre za tovrstno sredino dveh števil a, $b \in \mathbb{R}$; njuna aritmetična sredina je določena z vrednostjo izraza $(a+b)/2$.

ZGLED 3.3.1:

Tovarna je prodala tri izdelke po cenah 25 000, 31 000 in 34 000 denarnih enot. Kakšna je bila povprečna cena?

Po analogiji k obrazcu za sredino dveh števil je povprečna cena $(25\,000 + 31\,000 + 34\,000)/3 = 30\,000$ denarnih enot.

Na splošno lahko rečemo takole: če spremenljivka na N opazovanih enotah zavzame vrednosti $y_1, y_2, \ldots, y_N$, definiramo (navadno) **aritmetično sredino** teh vrednosti s predpisom

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad (1)$$

Besedo **povprečje** običajno uporabljamo namesto izraza "sredina". V splošnem je, kot bomo videli v nadaljevanju, koristno, če povemo, kako to število računamo; le v primeru, ko je jasno, da gre za navadno aritmetično sredino, lahko ta podatek izpustimo.

Primer posredno pokaže, kaj povprečna vrednost, kot jo razumemo v vsakdanjem govoru, pomeni: pove nam, kakšno vrednost bi zavzela spremenljivka na vsakem posameznem elementu, če bi vsoto dejanskih vrednosti te spremenljivke razdelili enakomerno na vse enote.

Če bi morali na kratko povedati, kaj je namen tega razdelka, bi rekli takole: pokazati želimo, da je številu, ki čepi točno na sredini med sosedoma, samo včasih (beri: pri določenih pogojih) pametno reči **srednja** ali **povprečna vrednost**.

S povprečji so nasploh težave in z njimi je povezanih morda največ napačnih razlag statističnih podatkov. Najprej opozorimo, da je računanje povprečnih vrednosti smiselno samo za številske spremenljivke.

ZGLED 3.3.2:

Izračunati "povprečno barvo osebnega avtomobila, registriranega v Sloveniji", gotovo nima nobenega smisla, čeprav bi se gotovo našli ljudje, ki bi jim to formalno uspelo. Vse barve na primer razvrstimo v barvno lestvico in jim priredimo zaporedne številke, ki jih zavzamejo v takem katalogu, nato pa iz te spremenljivke izračunamo povprečno vrednost, pogledamo spet v katalog...

Take neumnosti prepoznamo od daleč in so zato znatno manj nevarne od naslednjih.

ZGLED 3.3.3:

Tovarna je prodala 9 izdelkov po 25 000 d.e., 7 po 31 000 d.e. in 4 po 34 000 d.e.; kolikšna je bila povprečna cena?

Žal se še vedno najdejo ljudje, ki povprečno ceno tudi sedaj računajo natanko tako kot v zgledu 3.3.1 in dobijo enak, tokrat seveda napačen rezultat. Sklepajmo tako, kakor nam narekuje vsebina pojma "povprečna vrednost", kot smo ga opisali zgoraj: izkupiček od prodaje delimo s številom vseh prodanih proizvodov. Tak račun nam da

$$\mu = \frac{9\cdot 25\,000 + 7\cdot 31\,000 + 4\cdot 34\,000}{9+7+4} = \frac{578\,000}{20} = 28\,900$$

Lahko bi ravnali tudi drugače, izpisali bi vsako ceno tolikokrat, kolikorkrat se je pojavila pri prodaji (ali drugače, kolikor je bila njena frekvenca), nato pa bi opravili seštevanje in vsoto delili s številom izdelkov: rezultat bi bil seveda enak.

Odkar je človeštvo izumilo večkratnike, se ne ubada več s krutimi različicami, kakršno smo ponudili kot alternativo v našem primeru. Prav obratno, iz primera vidimo, kako računati aritmetično sredino, če imamo za spremenljivko r možnih vrednosti $y_1, y_2, \ldots, y_r$ in za vsako od njih frekvenco f_i, torej podatek, ki pove, kolikokrat ta vrednost dejansko nastopa. V takem primeru imamo v obrazcu za izračun aritmetične sredine v števcu vsoto produktov posameznih vrednosti z njim pripadajočimi frekvencami, v imenovalcu pa vsoto frekvenc, ki je seveda nujno enaka številu vseh elementov N, na katerih smo izmerili vrednosti. Tako je v takem primeru

$$\mu = \frac{f_1\cdot y_1 + f_2\cdot y_2 + \ldots + f_r\cdot y_r}{f_1 + f_2 + \ldots + f_r} = \frac{1}{N}\sum_{i=1}^{r} f_i\, y_i \qquad (2)$$

Zaradi vloge, ki jo imajo v tem obrazcu frekvence, jih imenujemo **uteži** ali tudi **ponderji**, po obrazcu (2) izračunani povprečni vrednosti pa rečemo **tehtana aritmetična sredina**.

ZGLED 3.3.4:

32 učencev 4. A razreda je imelo ob polletju naslednje ocene iz matematike: trije odlično (5), pet prav dobro (4), deset dobro (3), osem zadostno (2) in šest nezadostno (1). Izračunajmo povprečno oceno iz matematike v tem razredu!

Po obrazcu imamo (če rezultat zaokrožimo na dve decimalni mesti)

$$\mu = \frac{3\cdot 5 + 5\cdot 4 + 10\cdot 3 + 8\cdot 2 + 6\cdot 1}{3+5+10+8+6} = 2{\cdot}72$$

Na populaciji 32 učencev smo v tem zgledu opazovali spremenljivko, ki ima pet možnih vrednosti (ocene 5, 4, 3, 2, 1), pripadajoče uteži pa so bile seveda frekvence posameznih ocen. Iz povprečne ocene ne moremo prav nič povedati o tem, kakšne ocene imajo posamezniki iz proučevanega razreda, pove pa nam nekaj o populaciji kot celoti, denimo to, da je več učencev imelo oceno 1 ali 2, kot pa je bilo tistih, ki so prislužili 4 ali 5. Če poznamo povprečno oceno kakega drugega razreda in vemo, da je ta višja od povprečja v 4. A razredu, lahko govorimo o uspešnejšem razredu, še zmeraj pa imajo lahko nekateri posamezniki iz 4. A razreda boljšo oceno kot nekateri učenci v primerjalnem razredu. Na kratko, aritmetična sredina μ je **statistični parameter**, govori o populaciji kot celoti in sama po sebi zelo malo o tem, kako so porazdeljene individualne vrednosti statistične spremenljivke.

Tehtano aritmetično sredino lahko očitno nadomestimo z navadno aritmetično sredino, izračunano po obrazcu (1), samo v primeru, ko imajo vse možne vrednosti enake frekvence; v vseh drugih primerih pa nas take poenostavitve pripeljejo do napačnih rezultatov. (Zelo banalen zgled bi predstavljalo zanemarjanje uteži pri računanju povprečne ocene, saj bi potem vsi razredi v šoli imeli povprečno oceno 3.)

V prejšnjem razdelku smo se ukvarjali z grupiranjem podatkov v frekvenčne razrede. Kot vemo, nam frekvenca f_i v tem primeru pove, na koliko enotah je spremenljivka zavzela vrednost, ki sodi na i – ti podinterval. Tudi v tem primeru lahko uporabimo obrazec (2), če za vrednosti y_i vzamemo sredine omenjenih podintervalov oziroma frekvenčnih razredov. Seveda se ob tem zavedamo (glej okvir na strani 55!), da je tako izračunana aritmetična sredina samo približek za tisto vrednost, ki bi jo izračunali iz originalnih, to je negrupiranih podatkov.

ZGLED 3.3.5:

Za - že dobro znano - porazdelitev mas stotih učencev v naslednji tabeli izračunajmo aritmetično sredino!

Razred	Masa (kg)	Sredine (y_i)	f_i	$f_i \cdot y_i$
1	59·5 - 62·5	61	5	305
2	62·5 - 65·5	64	18	1 152
3	65·5 - 68·5	67	42	2 814
4	68·5 - 71·5	70	27	1 890
5	71·5 - 74·5	73	8	584
			100	6 745

Tabelo smo na desni strani že razširili s stolpcem za produkte $f_i\, y_i$, ki jih potrebujemo za izračun:

$$\mu = \frac{1}{N}\sum_{i=1}^{r} f_i\, y_i = \frac{6\,745}{100} = 67{\cdot}45$$

Razlaga rezultata je enaka kot v vseh dosedanjih zgledih z aritmetično sredino: Če bi vsi učenci imeli enake mase, bi bila ta masa 67,45 kg; toliko dobimo, če skupno maso vseh stotih učencev razdelimo na enake dele.

Bistvena značilnost tehtane aritmetične sredine dveh danih števil a in b je v dejstvu, da je na številski premici ustrezna točka sicer nekje na daljici, ki ima krajišči v točkah a in b, vendar ne nujno na polovici te daljice; na katero stran "zdrsne", je odvisno od uteži, ki jih imata števili a in b v obrazcu (2).

Poleg aritmetične sredine uporabljamo še dve srednji vrednosti. Posebej pomembna je **geometrijska sredina** ali geometrijsko povprečje.

Geometrijska sredina pozitivnih števil $y_1, y_2, \ldots, y_N$ je število

$$G = \sqrt[N]{y_1 \cdot y_2 \cdot \ldots \cdot y_N} \tag{3}$$

Kako pridemo do nje, je mogoče opisati še na bolj zapleten način, ki pa ima v statistiki posebno vsebinsko ozadje: če najprej logaritmiramo vrednosti $y_1, y_2, \ldots, y_N$, iz dobljenih logaritmov izračunamo aritmetično sredino in nato rezultat opisanega postopka še antilogaritmiramo, bomo natanko zadeli G. (Bralec bo izkoristil priložnost in bo to trditev tudi dokazal, da osveži poznanstvo z logaritmi še pred maturo...)

Geometrijsko sredino največkrat uporabljamo pri računanju povprečnega koeficienta rasti, zato se bomo k njej vrnili v razdelku 3.5, ko bomo govorili o indeksnih številih; tam nas čaka tudi praktičen zgled.

V statistiki srečamo tudi **harmonično sredino** oziroma harmonično povprečje števil. Za dana pozitivna števila $y_1, y_2, \ldots, y_N$ to sredino izračunamo po obrazcu

$$H = \frac{N}{\frac{1}{y_1} + \frac{1}{y_2} + \ldots + \frac{1}{y_N}} = \{\, \frac{1}{N}\sum_{i=1}^{N} \frac{1}{y_i} \,\}^{-1} \tag{4}$$

Način računanja enostavneje pojasnimo z desno obliko zapisa:

Harmonična sredina je enaka recipročni vrednosti aritmetične sredine recipročnih vrednosti danih števil.

Za tehtano harmonično sredino imamo analogen obrazec,

$$H = \{\, \sum_{i=1}^{r} N_i \,\} : \{\, \sum_{i=1}^{r} \frac{N_i}{y_i} \,\} \tag{5}$$

kjer so y_i kot običajno vrednosti spremenljivke, N_i pa pripadajoče uteži. Največkrat uporabljamo harmonično sredino pri računanju povprečij iz različnih koeficientov, po zelo naravni poti pa pridemo do nje tudi pri računanju povprečne hitrosti.

ZGLED 3.3.6:

V šolo pridemo tako, da se najprej peljemo 3 km z avtobusom, ki vozi po mestu s hitrostjo 40 km/h, nato pa prehodimo še 1 kilometer s hitrostjo 4 km/h. Kakšna je povprečna hitrost, ki jo dosežemo na poti v šolo?

Uporabimo oznake, domače iz fizike, torej s za pot in v za hitrost. Povprečno hitrost dobimo tako, da celotno pot $(s_1 + s_2)$ delimo s celotnim časom, porabljenim zanjo; porabili pa smo s_1/v_1 ure na avtobusu in s_2/v_2 ure pri pešačenju; tako je povprečna hitrost enaka

$$\frac{s_1 + s_2}{\frac{s_1}{v_1} + \frac{s_2}{v_2}} = \frac{3 + 1}{\frac{3}{40} + \frac{1}{4}}\,\mathrm{km/h} \approx 12{\cdot}31\,\mathrm{km/h}$$

Obrazec, do katerega smo prišli s fizikalnim premislekom, je natančno obrazec (5), če vzamemo hitrosti v_1 in v_2 za vrednosti proučevane statistične spremenljivke, dolžini poti s_1 in s_2, na katerih je bila posamezna hitrost dosežena, pa za pripadajoči uteži.

ZGLED 3.3.7:

Trgovina je prodajala tri izdelke, recimo jim A, B in C. Pri prvem je ob prometu 5 666 000 d.e. enot imela t.i. koeficient obračanja zaloge 6˙8, pri drugem je bil promet 1 585 000 d.e. in koeficient obračanja 13˙8, pri izdelku C sta ustrezna podatka 2 028 000 d.e. in 6˙0. Kakšen je bil povprečen koeficient obračanja zalog?

Ta naloga je zelo pomembna za trgovce, ki se jim mora zaloga čim hitreje obračati. Kako hitro se to dogaja, merijo ekonomisti s koeficientom obračanja zaloge, ki je definiran kot količnik med prometom v določenem obdobju (npr. letu) in povprečno velikostjo zaloge v tem obdobju.

Naj bodo torej P_i zneski prometa po posameznih izdelkih, k_i pripadajoči koeficienti obračanja zaloge in Z_i povprečna stanja zaloge v proučevanem obdobju; zadnjih podatkov sicer nimamo, vemo pa, da je koeficient obračanja za vsako blago posebej definiran kot količnik P_i/Z_i, tako da je $Z_i = P_i/k_i$. Za povprečni koeficient obračanja celotne zaloge imamo v števcu skupni promet in v imenovalcu skupno velikost povprečnih zalog,

$$k = \{\sum_{i=1}^{3} P_i\} : \{\sum_{i=1}^{3} \frac{P_i}{k_i}\}$$

to pa je natanko harmonična sredina koeficientov k_1, k_2, k_3 z zneski prometov P_i kot utežmi. Pri naših podatkih imamo

$$k = \frac{5\,666\,000 + 1\,585\,000 + 2\,028\,000}{\frac{5\,666\,000}{6{\cdot}8} + \frac{1\,585\,000}{13{\cdot}8} + \frac{2\,028\,000}{6{\cdot}0}} = 7{\cdot}2\ldots$$

Bralec se lahko prepriča, da bi z aritmetično sredino v obeh zgledih dobil (zelo) napačen rezultat.

Na koncu razdelka omenimo še dva parametra, ki na svoj način govorita o "povprečnem obnašanju" statistične spremenljivke na neki populaciji.

Mediana ali *središčnica* je število Me, ki na določen način razpolovi populacijo: tako število je, da spremenljivka na polovici enot zavzame manjšo, na polovici pa večjo vrednost od nje. Formalno rečemo takole: mediana bi se v naraščajočem zaporedju N vrednosti, ki jih zavzame statistična spremenljivka, uvrstila na mesto ("rang"), ki ga določa vrednost izraza $(N + 1)/2$. Če je N liho število, je vrednost tega izraza celo število in mediana je preprosto enaka tisti vrednosti, ki je natančno v sredini tako dobljene vrste. Kadar je vseh vrednosti sodo število, $(N + 1)/2$ ni celo število, mediano je treba pač postaviti "v sredino med srednji dve vrednosti", pa je.

Več dela je pri grupiranih podatkih, kjer s frekvencami določena relativna pomembnost posameznih vrednosti (ali sredin posameznih razredov) odloča o tem, kje leži mediana. Poglejmo si kar na zgledu, kako izračunamo mediano takrat, ko so vrednosti grupirane v razrede.

ZGLED 3.3.8:

Izračunajmo mediano za porazdelitev mas učencev, ki smo jo uporabljali že v prejšnjih zgledih. Iz kumulativne porazdelitve

Razred	Masa (kg)	Sredine (y_i)	f_i	F_i
1	59˙5 - 62˙5	61	5	0
2	62˙5 - 65˙5	64	18	5
3	65˙5 - 68˙5	67	42	23
4	68˙5 - 71˙5	70	27	65
5	71˙5 - 74˙5	73	8	92
			100	100

vidimo, da je spodnja meja 3. razreda zagotovo manjša od mediane (kumulativa je 23, kar je manj od polovice enot v populaciji), medtem ko je spodnja meja naslednjega razreda že večja od mediane, saj je kumulativa iz naslednjega razreda (65) večja od polovice populacije. Mediana se potemtakem skriva nekje v 3. razredu. Natančnejši položaj določimo z linearno interpolacijo. Sklepamo takole: mediana leži tako, da ima 50 učencev manjšo, ostalih 50 pa večjo maso. Predpostavimo, da je porazdelitev mas znotraj posameznega razreda enakomerna (kar sicer ni najbolj logično, vendar običajno nimamo boljše možnosti). Potem je razdalja od spodnje meje razreda do mediane v enakem razmerju do celotne širine intervala, kot je razmerje med potrebnim prirastkom kumulative (v našem primeru 50·5 - 23) in njenim celotnim prirastkom oziroma frekvenco tega razreda (42). Tako dobimo oceno

$$Me' = 65{\cdot}5 + 3{\cdot}0 \cdot \frac{50{\cdot}5 - 23}{42} \approx 67{\cdot}46$$

Če rezultat primerjamo z onim iz zgleda 3.3.5, vidimo, da je mediana za to porazdelitev samo za spoznanje večja od aritmetične sredine. Tako dobro ujemanje je značilno za porazdelitve, ki so vsaj približno simetrične.

V splošnem pa prikazano interpolacijo opravimo takole:

- izračunamo rang, ki ustreza mediani, to je $R_{Me} = (N+1)/2$; tu N kot običajno pomeni število vrednosti,
- pogledamo, kateri razred s svojo frekvenco f_{Me} poveča dotedanjo kumulativo frekvenc F_{Me} nad rang R_{Me},
- mediano izračunamo po obrazcu

$$Me' = y_{Me,s} + (y_{Me,z} - y_{Me,s}) \cdot \frac{R_{Me} - F_{Me}}{f_{Me}} \tag{6}$$

V obrazcu se poleg že opisanih oznak pojavljata še spodnja in zgornja meja razreda, v katerem leži mediana. Za vsak slučaj ponovimo opozorilo: iz grupiranih podatkov izračunana mediana je samo približek za dejansko vrednost tega parametra, ki bi jo dobili iz individualnih vrednosti statistične spremenljivke; odtod tudi črtica ob simbolu v obrazcu (6), ki opozarja, da gre za približek.

Zadnja srednja vrednost, ki jo omenjamo, je **modus** (tudi *gostiščna vrednost* ali *gostiščnica*). To je najpogostejša vrednost med vsemi vrednostmi, ki jih je zavzela opazovana statistična spremenljivka.

Dokler so na voljo izvorni negrupirani podatki, z modusom ni težav, pogledamo pač, katera vrednost se največkrat pojavi. Še najhujši problem je lahko v tem, da pri nekaterih porazdelitvah ni enolično določene najpogostejše vrednosti, ampak ima več vrednosti isto (največjo) frekvenco.

Pri podatkih, ki so grupirani v (enako široke) razrede, pa je spet potrebnega nekaj več dela. Najprej določimo **modalni razred**, to je razred z največjo frekvenco. Tudi tu se lahko zgodi, da obstajata dva razreda z enako, med vsemi največjo frekvenco; takrat modus ni enolično določen, za porazdelitev pravimo, da je *bimodalna*. Na pojav nas lepo opozori njen histogram, v katerem se enakopravno kosata za prvenstvo dva enako visoka stolpca. (Učeno ime za tako porazdelitev, ki se ponaša z enim samim izrazitim "vrhom", je v sorodu z brezalkoholnim pivom in ga bo bralec sestavil sam.)

Če je modalni razred enolično določen, izračunamo modus (bolje: približek zanj) z linearno interpolacijo po obrazcu

$$Mo' = y_{Mo,s} + (y_{Mo,z} - y_{Mo,s}) \cdot \frac{f_{Mo} - f_{Mo-1}}{2\,f_{Mo} - f_{Mo-1} - f_{Mo+1}} \tag{7}$$

Poleg frekvence f_{Mo}, ki pripada modalnemu razredu, se v obrazcu pojavljata tudi frekvenci obeh sosednjih razredov, tistega pred njim (f_{Mo-1}) in onega za njim (f_{Mo+1}).

ZGLED 3.3.9:

Za tolikokrat uporabljeno porazdelitev mas je modalni razred tretji, s frekvenco $f_{Mo} = 42$, spodnjo mejo $y_{Mo,s} = 65{\cdot}5$, širino $y_{Mo,z} - y_{Mo,s} = 3{\cdot}0$ in frekvencama obeh sosedov $f_{Mo-1} = 18$, $f_{Mo+1} = 27$. Po obrazcu je zato

$$Mo' = 65{\cdot}5 + 3{\cdot}0 \cdot \frac{42 - 18}{2 \cdot 42 - 18 - 27} \approx 67{\cdot}35$$

Spoznali smo kopico srednjih vrednosti in se sprijaznili s tem, da bolj ali manj dobro znano aritmetično povprečje ni edini parameter te vrste. Stvari pa šele postajajo prav zanimive: kakšna je izpovedna moč posameznega parametra, s katerim izražamo srednjo vrednost, kdaj je eden od njih bolj primeren od drugih, kakšen je numeričen odnos med njimi, kaj lahko iz teh odnosov povemo o porazdelitvi vrednosti statistične spremenljivke... to so vprašanja, ki presegajo naš okvir, po odgovore bo treba v resnejšo statistično literaturo.

V A J E

1. Katere vrste srednjih vrednosti poznaš? Po čem se razlikuje navadna aritmetična sredina od tehtane?
2. Izračunaj aritmetično, geometrijsko in harmonično sredino števil 2, 4, 7, 9, 11, 12, 15 in 17.
3. Naj bosta a in b poljubni pozitivni števili. Presodi, kdaj se lahko aritmetična in geometrijska sredina ujemata.

4. Pet strelcev je zadelo 8 krogov, štirje po 5 krogov, en strelec samo 4 kroge, šest strelcev 7 krogov in dva strelca sredino tarče (10 krogov). Izračunaj povprečen rezultat tega streljanja.

5. Nekdo je prodal 3 parcele po 400 m^2 po 35 d.e. za kvadratni meter in 7 parcel po 500 m^2 po 42 d.e. za kvadratni meter. Kolikšna je bila povprečna cena kvadratnega metra zemljišča?

6. Na žagi so imeli pri žaganju desk na predpisane dolžine precej odpadkov, zato so se odločili, da bodo podrobneje proučili dolžine ostankov. Na vzorcu, sestavljenem iz 40 enot, so izmerili dolžine (v cm):

148	155	185	164	177	165	160	174	172	169
176	173	139	168	145	170	155	183	156	164
162	167	196	167	152	155	162	158	160	170
165	193	146	178	184	181	166	188	166	158

Sestavi frekvenčno porazdelitev z razredi
138 – 146, 147 – 155, ..., 192 – 200,
nato pa izračunaj aritmetično sredino
a) iz negrupiranih podatkov;
b) iz dobljene porazdelitve v razrede.
Primerjaj rezultata. (Opomba: pridni učenci imajo polovico naloge že narejeno iz prejšnjega razdelka.)

7. Za frekvenčno porazdelitev iz prejšnje naloge izračunaj tudi modus in mediano.

8. Za vzorec 6 570 družin iz 5. naloge v prejšnjem razdelku izračunaj povprečno število otrok.

9. Izračunaj aritmetično sredino prvih n naravnih števil.

10. Denimo, da se v šolo peljemo z avtobusom 6 km in nato s hitrostjo 5 km/h hodimo še 1·5 kilometra daleč. Kako hitro mora voziti avtobus, da bo povprečna hitrost, ki jo dosežemo na celi poti, vsaj 20 km/h?

11. Sam ali z učiteljevo pomočjo ugotovi, kako pridemo do obrazca (7).

3.4 Mere variabilnosti

Koliko pove neka srednja vrednost (v nadaljevanju bomo zaradi enostavnosti govorili samo o aritmetični sredini) o populaciji, je odvisno od tega, kako so individualne vrednosti, ki jih je zavzela statistična spremenljivka, razpršene okrog tega povprečja. Bolj ko so individualne vrednosti razpršene, manjša je izpovedna moč srednje vrednosti.

Prva ideja, ki pride človeku na misel, ko poskuša meriti razpršenost podatkov, je najbrž naslednja: vzemimo njihovo aritmetično sredino, izračunajmo vse odklone individualnih vrednosti od nje in tako dobljene razlike seštejmo. Preprosto, toda žal neučinkovito: aritmetična sredina μ po definiciji leži na takem mestu, da se vsota negativnih odklonov natanko uravnovesi z vsoto pozitivnih, ali drugače,

Vsota odklonov individualnih vrednosti od aritmetične sredine je za vsako statistično spremenljivko enaka 0.

Če bi namesto odklonov sešteli njihove absolutne vrednosti, bi se opisane težave rešili, nakopali pa bi si kopico drugih, tako tehničnih kot vsebinskih. Zato se je kot merilo za razpršenost oziroma **variabilnost** podatkov uveljavila **varianca** (simbol σ^2), ki jo izračunamo kot povprečje kvadratov odklonov vrednosti y_i od aritmetične sredine μ; za negrupirane podatke izračun dobesedno sledi opisu,

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mu)^2 \tag{1}$$

pri podatkih, grupiranih v r razredov s sredinami y_i, pa seštevanje kot običajno skrajšamo z večkratniki (f_i so spet frekvence),

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{r} f_i(y_i - \mu)^2 \tag{2}$$

ZGLED 3.4.1:

Statistični znak (spremenljivka) Y je na neki populaciji zavzela vrednosti 12, 6, 7, 3, 15, 10, 18 in 5. Izračunajmo varianco!

Dobesedno po receptih gre račun takole:

$$\mu = \frac{12+6+7+3+15+10+18+5}{8} = 9{\cdot}5$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mu)^2 = \frac{1}{8}\{(12-9{\cdot}5)^2 + (6-9{\cdot}5)^2 \ldots$$
$$\ldots (18-9{\cdot}5)^2 + (5-9{\cdot}5)^2\} = 23{\cdot}75$$

Za obdelavo večjih količin podatkov je zelo koristna ugotovitev, da je mogoče osnovno obliko obrazcev (1) in (2) predelati v

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \mu^2 \tag{3}$$

oziroma

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{r} f_i\, y_i^2 - \mu^2 \tag{4}$$

Ker si večina ljudi obrazce bolje zapomni po vsebini kot po simbolih, povejmo navodilo: varianco dobimo tako, da povprečje kvadriranih vrednosti opazovane spremenljivke zmanjšamo za kvadrat njihovega običajnega povprečja μ.

Obrazca, do katerih pridemo s kvadriranjem dvočlenikov v (1) in (2) ter s poenostavljanjem tako dobljenih izrazov (vaja!), sta znatno ugodnejša za praktično računanje s skromnimi priročnimi (beri: žepnimi) sredstvi.

ZGLED 3.4.2:

Izračunajmo varianco za frekvenčno porazdelitev... se ve, katero.

Razred	Masa (kg)	y_i	f_i	$f_i \cdot y_i$	$f_i \cdot y_i^2$
1	59·5 - 62·5	61	5	305	18 605
2	62·5 - 65·5	64	18	1 152	73 728
3	65·5 - 68·5	67	42	2 814	188 538
4	68·5 - 71·5	70	27	1 890	132 300
5	71·5 - 74·5	73	8	584	42 632
			100	6 745	455 803

Tabelo, ki smo jo srečali pri računanju aritmetične sredine, smo na desni strani še razširili s stolpcem za produkte $f_i\, y_i^2$, ki jih potrebujemo za izračun variance. Aritmetično sredino za to porazdelitev že poznamo,

$$\mu = \frac{1}{N}\sum_{i=1}^{r} f_i\, y_i = \frac{6\,745}{100} = 67{\cdot}45$$

z uporabo podatka iz skrajnega jugovzhodnega vogala tabele pa po obrazcu (4) takoj dobimo še varianco:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{r} f_i\, y_i^2 - \mu^2 = \frac{1}{100}\cdot 455\,803 - 67{\cdot}45^2 = 8{\cdot}5275$$

Za praktično uporabo kaže rezultat zaokrožiti, denimo na 8·5, saj ne more biti natančnejši od vhodnih podatkov. Bralca pa vabimo, da varianco na vsak način izračuna tudi neposredno po osnovnem obrazcu (2), da bo znal ceniti prihranke, ki jih prinašata različici (3) in (4).

Kakorkoli pa že prispemo do variance, ostaja na njej grd madež: v našem konkretnem primeru je izražena v kvadratnih kilogramih, v splošnem ima za mersko enoto kvadrat tiste enote, v kateri izražamo vrednosti opazovane spremenljivke. To je dovolj nerodno, da imamo v praksi namesto nje dosti rajši njen kvadratni koren, ki mu rečemo **standardni odklon** (bolj učeno tudi *standardna deviacija*) in ga - razumljivo - označimo s simbolom σ. Posebnih računskih težav z njim ni, ko imamo v delovnem pomnilniku (česarkoli) izračunano varianco, je potrebno samo še korenjenje.

ZGLED 3.4.3:

Za porazdelitev iz prejšnjega zgleda je

$$\sigma = \sqrt{\sigma^2} = \sqrt{8{\cdot}5275} = 2{\cdot}92$$

Rezultat je enak tudi v primeru, da korenimo zaokroženo vrednost za varianco.

Poleg te, da ima enako mersko enoto kot izhodiščni podatki, pa ima standardni odklon še eno lepo lastnost. Za približno simetrične porazdelitve z izrazitim modalnim razredom (zdaj se ne moremo več sprenevedati: imenujemo jih *unimodalne*) lahko na intervalu, ki sega po en standardni odklon v levo in desno od aritmetične sredine μ, pričakujemo približno dve tretjini individualnih vrednosti statistične spremenljivke.

Varianca in standardni odklon sta samo dva - čeprav najpogosteje uporabljana - predstavnika parametrov, ki jih uporabljamo za merjenje variabilnosti podatkov. Analogno lahko definiramo povprečja različnih potenc odklonov od mediane ali modusa, povprečni absolutni odklon in tako dalje. Nenazadnje je najpreprostejše (in temu primerno precej nezanesljivo) merilo za variabilnost podatkov že t.i. **variacijski razmik** (tudi: *razmah, variacijski razpon*), definiran kot razlika med največjo in najmanjšo individualno vrednostjo statistične spremenljivke:

$$R_v = y_{max} - y_{min} \tag{5}$$

Za konec razdelka še tole. Če nekomu rečemo, da je v proučevanem podjetju standardni odklon pri plačah približno 10 000 SIT, bo najbrž takoj vprašal, kolikšna je povprečna plača, saj je pri danem standardnem odklonu variabilnost v bistvu tem manjša, čim večja je aritmetična sredina.

Da bi omogočil primerljivost variabilnosti za različne spremenljivke oziroma porazdelitve, je znameniti statistik K. Pearson že ob koncu prejšnjega stoletja predlagal **koeficient variacije**, ki je definiran kot razmerje med standardnim odklonom in aritmetično sredino dane porazdelitve, oziroma, če želimo parameter izraziti v odstotkih, kot stokratnik opisanega razmerja. Za tolikokrat uporabljeno porazdelitev mas stotih učencev je koeficient variacije

$$K_v = \frac{\sigma}{\mu} = \frac{2{\cdot}92}{67{\cdot}45} = 0{\cdot}04329... \approx 4{\cdot}3\,\%$$

in lahko bi rekli, da variabilnost ni prav izrazita.

VAJE

1. Izračunaj varianco in standardni odklon števil 8, 5, 4, 7 in 10.

2. Od 47 učencev je 7 učencev doseglo oceno odlično, 16 oceno prav dobro, 13 dobro, 5 zadostno in 6 nezadostno. Izračunaj standardni odklon in koeficent variacije za to porazdelitev.

3. Nadaljuj nalogo 8 iz prejšnjega razdelka (dolžine ostankov na žagi) in izračunaj še vse obravnavane parametre, s katerimi merimo variabilnost, tako za podatke pred grupiranjem kot za frekvenčno porazdelitev po razredih. Oceni, kakšna je v tem primeru napaka, če prave parametre nadomestimo s približki, ki jih dobimo iz grupiranih podatkov.

4. V podjetju, ki ima 138 delavcev, so preverjali, koliko časa porabijo za prevoz na delo in z dela. Iz naslednje tabele, kjer so y_i v minutah izraženi časi, izračunaj vse potrebne parametre in poskusi pojasniti, kakšna je porazdelitev transportnih časov:

Razred	y_i	f_i
1	0 - 30	70
2	31 - 60	47
3	61 - 90	12
4	91 - 120	4
5	121 - 150	3
6	151 - 180	2

5. Denimo, da ima podjetje "VARNOST" direktorja z zelo visoko plačo, snažilko z zelo majhno in 20 varnostnikov s približno enakimi dohodki. Kaj porečeš o povprečni plači v tem podjetju? Kako je s standardnim odklonom? Kaj nam o variabilnosti pove variacijski razmik?

3.5 Indeksna števila

Statistiki pravijo, da so **indeksi** taka relativna števila, ki jih dobimo z razmerjem dveh podatkov, ki se nanašata na istovrstna pojava.

Število, s katerim primerjamo (in ga zato postavimo v imenovalec), imenujemo osnova ali **baza**, standardna oznaka zanj je Y_0. Omenjeno razmerje običajno razširimo s 100, da je indeks izražen v odstotkih.

Ko govorimo o istovrstnih pojavih, običajno mislimo na vrednosti istega pojava v različnih časih ali na različnih geografskih območjih. Najpogosteje se ukvarjamo ravno s **časovnimi indeksi**, ki jih dobimo tako, da primerjamo vrednosti določenega statističnega znaka, ki jih je ta imel v dveh različnih časovnih trenutkih oziroma na dveh različnih časovnih intervalih. Glede na to, ali za primerjavo vedno uporabljamo isti podatek ali pa se osnova za primerjavo spreminja, govorimo o **indeksih s stalno osnovo** in o **indeksih s premično osnovo**.

Naj bodo $Y_0, Y_1, Y_2, \ldots, Y_n$ vrednosti proučevanega pojava v zaporednih časovnih obdobjih. Če vzamemo za osnovo prvi podatek, dobimo naslednje zaporedje **baznih indeksov**:

$$I_{k/0} = 100 \cdot \frac{Y_k}{Y_0} \qquad (k = 1, 2, \ldots, n) \tag{1}$$

ZGLED 3.5.1:

Iz tabele o gibanju proizvodnje nekega blaga v letih 1980 - 1989 izračunajmo indekse z osnovo v letu 1980!

LETO	PROIZVODNJA	INDEKS
1980	**5 188**	**100·0**
1981	5 322	102·6
1982	5 249	101·2
1983	5 150	99·3
1984	7 235	139·5
1985	7 942	153·1
1986	8 904	171·7
1987	9 276	178·8
1988	9 864	190·1
1989	11 544	222·5

Rezultate v zadnjem stolpcu smo zaokrožili.

Kaj nam povedo indeksi s stalno osnovo? Glede na obrazec (1) lahko rečemo, da indeks $I_{k/0}$ pove, koliko % celote Y_0 predstavlja vrednost pojava Y_k. Vrednost indeksa nad 100 pomeni, da je imel v proučevanem obdobju pojav večjo vrednost kot v baznem obdobju, vrednost indeksa pod 100 pa nasprotno priča, da je vrednost pojava za ustrezno število odstotkov – namreč za $(100 - I_{k/0})\%$ – manjša kot v baznem obdobju.

Za vsebinsko analizo dinamike pojavov je nerodno, če si za osnovo analize izberemo "patološko" obdobje, zato ni nujno, da je izhodišče prvo leto časovne vrste. Kako preračunati indekse na kako drugo osnovo?

Recimo, da namesto začetnega obdobja vzamemo za osnovo j – to obdobje. Potem velja naslednja ugotovitev:

$$I_{k/j} = 100 \cdot \frac{Y_k}{Y_j} = 100 \cdot \frac{100 \cdot (Y_k \,/\, Y_0)}{100 \cdot (Y_j \,/\, Y_0)} = 100 \cdot \frac{I_{k/0}}{I_{j/0}} \tag{2}$$

Predelava indeksov na novo osnovo torej zahteva samo eno deljenje in premik decimalne vejice ali pike. (Podrobna izpeljava nakazanega dokaza bi lahko bralcu prinesla obilo veselja s tako ljubimi dvojnimi ulomki!)

ZGLED 3.5.2:

Indekse iz prejšnjega zgleda preračunajmo tako, da bo namesto leta 1980 osnova leto 1984!

Po obrazcu (2) za vsako leto vzamemo razmerje med njegovim dosedanjim indeksom in indeksom leta 1984 glede na leto 1980 (to je 139·5) in dobljeni količnik pomnožimo s 100, kar nam da novo vrsto baznih indeksov:

LETO	PROIZVODNJA	INDEKS
1980	5 188	71·7
1981	5 322	73·5
1982	5 249	72·5
1983	5 150	71·2
1984	**7 235**	**100·0**
1985	7 942	109·7
1986	8 904	123·1
1987	9 276	128·2
1988	9 864	136·3
1989	11 544	159·5

Interpretacija je analogna prejšnji: indeks 136·3 pove, da je bila proizvodnja v letu 1988 za 36·3% večja od tiste v letu 1984. Do

enakega rezultata bi prišli tudi z neposrednim računanjem,

$$I_{88/84} = 100 \cdot \frac{Y_{88}}{Y_{84}} = 100 \cdot \frac{9\,864}{7\,235} = 136{\cdot}3$$

Pri posameznih letih (npr. 1989) lahko pride do malenkostne razlike, ki je posledica tega, da so prvotni indeksi že bili zaokroženi na eno decimalno mesto.

Od indeksov s premično osnovo so najpomembnejši **verižni indeksi**, pri katerih primerjamo vrednosti pojava v dveh zaporednih obdobjih, konkretno, za vrednost pojava v izbranem obdobju izračunamo, koliko odstotkov od vrednosti pojava v preteklem obdobju predstavlja. Iz vrste podatkov $Y_0, Y_1, Y_2, \ldots, Y_n$ torej računamo verižne indekse J_k po obrazcu

$$J_k = 100 \cdot \frac{Y_k}{Y_{k-1}} \qquad (k = 1, 2, \ldots, n) \tag{3}$$

Za prvo obdobje v časovni vrsti seveda verižnega indeksa ne moremo izračunati.

ZGLED 3.5.3:

Za časovno vrsto podatkov o proizvodnji v letih 1980 - 1989 iz prejšnjih zgledov izračunajmo še verižne indekse!

Izračun poteka neposredno po obrazcu, rezultati so v tabeli:

LETO	PROIZVODNJA	INDEKS
1980	5 188	–
1981	5 322	102·6
1982	5 249	98·6
1983	5 150	98·1
1984	7 235	140·5
1985	7 942	109·8
1986	8 904	112·1
1987	9 276	104·2
1988	9 864	106·3
1989	11 544	117·0

V primeru verižnih indeksov vrednost nad 100 pomeni povečanje, vrednost pod 100 pa zmanjšanje vrednosti pojava v primerjavi s preteklim obdobjem, vsakič za ustrezno število odstotkov (kolikor pač štrli preko 100 oziroma kolikor manjka do te vrednosti). Proizvodnja v letu 1982 je bila za 100 – 98·6 = 1·4% manjša od one v letu 1981, proizvodnja v letu 1981 pa je za 102·6 – 100 = 2·6% prekašala tisto v začetnem letu časovne vrste.

Kakšna je zveza med verižnimi in baznimi indeksi? Denimo, da imamo spet vrsto podatkov $Y_0, Y_1, Y_2, \ldots, Y_n$ in zapišimo bazni indeks I_k, nato pa ulomek ustrezno razširimo:

$$I_k = 100 \cdot \frac{Y_k}{Y_0} = 100 \cdot \frac{Y_k}{Y_{k-1}} \cdot \frac{Y_{k-1}}{Y_{k-2}} \cdot \frac{Y_{k-2}}{Y_{k-3}} \cdot \ldots \cdot \frac{Y_1}{Y_0}$$

V posameznih ulomkih na desni prepoznamo stotinke verižnih indeksov J_k, kar pomeni, da bazni indeks I_k dobimo tako, da zmnožimo stotinke vseh verižnih indeksov do vključno k – tega, produkt pa še pomnožimo s 100:

$$I_k = 100 \cdot \frac{J_k}{100} \cdot \frac{J_{k-1}}{100} \cdot \frac{J_{k-2}}{100} \cdot \ldots \cdot \frac{J_1}{100} \tag{4}$$

Če iz definicije verižnega indeksa izpustimo od procentnega računa podedovani faktor 100, dobimo **koeficient dinamike**, $K_i = Y_i/Y_{i-1}$. Obrazec (4) lahko potemtakem interpretiramo tudi tako, da je treba za izračun baznega indeksa zmnožiti vse koeficiente dinamike in dobljeni produkt dodatno še pomnožiti s 100. (Pri tem pazimo na natančnost računanja: zaokrožanju indeksov na eno decimalno mesto ustreza zaokrožanje koeficientov dinamike na tri decimalna mesta!)

Če definicijo koeficienta dinamike zapored uporabimo na vseh členih časovne vrste, vidimo, da je

$$Y_N = Y_0 \cdot K_1 \cdot K_2 \cdot \ldots \cdot K_N$$

Ta obrazec pravzaprav pokaže, kako vsako novo obdobje doda določeno rast (lahko tudi upadanje, če je $K_i < 1$!) na kumulativo vsega, kar se je dogajalo s pojavom v preteklih obdobjih. Kaj naj bi v tej luči pomenil **povprečni koeficient dinamike** $\overline{K}$? Gotovo takšen konstanten koeficient dinamike, da bi iz dane začetne vrednosti Y_0 z njegovo N – kratno uporabo prišli do enake končne vrednosti Y_N. To pa pomeni, da mora biti

$$Y_0 \cdot K_1 \cdot K_2 \cdot \ldots \cdot K_N = Y_0 \cdot \overline{K} \cdot \overline{K} \cdot \ldots \cdot \overline{K}$$

oziroma

$$\overline{K} = \sqrt[N]{K_1 \cdot K_2 \cdot \ldots \cdot K_N} \tag{5}$$

ali še drugače, **povprečni koeficient dinamike je geometrijska sredina posameznih koeficientov dinamike** za proučevano obdobje.

ZGLED 3.5.4:

Izračunajmo povprečni koeficient dinamike za podatke o rasti proizvodnje, ki smo jih proučevali v prejšnjem razdelku!

V zgledu 3.5.3 smo izračunali verižne indekse, ki jih takoj predelamo v koeficiente dinamike in imamo

LETO	PROIZVODNJA	K_i
1980	5 188	–
1981	5 322	1·026
1982	5 249	0·986
1983	5 150	0·981
1984	7 235	1·405
1985	7 942	1·098
1986	8 904	1·121
1987	9 276	1·042
1988	9 864	1·063
1989	11 544	1·170

Produkt vseh devetih koeficientov dinamike je 2·2241517... in deveti koren je (zaokroženo) 1·093, toliko znaša povprečni koeficient dinamike. Rezultat lahko interpretiramo še takole: proizvodnja bi morala (v povprečju) naraščati po 9·3% letno, da bi se od 5 188 enot v letu 1980 povečala na 11 544 enot v letu 1989.

Bralec, ki se bo sam lotil reševanja te naloge, bo ugotovil, da so zaokrožitvene napake relativno pomembne; če začetni obseg proizvodnje 5 188 pomnožimo z deveto potenco števila 1·093, dobimo 11 549·99, kar je približno pol promila več, kot je dejanski obseg v letu 1989.

Več o indeksih najdemo v statistični literaturi, zelo poučne pa so lahko tudi računalniške vaje s programi za obdelavo preglednic, kjer si lahko zelo enostavno definiramo obrazce za računanje indeksov in drugih relativnih števil, tako da kasneje vnašamo samo podatke iz časovne vrste, ostali postopki pa so avtomatizirani. Ustrezni programski paketi imajo tudi kopico grafičnih možnosti, s katerimi si lahko še bolj nazorno predočimo dinamiko proučevanega pojava.

V A J E

1. Kaj ti pravzaprav pomeni izraz "relativna števila"? Kaj so indeksi? Katere vrste indeksov poznaš? Katere vrste časovnih indeksov poznaš?
2. Za leta 1985 - 1991 imamo naslednje podatke o številu prijavljenih prometnih nesreč na nekem območju: 134, 152, 187, 171, 192, 198 in 182. Izračunaj
   a) bazne indekse (bazno leto 1985);
   b) bazne indekse (bazno leto 1991);

c) verižne indekse
in interpretiraj rezultate, še posebej pa odgovori na naslednja vprašanja:
- Kakšno je bilo na splošno stanje glede na leto 1985?
- V katerem letu je bila rast števila nesreč najhitrejša?
- Za koliko odstotkov je bilo število nesreč v letu 1991 večje od tistega v letu 1988?

3. Za podatke iz prejšnje naloge izračunaj, za koliko odstotkov se je v proučevanem obdobju povprečno vsako leto povečalo število prometnih nesreč.

4. Imamo naslednje podatke o prodaji v letih 1980 - 1993 in bazne indekse z letom 1980 kot osnovo:

LETO	PRODAJA	INDEKS
1980	**245**	**100'0**
1981	354	144'5
1982	453	184'9
1983	378	154'3
1984	424	173'1
1985	488	
1986	499	203'7
1987	512	209'0
1988		194'3
1989	523	213'5
1990		226'5
1991	511	208'6
1992	572	
1993	613	250'2

a) Izpolni praznine v tabeli.
b) Izračunaj tudi bazne indekse z osnovo v letu 1986.
c) Izračunaj verižne indekse.
č) Izračunaj povprečni koeficient dinamike.

5. Imamo naslednje podatke o nabavah materiala v letih 1980 - 1993 in pripadajoče verižne indekse:

LETO	NABAVA	VER. IND.
1980	160	–
1981	224	140'0
1982	255	113'8
1983	278	109'0
1984	312	112'2
1985	378	
1986	409	108'2
1987	512	125'2
1988		93'0
1989	523	109'9
1990		96'4
1991	511	101'4
1992	572	
1993	613	108'4

a) Izpolni praznine v tabeli.
b) Izračunaj bazne indekse z osnovo v letu 1980.
c) Izračunaj bazne indekse z osnovo v letu 1986.
č) Izračunaj povprečni koeficient dinamike.

6. Nekaj časa spremljaj gibanje borznih tečajev v dnevnem časopisju. Izberi si delnico kakega podjetja, grafično predoči spreminjanje njenega tečaja, izračunaj bazne in verižne indekse in povprečni koeficient dinamike za - denimo - štirinajstdnevno obdobje.

PREGLED POMEMBNEJŠIH OBRAZCEV

Kumulativna porazdelitev frekvenc

$$F_1 = 0\,, \quad F_k = F_{k-1} + f_{k-1}$$

Relativna frekvenca in kumulativna porazdelitev

$$f_k^o = \frac{f_k}{N}\,, \quad F_1^o = 0\,, \quad F_k^o = F_{k-1}^o + f_{k-1}^o$$

Navadna in tehtana aritmetična sredina

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i\,, \quad \mu = \frac{f_1 \cdot y_1 + \ldots + f_r \cdot y_r}{f_1 + \ldots + f_r} = \frac{1}{N}\sum_{i=1}^{r} f_i\, y_i$$

Geometrijska in harmonična sredina

$$G = \sqrt[N]{y_1 \cdot y_2 \cdot \ldots \cdot y_N}\,, \quad H = \{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{y_i}\}^{-1}$$

Varianca (negrupirani podatki)

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mu)^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \mu^2$$

Varianca (grupirani podatki)

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{r} f_i\,(y_i - \mu)^2 = \frac{1}{N}\sum_{i=1}^{r} f_i\, y_i^2 - \mu^2$$

Standardni odklon, variacijski razmik, koeficient variacije

$$\sigma = \sqrt{\sigma^2}\,, \qquad R_v = y_{max} - y_{min}\,, \qquad K_v = \frac{\sigma}{\mu}$$

Bazni in verižni indeksi

$$I_{k/0} = 100 \cdot \frac{Y_k}{Y_0}\,, \quad J_k = 100 \cdot \frac{Y_k}{Y_{k-1}} \qquad (k = 1, 2, \ldots, n)$$

Koeficient dinamike in povprečni koeficient dinamike

$$K_i = Y_i\,/\,Y_{i-1}\,, \qquad \overline{K} = \sqrt[N]{K_1 \cdot K_2 \cdot \ldots \cdot K_N} = \sqrt[N]{Y_n\,/\,Y_1}$$

4. VERJETNOSTNI RAČUN

Verjetnostni račun je v nekaj sto letih prehodil zanimivo pot od mističnih začetkov in "črne magije" do korektno postavljene matematične discipline in enega najlepših, pa tudi najbolj uporabnih matematičnih orodij. Verjetnostni modeli so postali prevladujoča govorica v različnih tehničnih vedah, enako velja za področja izven tehnike. Tako tudi sodobna ekonomija in management ne moreta več brez verjetnostnega računa; to ni le nepogrešljivi element, brez katerega je statistika samo pripovedovanje zgodbic, ampak se pojavlja tudi kot povsem samostojno orodje za analizo poslovnih tveganj in možnosti za njihovo zmanjševanje.

Kratek srednješolski učbenik sicer ne more pričarati vse lepote tega dela matematike, vendar lahko odstre zaveso, s pregledom najpomembnejših osnovnih gradnikov pa pripravi bralcu nekoliko lažjo pot naprej, k resnejši literaturi.

4.1 Poskusi in dogodki

Trije osnovni pojmi, ki jih bomo srečevali ves čas našega ukvarjanja z verjetnostnim računom, so
- **poskus**,
- **dogodek** in
- **verjetnost dogodka**.

Ker je tako pomembno, da jasno razumemo vloge teh treh gradnikov, jim bomo namenili več pozornosti, kot bi se morda zdelo potrebno bralcu, ki se prvič srečuje z verjetnostnim računom. Prosimo za potrpljenje: kasneje se vedno izkaže, da je bilo uvodnega sitnarjenja prej premalo kot preveč.

V verjetnostnem računu bomo z izrazom **poskus** označevali vsako hoteno dejanje, ki ga opravimo v natanko določenih pogojih, na primer:
- Po mizi zakotalimo običajno igralno kocko.
- Iz običajnega kompleta 32 kart izberemo eno karto.
- Kupimo eno od stotisoč loterijskih srečk.
- S puško ustrelimo proti tarči.

Vsak poskus torej pomeni realizacijo neke množice skupaj nastopajočih dejstev ali *kompleksa pogojev*. Bolj po domače: predpostavljamo, da se poskus vedno odvija pod enakimi, natančno določenimi pogoji.

Pojav, ki v to množico pogojev ne spada in se lahko v posameznem poskusu zgodi ali pa tudi ne, imenujemo **dogodek**. V omenjenih poskusih so možni na primer naslednji dogodki:

• Kocka se ustavi tako, da je na zgornji strani ploskev s šestimi pikami ('pade šestica'). Pade sodo število pik. Pade liho število pik. Padejo več kot štiri pike.
• Izvlečena karta je as. Izvlečena karta je rdeče barve. Izvlečena karta je pik. Izvlečena karta je figura.
• Kupljena srečka zadene glavni dobitek. Kupljena srečka zadene neki manjši dobitek. Kupljena srečka ne zadene ničesar.
• Tarča je zadeta. Tarča ni zadeta. Tarča je zadeta 1 cm od geometrijskega središča.

Dogodke označujemo z velikimi tiskanimi črkami z začetka abecede, npr. $A, B, C, \ldots$, če bo potrebno, jih tudi oštevilčimo, npr. $E_1, E_2, \ldots, E_n$.

Naj opozorimo, da je proučevanje nekega dogodka neločljivo povezano s poskusom, v okviru katerega se zanje zanimamo. Brž ko se spremeni samo eden od pogojev, je to že drug poskus.

Pri ponavljanju določenega poskusa obstajajo za neki dogodek iz tega poskusa (recimo mu dogodek A) v bistvu tri različne možnosti:

a) Vzemimo, da smo poskus velikokrat ponovili, dogodek A pa se je zgodil pri vsaki ponovitvi poskusa, zato na osnovi teh izkušenj sklepamo, da se bo zgodil v katerikoli nadaljnji ponovitvi tega poskusa. Tedaj rečemo, da je dogodek A v proučevanem poskusu **gotov dogodek**.

ZGLED 4.1.1:

Včasih je izkušnja edina osnova za trditev, da imamo opraviti z gotovim dogodkom (na primer: vsakič, ko smo vrgli kamen v zrak, je padel nazaj na zemljo, zato sklepamo, da je pri takem poskusu 'vrnitev kamna na zemljo' gotov dogodek, v drugih primerih lahko to logično utemeljimo (dogodek, da bo pri metu poštene igralne kocke padlo manj kot sedem pik, je v tem poskusu gotov dogodek, ker so pač ploskve normalne igralne kocke označene z eno, dvema, ..., šestimi pikami.

b) Denimo, da se dogodek ne zgodi v nobeni od zelo velikega števila ponovitev nekega poskusa. Potem rečemo, da je dogodek v tem poskusu **nemogoč**, ker sklepamo, da se ne more zgoditi tudi v nobeni od ponovitev poskusa, ki naj bi jih še naredili.

ZGLED 4.1.2:

Dogodek, da bi kdo s prostim očesom videl posamezne atome, je nemogoč. Enako velja za dogodek, da potegnemo belo kroglico iz posode, v kateri so same črne kroglice.

c) Tretja možnost je najpogostejša in za verjetnostni račun najzanimivejša. Vzemimo, da se dogodek A v nekaterih ponovitvah poskusa zgodi, v nekaterih pa ne, pri čemer pred posameznim poskusom ne moremo z gotovostjo napovedati, ali se bo zgodil ali ne. Če nismo pretirano natančni, lahko v takem primeru rečemo, da gre za **slučajen dogodek**.

ZGLED 4.1.3:

Slučajen je dogodek, da vržemo s kocko sodo število pik. Slučajen je dogodek, da naletimo pri slepem izbiranju ene Slovenke na tako, ki ima modre oči. Slučajen je dogodek, da pri slepem izbiranju točke v koordinatnem sistemu zadenemo ravno koordinatno izhodišče. Slučajen je dogodek, da na loteriji z eno samo srečko zadenemo glavni dobitek.

Slučajnost dogodka je odvisna od tega, v katerem poskusu ga opazujemo. Če na primer pokupimo vse srečke neke loterije, dogodek, da zadenemo glavni dobitek, ni več slučajen, ampak gotov.

Oglejmo si zdaj nekaj posebnih odnosov med dogodki in operacije, ki nam iz dveh ali več danih dogodkov dajo nove dogodke. Za ilustracijo bomo v večini primerov porabili kar enega najpreprostejših poskusov, ki ga vsi dobro poznamo, to je 'met igralne kocke' in dogodke, ki pri tem poskusu lahko nastopijo; dostikrat se bomo srečali tudi z drugo obliko hazarda, to je s kartami. Zaradi enostavnosti bomo ves čas delali s kompletom 32 kart, ki ima glede na barve po osem src, kar, pikov in križev, po vrednosti pa po štiri sedmice, osmice, devetice, desetice, fante, dame, kralje in ase. (Od tu dalje bomo – kot na začetku poglavja – takemu kompletu rekli "običajni komplet kart", ne da bi vsakič posebej pojasnjevali, kaj s tem mislimo.)

Zveza med igrami na srečo in verjetnostnim računom ima globoke korenine, ki dokazano segajo vsaj v sredino 17. stoletja, ko sta se dva vrhunska francoska matematika, nam že znani Blaise Pascal in Pierre Fermat, intenzivno ukvarjala s svetovalno dejavnostjo na področju kockanja. Šlo je za naslednji konkretni problem: Kaj je bolj verjetno, da pri štirih metih poštene igralne kocke vsaj enkrat pade šest pik, ali dogodek, da pri 24 metih dveh kock vsaj enkrat na obeh kockah pade šestica?

Ljudsko izročilo pa omenja kot enega najstarejših problemov s področja verjetnostnega računa vprašanje, kako je treba korektno razdeliti obljubljeno nagrado (ali denar od stav), če je igra predčasno prekinjena. Nalogo še danes srečujemo v učbenikih verjetnostnega računa, dostikrat nekoliko preoblečeno, na primer takole: Dva enako kvalitetna šahista, ki v medsebojnem dvoboju štejeta samo zmage, igrata na šest zmag. Kolikšna je verjetnost, da bo zmagal tisti, ki trenutno vodi s 4:2?

Tako lahko mirno rečemo, da so imele igre na srečo v zgodovini poleg polnjenja davčnih blagajn še eno dobro lastnost, namreč vpliv na razvoj verjetnostnega računa...

Zdaj pa k nekoliko dolgoveznemu spletu definicij.

Definicija: Dogodek A je **način** dogodka B (v znakih $A \subset B$), če se vsakič, ko nastopi dogodek A, zagotovo zgodi tudi dogodek B.

ZGLED 4.1.4:

a) Pri metu kocke je dogodek A, da pade šestica, način dogodka B, da pade sodo število pik.

b) Pri izboru ene karte iz običajnega kompleta je dogodek, da potegnemo srce, način dogodka, da je bila izbrana karta rdeče barve.

Definicija: Dogodka A in B sta **enaka**, če sočasno velja, da je dogodek A način dogodka B in dogodek B način dogodka A,

$$A = B \iff A \subset B \wedge B \subset A$$

V skladu z zapisano definicijo enakosti so med seboj enaki vsi gotovi dogodki danega poskusa, zato jim damo enotno oznako G. Podobno velja za vse nemogoče dogodke, označili jih bomo z N.

Definicija: Če se v nekem poskusu zgodi vsaj eden od dogodkov A ali B, rečemo, da se je zgodila **vsota** ali **unija dogodkov** A in B. Ta dogodek bomo označili z $A \cup B$.

ZGLED 4.1.5:

a) Pri metu kocke naj bo dogodek A, da padejo več kot 4 pike, dogodek B pa padec sodega števila pik. Vsota $A \cup B$ nastopi natanko takrat, ko pade bodisi 5 ali 6 pik (dogodek A) ali če padejo 2, 4 ali 6 pik (dogodek B).

b) Iz množice $\mathbb{N}_{90} = \{1, 2, \ldots, 89, 90\}$ na slepo izberemo eno od števil. Naj bo dogodek A, da je izbrano število deljivo s 3, B pa dogodek, da je izbrano število sodo. Vsota dogodkov nastopi, kakor hitro dobimo število, ki se ponaša vsaj z eno od obeh lastnosti, ali drugače, vsota ne nastopi v primeru, ko izbrano število ni deljivo niti z 2 niti s 3.

Definicija: Produkt dogodkov (tudi presek dogodkov) A in B se zgodi natanko takrat, kadar v nekem poskusu hkrati nastopita dogodka A in B.

Izraz "hkrati" tu ne pomeni nujno popolne sočasnosti, ampak dejstvo, da oba dogodka nastopita pri isti ponovitvi poskusa. Po analogiji z vsoto dogodkov njun produkt pogosto označujemo z $A \cap B$, kadar pomote niso možne, pa se odločimo za enostavnejšo pisavo AB.

ZGLED 4.1.6:

Imejmo poskus in dogodka A in B iz zgleda 4.1.5 (a).

a) Če naj se zgodi produkt AB, morata nastopiti A in B hkrati, kar se očitno zgodi natanko takrat, ko pade šest pik.

b) Produkt se zgodi, ko je število sočasno deljivo z dva in s tri, torej natanko takrat, ko je deljivo s šest. V dani množici so taka števila 6, 12, 18, 24, 30, ..., 84 in 90.

Iz zadnjih dveh zgledov se vidi, da je produkt dveh dogodkov način njune vsote, kar očitno velja čisto splošno. Še več, produkt kateregakoli para dogodkov je način vsakega od njiju:

$$A \cap B \subset A, \quad A \cap B \subset B$$

Definiciji vsote in produkta dogodkov lahko razširimo z dveh dogodkov na večje število le-teh: vsota poljubnega števila dogodkov se zgodi, brž ko nastopi vsaj eden od njih, produkt teh dogodkov pa nastopi edino v primeru, če se zgodijo vsi hkrati.

Definicija: V nekem poskusu nastopi **negacija** dogodka A, če ne nastopi dogodek A. Negacijo dogodka A označujemo s simbolom $\overline{A}$, še rajši z A', po domače pa ji rečemo kar **nasprotni dogodek** dogodka A.

ZGLED 4.1.7:

Za dogodek A, da padejo pri metu kocke več kot 4 pike, je negacija dogodek, da ne padejo več kot 4 pike, z drugimi besedami, da padejo kvečjemu štiri pike.

POZOR: Relativno veliko je ljudi, ki imajo težave z opisi "največ", "kvečjemu", "ne več kot..." in podobnimi. To je lahko pri nalogah iz verjetnostnega računa začetek napačne poti.

Definicija: Dogodka A in B sta v danem poskusu **združljiva**, če lahko nastopita oba hkrati. Če se to ne more zgoditi, rečemo, da sta **nezdružljiva**.

Produkt nezdružljivih dogodkov A in B je očitno nemogoč dogodek, $AB = N$.

ZGLED 4.1.8:

Pri metu kocke sta dogodka A, da pade šestica, in B, da pade liho število pik, nezdružljiva, ker ne moreta nastopiti oba hkrati. Dogodka C, da pade sodo število pik, in D, da padejo manj kot tri pike, pa lahko nastopita v isti ponovitvi poskusa (kadar padeta dve piki), zato sta združljiva.

Poljuben dogodek in njegova negacija sta vedno nezdružljiva, poleg tega pa se v vsaki ponovitvi poskusa eden od njiju zagotovo zgodi, zaradi česar je njuna vsota gotov dogodek. Če se enkrat za vselej dogovorimo, da bomo pri nezdružljivih dogodkih vsoto dogodkov označevali kar z običajnim znakom za seštevanje (+), lahko pravkar povedano na kratko zapišemo takole:

$$A \cap A' = N \quad \text{in} \quad A + A' = G$$

Definicija: Če lahko dogodek A izrazimo kot vsoto vsaj dveh med seboj nezdružljivih, sicer pa mogočih dogodkov, na primer $A = B + C$, pravimo, da je A **sestavljen dogodek**. Dogodku, ki ni sestavljen, rečemo **elementaren dogodek** ali **izid**.

ZGLED 4.1.9:

Pri metu kocke imamo šest elementarnih dogodkov, ki po vrsti pomenijo padec ene, dveh, ..., šestih pik. Sestavljeni dogodek pa je na primer padec sodega števila pik.

ZGLED 4.1.10:

Iz množice črk slovenske abecede na slepo izberemo eno črko. Slučajni dogodek, da izberemo črko Ž, je elementaren, slučajni dogodek, da je izbran samoglasnik, pa je sestavljen iz petih izidov.

Definicija: Za izide, iz katerih je neki dogodek sestavljen, rečemo, da so zanj **ugodni**, ali še drugače, da so to **načini**, na katere se ta dogodek lahko zgodi.

ZGLED 4.1.11:

Sočasno vržemo rdečo in črno igralno kocko. Dogodek A, da je vsota obeh pik na kockah enaka 7, se lahko zgodi na naslednje načine:

1 pika na rdeči kocki in 6 pik na črni;

2 piki na rdeči kocki in 5 pik na črni;

3 pike na rdeči kocki in 4 pike na črni;

4 pike na rdeči kocki in 3 pike na črni;

5 pik na rdeči kocki in 2 piki na črni;

6 pik na rdeči kocki in 1 pika na črni.

Za dogodek A je tedaj ugodnih 6 izidov od 36 možnih. Da je vseh možnih izidov res 36, pove osnovni izrek kombinatorike: na rdeči kocki je 6 možnih izidov, neodvisno od tega je na črni kocki 6 možnih izidov, skupaj torej toliko, kot znaša produkt obeh števil, namreč 36.

Z dosedanjimi definicijami smo povedali, kako računamo v množici vseh dogodkov, ki jih lahko pripišemo določenemu poskusu. Sedaj postavimo povedano v nekoliko splošnejši okvir.

Imejmo poljuben poskus in naj bodo $E_1, E_2, ..., E_n$ vsi možni elementarni dogodki tega poskusa. Vsakemu dogodku A iz tega poskusa priredimo natanko določeno podmnožico množice $\{E_1, E_2, ..., E_n\}$, ki jo sestavljajo elementarni dogodki, ugodni za nastop dogodka A, ali drugače rečeno, iz katerih je le-ta sestavljen. Če je A eden od možnih elementarnih dogodkov

E_i, ima ta podmnožica pač en sam element (namreč ta izid), pri sestavljenih dogodkih pa premore dva ali več elementov. Vseh možnih dogodkov (elementarnih in sestavljenih) v danem poskusu je zato natanko toliko, kolikor različnih podmnožic lahko naredimo iz množice elementarnih dogodkov. Če je teh n, je potem vseh možnih dogodkov natanko 2^n, tolikšna je namreč moč potenčne množice, če ima osnovna množica moč n. (Rezultat poznamo iz kombinatorike!)

V verjetnostnem računu bomo v skladu z njeno vlogo dali tej potenčni množici posebno ime, rekli ji bomo **algebra dogodkov**. Med njenimi elementi sta posebej odlikovana nemogoči dogodek (ker zanj ni ugoden nobeden od dogodkov E_i, lahko nemogoči dogodek enačimo s prazno podmnožico množice elementarnih dogodkov) ter gotovi dogodek, za katerega so ugodni prav vsi elementarni dogodki E_i, zaradi česar ga lahko zapišemo kot njihovo unijo,

$$G = E_1 \cup E_2 \cup ... \cup E_n = \bigcup_{i=1}^{n} E_i$$

oziroma v obliki

$$G = \{ E_1, E_2, ..., E_n \}$$

če ostanemo pri dogovorjenem enačenju sestavljenega dogodka z množico izidov, ki so zanj ugodni.

Uporabljali bomo oba zapisa, odvisno od tega, kateri nam bo prišel bolj prav. Algebra dogodkov je v skladu z drugim zapisom potenčna množica množice G, zato jo bomo označili s $\mathcal{P}\,G$.

ZGLED 4.1.12:

Pri metu kocke imamo šest elementarnih dogodkov,

$G = \{E_1, E_2, ..., E_5, E_6\}$,

zato ima ustrezna algebra dogodkov $\mathcal{P}\,G$ natanko 64 $(= 2^6)$ dogodkov, pri čemer sta seveda všteta nemogoči dogodek ter gotovi dogodek.

ZGLED 4.1.13:

Trikrat vržemo kovanec. Opiši in preštej elementarne dogodke ter izračunaj moč algebre dogodkov!

Denimo, da poskus dejansko opravimo in beležimo, kaj se zgodi pri posameznem metu: če pade številka, zapišemo Š, če pade grb ("mož"), zapišemo M. Vseh možnih izidov je potemtakem toliko, kolikor je nizov dolžine 3, ki imajo na vsakem od treh mest bodisi Š ali M. Po osnovnem izreku kombinatorike (ali kar po obrazcu za število variacij reda 3 z dovoljenim ponavljanjem iz 2 elementov) je takih nizov $2 \cdot 2 \cdot 2 = 8$. Brez težav jih tudi naštejemo, urejene po abecedi:

MMM, MMŠ, MŠM, MŠŠ, ŠMM, ŠMŠ, ŠŠM, ŠŠŠ.

Pri preštevanju izidov je v takšnih primerih treba paziti. Dogodek "v treh metih pade enkrat grb, dvakrat pa številka" na primer ni elementaren, kot bi kdo pomislil na prvi pogled, sestavljen je iz treh elementarnih dogodkov MŠŠ, ŠMŠ, ŠŠM. Ker ima množica izidov moč $n = 8$, ima algebra dogodkov v našem primeru moč $2^8 = 256$.

Z uporabo enačenja dogodkov z množicami zanje ugodnih izidov dosežemo, da lahko vsa računanja z dogodki, ki smo jih spoznali v začetku tega razdelka, prevedemo na operacije med ustreznimi množicami, iz teorije množic pa si potem izposodimo nekatera znana računska pravila. Vsoti dogodkov ustreza unija pripadajočih množic, produktu njun presek, negaciji dogodka komplementarna množica, relacija "A je način dogodka B" pa pove, da je množica za dogodek A ugodnih izidov podmnožica v množici izidov, ki so ugodni za dogodek B. Kako je s sorodstvenimi vezmi nemogoč dogodek - prazna množica in gotov dogodek - univerzalna množica, tudi že vemo, zato ne bo težko premisliti o pravilnosti naslednjih trditev:

$$A \cup N = A \qquad A \cap N = N \qquad (1)$$
$$A \cup G = G \qquad A \cap G = A \qquad (2)$$
$$A \cup B = B \cup A \qquad A \cap B = B \cap A \qquad (3)$$
$$A \cup (B \cup C) = (A \cup B) \cup C \qquad A \cap (B \cap C) = (A \cap B) \cap C \qquad (4)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \qquad (A \cup B) \cap C = (A \cap C) \cup (B \cap C) \qquad (5)$$
$$(A \cup B)' = A' \cap B' \qquad (A \cap B)' = A' \cup B' \qquad (6)$$
$$(A')' = A \qquad (7)$$
$$N' = G \qquad G' = N \qquad (8)$$

Za vajo je koristno te obrazce premisliti na dva načina, enkrat v jeziku algebre dogodkov, drugič v najbrž bolj domačem jeziku algebre množic.

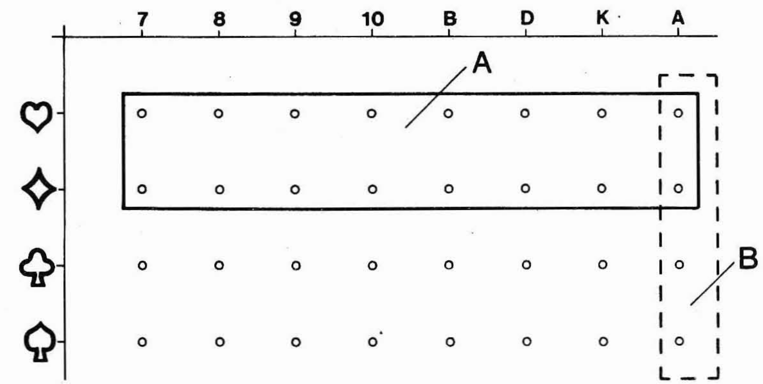
Iz teorije množic povzemimo še koristno navado, da problem predstavimo tudi grafično. Če vsakemu elementarnemu dogodku E_i priredimo točko v ravnini učbenika (zvezka, table,...), dobimo t.i. **vzorčni prostor**. V njem tako kot v običajnih Vennovih diagramih s sklenjenimi krivuljami omejimo množice točk, ki ustrezajo posameznim sestavljenim dogodkom, da lahko nato nazorno prikažemo operacije med njimi.

ZGLED 4.1.14:

Iz običajnega kompleta kart izvlečemo eno karto. Prikaži vzorčni prostor za ta poskus in v njem ponazori dogodka A - izvlečena karta je rdeče barve (srce ali karo); B - izvlečena karta je as!

Vzorčni prostor je prikazan na sliki 4.1. Ob njem lahko spet premisliš vsaj nekatere od obrazcev (1) - (8).

Zbirki novih pojmov dodajmo še enega.



Slika 4.1: Vzorčni prostor za poskus iz zgleda 4.1.14

Definicija: Množico poljubnih dogodkov

$$S = \{A_1, A_2, \ldots, A_n\}$$

imenujemo **popoln sistem dogodkov**, če se v vsaki ponovitvi poskusa zgodi natanko eden od njih.

Definicija pove, da so dogodki iz množice S med seboj paroma nezdružljivi, njihova vsota pa je gotov dogodek,

$$A_i \cap A_j = N \quad (i \neq j), \quad A_1 \cup A_2 \cup \ldots \cup A_n = G$$

Pri tem so dogodki iz popolnega sistema lahko elementarni ali sestavljeni (ali pa je celo vsakih nekaj), važno je le to, da zadoščajo obema pogojema.

ZGLED 4.1.15:

Popoln sistem dogodkov za poskus "met igralne kocke" lahko sestavimo na različne načine, na primer a) iz že večkrat omenjenih elementarnih dogodkov E_i;

b) iz dogodkov "padec sodega števila pik" in "padec lihega števila pik";

c) iz dogodkov "padejo kvečjemu tri pike" in "padejo vsaj štiri pike";

č) iz treh dogodkov E_1, E_2 in dogodka "padejo vsaj tri pike";

ali še kako drugače.

Iz razlogov, ki bodo postali jasni v naslednjem razdelku, v večini primerov najraje delamo s popolnim sistemom, ki ga sestavljajo vsi elementarni dogodki danega poskusa.

△△ Ko smo začeli razmišljanje o algebrah dogodkov in popolnih sistemih, smo potihem privzeli, da je v danem poskusu možno samo končno število izidov. To seveda v praksi sploh ni nujno. Če mečemo kocko toliko časa, da prvič pade šestica in je E_i zdaj dogodek, da se to zgodi v natanko $i-$tem poskusu, popoln sistem dogodkov ni končen, ampak števno neskončen (ima enako moč kot množica vseh naravnih števil), saj ni nujno, da bi se šestica sploh pojavila v še tako velikem končnem številu metov. Pri slepem izbiranju točke z danega intervala na realni osi pa je elementarnih dogodkov očitno toliko, kolikor točk premore ta interval, zato popolni sistem dogodkov ni niti števen, ampak neštevno neskončen. V tem učbeniku se bomo največ ukvarjali s poskusi, pri katerih je popolni sistem dogodkov končen. Kadar bomo imeli opraviti z bogatejšimi popolnimi sistemi, bomo na to posebej opozorili in po potrebi popravili ustrezne definicije.

V A J E

1. Naj bo
- A - dogodek, da vržemo s kocko 6 pik;
- B - dogodek, da vržemo liho število pik;
- C - dogodek, da vržemo 3 pike;
- D - dogodek, da vržemo sodo število pik;
- E - dogodek, da vržemo manj kot 5 pik.

Med katerimi dogodki velja relacija: "je način"?

2. Naj bo dan poljuben dogodek A iz nekega poskusa. Kaj meniš o $A \cap A$ in $A \cup A$?

3. Na izpitu so možne ocene od 1 do 10, najnižja pozitivna ocena je 6. Naj bodo dani dogodki:
- A - študent uspešno opravi izpit;
- B - študent ne opravi izpita;
- C - študent dobi pri izpitu oceno 6;
- D - študent dobi pri izpitu vsaj 6;
- E - študent dobi pri izpitu kvečjemu 5;
- F - študent dobi pri izpitu kvečjemu 6;
- G - študent dobi pri izpitu več kot 5;
- H - študent dobi pri izpitu več kot 4.

Z besedami opiši nekaj odnosov med posameznimi dogodki.

4. Naj bo poskus met igralne kocke in naj bodo A, B, C, D in E dogodki, definirani v 1. vaji. Z besedami opiši njihove negacije oziroma njim nasprotne dogodke.

5. Naj bodo A_1, A_2, A_3 dogodki iz poljubnega poskusa. Z njimi izrazi naslednje dogodke:

- B - zgodi se A_1, ne pa tudi ostala dva dogodka;
- C - v poskusu nastopijo vsi trije dogodki A_1, A_2, A_3;
- D - nastopi natanko eden od teh treh dogodkov;
- E - nastopi vsaj eden od teh treh dogodkov;
- F - nastopi kvečjemu eden od teh treh dogodkov.

Pripravi si ustrezno grafično ponazoritev (vzorčni prostor) in pokaži, kaj ustreza dogodkom B, C, D, E in F.

6. Sistem ima dva elementa tipa A in tri elemente tipa B, za delovanje mora biti brezhiben vsaj eden od elementov prve vrste (dogodka, da je element brezhiben, označimo z A_1, A_2) in vsaj dva elementa druge vrste B_1, B_2, B_3. Z navedenimi dogodki izrazi dogodek C, da je sistem sposoben za delovanje.

7. Strelec strelja proti cilju štirikrat, pri tem beležimo zadetke in zgrešene strele (oznaka 1 oziroma 0). Zapiši množico vseh možnih izidov in predstavi naslednje dogodke:
- A - da je strelec prvič zgrešil;
- B - da so bili vsi štirje streli enako uspešni;
- C - da je bil cilj zadet natanko dvakrat;
- D - da je bil cilj zadet vsaj dvakrat.

8. V posodi imamo štiri krogle, oštevilčene od 1 do 4. Iz posode potegnemo kroglo, jo vrnemo in ponovno izvlečemo eno kroglo. Zapiši množico vseh elementarnih dogodkov tega poskusa, ugotovi njeno moč in moč pripadajoče algebre dogodkov, nato pa z izidi izrazi dogodka
- A - obe krogli nosita lihi števili;
- B - vsota števil na kroglah je deljiva s 3.

9. Trikrat vržemo kovanec. Zapiši elementarne dogodke ter izrazi z njimi naslednje sestavljene dogodke:
- A - grb pade vsaj dvakrat;
- B - vsaj enkrat pade številka;
- C - v tretjem metu pade številka;
- D - kovanec pade vsaj dvakrat zapored na isto stran.

10. Naj bo A dogodek, da pri izbiranju ene karte iz običajnega kompleta dobimo asa, K dogodek, da dobimo pri tem kralja, in S dogodek, da je izbrana karta srce. Z besedami opiši naslednje dogodke in za vsakega od njih ugotovi, koliko elementarnih dogodkov je ugodnih zanj:

a) $A \cup K$	b) $A \cup S$
c) $K \cup S$	č) $A \cup K \cup S$
d) $A \cup S'$	e) $A \cap K$
f) $K \cap S$	g) $A \cap S$
h) $A \cup (K \cap S')$	i) $(K \cap S) \cup (A \cap S')$
j) $(A' \cap K')'$	k) $(A' \cap S')'$

11. Iz posode, v kateri so 1 bela, 2 modri in 3 rdeče kroglice, potegnemo sočasno dve kroglici. Naj bodo dogodki A, B in C definirani takole:
A - izvlečeni kroglici sta enake barve;
B - ena od kroglic je bela;
C - izvlekli smo modro in rdečo kroglico.
Presodi, ali so ti dogodki paroma nezdružljivi in ali tvorijo popoln sistem dogodkov.

12. $\triangle$ Naj bosta A in B poljubna dogodka iz nekega poskusa. Dokaži, da za ta poskus dogodki A, $A' \cap B$ in $(A \cup B)'$ tvorijo popoln sistem dogodkov.

13. Naj bodo A, B in C dogodki iz istega poskusa. Kaj lahko sklepaš iz enakosti
a) $A \cup B \cup C = A$ oziroma b) $A \cap B \cap C = A$?

14. Dvakrat zapored vržemo igralno kocko. Ugotovi, kakšen je popolni sistem elementarnih dogodkov, predstavi ga z vzorčnim prostorom (predlog: 6×6 urejenih parov v kvadratni shemi) in v njem ponazori dogodke:
A - v obeh metih pade enako število pik;
B - v obeh metih pade liho število pik;
C - v prvem metu pade večje število pik kot v drugem;
D - v prvem metu pade vsaj toliko pik kot v drugem;
E - obakrat padejo vsaj tri pike;
F - vsota pik v obeh metih je 7;
G - obakrat pade šestica;
H - vsaj enkrat pade šestica;
I - kvečjemu enkrat pade šestica.

15. Kovanec mečemo toliko časa, dokler ena stran ne pade dvakrat po vrsti. Opiši množico izidov in dogodke:
A - da v četrtem metu pade grb;
B - da v prvem metu pade številka;
C - da grb pade večkrat kot številka;
D - pri prvem metu pade številka in številka pade trikrat;
E - številka pade trikrat;
F - prvič pade številka ali v četrtem metu pade grb.

16. Kovanec mečemo toliko časa, da prvič pade grb ali pa da dobimo pet številk po vrsti. Ugotovi, kakšen je popolni sistem dogodkov.

4.2 Verjetnost slučajnega dogodka

Izraz **verjetnost** in izpeljanke iz njega so nam iz vsakodnevnega življenja dobro znani. Verjetnost slučajnega dogodka povezujemo s tem, kako pogosto proučevani dogodek nastopi, če poskus velikokrat ponavljamo.

Za dogodke, ki se pri tem le redko zgodijo, rečemo, da so **malo verjetni**, za dogodke, ki se tudi pri velikem številu ponovitev zgodijo v večini primerov, pa pravimo, da so **zelo verjetni**; tak je – žal - na primer dogodek, da s kupljeno srečko ne dobimo niti najmanjšega dobitka.

S takimi ohlapnimi izrazi seveda v matematiki ne moremo početi nič pametnega, navajajo pa nas na misel, da bi verjetnost dogodka na določen način merili s pomočjo razmerja med številom tistih ponovitev, v katerih je dogodek dejansko nastopil, in številom vseh ponovitev poskusa.

ZGLED 4.2.1:

Vzemimo za zgled zaporedje metov igralne kocke, ki smo ga srečali v razdelku 3.2:

5, 2, 1, 6, 2, 3, 1, 2, 5, 4, 6, 3, 5, 6, 1, 2, 4, 4, 3, 6, 4, 3, 2, 4, 3, 3, 5, 1, 3, 4, 6, 1, 2, 6, 3, 1, 2, 4, 5, 3, 5, 6, 6, 3, 1, 4, 4, 5, 6, 2, 4, 6, 3, 2, 5, 3, 2, 1, 6, 5, 4, 1, 3, 2, 4, 6, 3, 2, 6, 1, 4, 5, 5, 6, 1, 3, 2, 5, 1, 4.

Naj bo dogodek A padec šestih pik; v tem zaporedju poskusov se je A zgodil štirinajstkrat; kot že vemo, ta podatek imenujemo **frekvenca** dogodka A in označimo $f(A)$. Če frekvenco dogodka A delimo s številom vseh poskusov n, dobimo **relativno frekvenco**,

$$f^o(A) = \frac{f(A)}{n} \tag{1}$$

V našem primeru je relativna frekvenca $14/80 = 0{\cdot}175$.

Pomen izraza "relativna frekvenca" ne odstopa od tistega, ki smo ga srečali v razdelku 3.2; število poskusov je seveda enako številu podatkov, ki jih imamo o izidih, edino, kar se na pogled loči, je - tradiciji na ljubo - uporaba male črke v imenovalcu.

Relativno frekvenco dogodka lahko izračunamo po vsaki ponovitvi poskusa, s tem dobimo ustrezno zaporedje relativnih frekvenc. S poskusi so ugotovili zanimivo značilnost tega zaporedja: relativna frekvenca, ki se spočetka dokaj neurejeno spreminja, se sčasoma (pri dovolj velikem številu ponovitev poskusa) ustali pri nekem številu in od njega le malo odstopa, tudi če še podaljšujemo zaporedje poskusov. Ta ugotovitev pomeni eno od najvažnejših verjetnostnih zakonitosti. Preverjali so jo empirično na različne načine; iz zgodovine so najbolj znani različni poskusi z dolgimi zaporedji metov kovanca. Pri tem so ugotovili, da je ustalitev (stabilizacija) relativne frekvence v dovolj velikem številu poskusov splošna zakonitost, zato se zdi smiselna t.i.

Statistična definicija verjetnosti:

Verjetnost dogodka A v proučevanem poskusu je število $P(A)$, pri katerem se navadno ustali relativna frekvenca dogodka A v dovolj velikem številu ponovitev tega poskusa.

ZGLED 4.2.2:

Ker se relativna frekvenca grba v velikem številu metov ustali (vsaj približno) pri $1/2$, rečemo, da je verjetnost za padec grba pri metu običajnega kovanca 0·5.

Statistična definicija verjetnosti dogodka skriva vsaj tri pasti:
- govori o nekem (spočetka neznanem) številu $P(A)$,
- pri katerem se "navadno" ustali relativna frekvenca,
- če smo naredili "dovolj" poskusov.

Včasih pa skušamo oceniti verjetnost slučajnega dogodka, ne da bi v resnici izvedli poskuse. Pri metu kovanca na primer razmišljamo takole: "Če je kovanec simetričen in narejen iz homogene zlitine, ni nobenega razloga, da bi grb padel večkrat (ali manjkrat) kot številka, torej v polovici vseh poskusov. Zato pričakujemo, da bo njegova relativna frekvenca vsaj približno enaka $1/2$, zato tudi verjetnost tega dogodka ne more biti različna od te vrednosti."

Običajno že spočetka pozabimo še na frekvence in relativne frekvence ter pravimo: "Padec grba je ena od dveh enakovrednih možnosti, verjetnost za ta dogodek je $1/2$."

ZGLED 4.2.3:

Kolikšna je verjetnost, da pri slepem izboru ene karte iz običajnega kompleta dobimo pika?

Če ravnamo po prejšnjem premisleku, lahko rečemo takole: na voljo so karte štirih barv, pikov je natanko toliko kot vsake od preostalih barv in zaradi izbiranja na slepo so vse barve enakovredne, vsaka barva ima enake možnosti, da bo izbrana, zato je verjetnost za pika $1/4$.

Postopek lahko posplošimo in rečemo takole: V primerih, ko so posamezni izidi enakovredni glede možnosti, da se zgodijo, verjetnost dogodka izračunamo tako, da "število ugodnih možnosti" delimo s "številom vseh možnosti" v proučevanem poskusu. Bolj formalno to pove t.i.

Klasična definicija verjetnosti:

Naj bo v nekem poskusu popolni sistem dogodkov sestavljen iz n elementarnih dogodkov in naj bo dogodek A vsota katerihkoli m od teh izidov.

Potem definiramo verjetnost dogodka A s predpisom

$$P(A) = \frac{m}{n} \tag{2}$$

Klasično definicijo smo precej svobodno izbrali za recept, po katerem računamo verjetnosti dogodkov. Ali je postopek smiseln, pa je odvisno od proučevanega poskusa. Če je namreč dogodek A kar eden izmed elementarnih dogodkov iz proučevanega poskusa, mora biti njegova verjetnost enaka $1/n$: kakor hitro predpostavimo, da verjetnosti računamo po klasični definiciji, s tem avtomatično pripišemo vsem elementarnim dogodkom enako verjetnost $1/n$, predpostavljamo torej, da gre za **simetričen popoln sistem dogodkov**. Logični obrat te ugotovitve pa pove:

Brž ko popolni sistem elementarnih dogodkov v nekem poskusu ni simetričen, je klasična definicija verjetnosti nesmiselna.

O simetričnem popolnem sistemu lahko govorimo v primeru, ko so pogoji poskusa takšni, da ne dajejo prednosti nobenemu od možnih izidov. Pri nekaterih poskusih zahteva pomeni kar simetričnost v običajnem geometrijskem pomenu, odtod tudi ime. Če kocka ni simetrična (ali če ni homogena), bodo nekateri izidi imeli prednost pred drugimi. Z izrazom "poštena igralna kocka" običajno povemo, da predpostavljamo simetričnost ustreznega popolnega sistema elementarnih dogodkov. V zgledu s kartami smo imeli pravico sklepati, da je verjetnost za pika $1/4$ zato in samo zato, ker smo se nekoč že dogovorili, da z izrazom "običajni komplet" označujemo tak nabor kart, da je v njem enako število kart posamezne barve.

Simetričnost popolnega sistema dogodkov dostikrat zagotavljamo z izbiranjem na slepo; takšno izbiranje je definirano ravno z zahtevo, da morajo imeti vsi elementi iz dane množice enake možnosti, da bodo izbrani.

ZGLED 4.2.4:

V posodi imamo 10 enakih kroglic, od katerih so 4 črne, ostale pa bele. Na slepo izvlečemo eno kroglico. Kolikšna je verjetnost dogodka B, da je izvlečena kroglica bela?

Ker je v posodi 10 kroglic, imamo $n = 10$ elementarnih dogodkov, od teh jih je za dogodek B ugodnih $m = 6$; verjetnost tega dogodka je zato po obrazcu enaka $P(B) = 6/10 = 0{\cdot}6$.

Verjetnost za izbor konkretnega elementa pri izbiranju na slepo je potem enaka recipročni vrednosti števila elementov v tej množici. V splošnem pa je v praksi težko zagotoviti simetričnost popolnega sistema dogodkov, dostikrat pa ne moremo niti presoditi, ali je popolni sistem tak, kot želimo. Zato tudi klasična definicija verjetnosti ne more biti univerzalna osnova za izgradnjo verjetnostnega računa kot matematične discipline, čeprav ustreza nekaterim praktičnim problemom. Za matematično korektno definicijo bomo zato uporabili samo tri skupne značilnosti, ki jih lahko izpeljemo tako iz statistične kot iz klasične definicije in se v celoti ujemajo z našimi predstavami o verjetnosti slučajnega dogodka.

Imejmo poljuben poskus. Potem velja:

D1. Verjetnost poljubnega dogodka A je nenegativno število,

$$P(A) \geq 0 \tag{3.a}$$

D2. Verjetnost gotovega dogodka je enaka 1,

$$P(G) = 1 \tag{3.b}$$

D3. Naj bosta A in B poljubna nezdružljiva dogodka iz istega poskusa. Potem je verjetnost njune vsote enaka vsoti posameznih verjetnosti,

$$AB = N \implies P(A+B) = P(A) + P(B) \tag{3.c}$$

Naštete lastnosti so značilne tako za klasični pristop k verjetnostnemu računu kot za verjetnost, ki jo definiramo s statistično definicijo. O tem se lahko hitro prepričamo.

(a) V zaporedju relativnih frekvenc proučevanega dogodka je najmanjša možna vrednost 0 (če se nekaj časa dogodek sploh ne zgodi), zato tudi število $P(A)$, pri katerem se zaporedje relativnih frekvenc (navadno) ustali, ne more biti negativno. Če se držimo klasične definicije, pa recimo takole: za dogodek A ne more biti ugodnih manj kot nič izidov od n možnih, zato je tudi število $P(A)$ nenegativno.

(b) Gotov dogodek nastopi v vsaki ponovitvi poskusa, v zaporedju relativnih frekvenc so potemtakem vsi členi enaki 1, zato se tudi to zaporedje ne more ustaliti pri kaki drugi vrednosti. Klasični premislek: za nastop nekega dogodka so ugodni vsi možni izidi iz popolnega sistema, $m = n$ in zato ugotovitev velja.

(c) Ker sta dogodka A in B po predpostavki nezdružljiva, ne moreta nastopiti v isti ponovitvi poskusa oba hkrati. Zato je v poljubnem številu ponovitev poskusa frekvenca njune vsote enaka vsoti frekvenc $f(A)+f(B)$. (Ugotovitev je v sorodu s pravilom iz teorije množic o moči unije dveh tujih množic!) Relativna frekvenca vsote dogodkov $A+B$ ("plus" si dovolimo zaradi nezdružljivosti obeh dogodkov, glej str. 87) se zato ustali natanko tam kot $f^o(A) + f^o(B)$, torej pri $P(A) + P(B)$, verjetnost vsote nezdružljivih dogodkov je enaka vsoti njunih verjetnosti, če izhajamo iz statistične definicije.

Kadar nam okoliščine dopuščajo uporabo klasične definicije verjetnosti, pa rečemo takole: če je za dogodek A ugodnih m_A izidov in za dogodek B

natanko m_B izidov (od n možnih), jih je za vsoto $A+B$ zaradi nezdružljivosti ugodnih natanko $m_A + m_B$. Potem je po definiciji

$$P(A \cup B) = \frac{m_A + m_B}{n} = \frac{m_A}{n} + \frac{m_B}{n} = P(A) + P(B)$$

Lastnosti (3.a), (3.b) in (3.c) so torej skupno bistvo pojma "verjetnost", kot ga razumemo v vsakodnevni praksi. Kakršnakoli korektna definicija verjetnosti slučajnega dogodka mora torej zagotavljati najmanj to, da za število $P(A)$ veljajo omenjene tri lastnosti. Izkazalo se je, da k tem trem lastnostim ni treba privzeti nobenih bistveno novih zahtev, da bi zgradili logično zaokroženo matematično teorijo.

$\square$

Lepota matematike je tudi v tem, da lahko dostikrat pot, ki smo jo prehodili od konkretnega k splošnemu, z dodatnim užitkom premerimo tudi v obratni smeri, postavimo bolj abstrakten model in se nato razveselimo, ko mu v naravi ("vsakdanjem življenju") najdemo konkretno podobo, eno ali več.

Imejmo neko končno, sicer pa povsem poljubno množico

$$G = \{E_1, E_2, \dots, E_n\}$$

in jo do preklica proglasimo za univerzalno množico, torej za področje našega proučevanja.

Njene elemente imenujmo **elementarni dogodki**, bogatejšim podmnožicam - torej takim, ki vsebujejo več kot en element iz G, - recimo **sestavljeni dogodki**. Z vsakim parom dogodkov A in B sta v potenčni množici $\mathcal{P}\,G$ tako komplementa A' in B' (nasprotna dogodka dogodkov A oziroma B) kot tudi njun presek (produkt dogodkov) $A \cap B$. To je zadostno opravičilo za to, da $\mathcal{P}\,G$ (spet, glej str. 89) imenujemo **algebra dogodkov**. Med njenimi elementi sta **gotovi dogodek** G in **nemogoči dogodek** N, ki ustreza prazni podmnožici množice G, torej $N = \emptyset$.

Preslikavo P iz algebre dogodkov $\mathcal{P}\,G$ v množico realnih števil imenujemo **verjetnost** na algebri dogodkov $\mathcal{P}\,G$, če zadošča naslednjim trem aksiomom:

1. NENEGATIVNOST $\quad (\forall A)(A \in \mathcal{P}\,G \implies P(A) \geq 0)$
2. NORMIRANOST $\quad P(G) = 1$
3. ADITIVNOST $\quad A \cap B = N \implies P(A \cup B) = P(A) + P(B)$

Navedene aksiome najdemo v literaturi pod imenom **aksiomi Kolmogorova**, saj jih je ta ruski (tedaj pač sovjetski) matematik v tridesetih letih tega stoletja (1933) prvi zapisal v tako splošni obliki in hkrati pokazal, da te tri predpostavke o lastnostih preslikave P povsem zadoščajo za postavitev logično neoporečne matematične teorije.

Iz tretjega aksioma in asociativnosti unije množic lahko ugotovimo, da velja pravilo za seštevanje tudi za več kot dva nezdružljiva dogodka:

Verjetnost vsote poljubnega končnega števila nezdružljivih dogodkov je enaka vsoti verjetnosti posameznih dogodkov.

Zato lahko za preslikavo P, ki zadošča aksiomom Kolmogorova, izračunamo vrednost na poljubnem elementu A, brž ko poznamo njene vrednosti na elementarnih dogodkih, ki sestavljajo A kot podmnožico množice G; verjetnost $P(A)$ je pač vsota verjetnosti vseh "sestavnih delov". Nerodnost je v tem, da aksiomi ne povedo prav ničesar o tem, kako je verjetnost porazdeljena po popolnem sistemu elementarnih dogodkov, zato je to porazdelitev treba opisati v vsakem primeru posebej. Lahko je "klasična" (vsi izidi so enako verjetni, $P(E_i) = 1/n$), ali pa nesimetrična (izidi imajo različne verjetnosti), da je le vsota nenegativnih števil $P(E_i)$ (1. aksiom!) enaka 1, saj je vsota vseh izidov gotov dogodek, katerega verjetnost mora biti enaka 1 po 2. aksiomu.

ZGLED 4.2.5:

Pri metu poštene igralne kocke (homogena, simetrična) lahko predpostavimo klasično porazdelitev verjetnosti po elementarnih dogodkih. Če kocki premaknemo težišče iz geometrijskega središča, dobimo kakšno drugačno porazdelitev verjetnosti po popolnem sistemu elementarnih dogodkov.

ZGLED 4.2.6:

Naj bo $G = \{E_1, E_2, E_3, E_4\}$ in P verjetnost na algebri dogodkov $\mathcal{P}G$. Izračunaj, kakšna je porazdelitev verjetnosti, če za sestavljene dogodke

$A = \{E_2, E_3\}$, $B = \{E_3, E_4\}$ in $C = \{E_1, E_2, E_4\}$

velja $P(A) = 5/12$, $P(B) = 1/3$ in $P(C) = 3/4$!

Ker lahko pišemo

$A = E_2 + E_3$, $B = E_3 + E_4$, $C = E_1 + E_2 + E_4$,

je zaradi nezdružljivosti elementarnih dogodkov po tretjem aksiomu tudi

$$P(A) = P(E_2) + P(E_3)$$
$$P(B) = P(E_3) + P(E_4)$$
$$P(C) = P(E_1) + P(E_2) + P(E_4)$$

oziroma v skladu z našimi podatki

$$P(E_2) + P(E_3) = \frac{5}{12}$$
$$P(E_3) + P(E_4) = \frac{1}{3}$$
$$P(E_1) + P(E_2) + P(E_4) = \frac{3}{4}$$

K temu sistemu enačb sodi še ena, ki pove, da je vsota verjetnosti vseh izidov 1 (zahtevi običajno rečemo **normirni pogoj**),

$$P(E_1) + P(E_2) + P(E_3) + P(E_4) = 1$$

Iz sistema teh štirih linearnih enačb dobimo iskane verjetnosti $P(E_i)$ (lepa vaja v reševanju sistemov linearnih enačb!):

$P(E_1) = 1/2$, $P(E_2) = 1/6$, $P(E_3) = 1/4$, $P(E_4) = 1/12$.

Dejstvo, da je v okviru aksiomov Kolmogorova možnih več porazdelitev verjetnosti po popolnem sistemu elementarnih dogodkov, nas ne skrbi. Bistvena naloga verjetnostnega računa kot matematične discipline namreč ni preverjanje, kakšna je ta porazdelitev, ampak ugotavljanje, kako iz danih verjetnosti posameznih (elementarnih) dogodkov računati verjetnosti dogodkov, ki so iz njih sestavljeni; takšna bo tudi naša naloga v naslednjih razdelkih.

V A J E

1. Vrzi stokrat kovanec in pri tem proučuj zaporedje relativnih frekvenc. Primerjaj rezultate s pričakovanimi. Grafično prikaži, kako se je s podaljševanjem zaporedja poskusov obnašala relativna frekvenca.

2. Iz običajnega kompleta kart stokrat na slepo izvleci po eno karto in zabeleži, katere barve (srce, kara, križ ali pik) je karta, nato jo vrni v komplet in ga pred naslednjo izbiro dobro premešaj. Spremljaj obnašanje relativnih frekvenc za vsako barvo posebej. Kaj pričakuješ v dovolj velikem številu poskusov?

3. △ Če so se pri prejšnjih dveh vajah roke utrudile bolj kot glava, napiši računalniški program(ček), ki bo namesto tebe metal kovanec in izbiral karte (pa prešteval izide, pa naredil tabelo in morda tudi grafične prikaze, pa... pustimo domišljiji prosto pot!).

4. Naj bo

A	-	dogodek, da vržemo s kocko 6 pik;
B	-	dogodek, da vržemo liho število pik;
C	-	dogodek, da vržemo 3 pike;
D	-	dogodek, da vržemo sodo število pik;
E	-	dogodek, da vržemo manj kot 5 pik.

Izračunaj verjetnosti teh dogodkov, če smemo predpostaviti, da je kocka poštena in da lahko sledimo klasični definiciji verjetnosti.

5. △ Predpostavimo, da je mogoče kocko obtežiti tako, da bo verjetnost posameznega izida E_i premo sorazmerna številu pik. Določi porazdelitev verjetnosti, nato pa izračunaj, kakšne bi bile v tem primeru verjetnosti dogodkov iz prejšnje naloge.

6. V škatli so štiri kroglice, ki so oštevilčene z 1, 3, 8 in 9. Na slepo izbiramo po eno kroglico in je ne vračamo v škatlo, ampak jo vsakokrat odložimo na mizo. Izračunaj verjetnost, da dobimo zapisano letnico 1983.

7. Kolikšna je verjetnost dogodka iz prejšnje naloge, če po vsakem izbiranju kroglico vrnemo v škatlo (in si njeno številko samo zabeležimo)?

8. Reši nalogo še za primer, ko imamo na kroglicah številke 1, 9, 9 in 3 ter nas zanima letnica 1993.

9. Tombola ima 90 ploščic s števili od 1 do 90. Na slepo izvlečemo eno ploščico. Kolikšne so verjetnosti naslednjih dogodkov:

A - število na ploščici je deljivo s 3;
B - število na ploščici je sodo;
C - število na ploščici je deljivo s 3 ali sodo?

10. Loterija ima 5 000 srečk, od tega jih 100 zadene. Izračunajte verjetnosti naslednjih dveh dogodkov:

A - slučajno izbrana srečka zadene;
B - od dveh slučajno izbranih srečk natanko ena zadene.

11. V škatli imamo 5 belih, 6 rdečih in 9 črnih (sicer pa enakih) krogel. Hkrati na slepo potegnemo tri krogle iz škatle. Izračunajte verjetnosti naslednjih dogodkov:

A - vse tri krogle so bele;
B - ena krogla je rdeča, dve pa črni;
C - nobena krogla ni bela.

12. Iz lesa naredimo kocko, $a = 10$ cm, pobarvamo jo s črno barvo in razžagamo na 1 000 enakih kockic. Te premešamo, nato pa na slepo izberemo eno od njih. Kolikšna je verjetnost dogodka, da ni na izbrani kockici pobarvana nobena ploskev, in kolikšna verjetnost, da sta pobarvani natanko dve ploskvi?

13. V besedi MATEMATIKA na slepo prečrtamo eno črko. Kolikšna je verjetnost, da je to samoglasnik in kolikšna, da je to ravno črka T?

14. Nepismeni osebi damo listke s črkami A, N, A, N, A, S in jo prosimo, da jih postavi v vrsto. Kolikšna je verjetnost, da pri tem nastane beseda ANANAS? (Ali lahko pri taki nalogi obstaja kak pomislek glede uporabe klasične definicije verjetnosti?)

15. Dvakrat zapored vržemo igralno kocko. Pomagaj si z vzorčnim prostorom iz naloge 14 v prejšnjem razdelku in izračunaj verjetnosti naslednjih dogodkov:

A - v obeh metih pade enako število pik;
B - v obeh metih pade liho število pik;
C - v prvem metu pade večje število pik kot v drugem;
D - v drugem metu pade vsaj toliko pik kot v prvem;
E - obakrat padejo vsaj tri pike;
F - vsota pik v obeh metih je 7;
G - obakrat pade šestica;
H - vsaj enkrat pade šestica;
I - kvečjemu enkrat pade šestica.

16. Kolikšna je verjetnost,

A - da pri metu 5 kock pade 5 različnih vrednosti;
B - da pri metu 12 kock vsako število pik pade po dvakrat?

17. Na petih listkih so zapisane številke 1, 2, 3, 4 in 5. Na slepo vzamemo tri listke in jih položimo od leve proti desni. Izračunaj verjetnost dogodka, da bo trimestno število, ki pri tem nastane, sodo.

18. Učenci se učijo angleško (24), nemško (23) in francosko (15); pri tem je 7 takih, ki se sočasno učijo angleško in nemško, 8 takih, ki kombinirajo angleščino s francoščino, 10 se jih uči nemško in francosko, trije pa se mučijo z vsemi tremi jeziki. Kolikšne so verjetnosti dogodkov, da se iz te populacije na slepo izbrani učenec

A - uči samo angleško;
B - uči samo nemško;
C - uči samo francosko;
D - uči francosko, ne pa tudi angleško;
E - da se uči oba germanska jezika, ne pa tudi romanskega?

19. V škatli imamo 4 bele, 2 rdeči in 3 modre knjige. Na slepo izbiramo po eno knjigo, dokler jih ne zmanjka. Izvlečene knjige postavljamo v vrsto na polico. Kolikšna je verjetnost, da bodo na koncu knjige enake barve stale skupaj?

20. △ Množico $A = \{a_1, a_2, ..., a_n\}$ preslikamo vase tako, da vsakemu elementu $a_i \in A\,(i = 1, 2, ..., n)$ priredimo na slepo izbran (pa seveda enolično določen) element iste množice. Izračunaj verjetnost dogodka, da smo s tem definirali neko bijektivno preslikavo množice A nase.

21. △ Podobno kot v 3. nalogi napiši računalniški program, ki bo opravil predpisano število žrebanj v loteriji, opisani v 9. vaji. Spremljaj relativne frekvence posameznih dogodkov iz tega poskusa in primerjaj odstopanja od teoretičnih verjetnosti teh dogodkov.

22. △ Napiši program, ki bo simuliral obnašanje goljufive kocke, pri kateri je verjetnost za šestico 1/2, verjetnosti ostalih izidov pa po 1/10.

4.3 Lastnosti in računanje verjetnosti

V tem razdelku bomo spoznali najpomembnejše posledice obrazcev (3.a) - (3.c) iz prejšnjega razdelka (oziroma aksiomov Kolmogorova) in s tem dobili nekaj priročnih orodij za lažje računanje verjetnosti.

Poudarimo še enkrat, da obrazci, ki jih bomo izpeljali, veljajo neodvisno od tega, kakšna je porazdelitev verjetnosti po (končnem, če ne bo posebej povedano drugače) popolnem sistemu elementarnih dogodkov; od porazdelitve je odvisno samo računanje verjetnosti posameznih sestavljenih dogodkov. V splošnem primeru je treba za verjetnost dogodka A v potu lastnega obraza sešteti verjetnosti vseh izidov, ki so za A ugodni, v izjemno lepi klasični različici, ko lahko predpostavimo simetričnost popolnega sistema, si pomagamo z večkratniki, saj seštevamo enake verjetnosti; tako je znameniti obrazec $P(A) = m/n$ samo posledica poenostavitev v splošnem postopku:

$$P(A) = P(E_{i_1} + E_{i_2} + ... + E_{i_m}) = P(E_{i_1}) + P(E_{i_2}) + ... + P(E_{i_m}) \Longrightarrow$$

$$\Longrightarrow P(A) = \frac{1}{n} + \frac{1}{n} + ... + \frac{1}{n} = m \cdot \frac{1}{n}$$

Zdaj pa k obljubljenim lastnostim.

L1. *Če je dogodek A način dogodka B, verjetnost dogodka A ne presega verjetnosti dogodka B:*

$$A \subset B \Longrightarrow P(A) \le P(B) \tag{1}$$

Pravilo lahko premislimo na vse tri načine (statistična definicija, klasična definicija, aksiomi Kolmogorova).

a) Ker je A način dogodka B, se po definiciji relacije "je način" B zgodi vsaj tolikokrat kot A. Pri poljubnem zaporedju poskusov je relativna frekvenca dogodka B vsaj tako velika kot ona, ki pripada dogodku A, zato tudi odnos med verjetnostima ne more biti drugačen.

b) Če je A način dogodka B, to pomeni, da so vsi izidi, ki so ugodni za A, avtomatično ugodni tudi za B, lahko pa je za slednjega ugoden še kak izid, pri katerem A ne nastopi; na vsak način je $m_B \ge m_A$ in zato velja enaka ocena - po deljenju s številom vseh izidov n - tudi za pripadajoči verjetnosti.

c) Podrobnejši premislek prepuščamo bralcu: začel bo podobno kot v točki b), upošteval aditivnost verjetnosti (3. aksiom) in nenegativnost prispevkov verjetnosti posameznih izidov (1. aksiom).

L2. *Za poljuben dogodek A je vsota njegove verjetnosti in verjetnosti nasprotnega dogodka (negacije) enaka 1,*

$$P(A) + P(A') = 1 \tag{2}$$

Tokrat utemeljimo pravilo samo na en način (ostalo - vaja!): iz definicije nasprotnega dogodka sledi, da si dogodek A in nasprotni dogodek A' vse elementarne dogodke proučevanega poskusa razdelita tako, da je vsak od izidov ugoden natanko za enega od njiju. Dogodka sta nezdružljiva, $A \cap A' = N$, njuna vsota pa je gotov dogodek, $A \cup A' = G$. Zato je

$$1 = P(G) = P(A + A') = P(A) + P(A') \tag{$*$}$$

Pri tem je prvi enačaj iz drugega aksioma, zadnji pa iz tretjega. Rep in glava kače (*) pa prineseta (2), kar smo hoteli dokazati.

Enačbo (2) po potrebi razrešimo bodisi na $P(A)$ ali na $P(A')$, torej

$$P(A) = 1 - P(A') \tag{2.a}$$

ali

$$P(A') = 1 - P(A) \tag{2.b}$$

Prvi obrazec dostikrat uporabljamo v primerih, ko je verjetnost negacije laže izračunati kot verjetnost proučevanega dogodka.

ZGLED 4.3.1:

V posodi imamo 10 kroglic, od tega 6 belih in 4 zelene. Na slepo sočasno izvlečemo 4 kroglice. Izračunaj verjetnost dogodka, da je med izvlečenimi kroglicami vsaj ena zelena!

Dogodek A je sestavljen iz štirih med seboj nezdružljivih dogodkov, da je med izvlečenimi kroglicami natanko ena kroglica zelena, da sta takšni natanko dve, da so natanko tri oziroma da so vse štiri kroglice zelene. $P(A)$ bi torej lahko izračunali tako, da bi sešteli verjetnosti teh štirih načinov dogodka A, kar pa zahteva precej dela. Do rezultata pridemo hitreje tako, da izračunamo verjetnost negacije A' in jo odštejemo od 1:

$$P(A) = 1 - P(A') = 1 - \frac{\binom{6}{4}\binom{4}{0}}{\binom{10}{4}} \approx 0{\cdot}93$$

Upoštevali smo, da je vseh vzorcev toliko, na kolikor načinov lahko izmed 10 kroglic izberemo štiri, ugodnih za proučevani dogodek pa je med vsemi vzorci toliko, kolikor je načinov, na katere lahko izberemo tako velik vzorec iz samih belih kroglic.

Ob tem primeru lahko preverimo, ali pravilno razumemo rezultate, do katerih pridemo v verjetnostem računu. Dejstvo, da je verjetnost $P(A')$ približno 0·07, za posamezen poskus ne pove veliko, brez nadaljnjega se lahko zgodi tudi večkrat po vrsti, da so vse izvlečene kroglice bele. Vendar pa v dovolj velikem številu poskusov delež takšnih poskusov ne bo prav velik, sukal se bo okrog 0·07.

Ob tem rezultatu bi se kak proizvajalec zamislil, saj lahko nalogo postavimo tudi drugače: med desetimi izdelki smo kupcu poslali šest dobrih in štiri slabe; kolikšna je verjetnost, da bo pri slepem izboru štirih izdelkov naletel na same dobre? Odgovor poznamo – okrog 0'07. Če je ta dogodek pogoj, da našo pošiljko sprejme, imamo malo možnosti za uspeh...

L3. *Verjetnost dogodka ne more biti večja od 1,*

$$P(A) \leq 1 \tag{3}$$

Komentar k temu pravilu skoraj ni potreben, a vendar utemeljimo na najkrajši način: vsota dveh nenegativnih verjetnosti $P(A)$ in $P(A')$ v enačbi (2) je enaka 1, zato nobena od njiju ne more preseči tega števila.

POMNIMO: Če z mirno vestjo zapišemo kot verjetnost kakršnegakoli dogodka število, ki je večje od 1, se nemudoma vrnemo nekaj strani nazaj v učbeniku.

Ker je nemogoči dogodek nasprotni dogodek h gotovemu dogodku G in ker slednji "pobere vse", iz obrazca (2) sledi pravilo

L4. *Verjetnost nemogočega dogodka je enaka 0,*

$$P(N) = 0 \tag{4}$$

Eden od najpomembnejših je obrazec za verjetnost vsote dogodkov:

L5. *Pri poljubnih dogodkih A in B velja za verjetnost njune vsote obrazec*

$$P(A \cup B) = P(A) + P(B) - P(AB) \tag{5}$$

Če sta A in B nezdružljiva, nam (5) ne pove nič novega: njun produkt je tedaj nemogoči dogodek, ki ima po (4) verjetnost 0 in (5) je samo nov zapis tretjega aksioma Kolmogorova (ali obrazca 3.c s strani 98 za tiste, ki s pravokotnikom označeno branje ignorirajo). Za združljiva dogodka pa obrazec premislimo takole: da dobimo $P(A \cup B)$, moramo sešteti verjetnosti izidov, ki so za to vsoto ugodni. To so natanko tisti izidi, ki so ugodni za A ali za B, žal pa se nam pri seštevanju lahko zgodi, da kako verjetnost prištejemo dvakrat, enkrat na račun dogodka A in drugič na račun dogodka B. V mislih imamo seveda verjetnosti tistih izidov, ki so hkrati ugodni za oba dogodka, to pa so natanko tisti, ki s svojim nastopom pomenijo, da se je zgodil produkt $A \cap B$. Zato moramo od vsote verjetnosti posameznih dogodkov odšteti verjetnost njunega produkta.

ZGLED 4.3.2:

S prijateljem smo stavili, da bomo pri slepem izboru ene karte iz običajnega kompleta dobili srce ali asa. Kolikšna je verjetnost, da to stavo dobimo?

Z A označimo dogodek, da je na slepo izbrana karta srce, z B dogodek, da je izvlečena karta as, izračunati pa moramo verjetnost vsote teh dveh dogodkov. Če bi pozabili na to, da je treba odšteti verjetnost produkta dogodkov, bi to v praksi pomenilo, da smo srčnega asa dvakrat šteli med ugodne izide za nastop $A \cup B$. Tako pa pravilno dobimo

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{8}{32} + \frac{4}{32} - \frac{1}{32} = \frac{11}{32} \approx 0{'}344$$

ZGLED 4.3.3:

Dobavitelj nam je v pošiljko 100 izdelkov podtaknil 10 pokvarjenih. V našem podjetju pošiljko kontroliramo tako, da sočasno na slepo izberemo iz serije 5 izdelkov in jih preizkusimo. Pošiljko sprejmemo, če v izbranem vzorcu ni več kot en pokvarjen proizvod. Kolikšna je verjetnost, da bo opisana pošiljka sprejeta?

Glede na opisani način sprejemanja pošiljk je dogodek A "Pošiljka je sprejeta" vsota dveh nezdružljivih dogodkov, da je izbrani vzorec brez slabih izdelkov in da je v vzorcu natanko en pokvarjen izdelek. Izidov je v obeh primerih toliko, kolikor podmnožic (vzorcev) po 5 elementov lahko sestavimo iz množice s 100 elementi; za prvi dogodek so ugodni tisti izidi, pri katerih izberemo 5 izdelkov med 90 kakovostnimi (in - formalno gledano - 0 izdelkov med 10 slabimi), za drugi dogodek pa tisti, pri katerih izberemo 4 izdelke med 90 kakovostnimi, en izdelek pa med 10 slabimi. Tako je

$$P(A) = \frac{\binom{90}{5}\binom{10}{0}}{\binom{100}{5}} + \frac{\binom{90}{4}\binom{10}{1}}{\binom{100}{5}} \approx 0{'}92$$

Rezultat v bistvu pove, da bomo v povprečju sprejeli kar 92% serij, ki vsebujejo po 10% nekakovostnih izdelkov, zato bo verjetno treba poostriti kontrolo, saj je sicer tveganje preveliko.

ZGLED 4.3.4: △

Radi bi ugotovili, koliko rib je v nekem jezeru. Da bi ocenili to število, ulovimo m rib, jih označimo in spustimo nazaj v vodo. Po določenem času (ko se ribe, ki smo jih označili, pomešajo z ostalimi) ulovimo n rib in ugotovimo, da je med njimi natanko k označenih. Izračunaj najverjetnejše število rib v jezeru!

To neznano število bomo označili z N. V jezeru je skupno m označenih in $N - m$ neoznačenih rib. Verjetnost dogodka, da je v vzorcu, v katerega smo na slepo izbrali n rib, natanko k rib označenih, je tedaj enaka

$$P(N) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \tag{6}$$

Z obliko zapisa smo označili, da je ta verjetnost odvisna od neznanega skupnega števila rib v jezeru. Število N bomo določili tako, da bo pripadajoča verjetnost maksimalna, torej večja tako od $P(N+1)$ kot od $P(N-1)$. Če (6) primerjamo z ustreznima obrazcema, v katerih namesto N vstavimo $N-1$ oziroma $N+1$, ugotovimo, da mora število N ležati med vrednostima $(nm/k)-1$ in nm/k, kar pomeni, da je najverjetnejše število N enako celemu delu števila nm/k. (Do ugotovitve, da je uporabljeno sklepanje pravilno, nas pripelje lepa vaja v premetavanju binomskih simbolov, ki je pogumnejši ne bodo izpustili...)

Če smo denimo označili $m = 20$ rib in nato v vzorcu velikosti $n = 10$ odkrijemo $k = 3$ ribe, je $nm/k = 10 \cdot 20/3 = 66\dot{\overline{6}}$, zato bi po našem receptu število rib v jezeru ocenili na 66.

Bralec naj (obvezno) izračuna $P(66)$ in rezultat primerja s $P(65)$ oziroma $P(67)$ ter pri tem (neobvezno) razmišlja o zanesljivosti takega načina ocenjevanja. Ko se mu bo zdelo, da se muči z binomskimi simboli, pa naj se spomni, da je v spodobnih jezerih rib mnogo več, zato je pri korektnem vzorčenju imenovalec k znatno manjši od produkta $n \cdot m$, ustrezni količnik temu ustrezno večji in binomski simboli - da ne omenjamo!

V A J E

1. Vse lastnosti verjetnosti, ki smo jih dokazali samo na enega od načinov, premisli tudi na ostale.

2. Nalogo iz zgleda 4.3.1 reši tudi z neposrednim računom, brez pomoči obrazca (2.a).

3. Med 50 izdelki so 4 pokvarjeni. Na slepo izberemo v vzorec (istočasno) štiri izdelke. Kolikšna je verjetnost, da sta v izbranem vzorcu natanko dva izdelka pokvarjena?

4. Loterija je dala v prodajo $M + N$ srečk, od katerih jih zadene natanko N. Kolikšna je verjetnost, da med k kupljenimi srečkami zadene natanko q srečk?

5. Med 10 čipi je 8 brezhibnih. Na slepo sočasno izberemo polovico od njih. Kolikšna je verjetnost, da smo dobili same brezhibne čipe?

6. Sočasno vržemo dve pošteni igralni kocki. Kolikšna je verjetnost, da vsaj na eni pade šestica? (Morda poskusiš s primerno dolgim zaporedjem poskusov - ali računalniško simulacijo - dobiti eksperimentalni približek za iskano verjetnost...)

7. △ Zgodovini na ljubo naslednja "naloga o točkah", originalno "problème des points":

Kaj je bolj verjetno, da pri štirih metih poštene igralne kocke vsaj enkrat pade šest pik, ali dogodek, da pri 24 metih dveh kock vsaj enkrat na obeh kockah pade šestica? (Problem najdemo v literaturi tudi pod oznako "paradoks viteza de Méréja", po znamenitem svetovljanu, ki ga je odgovor na postavljeno vprašanje zanimal iz povsem neznanstvenih razlogov; primerjaj tudi str. 85!)

8. Iz običajnega kompleta kart na slepo sočasno izvlečemo tri karte. Kolikšne so verjetnosti dogodkov

- A - vse tri karte so dame;
- B - vse tri karte so srca;
- C - natanko dve karti sta pika;
- D - vse tri karte so iste barve (črne ali rdeče);
- E - vsaj ena od kart je dama?

9. V posodi imamo 3 rdeče in 7 belih kroglic. Na slepo eno za drugo izvlečemo iz posode vse kroglice razen ene. Kolikšna je verjetnost, da je preostala kroglica bela?

10. Iz serije proizvodov, v kateri je N dobrih in M defektnih, smo za kontrolo na slepo sočasno izbrali K izdelkov. Pri pregledu prvih k izdelkov iz vzorca smo naleteli na same dobre izdelke. Kolikšna je - pri takih vrednostih števil N, M, K in k, ki naredijo nalogo smiselno - verjetnost, da je tudi naslednji proizvod iz vzorca dober?

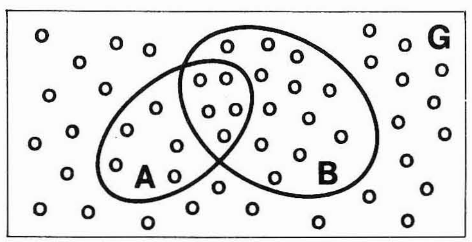
11. △ Izpelji obrazec za verjetnost vsote treh dogodkov. (Nasvet: če si zadovoljen s klasičnim pristopom, si lahko pomagaš z načelom vključitev in izključitev, da prešteješ ugodne izide; če želiš splošnejšo izpeljavo, pa za $P(A \cup B \cup C)$ dovoljkrat uporabi asociativnostni in distributivnostni zakon za računanje z dogodki, to je obrazca (4) in (5) s strani 90.)

4.4 Pogojna verjetnost in verjetnost produkta

Vzemimo neki poskus, za katerega bomo vsaj spočetka zaradi enostavnosti predpostavili, da je pripadajoči popolni sistem elementarnih dogodkov simetričen. Zanimala nas bo verjetnost dogodka $A \in \mathcal{P}\,G$ pri pogoju, da se je zgodil nek drug dogodek, recimo B, seveda iz istega poskusa; tej verjetnosti bomo rekli **pogojna verjetnost** dogodka A glede na dogodek B in jo bomo označili s $P(A/B)$.

Za prakso je kakopak zanimiv samo primer, ko sta v proučevanem poskusu dogodka A in B združljiva, saj je v nasprotnem primeru pogojna verjetnost $P(A/B)$ nujno enaka 0.

Razmišljanje o pogojni verjetnosti si bomo najlaže ponazorili z vzorčnim prostorom.



Slika 4.2: Vzorčni prostor - pogojna verjetnost

Ker nas zanima verjetnost dogodka A pri dodatnem pogoju, da se je zgodil dogodek B, proučujemo samo tisti del vzorčnega prostora, ki ustreza dogodku B, torej tistih m_B izidov (od skupno n možnih), ki sestavljajo dogodek B. Za dogodek A je od teh izidov ugodnih natanko tistih m_{AB} izidov, ki so sočasno načini dogodka A in dogodka B, torej tisti izidi, ki so ugodni za produkt AB. Verjetnost dogodka A pri pogoju B je tedaj po klasični definiciji enaka razmerju med številoma m_{AB} in m_B. Iz ulomka m_{AB}/m_B naredimo dvojni ulomek tako, da števec in imenovalec delimo s številom vseh elementarnih dogodkov tega poskusa, to je n; tako dobimo

$$P(A/B) = \frac{m_{AB}/n}{m_B/n} = \frac{P(A\,B)}{P(B)} \qquad (1)$$

Pogojno verjetnost dogodka A pri pogoju B torej izračunamo tako, da verjetnost produkta teh dveh dogodkov delimo z verjetnostjo pogoja B. S tem pravilom definiramo pogojno verjetnost tudi v primeru, ko verjetnost ni enakomerno porazdeljena po popolnem sistemu izidov in ne moremo uporabiti klasične definicije verjetnosti, seveda pa mora biti $P(B) > 0$, če naj prijem deluje.

Obrazec za pogojno verjetnost (1) lahko razrešimo na verjetnost produkta $P(A\,B)$,

$$P(A\,B) = P(B)\,P(A/B) \qquad (2)$$

Po drugi strani pa z analognim premislekom pridemo do pogojne verjetnosti dogodka B pri pogoju A (ne pozabimo, da je produkt dogodkov komutativen in da zato ni treba skrbeti za vrstni red dogodkov v produktu),

$$P(B/A) = \frac{P(A\,B)}{P(A)} \qquad (3)$$

odkoder pa dobimo drugo obliko izraza za verjetnost produkta,

$$P(A\,B) = P(A)\,P(B/A) \qquad (4)$$

Končno združimo obrazca (2) in (4) v

$$P(A\,B) = P(A)\,P(B/A) = P(B)\,P(A/B) \qquad (5)$$

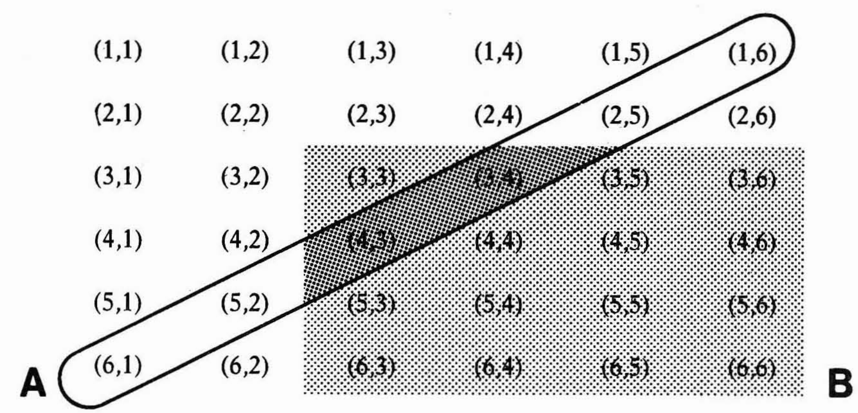
in pojasnimo ta obrazec še z besedami oziroma z njegovo vsebino (kar daje več upanja, da si ga tudi zapomnimo):

Verjetnost produkta poljubnih dogodkov A in B izračunamo tako, da verjetnost kateregakoli od njiju pomnožimo s pogojno verjetnostjo drugega dogodka pri pogoju, da se je prvi zgodil.

ZGLED 4.4.1:

Sočasno vržemo dve pošteni igralni kocki. Kolikšna je verjetnost, da je vsota pik 7, če vemo, da so na vsaki kocki padle vsaj tri pike?

Pri računanju si bomo pomagali z vzorčnim prostorom, kakršnega smo bralcu že predlagali v 14. vaji razdelka 4.1 (str. 94):



Slika 4.3: Vzorčni prostor za met dveh igralnih kock

Z A smo označili dogodek, da je vsota pik 7, z B pa dogodek, da so na vsaki kocki padle vsaj po 3 pike. Rwačunamo potemtakem $P(A/B)$. Vseh možnih izidov, prikazanih v vzorčnem prostoru na sliki 4.3, je 36; od teh jih je za produkt AB ugodnih toliko, kolikor točk je v preseku ustreznih množic vzorčnega prostora, ki pripadata dogodkoma. Če prvo število v urejenem paru (i,j) pomeni število pik na prvi kocki, drugo pa število pik na drugi kocki, sta torej ugodna dva izida, $(3,4)$ in $(4,3)$. Za pogoj B je ugodnih 16 od skupaj 36 izidov, zato je $P(AB) = 2/36$ in $P(B) = 16/36$, zaradi česar je po obrazcu (1) iskana pogojna verjetnost

$$P(A/B) = \frac{P(AB)}{P(B)} = \frac{\frac{2}{36}}{\frac{16}{36}} = \frac{1}{8}$$

Verjetnost $P(A/B)$, ki smo jo izračunali v tem zgledu, ni enaka "običajni" verjetnosti dogodka A v proučevanem poskusu. Če ne postavljamo dodatnih pogojev, je za dogodek A ugodnih 6 izidov od 36 možnih, torej je $P(A) = 6/36 = 1/6$. Namesto "običajna" bomo rajši rekli **brezpogojna verjetnost** dogodka A. Kadar se (tako kot v zgledu 4.4.1) brezpogojna verjetnost dogodka A razlikuje od njegove pogojne verjetnosti $P(A/B)$, rečemo, da je dogodek A odvisen od dogodka B. Zakaj tako poimenovanje, je jasno: dejstvo, da velja $P(A) \neq P(A/B)$, pove, da nastop dogodka B spremeni pogoje za nastop dogodka A.

V takšnem primeru se hitro vidi, da je tudi $P(B)$ različna od pogojne verjetnosti $P(B/A)$, zato je odvisnost dveh dogodkov vedno vzajemna in lahko preprosto rečemo, da sta dogodka odvisna.

Obratno pa definiramo:

$$A \text{ in } B \text{ neodvisna} \iff P(A) = P(A/B)\,, \quad P(B) = P(B/A)$$

Če sta dogodka A in B neodvisna, se obrazec (5), s katerim računamo verjetnost produkta, bistveno poenostavi:

$$A \text{ in } B \text{ neodvisna} \implies P(AB) = P(A)\,P(B) \tag{6}$$

Vzemimo obratno, da velja za dogodka A in B enakost $P(AB) = P(A)P(B)$. Potem imamo iz obrazca (1) takoj

$$P(A/B) = \frac{P(AB)}{P(B)} = \frac{P(A)\,P(B)}{P(B)} = P(A)$$

Izjavi "A in B sta neodvisna" in "za dogodka A in B velja enakost $P(AB) = P(A)P(B)$" sta torej ekvivalentni in lahko rečemo:

Dogodka A in B sta neodvisna natanko takrat, ko je verjetnost njunega produkta enaka produktu njunih (brezpogojnih) verjetnosti.

POMNI: Pojmov "neodvisna dogodka" in "nezdružljiva dogodka" nikakor ne smemo zamenjevati. Nezdružljiva dogodka sta v nekem smislu "izjemno odvisna", saj nastop enega onemogoči nastop drugega in obratno.

ZGLED 4.4.2:

Iz posode, v kateri imamo 4 črne in 6 belih kroglic, dvakrat na slepo izberemo po eno kroglico. Kolikšna je verjetnost dogodka, da je prva izvlečena kroglica bela, druga pa črna,

a) če po prvem izbiranju izvlečeno kroglico vrnemo v posodo;

b) če drugo izberemo, ne da bi prej vrnili prvo?

a) Verjetnost, da prvič potegnemo belo kroglico, je $P(B) = 6/10$. Ker kroglico nato vrnemo v posodo, imamo pred drugim izbiranjem v posodi spet 4 črne in 6 belih kroglic; dogodek, da potegnemo črno kroglico, ima verjetnost $P(C) = 4/10$ in je neodvisen od dogodka B, zato je po obrazcu (6) verjetnost produkta enaka produktu posameznih vrednosti, $P(BC) = P(B)P(C) = (6/10)(4/10) = 24/100 = 0{\cdot}24$.

b) Verjetnost, da v prvem poskusu izvlečemo belo kroglico, je spet $P(B) = 6/10$. Ker kroglice ne vrnemo, imamo za drugo izbiranje (pri pogoju, da se je zgodil dogodek B) v posodi 5 belih in 4 črne kroglice, zaradi česar je $P(C/B) = 4/9$, po obrazcu (5) pa dalje $P(BC) = P(B)P(C/B) = (6/10)(4/9) = 24/90 = 0{\cdot}2\overline{6}$.

Če imamo kroglice v posodi za neko populacijo, dve kroglici, ki ju na slepo izberemo iz nje, pa za slučajen vzorec, lahko rečemo, da smo v zadnjem zgledu opisali dve v osnovi različni možnosti za nastanek takšnega vzorca. V primeru a) smo naredili vzorec tako, da smo posamezne elemente vračali v populacijo, v primeru b) pa je vzorec nastal brez vračanja. V prvem primeru se v vzorcu posamezen element iz populacije lahko večkrat ponovi, zato rečemo, da gre za **vzorce s ponavljanjem**; v drugem primeru imamo **vzorce brez ponavljanja**, saj se poljuben element iz populacije lahko kvečjemu enkrat znajde v vzorcu.

Izkazalo se je, da je imel vzorec s ponavljanjem, dobljen z vračanjem elementov, lepo lastnost: ker z vračanjem elementov v populacijo pred vsakim izbiranjem obnovimo prvotno stanje, so dogodki, da proučevana statistična spremenljivka (v našem primeru barva kroglice) zavzame na elementih vzorca določene vrednosti, med seboj neodvisni, kar bistveno olajša nadaljnje delo. Pri vzorčenju brez ponavljanja takšne neodvisnosti ni: kakšen je naslednji element v vzorcu, je odvisno (tudi) od tega, kateri elementi so bili izbrani pri predhodnih izbiranjih. V našem primeru je dogodek, da v drugem poskusu izvlečemo črno kroglico, odvisen od tega, kakšno kroglico smo izvlekli pri prvem izbiranju.

ZGLED 4.4.3:

Dva strelca, ki zadevata svoj cilj (tarčo) z verjetnostjo $P(C_1) = 4/5$ oziroma $P(C_2) = 3/5$, sta po enkrat ustrelila proti tarči. Kolikšna je verjetnost, da je bila pri tem tarča dvakrat zadeta, če je bila zagotovo zadeta vsaj enkrat?

Naj bo A dogodek, da je tarča zadeta dvakrat, B pa dogodek, da je zadeta vsaj enkrat; iščemo $P(A/B) = P(AB)/P(B)$. Ker je v našem primeru dogodek A način dogodka B, se obrazec nekoliko poenostavi, $P(AB) = P(A)$ in $P(A/B) = P(A)/P(B)$. Iz besedila sledi, da je $P(A) = P(C_1 C_2)$. Predpostavili bomo, da sta dogodka C_1 in C_2 neodvisna, tedaj $P(A) = P(C_1)\,P(C_2)$. Verjetnost dogodka B lahko izračunamo neposredno, kot vsoto verjetnosti treh nezdružljivih načinov, ki ga sestavljajo (katerih?), veliko hitreje pa gre posredno, če od 1 odštejemo verjetnost njegove negacije, torej dogodka, da sta oba strelca zgrešila tarčo:

$$P(B) = 1 - P(B') = 1 - P(C_1' C_2') = 1 - P(C_1')\,P(C_2') =$$
$$= 1 - (1 - P(C_1))(1 - P(C_2)) = 1 - (1 - \frac{4}{5})(1 - \frac{3}{5}) =$$
$$= 1 - \frac{1}{5} \cdot \frac{2}{5} = \frac{23}{25} = 0{\cdot}92$$

V tem računu smo upoštevali, da sta pri neodvisnih dogodkih C_1 in C_2 neodvisni tudi njuni negaciji. Tako je končno

$$P(A/B) = P(A)/P(B) = \frac{0{\cdot}48}{0{\cdot}92} \approx 0{\cdot}52$$

Za množico več kot dveh dogodkov velja posplošitev obrazca (5),

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)\,P(A_2/A_1)...P(A_n/(A_1 \cap A_2 \cap ... \cap A_{n-1})) \quad (7)$$

Tudi ta obrazec je vsebinsko neobčutljiv na spremembe vrstnega reda faktorjev na levi strani; začnemo z brezpogojno verjetnostjo kateregakoli faktorja, jo pomnožimo s pogojno verjetnostjo kateregakoli drugega faktorja (glede na tistega, ki je bil prvi uporabljen) in tako dalje, dokler nam ne ostane en sam faktor, pri katerem je treba v obrazcu upoštevati pogojno verjetnost glede na produkt vseh ostalih, pred njim uporabljenih faktorjev.

ZGLED 4.4.4:

Iz posode, v kateri imamo 4 črne in 6 belih kroglic, trikrat po vrsti izvlečemo po eno kroglico, ne da bi izvlečeno kroglico vračali v posodo. Izračunaj verjetnost dogodka, da je prva izvlečena kroglica bela, druga črna in tretja spet bela!

Označimo omenjene dogodke z B_1, C_2 in B_3. Računali bomo $P(B_1 \cap C_2 \cap B_3)$. Po obrazcu (7) imamo

$$P(B_1 \cap C_2 \cap B_3) = P(B_1)\,P(C_2/B_1)\,P(B_3/(B_1 \cap C_2)) = \frac{6}{10} \cdot \frac{4}{9} \cdot \frac{5}{8} = \frac{1}{6}$$

Če izvlečeno kroglico vsakič vrnemo, so dogodki neodvisni, iskana verjetnost je v tem primeru enaka

$(6/10)(4/10)(6/10) = 144/1000$

Če sta dogodka A in B neodvisna, je verjetnost produkta kar enaka produktu $P(A)P(B)$. Če hočemo na podoben način poenostaviti obrazec (7) za verjetnost produkta več kot dveh dogodkov, moramo zahtevati kar precej: poljuben dogodek iz množice dogodkov $\{A_1, A_2, ..., A_n\}$ mora biti neodvisen ne le od vsakega drugega dogodka iz te množice, ampak tudi od kakršnegakoli produkta, ki ga lahko sestavimo iz preostalih dogodkov. Če je tak pogoj izpolnjen, pravimo, da so dogodki $A_1, A_2, ..., A_n$ *v celoti neodvisni.* Za takšne dogodke je

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)\,P(A_2)...P(A_n) \quad (8)$$

Naslednji zgled bo pokazal, da dogodki, ki so paroma neodvisni (torej po dva in dva), niso nujno tudi v celoti neodvisni.

ZGLED 4.4.5:

Pravilen tetraeder iz homogene snovi ima na eni stranski ploskvi rdeče pike, na drugi modre, na tretji bele, na četrti stranski ploskvi pa najdemo pike vseh treh barv. Tetraeder zakotalimo (kot igralno kocko). Naj bo A dogodek, da je, ko se tetraeder ustavi, na spodnji ploskvi kakšna rdeča pika, B dogodek, da je na spodnji ploskvi kakšna modra pika, in C dogodek, da je na spodnji ploskvi kakšna bela pika. Zaradi predpostavk o barvah na tetraedru je

$$P(A) = P(B) = P(C) = \frac{2}{4} = \frac{1}{2}$$

Enako lahko ugotovimo, da velja

$$P(A/B) = P(A/C) = P(B/C) = \frac{1}{2}$$

(rezultat utemelji z besedami za vsako od pogojnih verjetnosti posebej!). Zato so po definiciji dogodki A, B in C paroma neodvisni, saj je brezpogojna verjetnost vsakega od njih enaka pogojni verjetnosti glede na vsakega od preostalih dveh dogodkov. Na drugi strani pa recimo dogodek A ni neodvisen od produkta BC. Ta produkt nastopi edino v primeru, da tetraeder obstane na "pisani" stranski ploskvi, kjer so pike vseh treh barv. V tem primeru so med njimi zagotovo tudi rdeče pike in je zato

$$P(A/(B \cap C)) = 1 \neq P(A) = \frac{1}{2}$$

Dogodki A, B in C iz tega poskusa so torej paroma neodvisni, v celoti neodvisni pa niso.

Kadar bomo govorili o neodvisnosti dogodkov, bomo v splošnem predpostavljali, da so v celoti neodvisni, ali drugače, da velja obrazec (8); v posebnih primerih pa bomo opozorili na odstopanja.

ZGLED 4.4.6:

Izdelki imajo lahko dve vrsti napak, napako A in napako B. Prvo ima v povprečju 5% vseh izdelkov, drugo pa 10% vseh izdelkov. Znano je, da je med izdelki, ki imajo napako B, 8% tistih, ki imajo tudi napako A. V pošiljko so na slepo izbrali 30 izdelkov ter pri pregledu ugotovili, da sta med njimi 2 z napako A. Kolikšna je verjetnost dogodka, da nobeden od izdelkov v tej pošiljki nima napake B?

Ker lahko predpostavimo, da so napake nekega izdelka neodvisne od napak drugih izdelkov, je iskana verjetnost enaka produktu tridesetih faktorjev, pri čemer osemindvajsetkrat nastopi verjetnost, da izdelek brez napake A nima napake B, dvakrat pa verjetnost, da je brez napake B izdelek, ki ima napako A. Torej je ta verjetnost enaka $\{P(B'/A')\}^{28} \cdot \{P(B'/A)\}^2$. Pogojni verjetnosti bomo izračunali s pomočjo brezpogojnih verjetnosti $P(A) = 0{\cdot}05$ in $P(B) = 0{\cdot}10$ ter pogojne verjetnosti $P(A/B) = 0{\cdot}08$. Tako izpolnimo naslednjo preglednico, v kateri imamo na križiščih vrstic in stolpcev zapisane verjetnosti produktov ustreznih dogodkov:

	Z napako B	Brez napake B	
Z napako A	0·008	0·042	0·05
Brez napake A	0·092	0·858	0·95
	0·10	0·90	1·00

Najprej smo dobili

$$P(AB) = P(BA) = P(B)P(A/B) = 0{\cdot}10 \cdot 0{\cdot}08 = 0{\cdot}008$$

Nato pa je

$$P(A) = P(AG) = P(A(B+B')) = P(AB) + P(AB')$$

in zato

$$P(AB') = P(A) - P(AB) = 0{\cdot}05 - 0{\cdot}008 = 0{\cdot}042$$

podobno pa izračunamo še $P(A'B) = 0{\cdot}092$ in $P(A'B') = 0{\cdot}858$.

S takšno tabelo si dostikrat pomagamo; vsota verjetnosti v posamezni vrstici mora biti enaka pripadajoči brezpogojni verjetnosti, enako velja za stolpce. Vsota verjetnosti vseh možnih produktov mora biti enaka 1. Dalje gre račun hitro:

$$P(B'/A') = \frac{P(B'A')}{P(A')} = \frac{0{\cdot}858}{0{\cdot}95} \approx 0{\cdot}9032$$

$$P(B'/A) = \frac{P(B'A)}{P(A)} = \frac{0{\cdot}042}{0{\cdot}05} \approx 0{\cdot}84$$

in iskana verjetnost je

$$0{\cdot}9032^{28} \cdot 0{\cdot}84^2 \approx 0{\cdot}041$$

Prav malo verjetno je torej, da v pošiljki ne bi bilo nobenega izdelka z napako tipa B.

V A J E

1. Dogodka A in B sta nezdružljiva, $P(A) > 0$ in $P(B) > 0$. Pokaži tudi s formalnim računom, da sta dogodka odvisna.

2. Za dogodka A in B iz nekega poskusa velja
$P(A) = 3/8$, $P(B) = 5/8$, $P(A \cap B) = 1/4$.
Izračunaj pogojni verjetnosti $P(A/B)$ in $P(B/A)$.

3. Naj bodo A, B in C med seboj neodvisni dogodki in $P(A) = 1/2$, $P(B) = 1/3$, $P(C) = 2/3$. Izračunajte verjetnosti dogodkov $A \cap B$, $A \cup B$ in $A \cap B' \cap C'$.

4. Verjetnost, da prvi strelec zadene tarčo, je $P(Z_1) = 0{\cdot}6$, za drugega pa $P(Z_2) = 0{\cdot}5$. Izračunajte verjetnosti dogodkov:
   - A - da tarča ni bila zadeta, če je prvi strelec zgrešil;
   - B - oba strelca zadeneta, če je bila tarča vsaj enkrat zadeta;
   - C - prvi strelec je zgrešil, pri pogoju, da tarča ni zadeta;
   - D - drugi je zgrešil, če je tarča zadeta vsaj enkrat;
   - E - prvi je zadel, če je tarča zadeta natanko enkrat.

   (Opomba: Pri tej in vseh sorodnih nalogah predpostavljamo, da sta dogodka Z_1 in Z_2 neodvisna.)

5. Iz škatle s 3 belimi in 7 rdečimi kroglami vlečemo po eno kroglo in izvlečenih krogel ne vračamo. Izračunajte verjetnosti dogodkov
   - A - izvlečemo štiri rdeče krogle zapored;
   - B - izvlečemo najprej dve beli, nato dve rdeči krogli.

6. V škatli je 5 belih in 10 rdečih kroglic. Na slepo štirikrat potegnemo po eno kroglico. Kolikšna je verjetnost, da se šele v četrtem poskusu pokaže rdeča kroglica,
   a) če izvlečeno kroglico vsakič vrnemo v škatlo;
   b) če kroglic ne vračamo v škatlo?

7. Kolikšna je verjetnost, da je pri sočasnem metu dveh poštenih igralnih kock padla vsaj ena trojka, če je bila vsota pik na obeh kockah 6?

8. Trikrat zapored vržemo pošteno igralno kocko. Kolikšna je verjetnost dogodka, da prvič pade sodo število pik, drugič šestica in tretjič manj kot 5 pik?

9. Kolikšna je verjetnost, da na dveh kockah v prvem metu dobimo vsoto 9 ali – če se to ni zgodilo – v ponovljenem metu vsoto 7?

10. Petdesetkrat ustrelimo proti tarči, verjetnost zadetka pri posameznem strelu je $P(Z) = 0{\cdot}08$ in se ne spreminja. Kolikšna je verjetnost dogodka A, da je v teh 50 strelih tarča vsaj enkrat zadeta?

11. △ Kolikšna mora biti najmanj verjetnost zadetka pri posameznem strelu, da lahko pri petih strelih z verjetnostjo večjo od 0·99 pričakujemo vsaj en zadetek?

12. Dva igralca mečeta kovanec drug za drugim. Zmaga tisti, pri katerem se prej pojavi grb. Zapiši izraz za verjetnost, da zmaga igralec, ki je igro začel.

13. Reši enako nalogo za primer, ko igralca mečeta kocko in zmaga tisti, ki prvi dobi šestico.

14. Izdelke istega tipa izdelujeta dve tovarni, prva 60% in druga 40% od celotne proizvodnje. Med izdelki prve tovarne je 60% I. in 40% II. kvalitete; med izdelki druge tovarne je prvovrstnih 80%, ostali so II. kvalitete. Naj pomeni A dogodek, da je na slepo izbran izdelek izdelan v prvi tovarni, B pa dogodek, da je na slepo izbran izdelek I. kvalitete. Izračunaj verjetnosti dogodkov

$A,\ B,\ A',\ B', AB,\ A'B,\ AB',\ A'B',\ A/B,\ A/B',\ B/A,\ B/A'$

še prej pa vsakega od njih podrobno opiši z besedami.

15. V nekem razredu imajo učenci negativne ocene samo iz matematike ali tujega jezika (ali obojega). 20% učencev ima negativno oceno iz matematike (dogodek M), 10% iz tujega jezika (dogodek T), 5% iz obojega. Na slepo izberemo enega učenca. Izračunaj verjetnosti dogodkov:

a) - izbrani učenec ima vsaj eno negativno oceno;

b) - ima negativno oceno iz matematike, če ima zagotovo negativno oceno iz tujega jezika;

c) - nima negativne ocene iz matematike, če zagotovo nima negativne ocene iz tujega jezika.

16. Verjetnosti, da posamezen strelec zadene tarčo, so po vrsti $1/6, 1/4$ in $1/3$. Ko so vsak po enkrat ustrelili proti tarči, je bila ta natanko enkrat zadeta. Izračunaj verjetnost dogodka, da jo je zadel prvi strelec.

17. Lovec strelja na lisico iz razdalje 25 metrov. Verjetnost, da jo zadene, je 0·60. Če zgreši, mu lisica uide na razdaljo 50 metrov, preden lahko spet strelja nanjo. Izračunaj verjetnost dogodka, da je lisica s prvim ali drugim strelom zadeta, če je pri teh oddaljenostih verjetnost za zadetek

a) obratno sorazmerna razdalji;

b) obratno sorazmerna kvadratu razdalje.

18. △ Iz kupa kart na slepo sočasno izvlečemo dve karti, eno od njiju obrnemo in ugotovimo, da je to kralj. Karti premešamo, nato pa spet obrnemo eno od njiju. Kolikšna je verjetnost, da je obrnjena karta

a) as;

b) kralj?

19. △ Na šoli je 8% učencev z okvarami vida in 12% učencev s slabo držo. Od tistih, ki imajo slabo držo, jih ima 10% tudi težave z vidom. V slučajnem vzorcu so od 32 učencev štirje imeli okvaro vida. Kolikšna je verjetnost, da v tem vzorcu ni nobenega učenca s slabo držo? (Nasvet: pomagaj si z zgledom 4.4.6.)

20. △ △ Napiši računalniški program, ki bo simuliral sočasen met dveh poštenih igralnih kock in poskusi dobiti približek za pogojno verjetnost, ki smo jo računali v zgledu 4.4.1!

4.5 Dvofazni poskusi in Bayesov obrazec

Privzemimo, da določen poskus poteka v dveh fazah oziroma stopnjah. V prvi fazi nastopi natanko eden od dogodkov iz popolnega sistema dogodkov $\{H_1, H_2, ..., H_n\}$, od tega, kateri od njih se je pripetil, pa so odvisni pogoji za drugo fazo poskusa; v zvezi z drugo fazo proučujemo slučajni dogodek A: zanima nas, kako izračunati njegovo verjetnost, če poznamo verjetnosti dogodkov iz popolnega sistema prve faze in pripadajoče pogojne verjetnosti dogodka A. V izpeljavi obrazca za iskano verjetnost $P(A)$ bomo upoštevali nekatera že znana dejstva:

- vsota (nezdružljivih) dogodkov iz popolnega sistema dogodkov je gotovi dogodek G,
- za poljuben slučajni dogodek A velja $G \cap A = A$,
- verjetnost vsote poljubnega števila nezdružljivih dogodkov je vsota njihovih verjetnosti.

Zaradi povedanega je

$$A = G \cap A = (H_1 \cup H_2 \cup ... \cup H_n) \cap A$$

kar pa lahko zaradi distributivnostnega zakona napišemo tudi takole:

$$A = (H_1 \cap A) \cup (H_2 \cap A) \cup ... \cup (H_n \cap A)$$

Ker so nezdružljivi dogodki iz prve faze, so takšni tudi produkti v izrazu na desni in je tako

$$P(A) = P(H_1 \cap A) + P(H_2 \cap A) + ... + P(H_n \cap A)$$

Za vsak sumand na desni strani enačaja uporabimo pravilo za verjetnost produkta (odvisnih) dogodkov,

$$P(H_i \cap A) = P(H_i)\, P(A/H_i) \quad (i = 1, 2, ..., n)$$

kar nam da

$$P(A) = P(H_1)\, P(A/H_1) + P(H_2)\, P(A/H_2) + ... + P(H_n)\, P(A/H_n) \quad (1)$$

ali krajše,

$$P(A) = \sum_{i=1}^{n} P(H_i)\, P(A/H_i) \quad (2)$$

Formuli (1) oziroma (2) imenujemo **obrazec za popolno verjetnost dogodka** A. Ime izvira iz dejstva, da je verjetnost dogodka A razčlenjena na vse možne načine, na katere se dogodek A zgodi skupaj (ali v nadaljevanju) z enim od dogodkov H_i iz popolnega sistema dogodkov prve faze tega poskusa.

ZGLED 4.5.1:

V prvi posodi imamo 7 belih in 3 črne krogle, v drugi 4 bele in 5 črnih. Iz prve posode na slepo izberemo eno kroglo in jo prenesemo v drugo posodo, nato pa iz slednje spet na slepo izvlečemo eno kroglo. Kolikšna je verjetnost, da je ta krogla bela?

V prvi fazi imamo popoln sistem iz dveh dogodkov,

H_1 - iz prve posode prenesemo v drugo posodo belo kroglo;

H_2 - iz prve posode prenesemo v drugo posodo črno kroglo.

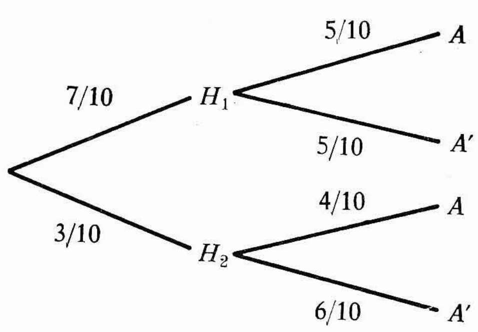
Verjetnosti teh dveh dogodkov sta po vrsti $P(H_1) = 7/10$ in $P(H_2) = 3/10$. Če nastopi v prvi fazi dogodek H_1, imamo pred izbiranjem iz druge posode 5 belih in 5 črnih krogel, zato je ustrezna pogojna verjetnost, da iz druge posode potegnemo belo kroglo, $P(A/H_1) = 5/10$. Če pa je v prvi fazi nastopil dogodek H_2, imamo v drugi posodi 4 bele in 6 črnih krogel, zato je $P(A/H_2) = 4/10$. Verjetnost dogodka A je po obrazcu (1)

$$P(A) = P(H_1)\, P(A/H_1) + P(H_2)\, P(A/H_2) =$$

$$= \frac{7}{10} \cdot \frac{5}{10} + \frac{3}{10} \cdot \frac{4}{10} = \frac{47}{100}$$

Pri računanju pogojnih verjetnosti $P(A/H_i)$ si prihranimo precej napak, če ob pisanju postopek glasno razlagamo v celih stavkih, na primer: "$P(A/H_2)$ je verjetnost, da iz druge posode izvlečemo belo kroglo, če smo v prvi fazi vanjo prenesli črno kroglo."

Manj izkušeni si lahko pomagajo tudi z drevesom, s katerim pregledamo vse načine dogodka A in pripadajoče verjetnosti. Kot primer narišimo drevo, ki ustreza podatkom iz zgleda 4.5.1.



Brez težav preverimo trditev: verjetnost $P(A)$ dobimo tako, da prehodimo vse poti od "korenin" do tistih vršičkov drevesa, kjer raste A; po poti množimo vse verjetnosti, na katere naletimo, zmnožke pa seštejemo. (Kdor verjame samo v pokončna drevesa, bo pač zavrtel učbenik za en pravi kot v pozitivnem smislu.)

Iz izpeljave obrazca za popolno verjetnost se vidi, da druga faza ni nujno tudi časovno za prvo, ampak lahko potekata sočasno. Bistveno je, da se proučevani dogodek A iz druge faze lahko zgodi vedno le z enim od dogodkov iz popolnega sistema dogodkov $\{H_1, H_2, ..., H_n\}$. Te dogodke imenujemo tudi **hipoteze**; če ne vemo, kateri od njih se je zgodil, imamo pač na voljo n različnih hipotez o tem, kako je nastopil dogodek A.

Poleg računanja popolne verjetnosti dogodka A iz druge faze nas lahko pri dvofaznih poskusih zanima še tole vprašanje:

Denimo, da se je dogodek A v drugi fazi zgodil. Kot smo že rekli, se to lahko zgodi na n različnih načinov, kot $H_1 \cap A$, $H_2 \cap A$,..., $H_n \cap A$. Zanima nas verjetnost dogodka, da je A nastopil skupaj z določenim dogodkom H_i iz prve faze, ali drugače, iščemo pogojne verjetnosti dogodkov H_i glede na dogodek A iz druge faze. Iz

$$P(A \cap H_i) = P(A)\,P(H_i/A) = P(H_i)\,P(A/H_i) \quad (i = 1, 2, ..., n)$$

dobimo z razreševanjem enačbe na $P(H_i/A)$ t.i. **Bayesov obrazec**

$$P(H_i/A) = \frac{P(H_i)\,P(A/H_i)}{P(A)} \qquad (i = 1, 2, ..., n) \tag{3}$$

ali v razviti obliki (upoštevaje, kako dobimo verjetnost $P(A)$ iz izhodiščnih podatkov)

$$P(H_i/A) = \frac{P(H_i)\,P(A/H_i)}{P(H_1)P(A/H_1) + ... + P(H_n)P(A/H_n)} \tag{4}$$

Obrazec jasno pove, kaj zmore:

Z Bayesovim obrazcem računamo verjetnosti posameznih dogodkov v prvi fazi (hipotez H_i), pri pogoju, da se je dogodek v drugi fazi zgodil.

ZGLED 4.5.2:

Izvedli smo dvofazni poskus, opisan v zgledu 4.5.1, in pri izbiranju iz druge posode ugotovili, da je izvlečena krogla bela. Izračunaj verjetnost, da je bila tudi krogla, ki smo jo prenesli iz prve posode v drugo, bela!

V zgledu 4.5.1 smo uporabljali naslednje oznake za dogodka iz prve faze:

H_1 - iz prve posode prenesemo v drugo posodo belo kroglo,

H_2 - iz prve posode prenesemo v drugo posodo črno kroglo.

Tudi potrebne podatke smo že pripravili:

$P(H_1) = 7/10$, $P(H_2) = 3/10$, $P(A/H_1) = 5/10$, $P(A/H_2) = 4/10$. Pogojnih verjetnosti niti ne potrebujemo, ker smo v zgledu 4.5.1 že izračunali popolno verjetnost $P(A) = 47/100$, zato lahko Bayesov obrazec uporabimo kar v obliki (3), tako da je

$$P(H_1/A) = \frac{P(H_1)\,P(A/H_1)}{P(A)} = \frac{0{\cdot}7 \cdot 0{\cdot}5}{0{\cdot}47} \approx 0{\cdot}745$$

ZGLED 4.5.3:

Znano je, da laboratorijski poskusi ne pokažejo vedno povsem natančno, ali je oseba dejansko okužena z neko boleznijo. Vzemimo takle (izmišljen!) primer: v neki populaciji imata 2% prebivalcev tuberkulozo, 8% prebivalcev ima lažja pljučna obolenja, ostali so (po tej plati) zdravi. Določen laboratorijski test, s katerim preizkušamo okuženost s tuberkulozo, da pozitiven rezultat pri 98% dejansko okuženih, žal pa nanj reagira tudi 8% tistih, ki imajo le lažja pljučna obolenja, in celo 3% zdravih prebivalcev. Na slepo izberemo enega od prebivalcev in ugotovimo, da je ta test dal pozitiven rezultat. Kolišna je verjetnost, da je ta prebivalec dejansko okužen s tuberkulozo?

Dogodki H_i naj pomenijo, da je slučajno izbrani prebivalec okužen s tuberkulozo, da je lažji pljučni bolnik oziroma da je zdrav; dogodek A nastopi, če je rezultat testa pozitiven. Iz besedila naloge razberemo, da je

$$\begin{array}{lcllcl} P(H_1) & = & 0{\cdot}02 & P(A/H_1) & = & 0{\cdot}98 \\ P(H_2) & = & 0{\cdot}08 & P(A/H_2) & = & 0{\cdot}08 \\ P(H_3) & = & 0{\cdot}90 & P(A/H_3) & = & 0{\cdot}03 \end{array}$$

Iščemo verjetnost dogodka H_1 ("okužen s TBC") pri pogoju A ("test pozitiven"). Po Bayesovem obrazcu je

$$P(H_1/A) = \frac{P(H_1)\,P(A/H_1)}{P(H_1)P(A/H_1) + P(H_2)P(A/H_2) + P(H_3)P(A/H_3)}$$

kar pri naših podatkih pomeni

$$P(H_1/A) = \frac{0{\cdot}02 \cdot 0{\cdot}98}{0{\cdot}02 \cdot 0{\cdot}98 + 0{\cdot}08 \cdot 0{\cdot}08 + 0{\cdot}90 \cdot 0{\cdot}03} \approx 0{\cdot}37$$

Rezultat je pravzaprav zelo poučen: med na slepo izbranimi prebivalci, za katere je laboratorijski preizkus dal pozitiven rezultat, jih ima v povprečju le 37% res tuberkulozo. Zato v praksi (dobri) zdravniki le izjemoma postavljajo diagnoze na osnovi rezultatov enega samega testa.

Pogojne verjetnosti $P(H_i/A)$, ki jih izračunamo po Bayesovem obrazcu (4), so računane "za nazaj", ko se je dogodek v drugi fazi poskusa že zgodil. Zato te verjetnosti imenujeno **aposteriorne verjetnosti** dogodkov iz popolnega sistema v prvi fazi. V splošnem se razlikujejo od apriornih verjetnosti $P(H_i)$, s katerimi ocenjujemo možnosti za nastop dogodkov H_i pred dejansko izvedbo poskusa. Med najpomembnejše uporabe Bayesovega obrazca sodi ravno "popravljanje" teh verjetnosti na osnovi podatkov o tem, kaj se dogaja v drugi fazi poskusa.

Dvofazne poskuse v praksi dostikrat srečamo, na primer v kontroli kvalitete, vendar si podrobnejše obravnave v tem elementarnem učbeniku ne moremo privoščiti.

V A J E

1. Imamo tri enake škatle: v prvi 2 črni in 1 belo kroglico, v drugi 3 bele in 2 črni, v tretji 1 belo in 3 črne kroglice. Na slepo sežemo v eno od škatel in na slepo izvlečemo eno kroglico. Kolikšna je verjetnost, da je ta kroglica črna?

2. Imamo iste škatle kot v prvi nalogi. Vržemo kocko; če padejo 1, 2 ali 3 pike, izberemo prvo posodo, če padejo 4 ali 5 pik, izberemo drugo posodo, če pade šestica, sežemo v tretjo posodo.

a) Izračunaj verjetnost, da bo izvlečena kroglica bela.

b) Predpostavimo, da smo pri poskusu dobili belo kroglico. Kolikšna je verjetnost, da je pri metu kocke padla šestica?

3. Iz škatle, v kateri je 5 belih in 7 črnih krogel, predenemo dve naključno izbrani krogli v škatlo z 2 belima in 5 črnimi kroglami, nato pa iz te na slepo izvlečemo eno kroglo. Kolikšna je verjetnost, da je ta krogla črna?

4. $\triangle$ Iz škatle, v kateri je 5 belih in 7 črnih krogel, predenemo na slepo dve krogli v škatlo z 2 belima in 5 črnimi kroglami, nato pa iz te na slepo izvlečemo dve krogli. Kolikšna je verjetnost, da sta obe izvlečeni krogli iste barve?

5. Pri poskusu iz prejšnje naloge smo pri izbiranju iz druge škatle dobili raznobarvni kroglici. Kolikšna je verjetnost, da sta bili tudi kroglici, ki smo ju preneseli iz prve škatle, raznobarvni?

6. Na neko gimnazijo se vpisujejo učenci s treh osnovnih šol, od tega 20% s prve, 45% z druge in 35% s tretje osnovne šole. Verjetnosti, da učenec s posamezne osnovne šole uspešno dokonča prvi letnik gimnazije, so po vrsti 0·95, 0·85 in 0·70. Izračunaj

a) verjetnost dogodka, da na slepo izbran učenec prvega letnika te gimnazije uspešno konča prvi letnik;

b) verjetnost dogodka, da je na slepo izbrani učenec, ki je uspešno končal prvi letnik, prišel na gimnazijo s prve osnovne šole.

(Predlog: razmišljaj o interpretaciji dobljenih rezultatov. Kaj za prakso pomenijo izračunane verjetnosti?)

7. Na izpit iz matematike je prišlo 50 študentov, od katerih jih je 38 preštudiralo poglavje o verjetnostnem računu. Na izpitu je tudi ena naloga iz verjetnostnega računa. Verjetnost, da to nalogo privede do pravilnega rezultata študent, ki je poglavje preštudiral, je 0·90; verjetnost, da ima pravilen rezultat študent, ki tega poglavja ni študiral, je 0·15 (prepisovanje, "trenutni navdih" ipd.).

a) Kolikšna je verjetnost, da slučajno izbrani študent pravilno reši nalogo iz verjetnostnega računa?

b) Na slepo izberemo enega študenta (izmed 50) in ugotovimo, da je pravilno rešil nalogo iz verjetnostnega računa. Kolikšna je verjetnost, da je pravilni rezultat dobil na nepošten način (da torej poglavja o verjetnostnem računu ni preštudiral)?

8. (Samo za vojake po srcu in duši!)

Na letalo ustrelimo dvakrat. Pri prvem strelu je verjetnost za zadetek 0·3, pri drugem strelu pa 0·6. Enkrat zadeto letalo se zruši z verjetnostjo 0·2, dvakrat zadeto pa z verjetnostjo 0·9.

a) Kolikšna je verjetnost, da s tema dvema streloma zrušimo letalo?

b) Denimo, da smo letalo uspešno sestrelili. Kolikšna je verjetnost, da se nam je to posrečilo že s prvim strelom?

9. V posodi imamo kroglico, za katero je znano samo to, da je lahko ali bela ali črna. V posodo dodamo eno belo kroglico, nato pa na slepo izberemo eno od dveh kroglic. Predpostavimo, da je izvlečena kroglica bela. Kolikšna je tedaj verjetnost, da je bila tudi kroglica, ki je bila pred poskusom v posodi, bela?

10. V posodo, v kateri smo imeli prvotno 6 belih in 4 črne kroglice, nam je padla kroglica, za katero ne vemo, ali je bila bela ali črna. Iz posode na slepo izberemo dve kroglici in ugotovimo, da sta obe beli. Kolikšni sta verjetnosti dogodkov, da je bila kroglica, ki je padla v posodo, bela oziroma črna?

11. Izmed števil 1, 2, 3, 4, 5, 6, 7, 8 in 9 na slepo izberemo dve števili. Izračunaj verjetnost dogodka A, da sta obe števili lihi, če je njuna vsota sodo število.

12. Skakalčevi uspehi pri skoku v daljavo so močno odvisni od vremena; če je lepo vreme, zmaga na državnem prvenstvu z verjetnostjo 0·75, če dežuje, je verjetnost za zmago le 0·50. V Celju, kjer je povprečno ena tretjina septembrskih dni deževna, je 25. septembra osvojil naslov državnega prvaka. Kolikšna je verjetnost, da je 25. septembra v Celju deževalo?

(Predlog: razmisli, v kolikšni meri so pri taki nalogi izpolnjene predpostavke, na katerih gradimo klasični model verjetnosti. Komentiraj in kritiziraj.)

4.6 Zaporedja poskusov in Bernoullijev obrazec

Pri praktičnih uporabah verjetnostnega računa velikokrat srečamo zaporedja poskusov, ki so na kak način povezani med sabo. V splošnem je proučevanje takih zaporedij za nas pretežka naloga. Zato se bomo omejili na t.i. **zaporedja neodvisnih poskusov**, pri katerih je poljuben dogodek iz nekega poskusa neodvisen od kateregakoli dogodka ali produkta dogodkov iz drugih poskusov. Delovno področje si za začetek še nekoliko omejimo in definirajmo:

Zaporedje neodvisnih poskusov imenujemo - po švicarskem matematiku Jakobu Bernoulliju (1654 - 1705) - **Bernoullijevo zaporedje**, *če je v vsakem poskusu iz tega zaporedja popolni sistem dogodkov sestavljen samo iz dogodka A s konstantno verjetnostjo $P(A) = p > 0$ in njegovega nasprotnega dogodka A' z verjetnostjo $P(A') = 1 - P(A) = 1 - p = q$.*

Takšno zaporedje imamo lahko kar za ponavljanje istega poskusa.

ZGLED 4.6.1:

Klasičen primer za Bernoullijevo zaporedje so zaporedni meti kovanca; rezultati posameznega poskusa so gotovo neodvisni od izidov v prejšnjih metih, v vsaki ponovitvi pa nastopi bodisi grb (dogodek A, z verjetnostjo $P(A) = p = 1/2$) ali številka kot prvemu nasprotni dogodek $(P(A') = 1 - p = 1/2 = q)$.

ZGLED 4.6.2:

Pri metu kocke ima popolni sistem dogodkov sicer šest elementov, vendar lahko zaporedne mete kljub temu obravnavamo kot Bernoullijevo zaporedje s $p = 1/6$, če se zanimamo za posamezen izid, vsi ostali izidi pa sestavljajo njemu nasproten dogodek z verjetnostjo $q = 5/6$ v posamezni ponovitvi poskusa, o neodvisnosti poskusov pa tudi ne more biti dvoma, dokler govorimo o običajnih igralnih kockah.

Najprej nas bo zanimalo, kolikšna je verjetnost, da v n ponovitvah poskusa dogodek A, ki ima v posamezni ponovitvi verjetnost $p(A) = p$, nastopi natanko k - krat $(k = 0, 1, ..., n)$.

Za ogrevanje poskusimo na konkretnem, pa zelo preprostem primeru. Petkrat vrzimo kocko in vprašajmo po verjetnosti dogodka A, da pri tem natanko dvakrat pade šestica. Padec šestice v posameznem metu označimo s simbolom S, nastop kakega drugega izida z D. Potem se proučevani dogodek lahko zgodi na naslednje - med seboj nezdružljive - načine:

SSDDD, SDSDD, SDDSD, SDDDS, DSSDD,
DSDSD, DSDDS, DDSSD, DDSDS, DDDSS

Vsak od teh načinov je produkt dveh dogodkov S in treh dogodkov $S' = D$.

Čeprav nastopajo "sestavni deli" v različnih vrstnih redih, imajo vsi našteti načini zaradi komutativnosti produkta dogodkov enake verjetnosti, takšne kot prvi način, ta pa ima zaradi neodvisnosti zaporedja poskusov verjetnost

$$P(S\,S\,D\,D\,D) = P(S) \cdot P(S) \cdot P(D) \cdot P(D) \cdot P(D) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6}$$

Ker je vseh načinov 10 (na toliko načinov lahko med petimi meti izberem tista dva, v katerih nastopi šestica, vse ostalo je s tem že določeno), je

$$P(A) = 10 \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = \frac{1\,250}{7\,776} \approx 0{\cdot}161$$

Zdaj pa se, podprti z izkušnjami, vrnimo k splošnemu primeru. Eden od načinov, na katere se lahko zgodi zapovedana frekvenca dogodka v n ponovitvah poskusa, je tale: najprej se k - krat zgodi dogodek A, nato pa $(n - k)$ - krat nastopi njegova negacija (nasprotni dogodek A'), kar simbolično pomeni, da nastopi dogodek

$$\underbrace{A \cap A \cap \ldots \cap A}_{k\,\text{faktorjev}} \cap \underbrace{A' \cap A' \cap \ldots \cap A'}_{n-k\,\text{faktorjev}} \qquad (1)$$

Zaradi neodvisnosti poskusov je verjetnost dogodka (1) enaka produktu verjetnosti posameznih faktorjev, torej $p^k(1-p)^{n-k}$. Enake verjetnosti imajo tudi drugi načini dogodka, da v n ponovitvah poskusa A nastopi natanko k - krat. Vseh takšnih načinov je natanko toliko, kolikor je možnosti za izbiro k poskusov, v katerih nastopi A, iz celotne množice poskusov, torej $\binom{n}{k}$. Ker so načini med seboj nezdružljivi, je verjetnost proučevanega dogodka enaka vsoti verjetnosti posameznih načinov, oziroma vsoti $\binom{n}{k}$ sumandov, ki so vsi enaki $p^k(1-p)^{n-k}$. Tako smo ugotovili:

Verjetnost $P(n; p; k)$, da dogodek A, ki ima v posamezni ponovitvi poskusa konstantno verjetnost $P(A) = p$, v n med seboj neodvisnih ponovitvah poskusa nastopi natanko k - krat, je enaka

$$P(n; p; k) = \binom{n}{k} p^k \,(1-p)^{n-k} \qquad (2)$$

Formulo (2) običajno imenujemo **Bernoullijev obrazec**. Pove nam, kakšne so verjetnosti posameznih možnih frekvenc $(k = 0, 1, ..., n)$ proučevanega dogodka v n ponovitvah poskusa.

ZGLED 4.6.3:

Kakšna je verjetnost dogodka B, da v osmih metih kovanca grb pade vsaj dvakrat?

Nalogo lahko neposredno rešimo tako, da seštejemo verjetnosti sedmih nezdružljivih načinov dogodka B, da grb pade natanko dvakrat, natanko trikrat, ..., natanko osemkrat. Manj dela pa imamo, če od 1 odštejemo vrednost nasprotnega dogodka B', da pade grb manj kot dvakrat, torej ničkrat ali enkrat,

$$P(B) = 1 - \left[\binom{8}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^8 + \binom{8}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^7\right] =$$

$$= 1 - \left[\frac{1}{256} + \frac{8}{256}\right] \approx 0'965$$

ZGLED 4.6.4:

V seriji N izdelkov je M pokvarjenih. Iz te serije na slepo izberemo drugega za drugim n izdelkov. Kolikšna je verjetnost dogodka A, da je med njimi natanko k pokvarjenih,

a) če po pregledu izdelek vrnemo v serijo;

b) če pregledani izdelek izločimo pred izbiranjem naslednjega in torej delamo vzorce brez ponavljanja?

V primeru a) z vračanjem izbranega izdelka v serijo zagotovimo, da so pred vsako ponovitvijo poskusa pogoji natanko takšni kot pri predhodnih, zato računamo verjetnost, da bo v n ponovitvah poskusa iz Bernoullijevega zaporedja poskusov dogodek "izdelek je pokvarjen" nastopil natanko k - krat. Verjetnost tega dogodka je $p = M/N$, zato je po Bernoullijevem obrazcu

$$P_1(A) = \binom{n}{k}\left(\frac{M}{N}\right)^k\left(1 - \frac{M}{N}\right)^{n-k} \qquad (3)$$

V drugem primeru, ko izdelkov ne vračamo, lahko poskus realiziramo tako, da izberemo vseh n izdelkov sočasno. Izidov je tedaj $\binom{N}{n}$, na toliko načinov lahko iz množice z N elementi izberemo podmnožico, ki jih premore n. Ugodni so tisti izidi, kjer je k izdelkov izbranih iz M pokvarjenih, preostalih $n-k$ izdelkov pa iz $N-M$ nepokvarjenih. Verjetnost proučevanega dogodka je zato

$$P_2(A) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \qquad (4)$$

S statističnega vidika imamo opraviti z vzorci s ponavljanjem oziroma z vzorci brez ponavljanja (drugi primer), zato je zanimivo vprašanje, kakšna je v konkretnih primerih razlika med verjetnostima, ki ju izračunamo po (3) ali (4).

Za $N = 20$, $M = 2$, $n = 5$, $k = 1$ dobimo na primer

$$P_1(A) = \binom{5}{1}\left(\frac{2}{20}\right)^1\left(1 - \frac{2}{20}\right)^4 = 0'32805$$

$$P_2(A) = \frac{\binom{2}{1}\binom{18}{4}}{\binom{20}{5}} = 0'39474$$

Povečajmo zdaj populacijo na $N = 100$, delež $p = M/N$ pa naj ostane nespremenjen, tako da je še vedno $M = 10$. Potem je

$$P_1(A) = \binom{5}{1}\left(\frac{10}{100}\right)^1\left(1 - \frac{10}{100}\right)^4 = 0'32805$$

$$P_2(A) = \frac{\binom{10}{1}\binom{90}{4}}{\binom{100}{5}} = 0'33939$$

V prvem primeru je bila druga verjetnost za več kot 20% večja od prve, pri povečanju populacije je ta razlika samo še dobre 3%. Čeprav posameznih primerov ne smemo prehitro in kar počez posploševati, pa tokrat le velja naslednja ugotovitev:

Če velikost populacije narašča preko vsake meje, delež M/N pa ostaja konstanten, se verjetnosti posameznih frekvenc, izračunane po obrazcu (4), vedno bolj ujemajo s tistimi, ki jih dobimo po obrazcu (3) za vzorce s ponavljanjem.

To za prakso pomeni, da je pri dovolj veliki populaciji (seriji izdelkov) in vzorcu, ki je dovolj majhen v primerjavi s celotno populacijo, za strukturo vzorca v bistvu vseeno, ali pri vzorčenju izbrane elemente vračamo ali ne. To ima koristne posledice za statistiko: obrazce izpeljujemo za vzorce s ponavljanjem, kjer je račun lažji zaradi neodvisnosti dogodkov, uporabljamo pa jih - vsaj kot približke - na vzorcih brez ponavljanja, ki so znatno bolj uporabni v praksi...

$\square$ Pri danem številu poskusov n in verjetnosti dogodka A v posamezni ponovitvi poskusa p je Bernoullijev obrazec (2) predpis, s katerim vsakemu številu k (frekvenci dogodka A) iz množice $\{0, 1, 2, ..., n\}$ priredimo natanko določeno število $P(n; p; k)$, ali drugače, z Bernoullijevim obrazcem je določena neka porazdelitev verjetnosti po posameznih frekvencah. Zanima nas, pri kateri vrednosti k je ta verjetnost največja. V ta namen primerjamo $P(n; p; k)$ in $P(n; p; k+1)$ ter ugotavljamo, do katerega k je druga še večja od prve. Z neposrednim računom (nekaj vmesnih korakov puščamo bralcu za vajo) se lahko prepričamo, da je

$$\frac{P(n; p; k+1)}{P(n; p; k)} = \frac{n-k}{k+1} \cdot \frac{p}{q} = \frac{np - kp}{kq + q}$$

(pri tem je $q = 1 - p$) in zato velja

$$\frac{P(n; p; k+1)}{P(n; p; k)} \geq 1 \iff np - kp \geq kq + q \iff np - q \geq k(p+q) = k \quad (5)$$

Pri tem je frekvenca k seveda celo število, za izraz $np-q$ pa to ni vedno res. Za $k < np-q$ je količnik na levi strani (5) večji od 1 in je $P(n;p;k+1) > P(n;p;k)$. Če $np-q$ ni celo število, je torej največja vrednost proučevane funkcije dosežena pri prvem večjem celem številu,

$k_o = [np-q] + 1$

(Z oglatim oklepajem okrog števila označimo njegov celi del.) Če pa je izraz $np-q$ slučajno enak nekemu celemu številu, je po (5) za frekvenco $k = np-q$ izpolnjena enačba $P(n;p;k+1) = P(n;p;k)$ in verjetnost je maksimalna za dve sosednji frekvenci, $k = np-q$ in $k = np-q+1 = (n+1)p$.

ZGLED 4.6.5:

Kovanec je tako deformiran, da je padec grba dvakrat verjetnejši od padca številke. Kolišna je verjetnost dogodka, da v dvajsetih metih natanko štirikrat pade številka? Katera frekvenca številke je najverjetnejša?

Za prvi del naloge dobimo pri n = 20, p = 1/3 in k = 4 po Bernoullijevem obrazcu

$$P(20;1/3;4) = \binom{20}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^{16} \approx 0{\cdot}091$$

Ker je pri naših podatkih $np-q = 6$, torej celo število, sta najverjetnejši frekvenci števila dve, namreč 6 in 7, pri obeh namreč dobimo enako verjetnost,

$$P(20;1/3;6) = \binom{20}{6}\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^{14} \approx 0{\cdot}182$$

$$P(20;1/3;7) = \binom{20}{7}\left(\frac{1}{3}\right)^7\left(\frac{2}{3}\right)^{13} \approx 0{\cdot}182$$

ZGLED 4.6.6:

Iz posode, v kateri imamo 3 bele in 7 črnih krogel, desetkrat na slepo izvlečemo po dve krogli hkrati in ju vsakič vrnemo v posodo. Kolikšna je najverjetnejša frekvenca dogodka "obe izvlečeni krogli sta črni"?

Ker krogle vračamo, gre za zaporedje neodvisnih poskusov, verjetnost za dve črni v posameznem poskusu pa je

$$p = \frac{\binom{7}{2}}{\binom{10}{2}} = \frac{7}{15}$$

zato je

$$q = 1-p = \frac{8}{15}, \quad np-q = 10\cdot\frac{7}{15}-\frac{8}{15} = 4\frac{2}{15}$$

Najverjetnejša frekvenca je prvo večje celo število, torej 5. V desetih ponovitvah opisanega poskusa je torej od vseh možnih frekvenc dogodka "obe izvlečeni krogli sta črni" najverjetnejša frekvenca $k_o = 5$.

K porazdelitvi verjetnosti po posameznih možnih frekvencah slučajnega dogodka v Bernoullijevem zaporedju poskusov se bomo še vračali v naslednjih razdelkih, zdaj pa si oglejmo še nekaj vprašanj, ki so v sorodu z Bernoullijevim obrazcem. Najprej poskusimo odgovoriti na naslednje:

Dogodek A ima v posameznem poskusu iz Bernoullijevega zaporedja poskusov konstantno verjetnost $P(A) = p$. Kolikšna je verjetnost, da se pri ponavljanju poskusa ta dogodek zgodi k - tič natanko v n - tem poskusu?

Dogodek, katerega verjetnost moramo izračunati, lahko opišemo kot produkt dveh dogodkov:

- v prvih $n-1$ poskusih $(k-1)$ - krat nastopi dogodek A;

- dogodek A nastopi v n-tem poskusu.

Verjetnost prvega faktorja je po Bernoullijevem obrazcu enaka

$$P(n-1;p;k-1) = \binom{n-1}{k-1}p^{k-1}(1-p)^{(n-1)-(k-1)} = \binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}$$

verjetnost drugega faktorja pa je seveda $P(A) = p$. Ker sta faktorja med seboj neodvisna (gre pač za Bernoullijevo zaporedje poskusov), je verjetnost, ki jo iščemo, kar produkt verjetnosti obeh faktorjev, torej

$$\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}\cdot p = \binom{n-1}{k-1}p^{k}(1-p)^{n-k} \qquad (6)$$

Ta obrazec, ki se od Bernoullijevega razlikuje samo po začetnem binomskem simbolu, imenujemo **Pascalov obrazec**. (Drži, gospoda že poznamo po znamenitem trikotniku, glej razdelek 2.5.) Po vsebini se seveda obrazca bistveno razlikujeta: Bernoullijev izda verjetnost, da se dogodek A v n poskusih zgodi k - krat, pri čemer so za to ugodne vse razporedbe uspešnih in neuspešnih poskusov (oziroma realizacij A in nasprotnega dogodka A'), Pascalov obrazec pa je v nekem smislu natančnejši, zahteva, da nastopi dogodek A k - tič natanko v n - tem, to je zadnjem poskusu.

ZGLED 4.6.7:

Strelec mora za odhod na tekmovanje doseči v 100 strelih 96 zadetkov. Kolikšna je verjetnost, da ujame normo z zadnjim strelom, če pri posameznem strelu doseže zadetek z verjetnostjo 0·90?

Pri $n = 100$, $p = 0{\cdot}90$ in $k = 96$ imamo po obrazcu (6) za iskano verjetnost

$$\binom{n-1}{k-1}p^{k}(1-p)^{n-k} = \binom{99}{95}0{\cdot}90^{96}\,0{\cdot}10^{4} \approx 0{\cdot}015$$

Za vajo lahko izračunaš še verjetnosti dogodkov, da doseže normo s šestindevetdesetim, sedemindevetdesetim, osemindevetdesetim ali devetindevetdesetim strelom. Vse, kar še manjka, je pač verjetnost dogodka, da norme sploh ne doseže, ker ima v 100 strelih manj kot 96 zadetkov. Ugotovil boš, da je zelo majhna verjetnost, da bi se naš strelec udeležil omenjenega tekmovanja...

S Pascalovim obrazcem se ne srečamo tako pogosto kot z Bernoullijevim, obstaja pa izjema, ki ni tako redka. Če v Pascalovem obrazcu pri dani verjetnosti za "uspeh" v posameznem poskusu (p) postavimo $k = 1$, dobimo za različne vrednosti n verjetnosti, da se nam poskus prvič "posreči" v prvem, drugem, tretjem... poskusu. Te verjetnosti so dane z obrazcem

$$P_n = \binom{n-1}{0} p^1 (1-p)^{n-1} = p\, q^{n-1}$$

torej

$$P_1 = p\,, \quad P_2 = p\,q\,, \quad P_3 = p\,q^2 \ldots$$

Tistim, ki se (že) spoznajo na zaporedja, se ne bo zdelo čudno, da so take verjetnosti značilne za porazdelitev, ki ji bomo kasneje rekli *geometrijska.*

Zdaj pa še k drugačni nalogi. Pri Bernoullijevem obrazcu smo začeli s predpostavko, da ima v vsaki ponovitvi poskusa popolni sistem dogodkov samo dva elementa, namreč dogodek A in njegovo negacijo A', pri čemer je njuno "razmerje moči" (beri: verjetnost $P(A) = p$) konstantno. Izpeljimo zdaj še t.i. **posplošeni Bernoullijev obrazec**, s katerim rešimo naslednji problem:

Opravimo n enakih, med seboj neodvisnih poskusov, pri katerih imamo končen popoln sistem izidov $\{E_1, E_2, ..., E_m\}$ s pripadajočimi verjetnostmi

$$P(E_1) = p_1\,, \quad P(E_2) = p_2\,, \quad \ldots\,, \quad P(E_m) = p_m$$

ki se od poskusa do poskusa ne spreminjajo. Kolikšna je verjetnost dogodka $A_n(k_1, k_2, ..., k_m)$, da se zgodi v teh n poskusih dogodek E_i natanko k_i - krat, $i = 1, 2, ..., m$?

Frekvence k_i so sicer poljubna nenegativna cela števila, njihova vsota pa mora biti enaka številu vseh poskusov n, ker pač ustrezni dogodki tvorijo popolni sistem dogodkov za proučevani poskus.

Premislek, ki nas pripelje do iskanega obrazca, je podoben kot pri njegovem skromnejšem sorodniku. Za dogodek, katerega verjetnost računamo, so ugodni vsi tisti produkti dogodkov E_i, v katerih vsak dogodek nastopa s predpisano frekvenco. Verjetnost vsakega od takih produktov je enaka
$p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_m^{k_m}$
Pri tem smo upoštevali, da gre za zaporedje neodvisnih poskusov, zaradi česar se ustrezne verjetnosti enostavno množijo. Za nastop dogodka ugodnih produktov pa je toliko, kolikor je permutacij nizov dolžine n, v katerih

nastopa m različnih elementov E_i, vsak s predpisano frekvenco k_i; po obrazcu za permutacije s ponavljanjem je to število $n!/(k_1! \cdot k_2! \cdot \ldots \cdot k_m!)$. Ker so produkti med seboj nezdružljivi, je iskana verjetnost enaka

$$P(n; p_1, p_2, ..., p_m; k_1, k_2, ..., k_m) = \frac{n!}{k_1! \cdot k_2! \cdot \ldots \cdot k_m!}\, p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_m^{k_m} \quad (7)$$

Obrazec je na prvi pogled morda nekoliko zapleten, si ga pa hitro zapomnimo, če obvladamo osnovni Bernoullijev obrazec in ga znamo logično razširiti.

ZGLED 4.6.8:

Kolikšna je verjetnost, da pri 12 metih poštene igralne kocke vsak od šestih možnih izidov nastopi natanko po dvakrat?

Do rezultata lahko pridemo tudi neposredno, tako da v klasični definiciji posebej izračunamo, koliko je vseh izidov in koliko je za naš dogodek ugodnih, vendar je račun z uporabo posplošenega Bernoullijevega obrazca bistveno preprostejši, z uporabo
$n = 12$,
$p_1 = p_2 = ... = p_6 = 1/6$,
$k_1 = k_2 = ... = k_6 = 2$
dobimo s pomočjo obrazca (7) rešitev

$$P(12;\, 1/6, 1/6, ..., 1/6;\, 2, 2, ..., 2) =$$
$$= \frac{12!}{2! \cdot 2! \cdot \ldots \cdot 2!} (1/6)^2 \cdot (1/6)^2 \cdot \ldots \cdot (1/6)^2 =$$
$$= \frac{12!}{2^6 \cdot 6^{12}} \approx 0\text{'}0034$$

Bernoullijev obrazec in njegovi sorodniki so koristna orodja pri računanju verjetnosti, kadar imamo opraviti z zaporedji neodvisnih poskusov. Vsaj kot približek jih lahko uporabljamo tudi takrat, ko sicer ne ponavljamo istega poskusa, ampak se vzporedno (sočasno) odvija več enakih procesov, ki ne vplivajo drug na drugega, tako da lahko predpostavimo neodvisnost med dogodki, ki pri tem (lahko) nastopijo. Tako z Bernoullijevim obrazcem na primer računamo verjetnosti,
- da se v nekem časovnem intervalu pokvari določeno število med seboj enakih strojev, ki se kvarijo neodvisno drug od drugega,
- da vzkali natanko k od vseh n enakih semen, če poznamo verjetnost, da vzkali posamezno zrno (praktični približek za to verjetnost je običajno kar relativna frekvenca "uspešnih semen" v prejšnji setvi)

in podobno.

Ti obrazci pa postanejo nekoliko naporni, ko število poskusov močno naraste, kar se v nekaterih praktičnih nalogah hitro zgodi. Obseg dela se zelo poveča tudi v primeru, ko moramo izračunati verjetnosti za večje število frekvenc, na primer verjetnost, da iz 1 000 posajenih semen vzkali od 800 do 900 rastlin. Kasneje bomo spoznali, da obstajajo približni obrazci, s katerimi si relativno udobno lahko pomagamo pri takšnih nalogah.

V A J E

1. Pri metanju kovanca je dvajsetkrat po vrsti padel grb. Kolikšna je verjetnost, da grb pade v enaindvajseti ponovitvi poskusa?

2. Desetkrat zapored vržemo kovanec. Kolikšne so verjetnosti, da

a) grb pade natanko dvakrat;

b) grb ne pade več kot dvakrat;

c) grb pade vsaj dvakrat?

3. V žari je enako število belih in črnih krogel. Osemkrat zaporedoma izvlečemo (na slepo) po eno kroglo in jo spet vrnemo v žaro. Kolikšna je verjetnost, da smo pri tem natanko šestkrat izvlekli črno kroglo?

4. Kateri dogodek ima večjo verjetnost?

A - štirje grbi pri sedmih metih kovanca ali

B - šest grbov pri devetih metih?

5. Podjetje ima štiri tovornjake, okvare posameznega tovornjaka so neodvisne od okvar ostalih tovornjakov. Verjetnost, da se v določenem časovnem intervalu posamezen tovornjak pokvari, znaša 0'04. Kolikšna je verjetnost dogodkov

A - da se pokvarijo vsi štirje tovornjaki;

B - da se ne pokvari vsaj en tovornjak;

C - da se pokvari vsaj en tovornjak?

(Predlog: Razmisli, kako bi rezultate te naloge uporabil pri načrtovanju potrebnih rezervnih delov... Zanesljivost sistema opazuj z dveh vidikov oziroma zahtev:
- sistem dela, če delajo vsi tovornjaki;
- sistem dela, kadar deluje vsaj še en tovornjak.)

6. V tekstilni tovarni je 80 enakih strojev; verjetnost okvare na posameznem stroju je za določeno časovno obdobje 0'06; stroji se kvarijo neodvisno drug od drugega; verjetnost, da bi se isti stroj v enem obdobju pokvaril dvakrat, je zanemarljivo majhna. Izračunajte verjetnost, da se v proučevanem obdobju pokvarijo natanko štirje stroji.

7. Kolikokrat moramo vreči kocko, da lahko z verjetnostjo najmanj 0'5 pričakujemo, da bo vsaj enkrat padla šestica?

8. Desetkrat vržemo po dve kocki. Izračunaj verjetnost, da je pri tem natanko dvakrat vsota pik enaka 7.

9. Tovarna proizvaja v povprečju 5% slabih proizvodov. Kolikšna je verjetnost, da so v škatli z (na slepo izbranimi) 20 proizvodi vsi brezhibni, in kolikšna verjetnost, da so v njej več kot trije slabi proizvodi?

10. Predpostavimo, da je dogodek, da se rodi deček, slučajen in neodvisen od spola ostalih otrok v družini, njegova verjetnost (v posameznem poskusu) bodi 0'518. Kolikšne so pri teh pogojih verjetnosti naslednjih dogodkov:

A - da so v družini med petimi otroki natanko trije dečki;

B - da ima oče s petimi otroki vsaj enega moškega potomca;

C - da bo oče, ki ima štiri hčerke, naslednjič dobil sina.

11. V aparatu je 20 elementov tipa A in 30 elementov tipa B, ki morajo vsi delovati, da je aparat uporaben. Proizvajalec sestavnih delov jamči, da je verjetnost za okvaro na posameznem elementu kvečjemu 0'001, kar velja za oba tipa elementov (elementi se kvarijo neodvisno drug od drugega). Kolikšen odstotek reklamacij zaradi okvar na aparatu lahko vendarle pričakujemo?

12. Katero število grbov je najbolj verjetno, če vržemo pošten kovanec

a) tridesetkrat; b) petintridesetkrat?

13. Vojak, ki zadeva z verjetnostjo 0,45, ima na razpolago 10 nabojev. Za pozitivno oceno potrebuje šest zadetkov. Kolikšna je verjetnost, da normo doseže ravno z zadnjim nabojem?

14. V škatli imamo 4 bele, 4 črne in 2 rdeči krogli. Šestkrat na slepo izvlečemo po eno kroglo in jo vsakič vrnemo v škatlo. Kolikšna je verjetnost, da se pri tem po dvakrat pojavi krogla posamezne barve?

15. Kolikšna je verjetnost, da se pri poskusu iz prejšnje naloge proučevani dogodek (po dvakrat krogla posamezne barve) zgodi tako, da se dvakrat ponovi zaporedje bela - črna - rdeča krogla?

4.7 Slučajne spremenljivke

Že v tretjem poglavju, ko smo se spogledovali z opisno statistiko, smo srečali pojem *statistična spremenljivka*, ki smo ga uporabili pri proučevanju nekega pojava na izbrani populaciji; vrednosti spremenljivke so bile določene z izmerjenimi vrednostmi tega pojava na posameznih elementih iz populacije.

Omenili smo, da dostikrat ne moremo (nočemo) proučevati populacije kot celote, ampak njene značilnosti ocenjujemo s pomočjo izbranega vzorca. Pri najpomembnejši vrsti vzorčenja na slepo izbiramo elemente v vzorec, kateri element bo izbran, je odvisno od slučaja in isto posledično velja za vrednosti spremenljivke, ki jo proučujemo.

Tako smo dobili spremenljivko, ki se bistveno razlikuje od vseh, s katerimi smo se srečali v matematiki: za nekatere (neodvisne) smo si svobodno izbirali vrednosti iz primerne množice, druge (odvisne) pa so s tem dobile trdno določene vrednosti. Sedaj pa sicer (lahko) poznamo množico vrednosti, med katerimi spremenljivka izbira, katero vrednost dejansko zavzame, pa je odvisno od slučaja. Zato ji rečemo **slučajna spremenljivka**.

Tudi naše siceršnje ukvarjanje s poskusi in slučajnimi dogodki lahko na novo opišemo. Za večino poskusov, s katerimi smo se do sedaj srečali, lahko izide neposredno ali vsaj posredno predstavimo v numerični obliki. Na primer:

Pri metu kocke je število pik 1, 2, 3, 4, 5 ali 6.

Pri streljanju v tarčo je izid npr. razdalja med zadetkom in sredino tarče.

Pri metu kovanca lahko obema možnima izidoma, ki sicer nista števili, pripišemo primerni vrednosti, na primer 1 za padec grba in 0 (ali recimo -1) za padec številke.

V naštetih - in mnogih drugih - primerih lahko poskusu priredimo spremenljivko, ki v vsaki ponovitvi poskusa sicer zavzame natanko eno od možnih vrednosti, vendar pred realizacijo poskusa ne moremo z gotovostjo napovedati, katera bo ta vrednost, ampak je ta vrednost odvisna od slučaja. Zato gre za slučajno spremenljivko.

Predstavitev slučajne spremenljivke

Množico vrednosti, ki jih lahko slučajna spremenljivka zavzame, bomo imenovali **zaloga vrednosti**. Ta je lahko različno bogata:

a) Če za slučajno spremenljivko proglasimo število pik pri metu kocke, je zaloga vrednosti končna množica $\{1,2,3,4,5,6\}$.

b) Kocko mečemo toliko časa, da prvič pade šestica, slučajna spremenljivka pa naj bo število metov, ki so za to potrebni. Ker načelno ne moremo izključiti (sicer malo verjetnega) zaporedja samih od 6 različnih vrednosti, je zaloga vrednosti te slučajne spremenljivke kar cela množica naravnih števil, torej števno neskončna množica.

c) Daljico dolžine a prelomimo v slučajno izbrani točki. Slučajna spremenljivka naj bo dolžina daljšega dela. Ta spremenljivka lahko zavzame poljubno vrednost z intervala $(0, a)$, njena zaloga vrednosti je zvezna množica.

Področje našega zanimanja bodo predvsem take slučajne spremenljivke, ki imajo **končno zalogo vrednosti**. Te sodijo skupaj s spremenljivkami, katerih zaloga vrednosti je števno neskončna, med **diskretne slučajne spremenljivke**.

Poleg zaloge vrednosti slučajne spremenljivke je treba poznati tudi njen **porazdelitveni zakon**, ki pove, kakšne so verjetnosti, da slučajna spremenljivka zavzame vrednosti iz nekega vnaprej predpisanega dela svoje zaloge vrednosti. V tem učbeniku se bomo največ ukvarjali s takimi slučajnimi spremenljivkami, kjer je porazdelitveni zakon mogoče opisati kar z verjetnostmi za posamezne točke iz zaloge vrednosti.

Slučajne spremenljivke bomo označevali z velikimi tiskanimi črkami, vrednosti, ki jih bodo zavzele, pa z enakimi malimi črkami. Tako je na primer $(X = x)$ dogodek, da slučajna spremenljivka X zavzame vrednost $x \in \mathbb{R}$; podobno lahko z besedami pojasnimo zapise $(X < x)$, $(X \geq x)$ in tako dalje.

Poznavanje porazdelitvenega zakona slučajne spremenljivke potem pomeni, da so znane verjetnosti $P(X = x)$, $P(X < x)$ in podobno.

Ker se bomo do preklica omejili na slučajne spremenljivke, pri katerih je zaloga vrednosti neka končna množica $\{x_1, x_2, ..., x_n\}$, tvorijo dogodki

$$(X = x_k) \qquad k = 1, 2, ..., n \tag{1}$$

popoln sistem dogodkov, vsota verjetnosti teh dogodkov je enaka 1. S predpisom

$$p_k = P(X = x_k) \qquad k = 1, 2, ..., n \tag{2}$$

je določena t.i. **verjetnostna funkcija**, ki pove, kako je verjetnost porazdeljena po posameznih vrednostih iz zaloge; z (1) in (2) je slučajna spremenljivka povsem opisana.

Enoten prikaz zaloge vrednosti in porazdelitvenega zakona za slučajno spremenljivko s končno zalogo vrednosti omogoča **verjetnostna shema**. To je tabela

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix} \tag{3}$$

ki ima v prvi vrstici naštete vse elemente x_k zaloge vrednosti, pod njimi pa zapisane pripadajoče verjetnosti; ker tvorijo dogodki (1) popoln sistem dogodkov, mora biti seveda vsota števil iz druge vrstice verjetnostne sheme enaka 1.

ZGLED 4.7.1:

Zapišimo verjetnostno shemo slučajne spremenljivke X, ki ustreza poskusu "met poštene igralne kocke", ki ga proučujemo - kot običajno - glede na število padlih pik.

Tu imamo šest enako verjetnih izidov in so vse verjetnosti p_k enake $1/6$, zato je verjetnostna shema enaka

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

Če slučajni spremenljivki X tako kot v zgledu 4.7.1 ustreza končen simetričen popoln sistem dogodkov z močjo n, zaradi česar so vse verjetnosti $P(X = x_k)$ enake $1/n$, imenujemo ustrezno porazdelitev verjetnosti **enakomerna diskretna porazdelitev**.

Ena najpomembnejših verjetnostnih porazdelitev pa je **binomska porazdelitev.** Takšno verjetnostno porazdelitev ima na primer frekvenca dogodka (še vemo, kaj izraz "frekvenca" pomeni?), ki ima v posamezni ponovitvi verjetnost p, v n realizacijah poskusa iz Bernoullijevega zaporedja. Binomska porazdelitev je tudi v splošnem določena z dvema podatkoma (parametroma), n in p; njena zaloga je (razmišljaj o frekvenci!) množica $\{0, 1, 2, ..., n-1, n\}$, posamezne verjetnosti pa računamo po Bernoullijevem obrazcu, zato je njena verjetnostna funkcija

$$P(X = k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0, 1, 2, ..., n) \tag{4}$$

ZGLED 4.7.2:

Štirikrat zaporedoma vržemo kovanec, slučajna spremenljivka X naj bo frekvenca grba v teh štirih ponovitvah poskusa. Določi njeno verjetnostno shemo!

Iz prej povedanega sledi, da je X porazdeljena binomsko s parametroma $n = 4$, $p = 1/2$; to simbolično zapišemo kot $X \sim b(4, 1/2)$, v splošnem torej $X \sim b(n,p)$. Spremenljivka X lahko zavzame vrednosti 0, 1, 2, 3 ali 4, pripadajoče verjetnosti računamo po obrazcu (4) z danima vrednostima $n = 4$, $p = 1/2$. Tako imamo (kar manjka v računu, bodi za vajo!)

$$\begin{array}{lclcl} P(X=0) & = & \binom{4}{0}(\frac{1}{2})^0(\frac{1}{2})^4 & = & \frac{1}{16} \\ P(X=1) & = & \binom{4}{1}(\frac{1}{2})^1(\frac{1}{2})^3 & = & \frac{4}{16} \\ \dots & \dots & \dots & \dots & \dots \\ P(X=4) & = & \binom{4}{4}(\frac{1}{2})^4(\frac{1}{2})^1 & = & \frac{1}{16} \end{array}$$

in verjetnostna shema je

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{1}{16} & \frac{4}{16} & \frac{6}{16} & \frac{4}{16} & \frac{1}{16} \end{pmatrix}$$

Nasvet: Kadar izračunaš verjetnostno shemo, vedno preveri, ali je vsota verjetnosti v drugi vrstici sheme enaka 1! (Zloba kraljice znanosti je v tem, da je to samo potreben, ne pa tudi zadosten pogoj za pravilnost verjetnostne sheme: dejstvo, da si delitev kilograma marmelade med n otrok končal s praznim kozarcem, še ne zagotavlja, da je vsak dobil natanko toliko, kolikor mu pripada...)

ZGLED 4.7.3:

Igralna kocka je obtežena tako, da v povprečju v tretjini poskusov pade šestica. Sestavi verjetnostno shemo za slučajno spremenljivko X, ki predstavlja frekvenco šestice v štirih metih kocke!

V skladu z besedilom lahko predpostavimo, da gre za Bernoullijevo zaporedje poskusov, v katerem ima dogodek A - "padec šestih pik" - v posameznem poskusu verjetnost $P(A) = p = 1/3$. Verjetnostno shemo izračunamo tako kot v prejšnjem zgledu, seveda s spremenjenim parametrom p, tako da je

$$P(X = k) \;=\; \binom{4}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{4-k} \quad (k = 0, 1, 2, ..., 4)$$

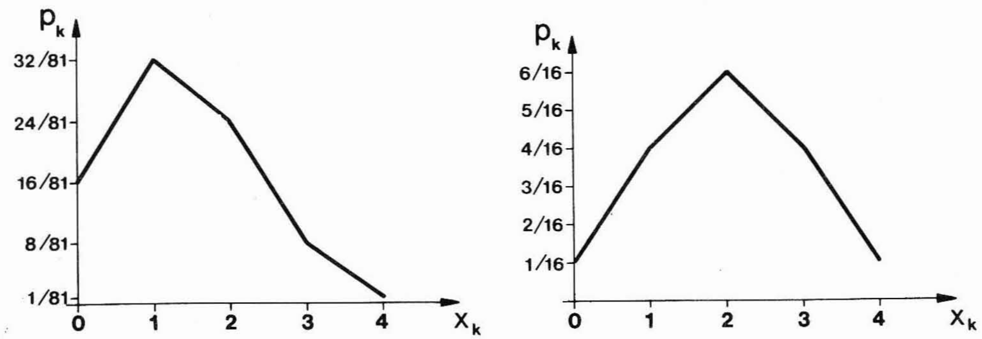
Kot boš preveril (!), je verjetnostna shema te slučajne spremenljivke

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{16}{81} & \frac{32}{81} & \frac{24}{81} & \frac{8}{81} & \frac{1}{81} \end{pmatrix}$$

Kakšna je porazdelitev verjetnosti po posameznih vrednostih slučajne spremenljivke, lahko lepo prikažemo tudi z grafom verjetnostne funkcije. Tak grafični prikaz v bistvu ustreza frekvenčnemu poligonu, ki smo ga spoznali v tretjem poglavju.

Ker gre za diskretno slučajno spremenljivko, je verjetnostna funkcija seveda definirana samo v tistih točkah realne osi, ki spadajo v zalogo vrednosti te slučajne spremenljivke; ustrezne točke, ki jih dobimo v koordinatnem sistemu (ordinata točke je pripadajoča verjetnost), povežemo med seboj z lomljeno črto samo zaradi večje nazornosti. Iz istega razloga si na ordinatni osi privoščimo večjo enoto, ker bi sicer bila slika nepregledna.

Iz obeh slik vidimo, da je porazdelitev v prvem primeru simetrična, v drugem pa ne, kar je posledica simetričnosti (nesimetričnosti) popolnega sistema dogodkov { padec grba, padec številke} v posamezni ponovitvi



Slika 4.4: Grafa verjetnostnih funkcij dveh binomskih porazdelitev

poskusa v prvem (drugem) primeru, ali kakor rečemo drugače, pri posamezni realizaciji prve (druge) slučajne spremenljivke.

Poleg enakomerne in binomske porazdelitve srečamo v praksi še več drugih pomembnih diskretnih porazdelitev, vendar jih tu ne utegnemo obravnavati v vseh podrobnostih, zato ostanimo samo pri slovarčku in zgledih.

ZGLED 4.7.4:

Iz škatle, v kateri imamo 10 belih in 5 črnih krogel, štirikrat na slepo izvlečemo po eno kroglo. Naj bo slučajna spremenljivka X frekvenca črne krogle v teh štirih poskusih. Izračunaj njeno verjetnostno shemo, če predpostavimo,

a) da izvlečenih krogel ne vračamo v škatlo oziroma

b) da izvlečeno kroglo vsakič vrnemo v škatlo!

a) V prvem primeru, ko krogel ne vračamo v posodo, lahko opis postopka nadomestimo s predpostavko, da vse štiri krogle izvlečemo sočasno; za tak primer že nekaj časa znamo izračunati verjetnosti, da je med izvlečenimi elementi iz neke populacije natanko predpisano število elementov z določeno lastnostjo. Slučajna spremenljivka X lahko zavzame vrednosti 0, 1, 2, 3 ali 4, verjetnostna funkcija pa je dana s

$$P(X=k) = \frac{\binom{5}{k}\binom{10}{4-k}}{\binom{15}{4}} \quad (k=0,1,2,3,4)$$

(Pri takih nalogah je treba vedno pogledati, ali je na voljo dovolj elementov predpisane vrste, da lahko frekvenca zavzame celotni spekter celoštevilskih vrednosti od $k=0$ do velikosti vzorca; če bi imeli v škatli manj kot 4 krogle, bi se morali tudi z vrednostjo k ustaviti pod velikostjo vzorca.)

Z uporabo gornjega obrazca dobimo vseh pet verjetnosti in lahko sestavimo verjetnostno shemo

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{42}{273} & \frac{120}{273} & \frac{90}{273} & \frac{20}{273} & \frac{1}{273} \end{pmatrix}$$

Glede na strukturo populacije (krogel v škatli) je bilo mogoče vnaprej pričakovati, da bosta najbolj verjetni frekvenci 1 in 2 in da je zelo majhna verjetnost, da bi bile vse štiri krogle v vzorcu črne.

b) Pri drugem načinu, ko izvlečene krogle vračamo v posodo, imamo opraviti z zaporedjem neodvisnih poskusov, kjer je v posamezni ponovitvi poskusa verjetnost za "uspeh" (v našem primeru črno kroglo) $p = 5/15 = 1/3$, zato je sedaj slučajna spremenljivka X porazdeljena binomsko s parametroma $n=4$ in $p=1/3$, torej natanko tako kot slučajna spremenljivka v zgledu 3. Primerjaj obe porazdelitvi in premisli, zakaj se pojavljajo razlike med njima.

Porazdelitev frekvence elementov s predpisano lastnostjo v vzorcu brez ponavljanja (kar je posledica tega, da ne vračamo izbranih elementov oziroma da jih izberemo sočasno) imenujemo **hipergeometrijska porazdelitev**. Njena verjetnostna funkcija je odvisna od velikosti populacije N, deležev elementov s predpisano lastnostjo (M) in komplementa te podmnožice ($N-M$) ter velikosti vzorca n. Predpostavljamo, da je vzorec (bistveno) manjši od moči obeh podmnožic, $n < M$, $n < N-M$. Potem je

$$P(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \qquad k=0,1,2,...,n \tag{5}$$

Opozorilo na obnašanje izraza na desni strani (4) pri naraščanju velikosti populacije, ki smo ga zapisali v razdelku 4.6 (str. 129), lahko sedaj povemo tudi takole:

Če velikost populacije narašča preko vsake meje, delež M/N pa ostaja konstanten, se hipergeometrijska porazdelitev (5) vedno bolje ujema z binomsko porazdelitvijo (4).

Od porazdelitev, ki imajo števno neskončno zalogo vrednosti, omenimo dve, ki ju pravzaprav že poznamo iz prejšnjega razdelka, to sta **Pascalova porazdelitev** in njena "posebna izvedba", to je **geometrijska porazdelitev**. Pri prvi slučajna spremenljivka pomeni število poskusov, potrebnih za k "uspehov" v Bernoullijevem zaporedju neodvisnih poskusov, in ima zato verjetnostno funkcijo

$$P(X=n) = \binom{n-1}{k-1} p^k (1-p)^{n-k} \qquad n=k,k+1,k+2,... \tag{6}$$

geometrijska porazdelitev pa je Pascalova porazdelitev za $k=1$, pomeni potemtakem porazdelitev števila poskusov, potrebnih za prvi "uspeh":

$$P(X=n) = p(1-p)^{n-1} = p\,q^{n-1} \qquad n=1,2,3,... \tag{7}$$

ZGLED 4.7.5:

Na začetku razdelka smo kot primer za slučajno spremenljivko s števno neskončno zalogo vrednosti navedli število metov, potrebnih, da pri metu poštene igralne kocke prvič pade šestica. Očitno gre za geometrijsko porazdelitev s $p = 1/6$ (in $q = 5/6$) in po obrazcu (7) imamo

$$\begin{aligned}
P(X=1) &= (\tfrac{1}{6})(\tfrac{5}{6})^0 = \tfrac{1}{6} \approx 0{\cdot}167\\
P(X=2) &= (\tfrac{1}{6})(\tfrac{5}{6})^1 = \tfrac{5}{36} \approx 0{\cdot}139\\
P(X=3) &= (\tfrac{1}{6})(\tfrac{5}{6})^2 = \tfrac{25}{216} \approx 0{\cdot}116\\
P(X=4) &= (\tfrac{1}{6})(\tfrac{5}{6})^3 = \tfrac{125}{1\,296} \approx 0{\cdot}096\\
P(X=5) &= (\tfrac{1}{6})(\tfrac{5}{6})^4 = \tfrac{625}{7\,776} \approx 0{\cdot}080\\
P(X=6) &= (\tfrac{1}{6})(\tfrac{5}{6})^5 = \tfrac{3\,125}{46\,656} \approx 0{\cdot}067\ldots
\end{aligned}$$

Matematično upanje slučajne spremenljivke

Denimo, da preučujemo slučajno spremenljivko X, ki ima zalogo vrednosti $\{x_k;\, k = 1, 2, ..., n\}$ in porazdelitveni zakon, opisan z verjetnostno funkcijo $P(X = x_k) = p_k$. Opravimo N - krat poskus, pri katerem nastopa ta slučajna spremenljivka in tako dobimo N realizacij te slučajne spremenljivke; pri tem zapisujemo, kolikokrat zavzame X vsako od možnih vrednosti x_k. Tako dobimo neko frekvenčno porazdelitev,

x_1	x_2	...	x_n
f_1	f_2	...	f_n

kjer so števila f_k zapisane frekvence posameznih vrednosti. Povprečno vrednost - tehtano aritmetično sredino - vseh realizacij znamo izračunati (glej III. poglavje!), enaka je

$$\mu = \frac{1}{N}\sum_{k=1}^{n} f_k\, x_k \tag{8}$$

Konstanto $1/N$ lahko postavimo tudi v samo vsoto in (8) prepišemo v

$$\mu = \sum_{k=1}^{n} \frac{f_k}{N}\, x_k \tag{9}$$

Vrednosti x_k so zdaj očitno pomnožene s svojimi relativnimi frekvencami. Če pustimo število vseh realizacij N naraščati prek vsake meje, se relativne frekvence običajno ustalijo pri verjetnostih dogodkov $(X = x_k)$, to je p_k, in

zato se pri velikem številu poskusov povprečna vrednost realizacij slučajne spremenljivke X običajno ustali pri nekem številu

$$E(X) = \sum_{k=1}^{n} p_k\, x_k \tag{10}$$

Število $E(X)$, definirano z obrazcem (10), imenujemo **matematično upanje**, matematično pričakovanje ali tudi kar povprečna vrednost slučajne spremenljivke X. Izraz (10) je pripravljen za spremenljivko s končno zalogo vrednosti in tako definirano število $E(X)$ vedno obstaja. Pri slučajnih spremenljivkah z bogatejšimi zalogami vrednosti pa ne moremo biti vnaprej prepričani, da matematično upanje obstaja. (Vprašanje bistveno presega okvire tega učbenika in se z njim ne bomo ukvarjali.)

Izraz "matematično upanje" izvira iz iger na srečo. Če je v loteriji n različnih dobitkov s frekvencami f_k, je povprečni dobitek na eno srečko dan z obrazcem (8); toliko lahko v povprečju pričakujemo (na toliko upamo) pri nakupu ene srečke.

ZGLED 4.7.6:

Izračunaj matematično upanje slučajne spremenljivke X, ki pomeni frekvenco grba v 4 metih kovanca!

Iz zgleda 4.7.2 vemo, da je ta slučajna spremenljivka porazdeljena binomsko, $X \sim b(n,p)$, imamo pa tudi njeno verjetnostno shemo

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{1}{16} & \frac{4}{16} & \frac{6}{16} & \frac{4}{16} & \frac{1}{16} \end{pmatrix}$$

Po obrazcu (10) je

$$E(X) = \frac{1}{16}\cdot 0 + \frac{4}{16}\cdot 1 + \frac{6}{16}\cdot 2 + \frac{4}{16}\cdot 3 + \frac{1}{16}\cdot 4 = 2$$

Rezultat se ujema s pričakovanji oziroma z razumevanjem matematičnega upanja slučajne spremenljivke: $E(X) = 2$ pomeni, da bo povprečna vrednost realizacij te slučajne spremenljivke v velikem številu poskusov enaka 2 (ali vsaj blizu temu številu), ali drugače, da bomo v velikem številu ponovitev poskusa "štirikrat vržemo kovanec" v povprečju dobili po dvakrat grb (in dvakrat številko). To seveda ne izključuje poskusov, v katerih dobimo po štiri grbe ali številke, zato tudi poudarjamo izraz "v povprečju"! Moč verjetnostnega računa je pač v opisu množičnih pojavov (dolgih zaporedij nekega poskusa), ne pa posameznih poskusov...

V našem zgledu smo dobili za $X \sim b(4, 0{\cdot}5)$ vrednost $E(X) = 2$, iz premisleka ob koncu zgleda pa sklepamo, da bi to vrednost lahko izračunali kot produkt števila poskusov n in verjetnosti proučevanega dogodka v posamezni ponovitvi poskusa (p). Mogoče je dokazati (pa zaradi obsežnosti tega ne bomo storili), da velja za poljubno binomsko porazdeljeno slučajno spremenljivko naslednji izrek:

$$X \sim b(n,p) \Longrightarrow E(X) = np \tag{11}$$

Dodajmo še dva zgleda.

ZGLED 4.7.7:

Lokostrelec, ki pri posameznem strelu zadene tarčo z verjetnostjo 0'5, ima štiri puščice in strelja proti tarči toliko časa, dokler je ne zadene ali pa mu zmanjka puščic. Naj bo slučajna spremenljivka X število izstreljenih puščic. Izračunaj njeno matematično upanje!

Zaloga vrednosti slučajne spremenljivke X je množica $\{1,2,3,4\}$, verjetnostna funkcija pa ima naslednje vrednosti:

$$\begin{aligned}
p_1 &= P(X=1) = \tfrac{1}{2} &&= \tfrac{1}{2}\\
p_2 &= P(X=2) = \tfrac{1}{2}\cdot\tfrac{1}{2} &&= \tfrac{1}{4}\\
p_3 &= P(X=3) = (\tfrac{1}{2})^2\cdot\tfrac{1}{2} &&= \tfrac{1}{8}\\
p_4 &= P(X=4) = (\tfrac{1}{2})^3 &&= \tfrac{1}{8}
\end{aligned}$$

Pri računanju p_4 je treba posebej paziti: to je verjetnost dogodka, da lokostrelec porabi vse štiri puščice, kar se zgodi natanko takrat, ko trikrat zgreši, nič pa ni važno, kaj je z uspešnostjo zadnjega, to je četrtega strela.

Matematično upanje je po (10) enako

$$E(X) = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + \frac{1}{8}\cdot 4 = \frac{15}{8} = 1'875$$

Kako bi pravzaprav interpretirali ta rezultat? Lokostrelec porabi v povprečju nekaj manj kot dve puščici za zadetek. Za posamezno streljanje podatek nima posebnega pomena: prav gotovo se ne da izstreliti 1'875 puščice: dostikrat (približno v polovici primerov) bo zadel že s prvo, v povprečju bo dvakrat manj poskusov, kjer bosta potrebni po dve puščici, pa tudi verjetnosti za dogodka, da porabi tri oziroma štiri puščice, nista zanemarljivi. Če pa bi opravili dovolj veliko število takih poskusov, vsakič zabeležili število izstreljenih puščic in nato izračunali "povprečno porabo" na eno ponovitev poskusa, bi se skoraj zagotovo rezultat ne razlikoval dosti od 1'875.

(Kaj pomeni v verjetnostnem računu "skoraj zagotovo", bomo povedali v razdelku o zakonu velikih števil.)

ZGLED 4.7.8:

Janez in Miha se dogovorita takole: prvi bo iz posode, v kateri so 3 bele in 7 črnih kroglic, na slepo sočasno izbral dve kroglici. Če bosta kroglici raznobarvni, igra ne šteje in poskus ponovita, če bosta obe kroglici črni, pa Janez plača Mihu 10 tolarjev. Koliko mora Miha plačati Janezu, kadar sta obe kroglici beli, če naj bo igra poštena?

Ker je število kroglic posamezne barve v vzorcu slučajna spremenljivka, je slučajna spremenljivka tudi Janezov dobiček v posamezni igri. Če naj bo igra poštena, mora biti matematično upanje dobička enako 0; od nič različno matematično upanje dobička bi namreč pomenilo, da ima eden od igralcev na dolgi rok sistematično prednost pred drugim.

Janezov dobiček v opisani igri lahko predstavimo z naslednjo verjetnostno shemo:

$$X : \begin{pmatrix} x_1 & 0 & -10 \\ \frac{3}{45} & \frac{21}{45} & \frac{21}{45} \end{pmatrix}$$

Pri tem smo z x_1 označili neznani prejemek v primeru vzorca iz dveh belih krogel, z 0 stanje, ko ni "plačilnega prometa" med igralcema, z -10 pa Janezov izdatek (odtod minus) v primeru, ko sta obe kroglici črni. V drugi vrstici imamo seveda verjetnosti dogodkov, da sta v vzorcu dve beli, raznobarvni oziroma črni kroglici (vaja!). Po obrazcu (10) je

$$E(X) = \frac{3}{45}\cdot x_1 + \frac{21}{45}\cdot 0 + \frac{21}{45}\cdot(-10) = 0 \Longrightarrow x_1 = 70$$

Ker matematično upanje pomeni povprečno vrednost realizacij slučajne spremenljivke v velikem številu poskusov, lahko odgovor na postavljeno vprašanje takole: v posamezni igri ima Janez ali 10 tolarjev izgube ali 70 tolarjev dobička. Ker se prvo v povprečju dogaja sedemkrat pogosteje kot drugo, je pričakovana vrednost njegovega dobička na dolgi rok - natanko tako, kot smo zahtevali - enaka 0.

Disperzija in standardni odklon

Matematično upanje slučajne spremenljivke ima podobno slabost kot aritmetična sredina, ki jo poznamo iz statistike (da sta tudi sicer v sorodstvu, smo ugotovili že na začetku tega razdelka!). Število $E(X)$ sicer zagotovo leži nekje med najmanjšo in največjo možno vrednostjo slučajne spremenljivke X, vendar so te vrednosti lahko precej razpršene okrog njega in je zato informacija, ki jo daje $E(X)$ o slučajni spremenljivki X, lahko v splošnem precej skopa. Pri enaki vrednosti $E(X)$ je lahko porazdelitev dveh slučajnih spremenljivk še zelo različna.

ZGLED 4.7.9:

Slučajni spremenljivki X in Y, dani z verjetnostnima shemama

$$X : \begin{pmatrix} -100 & -10 & 0 & 20 \\ 0{\cdot}1 & 0{\cdot}2 & 0{\cdot}1 & 0{\cdot}6 \end{pmatrix}, \quad Y : \begin{pmatrix} -0{\cdot}01 & 0{\cdot}01 \\ 0{\cdot}50 & 0{\cdot}50 \end{pmatrix}$$

imata enaki matematični upanji $E(X) = E(Y) = 0$, vendar očitno ta podatek precej več pove o slučajni spremenljivki Y kot pa o spremenljivki X.

Kvaliteto informacije, ki jo daje matematično upanje, bomo ocenili tako, da bomo povedali, kako so posamezne vrednosti te slučajne spremenljivke razpršene okrog njenega matematičnega upanja. Razpršenost bomo merili z **disperzijo** slučajne spremenljivke, ki jo definiramo analogno, kot smo to naredili z varianco v opisni statistiki: pogledamo odklone individualnih vrednosti slučajne spremenljivke od njenega matematičnega upanja, jih kvadriramo in izračunamo pričakovano vrednost:

$$D(X) \;=\; E\{\,[X - E(X)\,]^2\,\} \tag{12}$$

Konkreten izračun je odvisen od tega, s kakšnim tipom slučajne spremenljivke imamo opraviti; pri spremenljivki s končno zalogo vrednosti gre izračun po popolnoma enaki logiki kot računanje variance iz danih statističnih podatkov, s to razliko, da vlogo (empiričnih) relativnih frekvenc prevzamejo verjetnosti kot števila, pri katerih se prej omenjene relativne frekvence navadno ustalijo v velikem številu realizacij slučajne spremenljivke. Tako imamo za takšno slučajno spremenljivko obrazec

$$D(X) \;=\; \sum_{i=1}^{n} p_i\,[\,x_i \;-\; E(X)\,]^2 \tag{13}$$

ki ga lahko z enakimi prijemi kot v statistiki predelamo v bolj praktično obliko

$$D(X) \;=\; \{\sum_{i=1}^{n} p_i\; x_i^2\,\} \;-\; [\,E(X)\,]^2 \tag{14}$$

ZGLED 4.7.10:

Izračunaj disperzijo slučajne spremenljivke X, porazdeljene po zakonu b(4,1/2)!

Kot vemo, je tako npr. porazdeljena frekvenca grba v 4 metih kovanca. Iz zgleda 4.7.6 poznamo njeno matematično upanje, $E(X) = 2\,(= np)$, uporabimo lahko tudi tam zapisano verjetnostno shemo in obrazec (13) ali - še bolje - obrazec (14):

$$D(X) = \frac{1}{16}\cdot 0^2 + \frac{4}{16}\cdot 1^2 + \frac{6}{16}\cdot 2^2 + \frac{4}{16}\cdot 3^2 + \frac{1}{16}\cdot 4^2 \;-\; 2^2 \;=\; 1$$

Bralec bo tudi tu lahko ugotovil, da je ta oblika obrazca prijaznejša do uporabnika (kot je moderno reči) od prvotne oblike, kjer bi morali računati takole:

$$D(X) \;=\; \frac{1}{16}\cdot(0-2)^2 + \frac{4}{16}\cdot(1-2)^2 + \frac{6}{16}\cdot(2-2)^2 + \\ + \frac{4}{16}\cdot(3-2)^2 + \frac{1}{16}\cdot(4-2)^2 \;=\; 1$$

Komur ta zgled ni dovolj za potrditev, lahko primerja obe različici na kaki spremenljivki, pri kateri matematično upanje ni celo število...

Za vajo izračunaj tudi disperzijo spremenljivk X in Y iz zgleda 4.7.9!

Tudi zdaj je žal merska enota, v kateri izražamo disperzijo, kvadrat merske enote slučajne spremenljivke same. Zato je namesto disperzije za praktično rabo bolj priročna standardna deviacija,

$$\sigma \;=\; \sqrt{D(X)} \tag{15}$$

Za slučajno spremenljivko, ki je porazdeljena po binomskem zakonu $b(n,p)$, je mogoče pokazati, da velja

$$D(X) \;=\; n\,p\,(1-p) \;=\; n\,p\,q \tag{16}$$

in zato po obrazcu (15)

$$X \sim b(n,p) \implies \sigma(X) \;=\; \sqrt{n\,p\,q} \tag{17}$$

Ta rezultat bomo potrebovali v naslednjih razdelkih.

△ Neenačba Čebiševa

Disperzija in standardna deviacija slučajne spremenljivke merita razpršenost njenih vrednosti okrog matematičnega upanja. Konkreten občutek o razpršenosti dobimo, če na primer povemo, koliko je v povprečju (v velikem številu poskusov) delež tistih realizacij slučajne spremenljivke, pri katerih ta zavzame vrednost z nekega intervala okrog matematičnega upanja, recimo $(E(X) - t, E(X) + t)$, pri čemer je t poljubno pozitivno realno število, ali drugače, kolikšna je verjetnost, da se to zgodi. Pri tem nam je v pomoč t.i. neenačba Čebiševa (ruski matematik, 1821-1894). Neenačba pravi, da velja za vsako slučajno spremenljivko, ki ima končno disperzijo $D(X)$, naslednja ocena:

$$t > 0 \implies P(E(X) - t < X < E(X) + t) \geq 1 - \frac{D(X)}{t^2} \tag{18}$$

Neenačbe ne bomo dokazovali, nekaj pa le povejmo o njej. Iz (16) lahko takoj preberemo, da je verjetnost nasprotnega dogodka, da slučajna spremenljivka zavzame vrednost, ki se vsaj za t loči od $E(X)$, kvečjemu $D(X)/t^2$; veliki odkloni vrednosti slučajne spremenljivke od njenega matematičnega upanja so malo verjetni, tem manj, čim manjša je njena disperzija. Vidimo tudi, da te verjetnosti padajo s "kvadratom razdalje": dvakrat večji odklon ima štirikrat manjšo verjetnost.

Na levi strani neenačbe (18) je zapisana verjetnost dogodka, da X zavzame vrednost, ki je za manj kot t oddaljena od $E(X)$. Ker razdaljo dveh števil na realni osi merimo z absolutno vrednostjo njune razlike, lahko namesto (18) pišemo tudi

$$P(\,|X - E(X)|\, < t) \geq 1 - \frac{D(X)}{t^2} \qquad (19.a)$$

Ta zapis je v literaturi zelo pogost, enako tudi analogna ocena za verjetnost nasprotnega dogodka:

$$P(\,|X - E(X)|\, \geq t) \leq \frac{D(X)}{t^2} \qquad (19.b)$$

ZGLED 4.7.11:

Z neenačbo Čebiševa oceni verjetnost, da je pri 4 500 metih poštene igralne kocke število šestic večje od 700 in manjše od 800!

Vemo, da je frekvenca dogodka v Bernoullijevem zaporedju neodvisnih poskusov porazdeljena binomsko s parametroma n in p, ki pomenita število poskusov oziroma verjetnost za nastop dogodka v posamezni ponovitvi, torej pri nas $n = 4\,500$, $p = 1/6$, $q = 1 - p = 5/6$, zato je za to spremenljivko
$E(X) = 750\,,\ D(X) = 625$
Dogodek, katerega verjetnost želimo oceniti, lahko opišemo drugače, namreč da X zavzame vrednost, ki se za manj kot $t = 50$ loči od njenega matematičnega upanja; po obrazcu (18) dobimo za iskano verjetnost oceno

$$P(700 < X < 800) = P(E(X) - 50 < X < E(X) + 50) \geq$$

$$\geq 1 - \frac{D(X)}{50^2} = 1 - \frac{625}{2\,500} = 0{\cdot}75$$

Dobljeno oceno lahko interpretiramo takole: če opravimo veliko število serij po 4 500 metov poštene igralne kocke, lahko pričakujemo, da bo vsaj za tri četrtine serij število šestic med 700 in 800.

Tudi neenačba Čebiševa nas seveda ne zavaruje pred individualnimi odstopanji vrednosti, vendar nas to ne moti več. Opozoriti pa moramo na neko

drugo pomanjkljivost te neenačbe: število $1 - D(X)/t^2$ je res samo spodnja meja verjetnosti dogodka, da slučajna spremenljivka zavzame vrednost z intervala, ki seže t v levo in t v desno od njenega matematičnega upanja; dobljena ocena se lahko precej razlikuje od natančne spodnje meje oziroma prave vrednosti; za zgled 4.7.11 dobimo na primer z natančnejšim računanjem, da je $P(700 < X < 800) \approx 0{\cdot}95$.

VAJE

1. Sočasno vržemo dve (pošteni) igralni kocki. Slučajna spremenljivka X naj bo vsota pik na obeh kockah. Izračunaj njeno verjetnostno shemo.

2. Poskus naj bo enak kot v nalogi 1, slučajna spremenljivka Y pa naj ima za vrednost večje od obeh števil pik. Določi njeno verjetnostno shemo.

3. Slučajna spremenljivka X ima za zalogo vrednosti prvih 9 naravnih števil in enakomerno porazdelitev verjetnosti. Zapiši njeno verjetnostno shemo.

4. Slučajna spremenljivka X naj bo frekvenca številke v
a) dveh metih;
b) treh metih kovanca.
Izračunaj ustrezni verjetnostni shemi. Kako je X porazdeljena v posameznem primeru?

5. V posodi imamo 4 bele in 6 črnih kroglic. Štirikrat na slepo izvlečemo po eno kroglo in jo vsakič vrnemo v posodo. Slučajna spremenljivka Z naj bo frekvenca bele krogle. Izračunaj njeno verjetnostno shemo. Kolikšna je verjetnost, da v opisanem poskusu vsaj enkrat izvlečemo belo kroglo?

6. V posodi imamo 4 bele in 6 črnih kroglic. Na slepo sočasno izvlečemo štiri kroglice. Slučajna spremenljivka Y naj bo število belih krogel v tako oblikovanem vzorcu. Določi njeno verjetnostno shemo in jo primerjaj z verjetnostno shemo slučajne spremenljivke Z iz naloge 5. Kolikšna je sedaj verjetnost, da je med izvlečenimi kroglami vsaj ena bela?

7. Strelec, ki v posameznem strelu zadene tarčo z verjetnostjo 0'8, ima 4 naboje. Strelja toliko časa, da zadene tarčo, ali pa mu zmanjka nabojev. Izračunaj verjetnostno shemo slučajne spremenljivke X, ki pomeni število porabljenih nabojev.

8. Delavec preizkuša 5 naprav, ki se pokvarijo (neodvisno druga od druge) z verjetnostjo 0'1. Če je naprava brezhibna, nadaljuje s pregledovanjem, v nasprotnem primeru se loti popravljanja. Slučajna spremenljivka X naj pomeni število naprav, ki jih pregleda, preden prvič začne s popravilom. Določi njeno verjetnostno shemo, če lahko predpostavimo, da delavec v primeru, ko je vseh pet naprav brezhibnih, ne začenja znova s pregledovanjem in je zato največja možna vrednost te spremenljivke 5.

9. Slučajna spremenljivka X zavzame vrednosti -2, -1, 0, 1 oziroma 2 z verjetnostmi 1/10, 2/10, 2/10, 4/10, 1/10. Zapiši podatke v obliki verjetnostne sheme, nato pa iz nje ugotovi, kakšna je verjetnost, da ta slučajna spremenljivka zavzame vrednost,

a) ki je pozitivna;

b) ki je nenegativna;

c) ki je negativna;

č) ki je nepozitivna;

d) ki leži na intervalu $[-1, 1]$;

e) ki leži na intervalu $(-1, 1)$.

10. Dva košarkarja izmenično mečeta na koš, dokler eden od njiju ne zadene; verjetnosti za zadetek pri posameznem metu sta 0·4 za prvega oziroma 0·6 za drugega. Slučajna spremenljivka X naj bo število metov, ki so potrebni za dokončanje igre. Ali ima ta slučajna spremenljivka končno zalogo vrednosti? Izračunaj nekaj prvih vrednosti verjetnostne funkcije $p_k = P(X = x_k)$, na primer $P(X = 1)$, $P(X = 2)$ in $P(X = 3)$.

V vseh naslednjih nalogah izračunaj matematično upanje, disperzijo in standardno deviacijo slučajne spremenljivke:

11. Število pik pri metu poštene igralne kocke.

12. Vsota pik pri metu dveh poštenih igralnih kock.

13. Število grbov pri metu treh kovancev.

14. $\triangle$ Število šestic v 18 000 metih poštene igralne kocke.

15. Število metov, ki so potrebni, da pri metu kovanca prvič pade grb ali pet številk zapovrstjo.

16. Dobiček igralca v naslednji igrici: Igralec vrže dva kovanca hkrati; če pade pri tem natanko en grb, dobi en tolar; če padeta dva grba, prejme dva tolarja; če ni nobenega grba, plača pet tolarjev.

17. Število poskusov, ki so potrebni, da prvič izvlečemo belo kroglo pri poskusu: iz posode s 3 črnimi in 7 belimi kroglami na slepo brez vračanja izbiramo po eno kroglo.

18. Frekvenco bele krogle v 400 ponovitvah poskusa: slepo izbiranje ene krogle iz posode, v kateri imamo 6 belih in 4 črne krogle, če krogle vračamo v posodo.

19. Število črnih kroglic, ki jih dobimo v vzorcu iz dveh kroglic, sočasno na slepo izbranih iz posode, v kateri imamo 2 črni in 3 bele kroglice.

20. Število defektnih izdelkov v vzorcu iz šestih na slepo izbranih izdelkov iz velike serije, v kateri je 10% defektnih izdelkov.

21. $\triangle$ Z neenačbo Čebiševa oceni verjetnost dogodka, da slučajna spremenljivka iz naloge 14 zavzame vrednost z intervala (2 900, 3 100). Z besedami razloži ta rezultat.

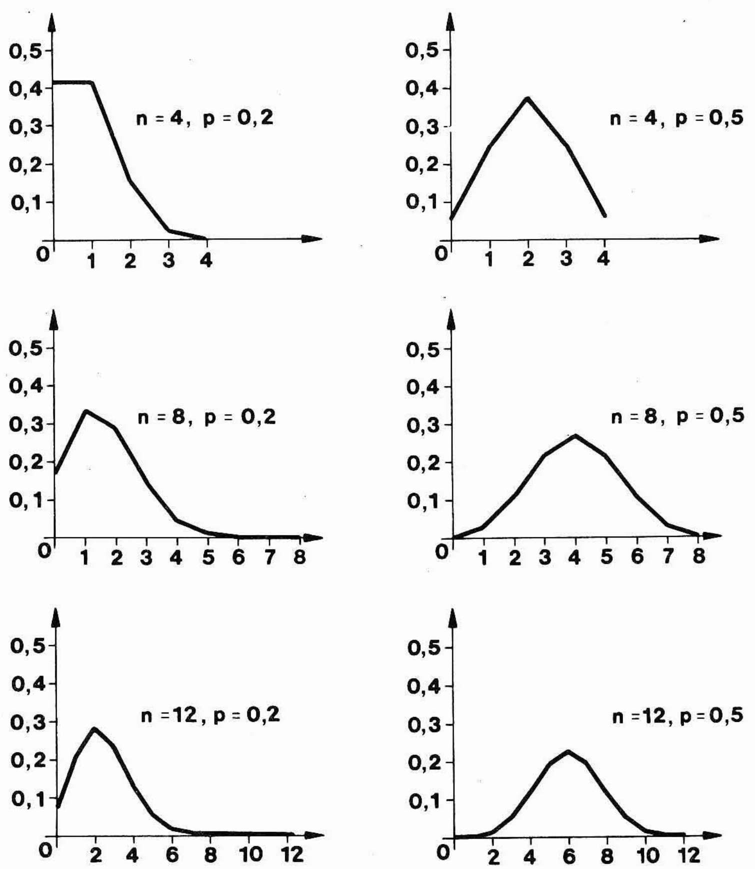
22. $\triangle$ Iz posode, v kateri imamo 4 rdeče in 6 belih krogel, štiristokrat na slepo izvlečemo po eno kroglo in jo vsakič vrnemo v posodo. Z neenačbo Čebiševa oceni verjetnost dogodka, da je pri tem frekvenca rdeče krogle od 140 do 180.

23. $\triangle$ Oceni verjetnost, da pri 10 000 metih kovanca pade manj kot 4 000 ali več kot 6 000 grbov.

4.8 Normalna porazdelitev

Do zdaj smo se ukvarjali samo z diskretnimi slučajnimi spremenljivkami, z nekaj izjemami celo samo s takimi, ki so imele končno zalogo vrednosti. Žal je naše matematično znanje preskromno, da bi se lahko zares lotili tudi zveznih porazdelitev verjetnosti, na kratko in na zelo poenostavljen način pa si bomo ogledali najpomembnejšo predstavnico tega razreda.

Binomska porazdelitev verjetnosti $b(n,p)$, ki smo jo srečali v prejšnjem poglavju, ima zanimivo lastnost, da pri povečevanju parametra n (števila poskusov) graf ustrezne verjetnostne funkcije postaja vedno bolj podoben neki gladki krivulji zvonaste oblike, tem hitreje, čim bliže 0'5 je parameter p. Za ilustracijo je na naslednji sliki nekaj primerov.



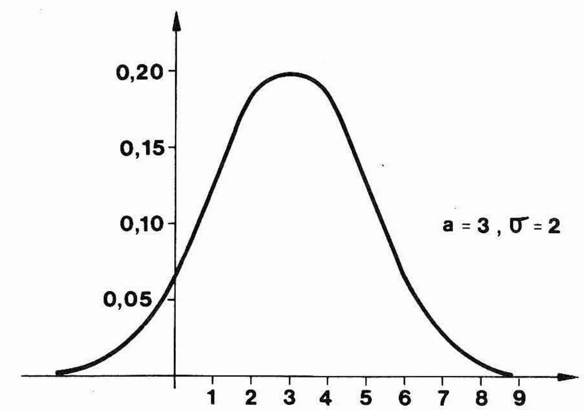
Slika 4.5: Verjetnostne funkcije binomske porazdelitve pri različnih n in p

Dovolimo zdaj, da n narašča preko vsake meje. Z nekaj truda je mogoče dokazati (dovolj preprost dokaz je v knjigi A. Vadnala *Elementarni uvod v verjetnostni račun*, Ljubljana, DZS 1972), da ima zvonasta krivulja, ki jo

pri tem dobimo iz verjetnostne funkcije, enačbo

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \qquad (1)$$

Za poseben primer $a = 3, \sigma = 2$ je krivulja narisana na naslednji sliki.



Slika 4.6: Graf funkcije (1) za $a = 3, \sigma = 2$

Krivuljo imenujemo **normalna krivulja**.[1]

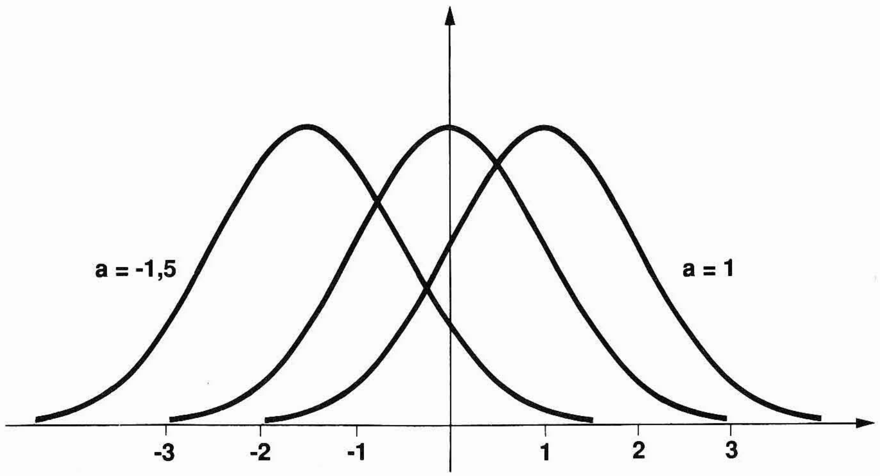
Ime "normalna" izvira iz dejstva, da z njo lepo opišemo, kako so porazdeljene frekvence posameznih izmerkov pri zaporednih merjenjih neke količine v "normalnih pogojih", torej takrat, ko pri merjenju ne delamo sistematičnih napak, ampak so napake povsem slučajne. Izmerki so pravzaprav vrednosti neke slučajne spremenljivke X, ki lahko (vsaj načelno) zavzame katerokoli realno vrednost in ne le celoštevilskih, zato imamo opraviti z zvezno slučajno spremenljivko oziroma z zvezno porazdelitvijo verjetnosti.

Normalno porazdelitev simbolično označimo z $N(a,\sigma)$, določena je namreč s tema dvema parametroma, namreč a in σ. Pokaže se, da je a ravno matematično upanje tako porazdeljene slučajne spremenljivke in σ njen standardni odklon. Iz funkcijske enačbe (1) tudi vidimo, da ta funkcija zavzame največjo vrednost $y = 1/(\sigma\sqrt{2\pi})$ pri $x = a$, torej pri abscisi, ki ustreza matematičnemu upanju slučajne spremenljivke X. Sprememba parametra a pomeni potemtakem pomik celotne krivulje v smeri abscisne osi. (Glej sliko 4.7!)

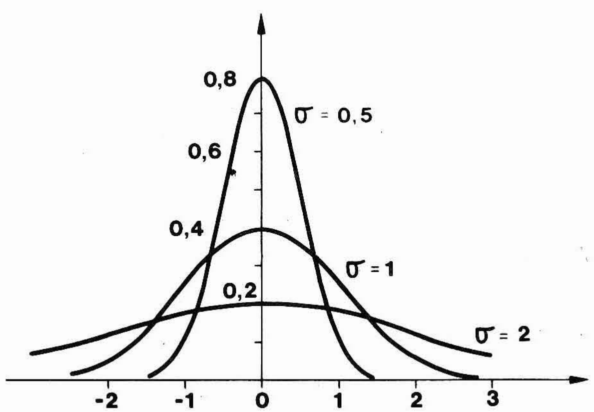
Drugi parameter σ (standardni odklon) meri razpršenost vrednosti slučajne spremenljivke okrog njenega matematičnega upanja, zato je pri večjih

[1] Mnogi jo imenujejo po nemškem matematiku Carlu Friedrichu Gaussu (1777 - 1855), ki jo je uporabil v analizi slučajnih napak pri merjenjih, Gaussova krivulja, čeprav jo je v bistvu v prvi polovici 18. stoletja odkril francoski matematik de Moivre.

σ normalna krivulja bolj sploščena (raztegnjena v smeri abscisne osi). Na sliki 4.8 je ta odvisnost prikazana s tremi krivuljami, ki imajo enako vrednost $a = 0$ in različne σ.



Slika 4.7: Odvisnost normalne porazdelitve od parametra a



Slika 4.8: Odvisnost normalne porazdelitve od parametra σ

Normalna porazdelitev je za prakso najpomembnejša verjetnostna porazdelitev, zato obstajajo zanjo posebne tabele. V splošnem moramo imeti za neko verjetnostno porazdelitev, ki je odvisna od parametrov, toliko različnih tabel, kolikor kombinacij teh parametrov nas zanima. Pri normalni porazdelitvi pa nam naslednji izrek pomaga, da shajamo z eno samo tabelo:

IZREK: Če je slučajna spremenljivka X porazdeljena normalno s parametroma a in σ, torej $X \sim N(a, \sigma)$, je tudi slučajna spremenljivka

$$Z = \frac{X - a}{\sigma} \tag{2}$$

porazdeljena normalno, vendar z matematičnim upanjem $a = 0$ in standardnim odklonom $\sigma = 1$, torej $Z \sim N(0, 1)$.

Porazdelitvi $N(0, 1)$ rečemo **standardizirana normalna porazdelitev**, prehodu (2) pa **standardizacija** slučajne spremenljivke. Za standardizirano normalno porazdelitev se funkcijska enačba (1) poenostavi v

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \varphi(x) \tag{3}$$

Tistim, ki že (še) obvladajo eksponentno funkcijo, tudi podrobnejša analiza te funkcije ne bi smela biti pretežka; za druge pač velja, da je poseben primer normalne krivulje. Ker je funkcija (3) soda (še vemo, kaj to pomeni?), je njen graf simetričen glede na ordinatno os. Funkcija ima največjo vrednost pri $x = 0$, ko x po absolutni vrednosti narašča, pa funkcijske vrednosti asimptotično padajo proti 0. Funkcija je povsod pozitivna in ima značilno zvonasto obliko. K splošni obliki (1) bi se lahko iz (3) vrnili s premikom in raztegom.

Kakšna je zveza med normalno krivuljo (1) - ali njenim posebnim primerom (3) - in verjetnostmi posameznih dogodkov, ki se lahko pripetijo normalno porazdeljeni slučajni spremenljivki? Najprej **opozorilo**: *Vrednost funkcije (1) v izbrani točki x* ***ni enaka verjetnosti*** , da za slučajno spremenljivko $X \sim N(a, \sigma)$ nastopi dogodek $(X = x)$.

Pač pa se pokaže (dokaza si z našim znanjem žal ne moremo privoščiti in bo treba verjeti na besedo), da je zveza med (1) in verjetnostmi kljub temu dokaj enostavna in da velja:

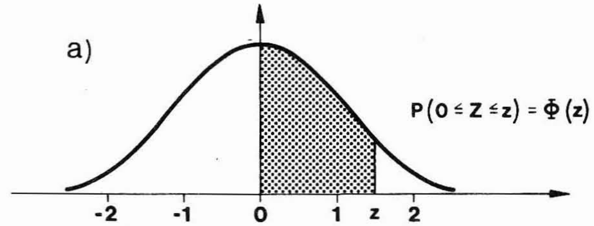
Verjetnost, da slučajna spremenljivka $X \sim N(a, \sigma)$ zavzame vrednost, ki leži na intervalu $[\alpha, \beta]$, je enaka ploščini lika, ki ga omejujejo abscisna os, navpičnici v $x = \alpha$ in $x = \beta$ ter graf funkcije (1).

To velja v bistvu za vsako zvezno porazdelitev verjetnosti, spreminja se le oblika funkcije. Odtod pa takoj sledi, da mora biti pri poljubni zvezni porazdelitvi verjetnost za vsako posamezno točko enaka 0, saj opisani lik v tem primeru degenerira v daljico, kako pa je s ploščino daljice, pa je splošno znano...

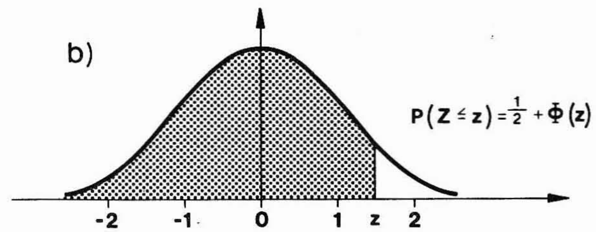
Pozitivna posledica tega - za nekatere šokantnega - dejstva pa je, da nam pri zveznih porazdelitvah - in posebej pri normalni porazdelitvi - ni treba posebej skrbeti za robne točke: verjetnost, da vrednost slučajne spremenljivke pade v interval (α, β), je natanko takšna kot za $[\alpha, \beta]$ in za oba "hibrida" $[\alpha, \beta)$ oziroma $(\alpha, \beta]$.

Posledica izreka o standardizaciji je med drugim tudi to, da bo dovolj, če bomo poznali ploščine pod krivuljo (3). Še več, zaradi sodosti funkcije in s tem povezane simetrije grafa je dovolj, če poznamo ploščino pod grafom (3) v mejah od 0 do poljubnega nenegativnega števila, recimo z, $z \geq 0$. To ploščino kot funkcijo zgornje meje intervala bomo označevali s $\Phi(z)$.

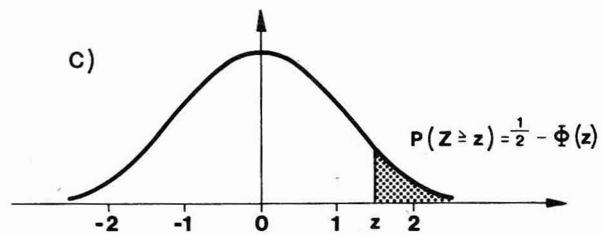
Verjetnost dogodka, da $Z \sim N(0,1)$ zavzame neko vrednost z intervala $[0, z]$.



Verjetnost dogodka, da $Z \sim N(0,1)$ ne preseže števila $z \geq 0$.



Verjetnost dogodka, da $Z \sim N(0,1)$ zavzame vrednost, ki je vsaj enaka $z \geq 0$.



Ker je $(-\infty < Z < +\infty)$ gotov dogodek z verjetnostjo 1, je tudi ploščina pod celotno krivuljo (nad celotno abscisno osjo) natanko 1, ploščina pod delom, ki leži v prvem kvadrantu, pa 1/2. Če k temu dodamo še dogovor, da bodi

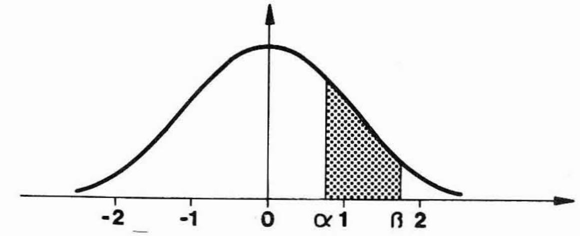
$$\Phi(-z) = -\Phi(z) \tag{4}$$

lahko z vrednostjo funkcije $\Phi(z)$ izrazimo verjetnosti, da standardizirano normalno porazdeljena slučajna spremenljivka zavzame vrednosti s poljubnega intervala na realni osi. (Tisti, ki se spoznajo na računanje ploščin z integrali, vedo, da se ni kaj dogovarjati, ampak je (4) neposredna posledica lastnosti določenega integrala.) Kako naj ravnamo, nam povesta tudi sliki na naslednji strani.

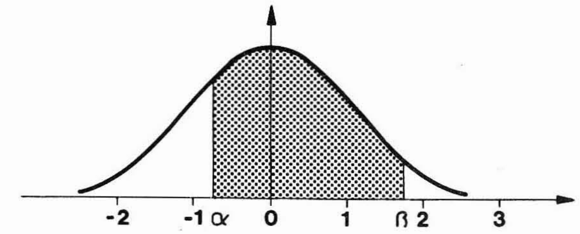
V prvem primeru imamo interval $[\alpha, \beta]$ s pozitivno spodnjo mejo; ploščino nad tem intervalom dobimo tako, da najprej vzamemo vse (ploščino od 0 do β), nato pa vrnemo tisto, kar nam ne pripada, namreč ploščino od 0 do α. Zato je

$$P(\alpha \leq Z \leq \beta) = \Phi(\beta) - \Phi(\alpha) \tag{5}$$

Verjetnost dogodka, da $Z \sim N(0,1)$ zavzame vrednost z intervala $[\alpha, \beta]$, $\alpha > 0$.



Verjetnost dogodka, da $Z \sim N(0,1)$ zavzame vrednost z intervala $[\alpha, \beta]$, $\alpha < 0$.



V drugem primeru je na prvi pogled drugače, saj je očitno treba k ploščini od 0 do β dodati ploščino med (negativnim) α in 0. Ker pa imamo (4), brez skrbi avtomatično uporabimo pravilo (5), saj je

$$P(\alpha \leq Z \leq \beta) = \Phi(\beta) - \Phi(\alpha) = \Phi(\beta) - (-\Phi(-\alpha)) = \Phi(\beta) + \Phi(-\alpha)$$

Ne pozabimo seveda, da je $-\alpha$ absolutna vrednost negativnega števila α in da zato obrazec res naroča seštevanje ustreznih ploščin.

Vse, kar nam še manjka, je tabela vrednosti funkcije $\Phi(z)$. Grobo preglednico, z dodanimi vrednostmi funkcije (3), najdemo na naslednji strani, natančnejšo tabelo pa najde bralec v zahtevnejših učbenikih verjetnostnega računa.

Kako ravnamo v splošnem primeru, ko normalna porazdelitev ni standardizirana? Po izreku s strani 154 jo lahko standardiziramo, ne da bi pokvarili normalnost; zato lahko računamo takole:

$$P(\alpha \leq X \leq \beta) = P\left(\frac{\alpha - a}{\sigma} \leq \frac{X - a}{\sigma} \leq \frac{\beta - a}{\sigma}\right) =$$

$$= P\left(\frac{\alpha - a}{\sigma} \leq Z \leq \frac{\beta - a}{\sigma}\right) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)$$

torej na kratko

$$P(\alpha \leq X \leq \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right) \tag{6}$$

z	Vrednosti funkcije $\varphi(z)$	$\Phi(z)$
0·0	0·3989	0·0000
0·1	0·3970	0·0398
0·2	0·3910	0·0793
0·3	0·3814	0·1179
0·4	0·3683	0·1554
0·5	0·3521	0·1915
0·6	0·3332	0·2257
0·7	0·3123	0·2580
0·8	0·2897	0·2881
0·9	0·2661	0·3159
1·0	0·2420	0·3413
1·5	0·1295	0·4332
2·0	0·0540	0·4772
2·5	0·0175	0·4938
3·0	0·0044	0·4987
3·5	0·0009	0·4998
...	...	...

Pred zgledi pa si poglejmo še neko izredno zanimivo in koristno lastnost normalne porazdelitve. Izrek pravi takole:

Verjetnost dogodka, da normalno porazdeljena slučajna spremenljivka zavzame vrednost, ki se kvečjemu za $k\,\sigma$ (torej za določeno število standardnih odklonov) razlikuje od njenega matematičnega upanja, je odvisna samo od števila k, nič pa od konkretnih vrednosti parametrov a in σ, ki določata porazdelitev.

Dokaz je povsem neposreden. Določiti je treba verjetnost, da slučajna spremenljivka $X \sim N(a, \sigma)$ zavzame vrednost, ki se kvečjemu za $k\sigma$ razlikuje od $E(X) = a$, povedano drugače, verjetnost, da ta slučajna spremenljivka zavzame vrednost z intervala $[a - k\sigma,\, a + k\sigma]$. Po obrazcu (6) je

$$P(a - k\sigma \le X \le a + k\sigma) \;=\; \Phi\left(\frac{(a+k\sigma)-a}{\sigma}\right) - \Phi\left(\frac{(a-k\sigma)-a}{\sigma}\right) \;\Rightarrow$$

$$\Rightarrow\; P(a - k\sigma \le X \le a + k\sigma) \;=\; \Phi(k) - \Phi(-k) \;=\; 2\,\Phi(k) \qquad (7)$$

in rezultat dejansko ni odvisen od parametrov, ki opisujeta normalno porazdelitev. Ta ugotovitev ima daljnosežne posledice. Če vzamemo samo celoštevilske mnogokratnike standardnega odklona, je zaradi

$$\Phi(1) = 0{\cdot}3413, \quad \Phi(2) = 0{\cdot}4772, \quad \Phi(3) = 0{\cdot}4987$$

po obrazcu (7)

$$\begin{aligned} P(a - \sigma \le X \le a + \sigma) &= 0{\cdot}6826 \\ P(a - 2\sigma \le X \le a + 2\sigma) &= 0{\cdot}9544 \\ P(a - 3\sigma \le X \le a + 3\sigma) &= 0{\cdot}9974 \end{aligned}$$

in to **za poljubno normalno porazdelitev**. Za vsako normalno porazdeljeno spremenljivko lahko torej v velikem številu poskusov, pri katerih to slučajno spremenljivko realiziramo, pričakujemo, da bodo približno v 2/3 vseh poskusov (68·26%) vrednosti ujete v interval, ki sega en standardni odklon v levo in en standardni odklon v desno od njenega matematičnega upanja. S "pastjo", ki ima premer 4σ (pazi: dva odklona v levo, dva v desno!), ujamemo normalno porazdeljeno spremenljivko v več kot 95% vseh poskusov, vrednosti, ki so več kot 3σ oddaljene od njenega matematičnega upanja, pa so pravzaprav izjemne, slučajna spremenljivka jih (v povprečju) zavzame približno enkrat na vsakih 400 ponovitev poskusa.

Zdaj pa še nekaj prakse.

ZGLED 4.8.1:

Za slučajno spremenljivko $Z \sim N(0, 1)$ določi verjetnosti dogodkov,

a) da zavzame poljubno vrednost z intervala $[0, 2]$;

b) da zavzame poljubno vrednost z intervala $[1, 2]$;

c) da zavzame poljubno vrednost, večjo od 2;

č) da zavzame poljubno vrednost, manjšo od 2;

d) da zavzame poljubno vrednost z intervala $[-1, 3]$!

a) Iščemo $P(0 \le Z \le 2) = \Phi(2) = 0{\cdot}4772$.

b) Nalogo lahko rešimo po obrazcu (5):

$P(1 \le Z \le 2) = \Phi(2) - \Phi(1) = 0{\cdot}4772 - 0{\cdot}3413 = 0{\cdot}1359$.

Opozorimo še enkrat, da dobimo popolnoma enak rezultat tudi za intervale $(1, 2)$, $[1, 2)$ in $(1, 2]$.

c) Po zadnjem vzorcu s strani 155 je

$P(Z > 2) = 0{\cdot}5000 - \Phi(2) = 0{\cdot}0228$

č) Upoštevaje, da je za normalno (in sploh za zvezno) porazdeljeno spremenljivko $P(Z < 2) = P(Z \le 2)$, imamo po drugem vzorcu s strani 155

$P(Z < 2) = 0{\cdot}5000 + \Phi(2) = 0{\cdot}5000 + 0{\cdot}4772 = 0{\cdot}9772$.

d) Podobno kot v dokazu izreka imamo zaradi (4) in (5)
$P(-1 \leq Z \leq 3) = \Phi(3) - \Phi(-1) = \Phi(3) - (-\Phi(1)) =$
$\Phi(3) + \Phi(1) = 0{\cdot}4987 + 0{\cdot}3413 = 0{\cdot}8400$

Ker tega že nekaj časa nismo storili, na zadnjem primeru ponovimo, kaj taki rezultati pomenijo: Če velikokrat opravimo poskus, pri katerem nastopa slučajna spremenljivka $Z \sim N(0,1)$, lahko pričakujemo, da bo približno v 84% poskusov zavzela vrednost med -1 in 3.

Edina razlika med končanim zgledom in splošnim primerom $X \sim N(a, \sigma)$ je v tem, da je treba pred iskanjem po tablicah standardizirati zgornjo in spodnjo mejo intervala, za katerega računamo verjetnost, zato puščamo take račune za vajo. (Glej obrazec (6) in dokazovanje izreka na strani 158.)

V razdelku 4.6 smo obljubili, da bomo pokazali, kako se da približno oceniti verjetnosti za frekvence dogodka v zaporedju neodvisnih poskusov, ki bi jih sicer morali računati po Bernoullijevem obrazcu. Da z naraščanjem števila poskusov binomska porazdelitev (če smo pozabili: tako so porazdeljene verjetnosti posameznih frekvenc!) postaja vedno bolj podobna normalni, smo zaslutili iz grafičnega prikaza na začetku tega razdelka. Seveda bi bilo to lahko slučajno, toda izrek, za katerega sta zaslužna dva francoska matematika, Gaussov sodobnik Pierre Simon Laplace in starejši (že omenjeni) Abraham de Moivre, pravi, da ni:

Z naraščanjem števila poskusov preko vsake meje postaja binomska porazdelitev $b(n,p)$ podobna normalni porazdelitvi z enakim matematičnim upanjem in enakim standardnim odklonom, torej $N(np, \sqrt{npq})$.

Matematično zahtevnejši bralec se bo zagotovo spotaknil ob uporabo nedefiniranega izraza *podobna*. Ker nam manjka orodij za korektno predstavitev izreka, naj zadošča, če rečemo takole: za verjetnost, da binomsko porazdeljena slučajna spremenljivka zavzame vrednost z nekega intervala, je pri dovolj velikem številu poskusov zelo dober približek verjetnost, da v ta interval pade vrednost normalno porazdeljene slučajne spremenljivke, ki se z originalno binomsko porazdeljeno spremenljivko ujema v matematičnem upanju in standardnem odklonu. Kot že vemo (str. 144, str. 147), ima porazdelitev $b(n,p)$ matematično upanje $E(X) = np$ in standardni odklon $\sigma(X) = \sqrt{npq}$. Če je število poskusov dovolj veliko, lahko torej namesto Bernoullijevega obrazca uporabljamo tablice za normalno porazdelitev...

Kaj pomeni "dovolj veliko"? Podrobnejše spoznavanje z izrekom pokaže, da smo že iz grafičnega prikaza pravilno uganili: približevanje binomske porazdelitve k normalni je tem hitrejše, čim manj se verjetnost za uspeh v posameznem poskusu (p) razlikuje od 1/2. To ne preseneča, saj vemo, da je pri $p = 1/2$ binomska porazdelitev simetrična, pri bistveno manjših ali večjih vrednostih parametra p pa je potrebnih znatno več poskusov, da porazdelitev postane vsaj približno taka. Praksa kaže, da je število poskusov zagotovo dovolj veliko, da uporabimo "normalno aproksimacijo" za binomsko porazdelitev, če je produkt $n\,p\,(1-p) = npq$ vsaj 9.

ZGLED 4.8.2:

Z normalno porazdelitvijo oceni verjetnost dogodka, da je pri 4 500 metih poštene igralne kocke število šestic večje od 700 in manjše od 800!

Računamo, kolikšna je verjetnost dogodka A, da leži pri $n = 4\,500$ poskusih iz Bernoullijevega zaporedja neodvisnih poskusov frekvenca dogodka "padec šestice", ki ima v posamezni ponovitvi verjetnost $p = 1/6$, nad 700 in pod 800. Kot vemo, dobimo to verjetnost tako, da seštejemo verjetnosti posameznih frekvenc $k = 701, 702, ..., 799$,

$$P(A) = \sum_{k=701}^{799} \binom{4\,500}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{4\,500-k}$$

Ljubiteljem neposrednega računanja želimo veliko veselja, mi pa se lotimo približnega, binomsko porazdelitev nadomestimo z ustrezno normalno. Matematično upanje binomske porazdelitve $b(4\,500, \frac{1}{6})$ je 750, standardni odklon pa $\sqrt{625} = 25$ (preveri!). Interval, za katerega računamo verjetnost, sega torej približno dva standardna odklona v levo in v desno od matematičnega upanja in pri normalni porazdelitvi to pomeni, da je verjetnost zanj (glej str. 159!) dobrih 95%.

Ker gre v vsakem primeru za približek, takšna natančnost zadošča. Če bi želeli biti natančnejši, bi bilo korektno, če bi kot meji intervala vzeli 700'5 in 799'5, saj nadomeščanje diskretne (binomske) porazdelitve z zvezno (normalno) porazdelitvijo pomeni, da kaže izolirane točke, ki predstavljajo diskretne celoštevilske vrednosti na realni osi, "prekriti" z intervali dolžine 1, ki imajo središča v teh točkah. Treba pa je reči, da je takšna natančnost v večini praktičnih primerov odveč.

Bralec se je najbrž spomnil, da smo enako nalogo reševali s pomočjo neenačbe Čebiševa v prejšnjem razdelku. Tam smo dobili oceno, da je iskana verjetnost vsaj 0'75 (glej zgled 4.7.11 na strani 148!); sedanji račun je neenačbo Čebiševa potrdil, spet pa opozoril, da z njo dobimo neko spodnjo mejo, ki pa je lahko še zelo daleč od natančne spodnje meje za verjetnost, da slučajna spremenljivka leži na nekem intervalu s središčem v svojem matematičnem upanju. Zato v primerih, ko je to mogoče, raje uporabimo približek, dobljen z normalno porazdelitvijo. Ker se izkaže, da (zelo ohlapno rečeno) z naraščanjem števila poskusov marsikaj postane podobno normalni porazdelitvi, lahko tak prijem relativno pogosto uporabimo.

V A J E

1. Za slučajno spremenljivko Z, ki je porazdeljena standardizirano normalno, izračunaj verjetnosti dogodkov,
   a) da zavzame vrednost, večjo od 0˙8;
   b) da zavzame vrednost, manjšo od 1˙5;
   c) da zavzame vrednost, večjo od -1;
   č) da zavzame vrednost, manjšo od 0˙5;
   d) da zavzame vrednost, manjšo od -0˙4;
   e) da zavzame vrednost z intervala $[0˙7, 1˙5]$;
   f) da zavzame vrednost z intervala $[-1, 1˙5]$.

2. Za slučajno spremenljivko $X \sim N(5, 2)$ izračunaj

   a) $P(X > 7)$ b) $P(X < 8)$ c) $P(X > 1)$
   č) $P(1 \leq X \leq 7)$ d) $P(0 < X < 5)$ e) $P(3 < X \leq 8)$

3. Življenjska doba žarnice je slučajna spremenljivka, porazdeljena po zakonu $N(1\,000, 50)$. Izračunaj verjetnost dogodka, da ta žarnica zdrži vsaj 1 050 časovnih enot.

4. Predpostavimo, da je debelina posameznih plošč, ki jih dobimo pri valjanju pločevine, normalno porazdeljena slučajna spremenljivka z $a = 4˙00$ mm in $\sigma = 0˙05$ mm. Kupčeva vhodna kontrola sprejme pločevino, če je njena debelina v mejah od 3˙95 mm do 4˙10 mm. Kolikšen odstotek plošč kupec v povprečju zavrne?

5. Podobno kot v zadnjem zgledu oceni verjetnost dogodka, da pri 10 000 metih kovanca več kot 5 100 - krat pade grb.

6. Za neko rastlino je značilno, da povprečno 10% semen ne vzkali. Kolikšna je verjetnost, da bo med 1 000 semeni te rastline kvečjemu 80 nekalivih?

7. Konkretno izračunaj verjetnost iz zgleda 4.8.2 in presodi, ali je bil sklep o zadostni natančnosti našega približka upravičen.

4.9 Zakon velikih števil

Pri razlagi rezultatov različnih nalog iz verjetnostnega računa, ki smo jih do zdaj rešili, smo sicer ves čas uporabljali pristavek "pri dovolj velikem številu poskusov", nismo pa uspeli preseči tega ohlapnega opisa. Tak položaj smo imeli že pri sami statistični definiciji verjetnosti, kjer smo lahko opazili, da se *pri dovolj dolgem* zaporedju poskusov relativna frekvenca dogodka *navadno* ustali pri *nekem* številu, ki smo ga nato proglasili za verjetnost dogodka. Čas je, da ta odnos med relativno frekvenco dogodka in njegovo verjetnostjo nekoliko podrobneje raziščemo.

Frekvenca dogodka je slučajna spremenljivka (označimo jo kar z X), za katero vemo, da je porazdeljena binomsko, s parametroma n in p, ki pomenita število poskusov in verjetnost za nastop dogodka v posamezni ponovitvi poskusa.

Če frekvenco dogodka delimo s številom poskusov, dobimo relativno frekvenco dogodka; za tako dobljeno slučajno spremenljivko $Y = (1/n)X$ lahko hitro izračunamo matematično upanje in standardni odklon. Kot vemo, je

$$X \sim b(n,p) \Rightarrow E(X) = np, \quad D(X) = np(1-p) \tag{1}$$

Kaj se zgodi z matematičnim upanjem in disperzijo, če slučajno spremenljivko pomnožimo z nekim številom? Če slučajna spremenljivka X z neko verjetnostjo zavzame vrednost x, slučajna spremenljivka αX z isto verjetnostjo zavzame vrednost αx, zato se tudi matematično upanje za isti faktor razlikuje od prvotnega:

$$E(X) = C \Rightarrow E(\alpha X) = \alpha C \tag{2}$$

Disperzija slučajne spremenljivke je matematično upanje kvadrata odklonov njenih individualnih vrednosti od matematičnega upanja. Ker se tako posamezne vrednosti kot matematično upanje pri množenju slučajne spremenljivke z α pomnožijo s tem istim faktorjem, se z njim pomnožijo tudi vsi odkloni, kar pa zaradi kvadriranja odklonov pomeni, da se disperzija pomnoži s faktorjem α^2:

$$D(X) = \sigma^2 \Rightarrow D(\alpha X) = \alpha^2 \sigma^2 \tag{3}$$

Za standardni odklon potemtakem velja, da se pri množenju slučajne spremenljivke z α pomnoži z absolutno vrednostjo tega števila. (Bralec naj sam premisli, zakaj absolutna vrednost!)

Iz obrazcev (1), (2) in (3) takoj dobimo vrednosti najpomembnejših parametrov za relativno frekvenco dogodka:

$$E(Y) = E(\frac{1}{n}X) = \frac{1}{n}E(X) = \frac{1}{n}np = p \tag{4}$$

Tu si lahko prvič oddahnemo: pričakovana vrednost relativne frekvence je verjetnost dogodka v posamezni ponovitvi poskusa; če bi bilo drugače, bi se nam podrl dobršen del dosedanje zgradbe...

Podobno izračunamo

$$D(Y) = D(\frac{1}{n} X) = (\frac{1}{n})^2 D(X) = (\frac{1}{n})^2 n p (1-p) = \frac{p(1-p)}{n} \quad (5)$$

Bistre oči tu opazijo novo lepoto: z naraščanjem števila poskusov se razpršenost relativne frekvence zmanjšuje. Boljšo povečavo pa nam pomaga doseči neenačba Čebiševa: ker je (obrazec (19.a) v razdelku 4.7!)

$$P(|X - E(X)| < t) \geq 1 - \frac{D(X)}{t^2}$$

je za relativno frekvenco $Y = X/n$ potem pri še tako majhnem $t > 0$

$$P\left(\left|\frac{X}{n} - p\right| < t\right) \geq 1 - \frac{p(1-p)}{n t^2} \quad (6)$$

Z naraščnjem števila poskusov n preko vsake meje postaja "popravek" na desni strani vedno manjši in verjetnost, zapisana na levi strani (6), se približuje 1. S tem pa smo dokazali t.i. **Bernoullijev zakon velikih števil**:

Naj ima slučajni dogodek A v posamezni ponovitvi poskusa verjetnost p, naj bo X njegova frekvenca v n ponovitvah poskusa in t poljubno pozitivno število. Potem s povečevanjem števila poskusov n preko vsake meje dosežemo, da se verjetnost dogodka

$$\left(\left|\frac{X}{n} - p\right| < t\right)$$

poljubno malo loči od 1.

Po domače povedano, *pri velikem številu poskusov je zelo malo verjetno, da bi se relativna frekvenca dogodka pomembneje razlikovala od njegove verjetnosti.* To nam še vedno ne zagotavlja, da se bistveni odkloni ne morejo pripetiti, pove pa, da so zelo redki: če bomo relativno frekvenco dogodka, izmerjeno v velikem številu poskusov, proglasili za njegovo verjetnost, bomo skoraj zagotovo dobro zadeli.

ZGLED 4.9.1:

Pri meritvah običajno ne dobimo povsem natančne vrednosti merjene količine (označimo jo z a), ampak se izmerki $\{x_1, x_2, ..., x_n\}$ od nje bolj ali manj razlikujejo zaradi različnih slučajnih vplivov. V izmerjenih vrednostih lahko vidimo vrednosti, ki jih je zavzela n - terica slučajnih spremenljivk X_k, $k = 1, 2, ..., n$. Če merska naprava ni pokvarjena (in če sami ne delamo sistematičnih napak),

lahko predpostavljamo, da imajo vse slučajne spremenljivke X_k isto matematično upanje $E(X_k) = a$, da torej v povprečju "zadenejo" pravo vrednost merjene količine. Ker je (razširitev (2)!)

$$E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \frac{1}{n}(a + a + ... + a) = a$$

ravnamo pametno, če za oceno prave vrednosti merjene količine vzamemo aritmetično sredino izmerkov x_1, x_2, ..., x_n. Po dosedanjih izkušnjah tudi vidimo, da je pri velikem številu merjenj napaka, ki jo pri tem zagrešimo, skoraj zagotovo zelo majhna.

Bernoullijev zakon je eden najpomembnejših predstavnikov izrekov, ki pojasnjujejo zakonitosti množičnih pojavov. Poleg natančnejših izrekov te vrste (to je o odnosu med relativno frekvenco dogodka in njegovo verjetnostjo) pa imamo še drugo skupino, katere predstavniki zaslužijo skupno ime **centralni limitni izrek**. Premalo znanja in prostora imamo, da bi se ga lotili na spodobni ravni, zato samo recimo, da gre za podobne izreke kot je bil Laplace - de Moivreov izrek o aproksimaciji binomske porazdelitve z normalno (glej str. 160!). Najbolj po domače je vsebina centralnega limitnega izreka naslednja: če lahko neki pojav opišemo kot posledico sočasnega delovanja velikega števila med seboj neodvisnih slučajnih dejavnikov, lahko pri določenih dodatnih pogojih pričakujemo, da bodo vrednosti tega pojava normalno porazdeljene. To je tudi neke vrste odgovor na vprašanje, zakaj je normalna porazdelitev v praksi tako pogosta.

V A J E

1. Dolžino smo merili stokrat. Varianca izmerkov ni večja od 2 cm^2. Kolikšna je verjetnost, da aritmetična sredina izmerkov ne odstopa od prave dolžine za več kot 0'3 cm?

2. Verjetnost dogodka A v posamezni ponovitvi poskusa znaša 5/8. Kolikšna je verjetnost, da se pri 6 000 realizacijah tega poskusa relativna frekvenca ne razlikuje od verjetnosti $P(A)$ za več kot 0'01?

3. $\triangle$ Dogodek A ima v posamezni ponovitvi poskusa verjetnost 0'71. Doseči želimo, da se z verjetnostjo vsaj 0'96 relativna frekvenca dogodka A razlikuje od njegove verjetnosti za manj kot 0'2. Koliko ponovitev poskusa moramo opraviti?

PREGLED POMEMBNEJŠIH OBRAZCEV

Relativna frekvenca dogodka $f^{o}(A) = \dfrac{f(A)}{n}$

Klasična definicija verjetnosti $P(A) = \dfrac{m}{n}$

Lastnosti verjetnosti - aksiomi Kolmogorova

NENEGATIVNOST	$(\forall A)(A \in \mathcal{P}G \Longrightarrow P(A) \geq 0)$
NORMIRANOST	$P(G) = 1$
ADITIVNOST	$A \cap B = N \Longrightarrow P(A \cup B) = P(A) + P(B)$

Posledice

$$A \subset B \Longrightarrow P(A) \leq P(B)$$

$$P(A) + P(A') = 1, \qquad P(A) = 1 - P(A')$$

$$P(A) \leq 1, \qquad P(N) = 0$$

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

Pogojna verjetnost, verjetnost produkta, neodvisnost

$$P(A/B) = \frac{P(AB)}{P(B)}, \quad P(AB) = P(B)\,P(A/B) = P(A)\,P(B/A)$$

$$A \text{ in } B \text{ neodvisna} \iff P(A) = P(A/B), \quad P(B) = P(B/A)$$

$$A \text{ in } B \text{ neodvisna} \iff P(AB) = P(A)\,P(B)$$

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)\,P(A_2/A_1) ... P(A_n/(A_1 \cap ... \cap A_{n-1}))$$

$$\text{V celoti neodvisni}: P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)\,P(A_2) ... P(A_n)$$

Obrazec za popolno verjetnost

$$P(A) = P(H_1)\,P(A/H_1) + P(H_2)\,P(A/H_2) + ... + P(H_n)\,P(A/H_n)$$

Bayesov obrazec

$$P(H_i/A) = \frac{P(H_i)\,P(A/H_i)}{P(A)} \qquad (i = 1, 2, ..., n)$$

Bernoullijev obrazec

$$P(n; p; k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad (k = 0, 1, 2, ..., n)$$

Najverjetnejša frekvenca v Bernoullijevem zaporedju

$$np - q \text{ ni celo število} \Longrightarrow k_o = [np - q] + 1$$

$$np - q \text{ je celo število} \Longrightarrow k_o = np - q \text{ in } k_o = np - q + 1$$

Pascalov obrazec

$$P(\text{frekvenca } k \text{ v } n - \text{tem poskusu}) = \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

Posplošeni Bernoullijev obrazec

$$P(n; p_1, p_2, ..., p_m; k_1, k_2, ..., k_m) = \frac{n!}{k_1! \cdot k_2! \cdot ... \cdot k_m!}\, p_1^{k_1} \cdot p_2^{k_2} \cdot ... \cdot p_m^{k_m}$$

Matematično upanje (končna zaloga vrednosti)

$$E(X) = \sum_{k=1}^{n} p_k\, x_k$$

Disperzija in standardni odklon

$$D(X) = E\{[X - E(X)]^2\}$$

$$D(X) = \sum_{i=1}^{n} p_i\,[x_i - E(X)]^2 = \{\sum_{i=1}^{n} p_i\, x_i^2\} - [E(X)]^2$$

$$\sigma = \sqrt{D(X)}$$

Binomska porazdelitev

$$X \sim b(n, p) \Longrightarrow E(X) = np, \quad \sigma(X) = \sqrt{npq}$$

Neenačba Čebiševa

$$P(|X - E(X)| < t) \geq 1 - \frac{D(X)}{t^2}$$

$$P(|X - E(X)| \geq t) \leq \frac{D(X)}{t^2}$$

Normalna porazdelitev

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Standardizacija

$$X \sim N(a,\sigma) \implies Z = \frac{X-a}{\sigma} \sim N(0,1)$$

Standardizirana normalna porazdelitev

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \varphi(x)$$

$$\Phi(z) = \text{ploščina pod } \varphi(x) \text{ od } x=0 \text{ do } x=z,\quad \Phi(-z) = -\Phi(z)$$
$$X \sim N(0,1) \implies P(\alpha \le Z \le \beta) = \Phi(\beta) - \Phi(\alpha)$$

Poljubna normalna porazdelitev

$$X \sim N(a,\sigma) \implies P(\alpha \le X \le \beta) = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right)$$

Verjetnosti za značilne intervale

$$X \sim N(a,\sigma) \implies \begin{aligned} P(a-\sigma \le X \le a+\sigma) &= 0{\cdot}6826 \\ P(a-2\sigma \le X \le a+2\sigma) &= 0{\cdot}9544 \\ P(a-3\sigma \le X \le a+3\sigma) &= 0{\cdot}9974 \end{aligned}$$

Bernoullijev zakon velikih števil

$$P\left(\left|\frac{X}{n} - p\right| < t\right) \ge 1 - \frac{p(1-p)}{n\,t^2}$$

X : frekvenca dogodka, p : njegova verjetnost, n : število poskusov

5. STATISTIČNO SKLEPANJE

Zadnje poglavje tega učbenika na določen način zaokroži dve predhodni v celoto in pokaže (pa čeprav samo na najpreprostejših primerih), kaj lahko statistik, ki obvlada verjetnostni račun, iz podatkov, dobljenih na primerno izbranem vzorcu, pove o populaciji kot celoti. Še enkrat začutimo, da lahko isti rezultat sočasno pripoveduje o blišču in o bedi statističnega sklepanja, odvisno od tega, s katere plati ga gledamo...

5.1 Vzorčenje

V poglavju o osnovah opisne statistike smo definirali najpomembnejše statistične pojme in povedali, da v praksi običajno proučujemo množične pojave tako, da izmerimo vrednosti ustreznih statističnih spremenljivk na primerno izbrani končni podmnožici celotne populacije, to podmnožico imenujemo **vzorec**, nato pa iz značilnosti, izračunanih iz vzorčnih podatkov, sklepamo na značilnosti populacije kot celote.

Omenili smo že, da je sklepanje najlažje, če je bil izbor elementov v vzorec opravljen povsem na slepo, če je torej šlo za **slučajni vzorec**. Izbor lahko natančneje opišemo z zahtevo, da morajo imeti pri vsakem odbiranju elementa v vzorec vsi elementi populacije enake možnosti, da bodo izbrani. V bistvu gre torej za izbiranje elementa z vračanjem, tako da je posamezen element lahko tudi večkrat izbran v vzorec, oziroma imamo **vzorce s ponavljanjem**. Pri dani velikosti populacije (N) in dani moči vzorca (n) najdemo toliko različnih vzorcev z vračanjem, kolikor je vseh kombinacij s ponavljanjem, to je $\binom{N+n-1}{n}$.

V praksi bi nas zahteva po vračanju elementov dostikrat ovirala, ker se v nekaterih primerih preizkušani element pač uniči (na primer žarnica pri preverjanju življenjske dobe). Na srečo se izkaže, da imamo lahko za slučajne tudi vzorce, ki jih oblikujemo brez vračanja elementov (vsak element iz populacije lahko v posameznem vzorcu nastopi kvečjemu enkrat), če je število elementov v vzorcu (n) majhno v primerjavi z močjo populacije (N); razlika med obema načinoma vzorčenja se zmanjšuje z večanjem osnovne populacije. To smo za poseben primer ugotovili že v pri primerjavi binomske in hipergeometrijske porazdelitve.

Vzorcev, ki nastanejo brez vračanja elementov - to je **vzorcev brez ponavljanja** - je toliko, kolikor je vseh kombinacij reda n iz N elementov (brez ponavljanja), torej $\binom{N}{n}$. Število je manjše kot pri prejšnjem načinu, a še vedno ogromno v vseh praktično zanimivih primerih.

Vzemimo, da imamo neko populacijo z N elementi, in naredimo vse možne vzorce po n elementov (s ponavljanjem ali brez njega). S tem smo dobili novo populacijo, ki ji bomo rekli **populacija vseh vzorcev**. Pri slučajnem vzorčenju, na katero se bomo v učbeniku omejili, dobimo najpomembnejše izreke o parametrih proučevane osnovne populacije iz zvez med osnovno populacijo in populacijo vseh možnih vzorcev, ki tej ustreza. Izreke bomo zapisali brez dokazov, ki so za nas ali pretežki ali predolgi ali oboje.

Imejmo torej neko populacijo z N elementi in statistično spremenljivko Y, ki jo želimo podrobneje proučiti, ker pač po našem mnenju dobro opisuje pojav, ki nas zanima. Izberimo si neki vzorec velikosti n in ugotovimo vrednost spremenljivke Y na vsakem od izbranih elementov. Zaradi slučajnega izbora elementov lahko v tako dobljenih številih $\{y_1, y_2, ..., y_n\}$ vidimo realizacije slučajnih spremenljivk $\{Y_1, Y_2, ..., Y_n\}$, ki so vse enako porazdeljene kot osnovna spremenljivka Y. Če se do preklica omejimo na vzorce s ponavljanjem, lahko rečemo tudi to, da so dogodki v množici $\{ (Y_i = y_i), i = 1, 2, ..., n \}$ v celoti neodvisni.

Izračunajmo aritmetično sredino vrednosti, ki jih proučevana spremenljivka zavzame na tem vzorcu, ali krajše in ohlapno rečeno, aritmetično sredino vzorca ali **vzorčno povprečje** $\bar{y}$. Kako računamo aritmetično sredino iz vrednosti $\{y_1, y_2, ..., y_n\}$, vemo iz razdelka 3.3; če se spomnimo še lastnosti matematičnega upanja slučajne spremenljivke, pa lahko rečemo, da je vzorčno povprečje število $\bar{y}$ realizacija (ena od $\binom{N+n-1}{n}$ ali $\binom{N}{n}$ možnih realizacij, kolikor je pač vzorcev) slučajne spremenljivke

$$\overline{Y} = \frac{1}{n} \sum_{k=1}^{n} Y_k \tag{1}$$

ki ji bomo rekli **vzorčno povprečje**.

Vzorčno povprečje ima nekaj lepih lastnosti:

1. *Če je na proučevani populaciji Y porazdeljena normalno, je tako porazdeljeno tudi vzorčno povprečje $\overline{Y}$.*

2. *Vzorčno povprečje $\overline{Y}$ se pri velikih vzorcih porazdeljuje približno normalno tudi v primeru, ko je spremenljivka Y na osnovni populaciji porazdeljena kako drugače.*

3. *Matematični upanji slučajnih spremenljivk Y in $\overline{Y}$ sta enaki.*

Prva lastnost ne potrebuje posebnega komentarja, kasneje bomo še povedali, kako je z razpršenostjo $\overline{Y}$; zaradi lastnosti 3 ima namreč Y enako matematično upanje kot osnovna spremenljivka Y. Zanimivejša je lastnost 2, ki je posledica v uvodu omenjenega centralnega limitnega izreka in jo je treba razumeti takole:

Naj bo Y porazdeljena kakorkoli, če povečujemo velikost slučajnega vzorca (n), postaja verjetnostna porazdelitev slučajne spremenljivke $\overline{Y}$ vedno bolj podobna normalni porazdelitvi (z istim matematičnim upanjem, kot

ga ima Y!). To pove, zakaj so tako pomembni t.i. **veliki vzorci**: pri njih se ni treba posebej zanimati za porazdelitev osnovne spremenljivke, ampak shajamo pri vseh nadaljnjih računih z normalno porazdelitvijo vzorčnega povprečja. V praksi je dostikrat že vzorec s 30 elementi "velik", povsem brez skrbi pa zaupamo lastnosti 2 pri $n \geq 100$. Pri malih vzorcih si tega ne smemo privoščiti in je treba za vsak primer posebej ugotavljati, kako se porazdeljuje $\overline{Y}$.

OPOZORILO: V tem učbeniku se bomo držali predpostavke, da je vzorec velik oziroma da je že osnovna spremenljivka porazdeljena normalno, tako da bo porazdelitev vzorčnega povprečja vedno normalna. Treba je povedati, da lastnosti 1-3 veljajo tako za vzorce z vračanjem kot za vzorce brez vračanja elementov.

Namesto v jeziku verjetnostnega računa lahko lastnost 3 opišemo še bolj konkretno. Imejmo populacijo z N elementi in statistično spremenljivko Y, ki na njej zavzame vrednosti $y_1, y_2, \ldots, y_N$. Naredimo vse možne vzorce po n elementov in za vsakega od njih izračunajmo ustrezno aritmetično sredino $\bar{y}$ (torej vrednost, ki jo je zavzelo vzorčno povprečje $\overline{Y}$). Aritmetična sredina teh aritmetičnih sredin je potem enaka aritmetični sredini spremenljivke Y na osnovni populaciji.

V praksi seveda ne naredimo vseh vzorcev, da bi proučevali dano spremenljivko Y, ker bi bilo potem enostavneje pregledati vrednosti te spremenljivke na celotni osnovni populaciji. Zato je dobljeno vzorčno povprečje (aritmetična sredina) $\bar{y}$ samo ena od vrednosti, ki jih lahko zavzame slučajna spremenljivka $\overline{Y}$. Vrednosti $\bar{y}$ se na splošno za posamezne vzorce odklanjajo od svoje aritmetične sredine $\mu_{\bar{y}} = E(\overline{Y})$ v levo ali desno, ker je pač $\overline{Y}$ kot slučajna spremenljivka bolj ali manj razpršena okrog svojega matematičnega upanja. Pri tem pa iz verjetnostnega računa vemo, da so veliki odkloni slučajne spremenljivke od njenega matematičnega upanja malo verjetni (neenačba Čebiševa!), za normalno porazdeljeno spremenljivko $\overline{Y}$ znamo verjetnosti posameznih odklonov tudi čisto natančno izračunati.

Zato bomo aritmetično sredino izbranega slučajnega vzorca (vrednost $\bar{y}$) uporabljali kot oceno za aritmetično sredino slučajne spremenljivke Y na osnovni populaciji, to je μ_Y. Iz tega razloga imenujemo vzorčno povprečje $\overline{Y}$ ***cenilka*** *aritmetične sredine μ_Y.*

Da bi lahko sklepali o zanesljivosti takšne ocene, moramo poznati razpršenost individualnih vrednosti, ki jih $\overline{Y}$ zavzame na posameznih vzorcih, okrog $\mu_{\overline{Y}} = E(\overline{Y})$. Kot običajno bomo razpršenost opisali z disperzijo oziroma standardnim odklonom slučajne spremenljivke $\overline{Y}$. Videli bomo, da je za razliko od lastnosti 1-3 tu pomembno, kako izbiramo vzorce.

4. Naj bo v populaciji N elementov in naj bo disperzija spremenljivke Y na tej populaciji σ_Y^2 (ali na kratko σ^2, če ni nevarnosti za pomoto). Potem je varianca vzorčnega povprečja $\overline{Y}$ pri vzorcih s ponavljanjem n-krat manjša *od disperzije spremenljivke Y na osnovni populaciji,*

$$\sigma^2_{\overline{Y}} = \frac{\sigma^2}{n} \tag{2.a}$$

Za vzorce brez ponavljanja pa ob nespremenjenih ostalih pogojih velja

$$\sigma^2_{\overline{Y}} = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} \tag{2.b}$$

Oznaka $\sigma^2_{\overline{Y}}$ poudarja, da gre v bistvu za varianco aritmetičnih sredin vseh možnih vzorcev z močjo n, izbranih iz dane populacije z N elementi. Iz primerjave obrazcev (2.a) in (2.b) lahko ugotovimo, da je ta varianca v drugem primeru sicer manjša , vendar postaja razlika zanemarljiva, če velikost populacije narašča preko vsake meje. Ker so lastnosti 1-3 neodvisne od tega, ali izbiramo vzorce z vračanjem ali brez njega, tako znova vidimo, da je pri velikih populacijah razlika med obema vrstama vzorcev zanemarljiva.

Iz obrazcev (2.a) oziroma (2.b) dobimo še ustrezna standardna odklona aritmetičnih sredin (oziroma slučajne spremenljivke $\overline{Y}$), namreč

$$\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}} \tag{3.a}$$

oziroma

$$\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \tag{3.b}$$

Ker je v praksi N vedno velik, v obrazcu (3.b) dostikrat zanemarimo odšteto enojko v imenovalcu; dobljeni izraz pod korenom pa ocenimo po znanem pravilu, da se za števila blizu 1 kvadratni koren razlikuje od 1 približno za polovico presežka, ki ga izkazuje število samo,

$$\sqrt{1+x} \approx 1 + \frac{x}{2}$$

kar nam da naslednjo kačo resnic:

$$\sqrt{\frac{N-n}{N-1}} \approx \sqrt{\frac{N-n}{N}} = \sqrt{1-\frac{n}{N}} \approx 1 - \frac{1}{2}\cdot\frac{n}{N} \tag{4}$$

Odtod vidimo, da je že pri vzorcih, ki so petdesetkrat manjši od populacije, napaka zaradi zanemarjenega "popravnega faktorja" za vzorce brez vračanja velikostnega reda 1%.

Zdaj lahko konkretneje predstavimo porazdelitev vzorčnega povprečja $\overline{Y}$: ta slučajna spremenljivka je porazdeljena približno po zakonu $N(a, \frac{\sigma}{\sqrt{n}})$ (za vzorce s ponavljanjem) oziroma $N(a, \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}})$ (za vzorce brez ponavljanja), pri čemer sta a in σ matematično upanje oziroma standardni odklon osnovne spremenljivke Y.

Vzorčno povprečje $\overline{Y}$ je torej primerna cenilka aritmetične sredine spremenljivke Y. Ker je $E(\overline{Y}) = \mu_y$, rečemo, da je **nepristranska**. Njene vrednosti so bolj zgoščene okrog skupne vrednosti $E(Y) = \mu_{\bar{y}} = \mu_y$ kot pa vrednosti osnovne spremenljivke Y; to lepo vidimo iz zapisov (3.a) oziroma (3.b). Zgoščenost je tem večja, čim večja je velikost vzorca n, saj se z večanjem vzorca zmanjšuje standardni odklon $\sigma_{\bar{y}}$.

Zaradi vloge, ki jo ima standardni odklon $\sigma_{\bar{y}}$, ga imenujemo **standardna napaka** ocene; mišljena je seveda ocena aritmetične sredine μ_y. Z večanjem vzorca lahko pri nespremenjenih ostalih pogojih naredimo to napako majhno, kot le želimo, seveda pa se pri tem povečajo tudi stroški vzorčenja in obdelave zbranih podatkov.

Na konkretnem zgledu bomo ilustrirali nekaj ugotovitev, ki smo jih zapisali v zvezi z normalno porazdelitvijo statističnih spremenljivk na populaciji vseh možnih vzorcev.

ZGLED 5.1.1:

Statistična spremenljivka Y je na populaciji z $N = 5$ elementi zavzela vrednosti 7, 8, 11, 13 in 16. Predstavi populacijo vseh možnih vzorcev velikosti $n = 2$ za vzorce z ponavljanjem in vzorce brez ponavljanja. Izračunaj aritmetično sredino in standardni odklon spremenljivke Y na osnovni populaciji, nato pa še aritmetično sredino porazdelitve vzorčnih aritmetičnih sredin in standardno napako ocene za vzorce z vračanjem oziroma brez vračanja. Ugotovi, ali se rezultati ujemajo s pričakovanimi vrednostmi!

Najprej izračunajmo parametra μ_y in σ za osnovno populacijo:

$$\mu_y = (7+8+11+13+16)/5 = 11{\cdot}0$$

$$\sigma^2 = (7^2+8^2+11^2+13^2+16^2)/5 - 11{\cdot}0^2 = 10{\cdot}8 \Rightarrow \sigma \approx 3{\cdot}29$$

Ker ima populacija $N = 5$ elementov in je velikost vzorca $n = 2$, imamo $\binom{N+n-1}{n} = \binom{6}{2} = 15$ različnih vzorcev z vračanjem. Pri računanju pa moramo upoštevati, da na primer vzorca (7;7) in (7;8) nista enakovredna v tem smislu, da se dve "sedmici" lahko pojavita samo enkrat, "sedmica in osmica" pa na dva načina, najprej 7 in potem 8 ali obratno. Zato je najvarneje, če vsaki od petih možnih vrednosti pridružimo spet katerokoli od teh vrednosti in vse možne vzorce s vračanjem predstavimo v naslednji preglednici:

(7; 7)	(7; 8)	(7;11)	(7;13)	(7;16)
(8; 7)	(8; 8)	(8;11)	(8;13)	(8;16)
(11; 7)	(11; 8)	(11;11)	(11;13)	(11;16)
(13; 7)	(13; 8)	(13;11)	(13;13)	(13;16)
(16; 7)	(16; 8)	(16;11)	(16;13)	(16;16)

Aritmetične sredine teh vzorcev so

7·0	7·5	9·0	10·0	11·5
7·5	8·0	9·5	10·5	12·0
9·0	9·5	11·0	12·0	13·5
10·0	10·5	12·0	13·0	14·5
11·5	12·0	13·5	14·5	16·0

Iz teh podatkov dobimo za njihovo aritmetično sredino vrednost

$$\mu_{\bar{y}} = (7{\cdot}0 + 7{\cdot}5 + \ldots + 14{\cdot}5 + 16{\cdot}0)/25 \;=\; 11{\cdot}0$$

ki je res enaka aritmetični sredini na osnovni populaciji, to je μ_y.

Varianca porazdelitve aritmetičnih sredin je enaka

$$\sigma_{\bar{y}}^2 = (7{\cdot}0^2 + 7{\cdot}5^2 + \ldots + 16{\cdot}0^2)/25 - 11{\cdot}0^2 = 5{\cdot}4$$

Ta varianca je natanko dvakrat manjša od variance spremenljivke Y na osnovni populaciji, kar se zaradi $n = 2$ ujema z obrazcem (3.a),

$$\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n} = \frac{10{\cdot}8}{2} = 5{\cdot}4$$

Za vzorce z vračanjem oziroma ponavljanjem elementov je potemtakem standardna napaka ocene $\sigma_{\bar{y}} = \sqrt{5{\cdot}4} \approx 2{\cdot}32$.

Ponovimo postopek še za vzorce brez ponavljanja. Ustrezni pari vrednosti spremenljivk so enaki kot v prvotni tabeli parov vrednosti, seveda brez njene "diagonale", ker sedaj pari (7;7), (8;8), ..., (16;16) pač niso možni. Iz parov vrednosti izračunamo vzorčne aritmetične sredine in aritmetično sredino njihove porazdelitve:

$$\mu_{\bar{y}} = (7{\cdot}5 + 9{\cdot}0 + \ldots + 14{\cdot}5)/10 = 11{\cdot}0$$

Nato izračunamo še varianco aritmetičnih sredin,

$$\sigma_{\bar{y}}^2 = (7{\cdot}5^2 + 9{\cdot}0^2 + \ldots + 14{\cdot}5^2)/10 - 11{\cdot}0^2 = 4{\cdot}05$$

Ker je pri naših pogojih

$$\frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} = \frac{10{\cdot}8}{2} \cdot \frac{5-2}{5-1} = 4{\cdot}05 = \sigma_{\bar{y}}^2$$

lahko rečemo, da se rezultat ujema s pričakovanji. Standardna napaka ocene za vzorce brez vračanja je v našem primeru $\sigma_{\bar{y}} = \sqrt{4{\cdot}05} \approx 2{\cdot}01$ in je precej manjša kot za vzorce s ponavljanjem.

Pripomnimo, da smo vse račune za vzorce brez vračanja opravili tako, da smo že vnaprej izpostavili skupni faktor 2, ker se vsak vzorec pojavlja v shemi natanko dvakrat.

Videli smo, da je $\bar{Y}$ nepristranska cenilka aritmetične sredine vrednosti spremenljivke Y na dani populaciji ali, kot rečemo krajše, aritmetične sredine dane populacije. Po analogiji bi morda pričakovali, da je potem vzorčna varianca

$$\frac{1}{n}\sum_{k=1}^{n}(Y_k - \overline{Y})^2 \tag{5}$$

tudi nepristranska cenilka variance σ^2 za osnovno populacijo. Vendar se izkaže, da je matematično upanje cenilke (5) enako $[(n-1)/n]\,\sigma^2$ in zato ni nepristranska cenilka za varianco σ^2. Iz lastnosti matematičnega upanja pa na srečo sledi, da jo lahko hitro popravimo tako, da bo nepristranska in bo njeno matematično upanje natanko enako σ^2, če namesto (5) vzamemo za cenilko

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(Y_k - \overline{Y})^2 \tag{6}$$

Trditev, da je $E(S^2) = \sigma^2$, velja za poljubne vzorce iz neskončne populacije ali za vzorce s ponavljanjem v primeru končne populacije. Če hočemo ohraniti nepristranost cenilke (6) tudi pri vzorcih, ki jih brez ponavljanja oblikujemo iz končne populacije, jo moramo pomnožiti še s "popravnim faktorjem" $(N-1)/N$. Ker je število elementov v populaciji navadno zelo veliko, lahko za našo natančnost računanja rečemo, da je ta faktor enak 1 in ga zanemarimo. Pri šolskih primerih, kakršen je na primer v zgledu 5.1.2, pa je razlika lahko znatna.

Pri računanju vrednosti cenilke S^2 je morda najbolje, če najprej za izbrani vzorec izračunamo varianco in jo nato pomnožimo s faktorjem $n/(n-1)$. Odtod pa dobimo še naslednji obrazec:

$$s^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{k=1}^{n} y_k^2 - (\bar{y})^2\right) = \frac{1}{n-1}\sum_{k=1}^{n} y_k^2 - \frac{n}{n-1}\left(\frac{1}{n}\sum_{k=1}^{n} y_k\right)^2$$

oziroma

$$s^2 = \frac{1}{n-1}\left(\sum_{k=1}^{n} y_k^2 - \frac{1}{n}\left(\sum_{k=1}^{n} y_k\right)^2\right) =$$

$$= \frac{1}{n(n-1)}\left(n\cdot\sum_{k=1}^{n} y_k^2 - (\sum_{k=1}^{n} y_k)^2\right) \tag{7}$$

Ta obrazec je najprimernejši za računanje vrednosti cenilke S^2 na danem vzorcu $y_1,\ y_2,\ \ldots,\ y_n$, če $\bar{y}$ še ne poznamo.

ZGLED 5.1.2:

Za podatke iz zgleda 5.1.1 preveri, da je $E(S^2) = \sigma^2$!

Pri reševanju naloge bomo upoštevali, kaj konkretno pomeni matematično upanje slučajne spremenljivke S^2. Ker so vsi vzorci enako verjetni, je matematično upanje slučajne spremenljivke S^2 enako aritmetični sredini vseh vrednosti, ki jih zavzame ta spremenljivka na posameznih vzorcih. Enak rezultat bomo dobili tako, da bomo izračunali aritmetično sredino vseh vzorčnih varianc in rezultat pomnožili z $n/(n-1)$.

V zgledu 5.1.1 smo že izračunali vse možne vzorce s ponavljanjem in njihove aritmetične sredine. Iz teh podatkov izračunamo vzorčne variance

0'00	0'25	4'00	9'00	20'25
0'25	0'00	2'25	6'25	16'00
4'00	2'25	0'00	1'00	6'25
9'00	6'25	1'00	0'00	2'25
20'25	16'00	6'25	2'25	0'00

Pri tem je na primer vrednost 20'25 iz prvega stolpca zadnje vrstice varianca za dvojico (16; 7), torej

$$\frac{(16^2+7^2)}{2} - \left(\frac{16+7}{2}\right)^2 = 20'25$$

Aritmetična sredina podatkov iz pravkar zapisane tabele je

$$\frac{(0'00 + 0'25 + 0'40 + \ldots + 6'25 + 2'25 + 0,00)}{25} = 5'40$$

Ker smo v vzorce izbirali po dva elementa (z dovoljenim ponavljanjem), je $n/(n-1) = 2/(2-1) = 2$ in tako končno

$$E(S^2) = 2 \cdot 5'40 = 10'80$$

Rezultat se ujema z varianco spremenljivke Y na osnovni populaciji, ki smo jo izračunali v zgledu 5.1.1; s tem smo trditev, da je $E(S^2) = \sigma^2$, preverili za vzorce s ponavljanjem (ali bolje, za ta konkretni zgled). Enako ravnamo pri vzorcih brez ponavljanja, kjer pa moramo rezultat še pomnožiti z $(N-1)/N = 4/5 = 0'8$. Ta del računa naj ostane za domačo vajo.

Z vrednostjo , ki jo cenilka S^2 zavzame na izbranem slučajnem vzorcu, bomo torej ocenjevali varianco σ^2 proučevane statistične spremenljivke, s pozitivnim kvadratnim korenom iz dobljene vrednosti s^2 pa standardni odklon te spremenljivke, čeprav taka ocena ni povsem nepristranska.

ZGLED 5.1.3:

Iz velike pošiljke cevi smo na slepo izbrali pet cevi in izmerili naslednje premere: 63'3 mm, 63'7 mm, 63'6 mm, 63'2 mm, 63'7 mm. Na osnovi tega vzorca oceni aritmetično sredino in varianco premerov za celotno pošiljko!

Aritmetično sredino ocenimo z realizacijo vzorčnega povprečja na izbranem vzorcu,

$$\bar{y} = \frac{63'3 + 63'7 + 63'6 + 63'2 + 63'7}{5} = 63'5\,(\mathrm{mm})$$

Varianco pa ocenimo z vrednostjo cenilke S^2 na tem vzorcu, pri čemer bomo uporabili obrazec, ki nas je pripeljal do (7),

$$s^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{k=1}^{n} y_k^2 - (\bar{y})^2\right)$$

ker je v sedanjem položaju, ko imamo aritmetično sredino $\bar{y}$ že izračunano, ugodnejši od končne oblike (7). Tako je

$$s^2 = \frac{5}{5-1}\left(\frac{1}{5}(63'3^2 + 63'7^2 + 63'6^2 + 63'2^2 + 63'7^2) - 63'5^2\right)$$

kar nam da rezultat

$$s^2 = 0'055\,(\mathrm{mm}^2) \Rightarrow \mathrm{s} = \sqrt{\mathrm{s}^2} = \sqrt{0'055} \approx 0'23\,(\mathrm{mm})$$

Podrobneje se z verjetnostno porazdelitvijo slučajne spremenljivke S^2 ne bomo ukvarjali, ker njeno obnašanje ni tako pohlevno kot pri $\overline{Y}$. Za vsak slučaj pa le opozorimo na tole dejstvo: S^2 je nepristranska cenilka variance, če uporabimo v (6) ocenjeno vrednost aritmetične sredine, torej $\bar{y}$ kot realizacijo vzorčnega povprečja Y. Kadar je znana prava aritmetična sredina populacije μ_y, ocenimo varianco spremenljivke kar z varianco izbranega slučajnega vzorca.

Načelno zdaj poznamo postopek za ocenjevanje parametrov:

1. izberemo slučajen vzorec;
2. ugotovimo vrednosti, ki jih na elementih vzorca zavzame proučevana statistična spremenljivka;
3. izračunamo vrednost ustrezne cenilke in
4. to vrednost proglasimo za oceno parametra celotne populacije.

Ostane nam odgovor na naslednje vprašanje: kako izbrati slučajen vzorec, da bo sploh zaslužil to ime?

V osnovi sestavljamo slučajne vzorce tako, da elemente vanje na ta ali oni način izžrebamo. Lahko gre za fizično izbiranje "na slepo", če je populacija takšna, da to dopušča (pri preverjanju kvalitete fižola, žita, ipd.

na ta način izberemo določen del populacije). Enostavneje je v večini primerov populacijo oštevilčiti in izžrebati zaporedne številke elementov, ki bodo izbrani v vzorec. Žrebanje lahko opravimo sproti ali pa si pomagamo s slučajnimi števili. V posebnih tabelah je zapisano zaporedje iz cifer 0, 1, 2, ..., 8, 9, ki je bilo generirano z zaporednim slepim izbiranjem iz množice $\{0, 1, \ldots, 9\}$. Glede na to, kako velika je osnovna populacija, uporabljamo po dve, tri,... cifre iz tabele, da tako dobimo toliko mestna števila, kolikor mestno je število N, torej moč populacije. Če pri tem dobimo število, ki je večje od N, ga pri pripravi vzorca ne upoštevamo, v nasprotnem primeru pa v vzorec vključimo enoto, katere zaporedna številka v "popisu populacije" je dano slučajno število. Za primer navajamo kratek izvleček iz tablic slučajnih števil.

51772	74640	42331	29044	46621	62898	93582	04186	19640
87056	24033	23491	83587	06568	21960	21387	76387	10836
97453	90581	54939	60173	52078	25424	11645	55870	...

Če bi recimo izbirali vzorec iz populacije 654 učencev, bi lahko jemali po tri cifre skupaj in izbrali v vzorec najprej 517. učenca, števila 727 ne bi upoštevali, ker je večje od N, nato bi izbrali 464. učenca itd. Ker lahko cifre združujemo tudi takole: 517, 177, 772, 727, 274, ... ali v obratnem vrstnem redu, nam že ta kratek izvleček iz tablic slučajnih števil zadošča za relativno obsežne primere. Seveda pa pri resnih statističnih poslih "žrebanja" prepuščamo računalnikom.

VAJE

1. Statistična spremenljivka Y ima na populaciji velikosti $N = 6$ vrednosti 4, 6, 7, 8, 3 in 2. Zapiši vse vzorce velikosti $n = 2$ s ponavljanjem elementov, izračunaj njihove aritmetične sredine in variance ter aritmetični sredini obojih vrednosti. Pravilnost preveri z neposrednim računanjem aritmetične sredine in variance Y ter z uporabo izrekov o zvezi med osnovno populacijo in populacijo vseh vzorcev.

2. Reši enako nalogo za vzorce brez ponavljanja elementov.

3. Za osnovno populacijo, ki ima 5 enot z vrednostmi 4, 6, 9, 8 in 3, izberi vse možne vzorce po 3 enote brez ponavljanja. Preveri pravilnost trditve, da sta $\mu_{\bar{y}} = E(\bar{y})$ in μ_y enaki in da velja obrazec (3.b). Kolikšna je standardna napaka ocene?

4. Na neki šoli je 300 učencev; njihove mase so porazdeljene normalno z matematičnim upanjem 68·0 kg in standardnim odklonom 3·0 kg. Denimo, da smo izbrali slučajen vzorec velikosti $n = 25$ učencev in

izračunali njegovo aritmetično sredino $\bar{y}$ ter ta poskus velikokrat ponovili. Kakšni vrednosti pričakuješ za aritmetično sredino in varianco tako dobljenih vzorčnih aritmetičnih sredin, če gre za vzorce
a) s ponavljanjem;
b) brez ponavljanja?
Koliko se spremenijo rezultati, če je v populaciji namesto 300 kar 3 000 učencev?

5. Za kolikšen delež vseh možnih vzorcev po $n = 25$ enot iz populacije z $N = 3\,000$ enotami lahko pri $Y \sim N(68; 3)$ pričakujemo, da bo vzorčno povprečje zavzelo vrednost med 66·8 in 68·3 kg, če gre za vzorce z vračanjem? Pri kolikšnem deležu vseh možnih vzorcev je vzorčno povprečje manjše od 65 kg?

6. Koliko odstotkov vseh možnih petindvajseteric iz 4. in 5. naloge se ne bi smelo peljati v dvigalu, za katero je dovoljena skupna masa tovora
a) 1 730 kg;
b) 1 700 kg?

7. Katerih 10 enot bi izbral v slučajen vzorec iz populacije z 9 000 enotami, če uporabljaš za izbiro tabelo slučajnih števil in jemlješ po vrsti skupine cifer po štiri skupaj tako, da se ne prekrivajo (od prve do četrte, od pete do osme,...)?

8. Skupini 50 učencev so izmerili telesno višino in dobili naslednje izmerke (v cm):

159	164	175	167	156	166	172	157	179	180
158	161	178	151	171	154	168	158	165	173
165	162	161	176	172	160	168	175	153	156
158	166	167	173	164	169	154	160	167	174
169	180	145	179	151	159	166	170	160	165

Na način, opisan v prejšnji nalogi, izberi iz tabele slučajnih števil 10 števil in sestavi ustrezen vzorec brez ponavljanja elementov. Oceni aritmetično sredino in varianco telesnih višin, izračunaj standardno napako ocene za μ_y, nato pa izračunaj iz gornjih podatkov še pravi vrednosti obeh parametrov ter ju primerjaj z ocenama. (Če meniš, da je ujemanje zelo slabo, ponovi postopek še z enim vzorcem.)

9. Na slepo izberi pet (deset) sošolcev in ugotovi njihove višine ter iz tega vzorca oceni aritmetično sredino višin celotnega razreda. Račun preveri neposredno. Kako je z zanesljivostjo ocene, kaj na splošno pričakuješ od večjega vzorca? Zakaj?

10. Na populaciji s sedmimi elementi ima Y aritmetično sredino 40 in standardni odklon 3. Izračunaj aritmetično sredino porazdelitve vzorčnih

varianc za vzorce s po petimi enotami
a) s ponavljanjem;
b) brez ponavljanja.

11. △ Podatki o vrednostih spremenljivke Y na proučevani populaciji so zbrani v vektorju $Y(I)$, $I = 1, 2, \ldots, N$. Sestavi naslednje računalniške programe:

a) program, ki bi naredil vse možne vzorce velikosti $n = 2$, s ponavljanjem oziroma brez ponavljanja elementov, in vzorce izpisal;

b) program, ki bo izračunal aritmetične sredine in variance vseh možnih vzorcev velikosti $n = 2$ brez ponavljanja in te vrednosti zapisal v vektorja $YS(I)$ in $YV(I)$;

c) program, ki bo na osnovi komponent vektorjev $YS(I)$ in $YV(I)$ izračunal aritmetični sredini vzorčnih sredin in vzorčnih varianc;

č) program, ki bo izračunal aritmetično sredino spremenljivke in njeno varianco na osnovni populaciji in izračunani vrednosti primerjal z ustreznima vrednostima za populacijo vseh vzorcev (točka c).

12. △ △ Podatki o vrednostih spremenljivke Y na proučevani populaciji so zapisani v vektorju $Y(I)$, $I = 1, 2, \ldots, N\,(= 500)$. Sestavi računalniški program, ki bo

a) izračunal aritmetično sredino in varianco spremenljivke Y (glej točko č pri prejšnji nalogi.);

b) s pomočjo generatorja slučajnih števil izbral slučajen vzorec velikosti $n\,(= 50)$ z vračanjem elementov (oziroma z dovoljenim ponavljanjem);

c) na osnovi dobljenega vzorca izračunal oceni za aritmetično sredino in varianco spremenljivke na osnovni populaciji ter dobljena rezultata primerjal s pravima vrednostima ter izračunal relativno napako ocene.

Program lahko še dopolniš, tako da tretji del večkrat ponovi in kot novo oceno izračuna aritmetični sredini dobljenih ocen za μ_y oziroma σ^2 ter ugotavlja, ali se je razlika med oceno in pravo vrednostjo parametra zmanjšala.

5.2 Ocenjevanje parametrov

Na primeru aritmetične sredine bomo spoznali intervalno ocenjevanje statističnih parametrov in premislili, kako je z zanesljivostjo takšnih ocen.

Z ocenjevanjem parametrov smo se srečali že v prejšnjem poglavju, kjer smo neznano aritmetično sredino oziroma varianco spremenljivke na osnovni populaciji ocenili z vrednostjo ustrezne cenilke na izbranem slučajnem vzorcu. Ker je bila ocena izražena z enim samim številom, rečemo taki oceni **točkovna ocena parametra**, postopek pa imenujemo točkovno ocenjevanje.

Takšno ocenjevanje z eno samo vrednostjo je pravzaprav zelo nezanesljivo, saj ne vemo, na katerega od možnih vzorcev smo naleteli, oziroma kje v porazdelitvi slučajnih spremenljivk $\overline{Y}$ ali S^2 leži dobljena vrednost. Zato si v praksi običajno postavimo nalogo nekoliko drugače: namesto točkovne ocene proučevanega parametra navedemo samo interval, ki z določeno (vnaprej predpisano) verjetnostjo pokriva pravo vrednost parametra.

Predpisano verjetnost imenujemo **stopnja zaupanja**, njene najpogostejše vrednosti so 0·95, 0·99 ali 0·999, torej 95%, 99% in 99·9%. Postopku rečemo ocenjevanje z intervali ali krajše **intervalno ocenjevanje**. Po domače bi lahko rekli, da pri tem določamo interval, ki s predpisano verjetnostjo "pokrije" pravo vrednost ocenjevanega parametra.

Ko zapišemo neki interval kot oceno za proučevani parameter, seveda tvegamo, da prava vrednost parametra leži izven tega intervala; **stopnja tveganja** je enaka verjetnosti, da se to zgodi. Ker sta "interval pokriva pravo vrednost" in "prava vrednost leži izven intervala" nasprotna dogodka, je med obema stopnjama preprosta zveza:

"stopnja tveganja" = 1 - "stopnja zaupanja"

To pomeni, da običajno postavljamo trditve s tveganjem 5%, 1% ali 0·1%.

Predpostavimo, da je statistična spremenljivka Y na proučevani populaciji porazdeljena po zakonu $N\,(a, \sigma_y)$, pri čemer parameter σ_y (na kratko σ) poznamo, parametra $a = \mu_y$ pa ne. Kot vemo, je v tem primeru cenilka $\overline{Y}$ porazdeljena normalno, s parametroma a in $\sigma/\sqrt{n}$, pri čemer n kot običajno pomeni velikost slučajnega vzorca. Iz te cenilke naredimo novo slučajno spremenljivko,

$$Z = \frac{(\overline{Y} - a)}{\frac{\sigma}{\sqrt{n}}}$$

oziroma

$$Z = \frac{\overline{Y} - a}{\sigma}\sqrt{n} \qquad (1)$$

Kot vemo iz verjetnostnega računa, je v tem primeru slučajna spremenljivka Z porazdeljena standardizirano normalno, $Z \sim N(0, 1)$. To je za nadaljnje delo zelo ugodno, ker znamo s pomočjo tabel za standardizirano normalno porazdelitev določiti verjetnost dogodka, da slučajna spremenljivka Z zavzame vrednost z danega intervala:

$$P(c \leq Z \leq d) = \Phi(d) - \Phi(c) \tag{2}$$

Pri ocenjevanju aritmetične sredine $\mu_y = a$ želimo doseči, da je ta verjetnost enaka vnaprej predpisani vrednosti, ki se za neko stopnjo tveganja loči od 1, torej

$$P(c \leq Z \leq d) = 1 - \alpha \tag{3}$$

Če upoštevamo definicijo slučajne spremenljivke Z, to pomeni, da naj bo

$$P(c \leq \frac{\overline{Y} - a}{\sigma}\sqrt{n} \leq d) = 1 - \alpha \tag{4}$$

oziroma (zaradi simetričnosti porazdelitve)

$$P(c \leq \frac{a - \overline{Y}}{\sigma}\sqrt{n} \leq d) = 1 - \alpha$$

Razrešimo izraz v oklepaju tako, da dobimo meji za parameter a:

$$P(\overline{Y} + c\frac{\sigma}{\sqrt{n}} \leq a \leq \overline{Y} + d\frac{\sigma}{\sqrt{n}}) = 1 - \alpha \tag{5}$$

Pri ocenjevanju je torej potrebno določiti števili c in d tako, da je izpolnjena enakost (5). Rešitev tega problema je neskončno mnogo, saj lahko k poljubnemu dovolj majhnemu c najdemo tak d, da velja (5). Na drugi strani pa naj bi bil interval, določen s (5), čim krajši, ker pač želimo "ujeti" pravo vrednost parametra a v čim ožji interval in s tem (pri dani stopnji tveganja) dobiti "konkretnejšo" oceno za a. Iz oblike normalne porazdelitve se vidi, da je od vseh možnih intervalov najkrajši tisti, ki ima središče v točki a. Z intervalom sežemo enako daleč v levo in v desno; razdalja med središčem in mejo intervala je seveda odvisna od stopnje tveganja, zato jo bomo označili s simbolom z_α. Potem moramo v (5) vzeti $c = -z_\alpha$ in $d = +z_\alpha$, pri čemer je vrednost z_α v bistvu določena z zahtevo (3),

$$P(-z_\alpha \leq Z \leq z_\alpha) = 1 - \alpha \tag{6}$$

kar pa zaradi simetričnosti normalne porazdelitve lahko nadomestimo z enostavnejšo (glej razdelek 4.8!),

$$P(0 \leq Z \leq z_\alpha) = \frac{1 - \alpha}{2} \tag{7}$$

Izraz na levi je natanko vrednost $\Phi(z_\alpha)$; pri predpisani vrednosti α, ki označuje stopnjo tveganja, moramo torej v tabeli funkcije $\Phi(z)$ odkriti, pri

kateri vrednosti neodvisne spremenljivke ima funkcija vrednost $(1 - \alpha)/2$. Ko to vrednost najdemo, lahko (5) prepišemo v

$$P(\overline{Y} - z_\alpha\frac{\sigma}{\sqrt{n}} \leq a \leq \overline{Y} + z_\alpha\frac{\sigma}{\sqrt{n}}) = 1 - \alpha \tag{8}$$

Interval

$$\left[\overline{Y} - z_\alpha\frac{\sigma}{\sqrt{n}}, \overline{Y} + z_\alpha\frac{\sigma}{\sqrt{n}}\right] \tag{9}$$

imenujemo interval zaupanja za parameter $a = \mu_y$ s stopnjo zaupanja (tudi: koeficientom zaupanja) $1 - \alpha$.

Interval (9) ni čisto takšen, kot smo jih bili doslej vajeni v matematiki: njegovi meji sta dejansko slučajni spremenljivki, ker se v izrazih za spodnjo mejo zaupanja $\overline{Y} - (z_\alpha\,\sigma)/\sqrt{n}$ oziroma za zgornjo mejo zaupanja $\overline{Y} + (z_\alpha\,\sigma)/\sqrt{n}$ pojavlja slučajna spremenljivka - cenilka aritmetične sredine. Gre torej za slučajen interval, ki ga lahko interpretiramo takole: Če slučajni vzorec $(Y_1, Y_2, \ldots, Y_n)$ velikokrat realiziramo, je delež tistih realizacij slučajnega intervala (9), ki pokrijejo dejansko vrednost parametra $a = \mu_y$, navadno (vsaj približno) enak $1 - \alpha$.

Vidimo, da je verjetnostna interpretacija intervalskega ocenjevanja takšna, kot smo je sicer vajeni od slučajnih spremenljivk. Čeprav iz konkretnega vzorca dobimo fiksni meji zaupanja (recimo c in d), je seveda nesmiselno zapisati, da je $P(c \leq a \leq d) = 1 - \alpha$; dogodek na levi strani je namreč ali gotov (če je neenačba za tri dana števila dejansko izpolnjena) ali pa nemogoč (če enačba ne drži), v nobenem primeru pa njegova verjetnost ni $1 - \alpha$, ampak bodisi 1 ali 0.

Pokažimo zdaj, kako poteka ocenjevanje v konkretnem primeru. V ta namen najprej pripravimo vrednosti z_α za najpogostejše stopnje tveganja $\alpha = 0{\cdot}05$, $\alpha = 0{\cdot}01$ in $\alpha = 0{\cdot}001$. V natančnejši tabeli funkcije $\Phi(z)$ (izvleček iz tabele smo uporabljali v razdelku 4.8) bi v skladu s (7) poiskali, pri katerih vrednostih z ima funkcija $\Phi(z)$ vrednost $(1 - \alpha)/2$, in za predpisane stopnje tveganja po vrsti dobili naslednje vrednosti:

$$\begin{array}{llll} \alpha = 0{\cdot}05 & : \Phi(z_{0{\cdot}05}) = \dfrac{1 - 0{\cdot}05}{2} = 0{\cdot}4750 & \Rightarrow z_{0{\cdot}05} & \approx 1{\cdot}96 \\ \alpha = 0{\cdot}01 & : \Phi(z_{0{\cdot}01}) = \dfrac{1 - 0{\cdot}01}{2} = 0{\cdot}4950 & \Rightarrow z_{0{\cdot}01} & \approx 2{\cdot}58 \\ \alpha = 0{\cdot}001 & : \Phi(z_{0{\cdot}001}) = \dfrac{1 - 0{\cdot}001}{2} = 0{\cdot}4995 & \Rightarrow z_{0{\cdot}001} & \approx 3{\cdot}29 \end{array} \tag{10}$$

Tako lahko v skladu z (9) zapišemo intervale zaupanja za dane stopnje tveganja, oziroma, kot rečemo krajše, pri danem tveganju:

$$\alpha = 0{\cdot}05 : \left[\overline{Y} - 1{\cdot}96\frac{\sigma}{\sqrt{n}}, \ \overline{Y} + 1{\cdot}96\frac{\sigma}{\sqrt{n}}\right] \tag{11.a}$$

$$\alpha = 0{\cdot}01 : \left[\overline{Y} - 2{\cdot}58\,\frac{\sigma}{\sqrt{n}},\ \overline{Y} + 2{\cdot}58\,\frac{\sigma}{\sqrt{n}}\right] \qquad (11.b)$$

$$\alpha = 0{\cdot}001 : \left[\overline{Y} - 3{\cdot}29\,\frac{\sigma}{\sqrt{n}},\ \overline{Y} + 3{\cdot}29\,\frac{\sigma}{\sqrt{n}}\right] \qquad (11.c)$$

Če želimo tveganje zmanjšati, mora biti "past" za neznani parameter širša; interval (11.a), ki ustreza v praksi najpogostejši stopnji tveganja $\alpha = 0{\cdot}05$, je takšen, da v povprečju v 95% vseh realizacij slučajnega vzorca pokrije pravo vrednost aritmetične sredine, pri 5% vseh možnih vzorcev pa izračunani meji zaupanja ne vključujeta prave vrednosti parametra a. Pri dani izbiri slučajnega vzorca seveda ne vemo, v katero od obeh skupin le-ta spada, odtod ravno tveganje, povezano z našo oceno.

Na drugi strani pa vidimo, da je širina intervala zaupanja odvisna ne le od stopnje tveganja a in z njo povezane kritične vrednosti z_a, ampak tudi od količnika $\sigma/\sqrt{n}$, v katerem prepoznamo natanko standardno napako ocene za aritmetično sredino μ_y. Oblika intervala zaupanja oziroma njegovih mej znova potrjuje že zapisano ugotovitev, da z večanjem števila enot v vzorcu (n) povečujemo natančnost ocene, saj s tem ožimo interval zaupanja (pri nespremenjeni stopnji tveganja). Res je, da to "ožanje" ni linearno odvisno od velikosti vzorca (za prepolovitev intervala potrebujemo - pri nespremenjenih drugih pogojih - štirikrat večji vzorec), a vendar.

O povezavi med velikostjo vzorca in zanesljivostjo ocene bomo še govorili, zdaj pa najprej številska zgleda.

ZGLED 5.2.1:

Statistična spremenljivka je porazdeljena po zakonu $N(a, 10)$, pri čemer parameter a (aritmetična sredina) ni znan. Iz populacije smo izbrali slučajen vzorec velikosti $n = 100$ enot in ugotovili, da je za ta vzorec $\overline{y} = 20$. Poišči interval zaupanja za parameter a pri stopnji tveganja $0{\cdot}05$!

Meji zaupanja dobimo za konkreten vzorec tako, da v (11.a) vstavimo vzorčno aritmetično sredino 20, $\sigma = 10$ in $n = 100$, kar nam da ob upoštevanju vrednosti $z_{0{\cdot}05} = 1{\cdot}96$ za iskani interval

$$\left[20 - 1{\cdot}96\,\frac{10}{\sqrt{100}},\ 20 + 1{\cdot}96\,\frac{10}{\sqrt{100}}\right] = [18{\cdot}04,\ 21{\cdot}96]$$

Rezultat pove, da je $[18{\cdot}04,\ 21{\cdot}96]$ tisti interval, ki z verjetnostjo $0{\cdot}95$ pokrije pravo vrednost parametra a.

ZGLED 5.2.2:

Kolikšen naj bo pri podatkih iz prejšnjega zgleda vzorec, da se pri stopnji tveganja $\alpha = 0{\cdot}05$ prava vrednost aritmetične sredine ne bo razlikovala od ocene za več kot $0{\cdot}5$ (navzgor ali navzdol)?

Vrednost cenilke $\overline{Y}$ na izbranem slučajnem vzorcu je točkovna ocena aritmetične sredine za celo populacijo; pri dani stopnji tveganja α sega interval zaupanja $(z_\alpha\sigma)/\sqrt{n}$ v levo oziroma desno. Če pri dani stopnji tveganja absolutno vrednost razlike med oceno in pravo aritmetično sredino omejimo na določeno vnaprej predpisano vrednost Δ (pri nas $\Delta = 0{\cdot}5$), to pomeni, da zahtevamo izpolnjenost pogoja

$$z_\alpha\,\frac{\sigma}{\sqrt{n}} \leq \Delta$$

Neenačbo kvadriramo in razrešimo na velikost slučajnega vzorca, torej n:

$$n \geq \frac{z_\alpha^2 \cdot \sigma^2}{\Delta^2} \qquad (12)$$

Pri naših podatkih ($z_{0{\cdot}05} = 1{\cdot}96$, $\sigma = 10$, $\Delta = 0{\cdot}5$) dobimo za velikost vzorca zahtevo

$$n \geq \frac{1{\cdot}96^2 \cdot 10^2}{0{\cdot}5^2} \Longrightarrow n \geq 1\,537$$

Ker smo želeli bistveno ožji interval zaupanja, kot pa smo ga dobili v zgledu 5.2.1, moramo uporabiti tudi bistveno večji slučajni vzorec. Pri tem je treba poudariti, da z večjim vzorcem sicer dobimo ožji interval zaupanja, verjetnost, da interval ne pokrije prave vrednosti ocenjevanega parametra, pa je še vedno enaka stopnji tveganja, v našem primeru potemtakem $0{\cdot}05$. Na drugi strani pa bi bil pri tako velikem vzorcu koeficient zaupanja $1 - \alpha$ za prvotni interval (izračunan v zgledu 5.2.1) bistveno večji od $0{\cdot}95 = 1 - 0{\cdot}05$. Za vajo lahko to vrednost tudi izračunaš.

Podatek o velikosti vzorca, ki je potrebna pri določeni raziskavi, je treba poznati vnaprej; dovoljeno odstopanje Δ je dostikrat dano kar v odstotkih od ocenjevanega parametra, recimo kot $(p\ \mu_y)/100$. V tem primeru ima obrazec (12) obliko

$$n \geq \left(\frac{100\, z_\alpha\, \sigma}{p\, \mu_y}\right)^2 \qquad (13)$$

Obrazec vsebuje parametra populacije, ki ju vnaprej ne poznamo. Zato velikost vzorca določimo na osnovi grobih ocen teh parametrov, ki jih strokovnjaki lahko dajo na osnovi splošnega poznavanja proučevane populacije.

V dosedanjih primerih smo se lotevali velikih populacij in vzorcev, ki so majhni v primerjavi z velikostjo populacije. V praksi včasih le naletimo na vzorce brez ponavljanja elementov, kjer velikost vzorca (n) ni zanemarljiva v primerjavi z velikostjo populacije (N). V tem primeru se namesto standardne napake ocene $\sigma/\sqrt{n}$, ki smo jo srečevali do sedaj, v intervalih zaupanja pojavlja standardna napaka, ki je značilna za vzorce brez ponavljanja elementov; v prvem razdelku tega poglavja smo jo zapisali v obrazcu (3.b):

$$\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \tag{14}$$

Tako ima za vzorce brez ponavljanja elementov pri dani stopnji tveganja α interval zaupanja za aritmetično sredino spremenljivke, ki je porazdeljena normalno z varianco σ^2, obliko

$$\left[\overline{Y} - z_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}},\ \overline{Y} + z_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}\right] \tag{15}$$

ZGLED 5.2.3:

200 učencev je pisalo preizkus matematičnega znanja. Za oceno uspeha je bil izbran slučajni vzorec 50 učencev, brez ponavljanja. Ugotovljeno je bilo, da so izbrani učenci v povprečju dosegli 65 točk od 100 možnih, varianca te spremenljivke pa je bila ocenjena na 100 (oziroma standardni odklon na 10 točk). Poišči 95% interval zaupanja za povprečno število točk v celotni populaciji!

Preden se lotimo računanja, nekaj pripomb. Ohlapno izražanje "95% interval zaupanja" je v praksi vendarle dovolj pogosto, da se ga splača razumeti: gre za interval zaupanja, pri katerem dopuščamo 5% tveganje, torej $\alpha = 0{\cdot}05$.

Statistične spremenljivke, kot je npr. uspeh pri šolskih preizkušnjah, se na velikih populacijah običajno porazdeljujejo vsaj približno normalno, enako velja za različne testne rezultate pri psiholoških raziskavah, za porazdelitev izmerkov določene količine, če meritve ponavljamo v enakih razmerah in ne delamo sistematičnih napak in podobno. Ne glede na to, ali lahko za spremenljivko iz našega zgleda predpostavimo normalno porazdeljenost na populaciji 200 učencev, pa je glede na velikost vzorca cenilka $\overline{Y}$ porazdeljena vsaj približno normalno, kot smo zapisali ob prvem izreku v prejšnjem razdelku. Interval zaupanja bomo potemtakem določali za neznano aritmetično sredino (μ_y oziroma a) normalne porazdelitve, za katero predpostavimo, da je varianca poznana.

Ker pomeni vzorec znaten del celotne populacije in je bil oblikovan brez ponavljanja elementov, moramo interval zaupanja določiti s pomočjo obrazca (15). Pri naših podatkih ($\alpha = 0{\cdot}05$, $z_\alpha = 1{\cdot}96$, $N = 200$, $n = 50$, $\sigma = 10$) je razlika med središčem intervala in spodnjo ali zgornjo mejo zaupanja enaka

$$z_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = 1{\cdot}96 \frac{10}{\sqrt{50}} \sqrt{\frac{200-50}{200-1}} \approx 2{\cdot}4 \tag{16}$$

tako da je v skladu s podatkom, da je $\bar{y} = 65$, ustrezni interval zaupanja pri tveganju $\alpha = 0{\cdot}05$ enak

$$[65 - 2{\cdot}4,\ 65 + 2{\cdot}4] = [62{\cdot}6,\ 67{\cdot}4]$$

Pri podatkih iz tega zgleda je vrednost faktorja, po katerem se razmik med središčem intervala in posamezno mejo pri vzorcih brez ponavljanja razlikuje od ustrezne vrednosti pri vzorcih s ponavljanjem elementov, enaka

$$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{200-50}{200-1}} = 0{\cdot}868\ldots$$

in faktorja ne smemo kar zanemariti. Ne glede na to, koliko se ta faktor razlikuje od 1, pa je od te vrednosti vedno manjši, zato je interval zaupanja pri vzorcih brez ponavljanja ob nespremenjenih ostalih pogojih ožji od ustreznega intervala za vzorce s ponavljanjem. To kaže, da so ocene z vzorci brez ponavljanja nekoliko zanesljivejše. Razlika med obema načinoma se seveda zmanjšuje, če se povečuje velikost populacije.

Oglejmo se še nekoliko drugačen problem; podobnega smo že predlagali za domačo nalogo.

ZGLED 5.2.4:

Pri podatkih iz prejšnjega zgleda smo ocenili, da interval [64, 66] pokriva aritmetično sredino števila točk. Kolikšno je pri takšni oceni tveganje?

Problem je sedaj obrnjen v tem smislu, da poznamo interval zaupanja, ne vemo pa, kateri stopnji tveganja ustreza. Ker je pri podatkih iz zgleda 5.2.3 aritmetična sredina izbranega vzorca enaka 65 (točk), je dani interval zaupanja takšen, da sega za 1 v levo oziroma v desno od aritmetične sredine ($\bar{y}$). Zaradi (15) mora torej biti

$$z_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = 1$$

odkoder izračunamo, da je

$$z_\alpha = \frac{\sqrt{n}}{\sigma} \sqrt{\frac{N-1}{N-n}} = \frac{\sqrt{50}}{100} \sqrt{\frac{200-1}{200-50}} = 0{\cdot}814\ldots$$

Tveganje pri ocenjevanju z intervalom je enako verjetnosti, da ustrezna slučajna spremenljivka zavzame vrednost izven tega intervala; v našem primeru verjetnosti, da standardizirano normalna spremenljivka zavzame vrednost izven intervala

$$[-z_\alpha,\ z_\alpha] = [-0{\cdot}814\ldots,\ 0{\cdot}814\ldots]$$

Ta verjetnost pa je enaka (glej spet razdelek 4.8!)

$$1 - 2\,\Phi(0{\cdot}814\ldots) = 1 - 0{\cdot}584\ldots \approx 0{\cdot}416$$

ali skoraj 42%, kar je seveda preveliko tveganje za kakršnokoli pametno uporabo ocene. Interval kaže razširiti, da bo tveganje manjše.

V praksi variance oziroma standardnega odklona proučevane spremenljivke običajno ne poznamo, ko ocenjujemo njeno aritmetično sredino. Pri velikih vzorcih si pomagamo z vrednostjo cenilke S^2 (oziroma s korenom iz te vrednosti) na izbranem slučajnem vzorcu kot oceno neznanega parametra; takšna uporaba približkov seveda povzroči, da je tudi dobljeni interval zaupanja za aritmetično sredino samo približen. Pravi interval bi dobili, če bi namesto kritičnih vrednosti normalne porazdelitve z_α vzeli ustrezne vrednosti t.i. *Studentove porazdelitve*, ki pa je v tem učbeniku ne utegnemo obravnavati.

V A J E

1. Statistična spremenljivka Y ima na populaciji velikosti $N = 6$ vrednosti 4, 6, 7, 8, 3 in 2. Izračunaj standardni odklon, nato pa izberi slučajen vzorec velikosti $n = 3$ (s ponavljanjem elementov) in oceni interval zaupanja za aritmetično sredino μ_y pri tveganju $\alpha = 0{\cdot}05$. Izračunaj še pravo vrednost tega parametra in ugotovi, ali jo je izračunani interval "pokril".

2. Ali imaš kake pomisleke v zvezi s 1. nalogo?

3. Na neki šoli je 300 učencev, njihove mase so porazdeljene normalno, $\mu_y = 68{\cdot}0$ kg in $\sigma = 3{\cdot}0$ kg. Ali je možno, da pri izboru slučajnega vzorca velikosti $n = 25$ (brez ponavljanja elementov) dobimo vzorčno povprečje $\bar{y} = 69{\cdot}78$ kg?

4. V 8. nalogi prejšnjega poglavja smo izračunali točkovno oceno aritmetične sredine μ_y telesnih višin za proučevano populacijo učencev, poznamo pa tudi pravo vrednost standardnega odklona, $\sigma = 8{\cdot}37$. Izračunaj ustrezne intervale zaupanja za običajne stopnje tveganja (0·05, 0·01 in 0·001) ob predpostavki, da je osnovna spremenljivka porazdeljena normalno. (Zakaj je ta predpostavka potrebna, če želimo biti natančni?)

5. Za telesne višine učencev vaše šole lahko predpostavimo, da so porazdeljene vsaj približno normalno. Izberi primeren slučajen vzorec in izračunaj intervalne ocene za μ_y pri različnih stopnjah tveganja.

6. Tovarna jamči, da teodolit pri merjenju kotov ne dela sistematičnih napak, za slučajne napake pa velja, da so normalno porazdeljene, pri čemer standardni odklon ne presega vrednosti 30″. Določen kot smo merili devetkrat in dobili naslednje vrednosti:

20°40′32″, 20°40′46″, 20°40′25″ ,

20°40′20″, 20°40′34″, 20°40′42″ ,

20°40′28″, 20°40′34″, 20°40′27″

Izračunaj interval zaupanja za pravo vrednost merjenega kota pri stopnji tveganja $\alpha = 0{\cdot}05$.

7. Izkušnje v podjetju kažejo, da je čas, ki ga strugarji porabijo za izdelavo nekega sestavnega dela, porazdeljen normalno s standardnim odklonom 13 minut. Kako širok bo pri 5% tveganju interval zaupanja za aritmetično sredino te spremenljivke, če ga ocenimo s slučajnim vzorcem brez ponavljanja enot, ki vsebuje

a) 16 strugarjev;

b) 64 strugarjev?

Vseh strugarjev je v tem podjetju 152. (Nalogo reši še ob predpostavki, da je strugarjev toliko, da ni treba upoštevati popravnega faktorja za vzorce brez ponavljanja.)

8. Določi interval zaupanja pri 5% tveganju za aritmetično sredino populacije, ki je normalno porazdeljena z varianco $\sigma^2 = 9$, če smo za vzorec s 100 elementi izračunali vzorčno povprečje (aritmetično sredino vzorca) $\bar{y} = 5$.

9. Vzorec s 225 elementi ima aritmetično sredino 7·26. Določi interval zaupanja pri 1% tveganju, če je osnovna spremenljivka Y porazdeljena po zakonu $N\,(a,\ 2{\cdot}4)$.

10. Predpostavimo, da je življenjska doba določenega avtomobilskega modela porazdeljena normalno. Na vzorcu 30 avtomobilov te vrste so ugotovili, da so prevozili povprečno 196 000 km, standardni odklon je ocenjen na 17 000 km. Oceni življenjsko dobo modela z 10% tveganjem.

11. Pri problemu iz 7. vaje so na vzorcu 16 strugarjev izmerili povprečni izdelavni čas $\bar{y} = 140$ minut. Določi 95% interval zaupanja za to spremenljivko, če je standardni odklon nespremenjen.

12. △ Kako velik vzorec moramo sestaviti pri problemu iz 7. vaje, če hočemo pri 5% tveganju dobiti interval zaupanja, ki bo krajši od 4 minut? Vzorec izbiramo

a) s ponavljanjem enot;

b) brez ponavljanja enot.

13. △ Kako velik mora biti vzorec s ponavljanjem, da je pri 5% tveganju dolžina intervala zaupanja enaka 1/5 standardnega odklona osnovne spremenljivke?

14. △ Podatki o vrednosti spremenljivke na osnovni populaciji so zapisani v vektorju $Y(I)$, $I = 1, 2, 3, \ldots, N$: Sestavi računalniški program, ki bo v slučajen vzorec

a) s ponavljanjem enot;

b) brez ponavljanja enot

izbral določen odstotek populacije, izračunal $\bar{y}$, s^2 in s ter pri predpisani stopnji tveganja izračunal interval zaupanja za aritmetično sredino M_y. Uporabnik naj ima možnost odločanja

a) za vzorce s ponavljanjem ali brez njega;

b) za relativno velikost vzorca (razmerje n/N v odstotkih);

c) za eno od treh najpogostejših stopenj tveganja (0'05, 0'01 ali 0'001).

Pri premajhnem vzorcu ($n < 30$) naj računalnik uporabnika opozori na možne posledice, ker osnovna spremenljivka ni nujno porazdeljena normalno.

5.3 Preizkušanje domnev

Splošni namen tega razdelka se ujema z namenom celotnega poglavja: še vedno ugotavljamo, kaj lahko na podlagi slučajnega vzorca povemo o statistični spremenljivki Y na celotni proučevani populaciji.

Problem bomo tokrat obravnavali konkretneje tako, da bomo postavili **domnevo** (statistično hipotezo) o aritmetični sredini proučevane spremenljivke, o njeni varianci ali kakem drugem parametru. Domneve so lahko tudi splošnejše, na primer o porazdelitvenem zakonu nasploh, vendar se z njimi ne moremo ukvarjati.

Domnevi ali hipotezi, ki jo preizkušamo na osnovi vzorčnih podatkov, rečemo **ničelna domneva** ali *ničelna hipoteza* in jo simbolično označimo s H_0. Pokazalo se bo, da jo pravzaprav preizkušamo zato, da bi jo po možnosti zavrnili in s tem sprejeli njej **nasprotno (alternativno) domnevo** H_1, ki je pravzaprav osnovni predmet našega zanimanja.

ZGLED 5.3.1:

Zanima nas pravilnost domneve, da je povprečna bruto plača v določeni občini različna od povprečja, ki velja za Slovenijo. Osnovna hipoteza je trditev "aritmetična sredina bruto plač v občini ni enaka aritmetični sredini bruto plač v RS", za ničelno domnevo, ki jo poskušamo zavrniti, pa postavimo nasprotno domnevo, da sta povprečni bruto plači enaki.

ZGLED 5.3.2:

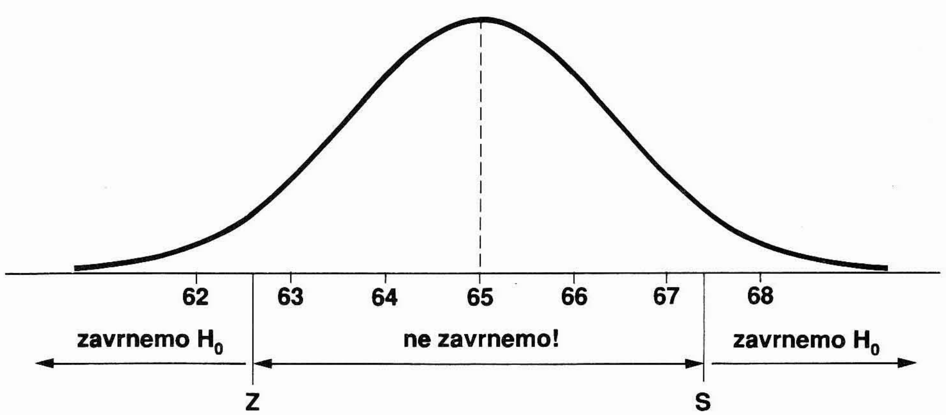
Pri problemu iz zgleda 5.2.3 v prejšnjem razdelku lahko na primer presojamo domnevo "aritmetična sredina rezultatov ni 65 točk" posredno tako, da testiramo ničelno hipotezo H_0: "aritmetična sredina rezultatov je 65 točk".

Domneve preizkušamo z ustreznimi testnimi pravili, ki jih izpeljemo iz značilnih zvez med vzorčnimi podatki in vrednostmi parametrov na celotni populaciji. Za naše potrebe bodo za preizkušanje domnev zadoščale intervalne ocene parametrov, ki smo jih spoznali v prejšnjem razdelku. Osnovni problem preizkušanja domnev bomo najbolje spoznali ob praktičnem primeru.

V zgledu 5.2.3 iz prejšnjega razdelka smo ugotavljali povprečno število točk pri preizkusu znanja in ocenili, da je to 65 točk. Predpostavimo, da ta preizkus znanja ponovimo. Kako bi na osnovi enako velikega vzorca (50 od skupaj 200 učencev, brez vračanja enot) presodili o tem, ali se je aritmetična sredina točk spremenila?

Za ničelno domnevo H_0 proglasimo nasprotno domnevo, da je aritmetična sredina nespremenjena, $H_0 : \mu_y = 65$. Na osnovi znanja, ki smo ga do zdaj pridobili, lahko rečemo tole: če je hipoteza H_0 pravilna, imajo aritmetične sredine vseh možnih vzorcev po $n = 50$ enot, izbranih brez vračanja elementov, aritmetično sredino $\mu_{\overline{Y}}$ enako aritmetični sredini osnovne spremenljivke, to je 65, standardni odklon teh aritmetičnih sredin pa je za faktor $\sqrt{(N-n)/(n\,(N-1))}$ manjši od standardnega odklona osnovne spremenljivke. V omenjenem zgledu smo pri 5% tveganju ($\alpha = 0{\cdot}05$) izračunali, da je v tem primeru interval zaupanja [62·6; 67·4]. Predpostavili bomo, da se standardni odklon spremenljivke ni spremenil. Potem v primeru pravilnosti hipoteze H_0 velja, da v ta interval padejo aritmetične sredine 95% vseh možnih vzorcev opisanega tipa.

Če se omejimo na posamezen slučajen vzorec iz populacije vseh možnih vzorcev predpisane velikosti iz te populacije, lahko zadnjo izjavo povemo drugače, namreč takole: verjetnost, da $\bar{y}$, ki jo izračunamo za ta vzorec, leži na danem intervalu zaupanja, je enaka 0·95. To dejstvo, ki ga že dobro poznamo, nam ponuja naslednje testno pravilo za obravnavani zgled: izberemo slučajen vzorec, izračunamo njegovo aritmetično sredino $\bar{y}$ in domnevo $H_0 : \mu_y = 65$ sprejmemo, če izračunana aritmetična sredina vzorca leži na intervalu zaupanja, torej če je $62{\cdot}6 \le \bar{y} \le 67{\cdot}4$; v nasprotnem primeru pa bi hipotezo H_0 zavrnili. Porazdelitev vseh vzorčnih in aritmetičnih sredin in način odločanja sta prikazana tudi na naslednji sliki.

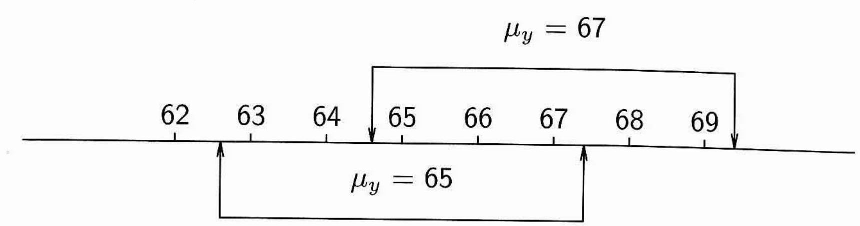


Slika 5.1: Preizkušanje ničelne domneve H_0

Pravilo, ki smo si ga pripravili, je dokaj logično, vendar skriva v sebi dve nevarnosti:

1. Čeprav je prava vrednost aritmetične sredine $\mu_y = 65$, lahko vzorčna aritmetična sredina $\bar{y}$ brez nadaljnjega pade izven intervala [62·6, 67·4]. Verjetnost, da se to zgodi, je sicer majhna (enaka stopnji tveganja, to je 0·05), vendar te možnosti ne smemo prezreti. Če se nam to pripeti, bi po našem pravilu morali zavrniti hipotezo H_0, čeprav bi bila v tem primeru pravilna; zagrešili bi tako imenovano **napako I. vrste**. Enostavno lahko rečemo, da napako I. vrste naredimo, če zavrnemo hipotezo, ki je sicer pravilna.

2. Lahko pa se nam pripeti še bolj nerodna reč, ko bi morali hipotezo zaradi napačnosti zavrniti, pa jo sprejmemo. Vzemimo na primer, da prava vrednost μ_y (pri nespremenjenih ostalih podatkih) ni 65, ampak 67, mi pa še vedno testiramo hipotezo $H_0 : \mu_y = 65$. Širina intervala zaupanja ni odvisna od tega, kje leži središče le-tega, zato je tudi pri $\mu_y = 67$ enaka 2·4, kakor smo jo izračunali na strani 39. Zato je pri stopnji tveganja $\alpha = 0{\cdot}05$ in $\mu_y = 67$ ustrezen interval enak [64·6, 69·4]. Iz naslednje slike vidimo, da se oba intervala zaupanja v dobršnem delu prekrivata.



Slika 5.2: Možnost za napako II. vrste

V skladu z našim testnim pravilom bomo pri vrednostih cenilke $\overline{Y}$, ki ležijo na intervalu [62·6; 67·4], hipotezo H_0 sprejeli. Kot vidimo iz slike, tudi pri $\mu_y = 67$ velikokrat naletimo na vzorec, ki da vrednost $\bar{y}$ z omenjenega intervala in tako narekuje sprejem napačne domneve. (Če obvladaš delo z normalno porazdelitvijo, ti ne bo pretežko izračunati verjetnosti take napake.) Če sprejmemo napačno hipotezo, smo naredili **napako II. vrste**.

Napake II. vrste so v praksi običajno nevarnejše, zato se jih bomo znebili tako, da bomo nekoliko spremenili naše pravilo:

Hipotez v nobenem primeru ne bomo sprejemali (da ne bi po nesreči sprejeli napačne), ampak jih bomo samo zavračali, če bodo vrednosti vzorčnih parametrov padle izven ustreznega intervala zaupanja.

V tem primeru bomo rekli, da se eksperimentalni podatki **značilno razlikujejo** (ali tudi *signifikantno* razlikujejo) od predpostavljenih, vsebovanih v hipotezi H_0, zato takrat hipotezo H_0 zavrnemo. V nasprotnem primeru, ko so vzorčni podatki taki, da izračunani vzorčni parameter leži v ustreznem intervalu zaupanja, pa rečemo, da razloček ni značilen pri dani stopnji tveganja, ali drugače, da hipoteza H_0 ni v nasprotju z eksperimentalnimi podatki in je zato ne zavrnemo.

ZGLED 5.3.3:

Če bi pri podatkih iz zgleda 5.3.2 dobili vrednost $\bar{y} = 69$, ki leži izven ustreznega intervala [62·5, 67·4], bi hipotezo $H_0 : \mu_y = 65$ na stopnji tveganja 0·05 zavrnili. Če pa bi iz vzorca za vrednost $\bar{y}$ dobili število z omenjenega intervala, bi rekli, da hipoteze H_0 na tej stopnji tveganja ne moremo zavrniti.

Namesto o stopnji tveganja govorimo pri preizkušanju domnev rajši o **stopnji značilnosti**. Če na primer tako kot v zgledu 5.3.3 preverjamo domnevo, da je $\mu_y = a$ (kjer je a neko realno število) in izračunamo pri dani vrednosti α, da vzorčna aritmetična sredina leži na intervalu zaupanja, rečemo, da razlika med vzorčnimi podatki in hipotezo na tej stopnji **ni značilna** (signifikantna) in hipoteze ne zavrnemo. Če bi v zgledu 5.3.3 dobili vrednost $\bar{y} = 68{\cdot}0$, bi rekli, da so razlike **značilne na stopnji** 0·05 in moramo hipotezo $H_0 : \mu_y = 65$ zavrniti. Razumljivo je, da se z manjšanjem α povečuje pripadajoča vrednost z_α in s tem tudi širi interval zaupanja. To pomeni naslednje:

Z manjšanjem tveganja, ki ga še dopuščamo, se širi interval, pri katerem ne moremo odločiti o ničelni hipotezi.

Seveda lahko vplivamo na širino intervala v nasprotni smeri tako, da uporabimo večji vzorec, kot smo ugotovili v prejšnjem poglavju.

Tako kot pri ocenjevanju parametrov z intervali se bomo tudi pri preizkušanju domnev omejili na delo z velikimi vzorci, kjer so cenilke, ki jih uporabljamo, porazdeljene normalno. Dejstvo je, da moramo v praksi velikokrat preizkušati domneve z malimi vzorci; takšnih primerov tu žal ne moremo obravnavati, ker ne poznamo verjetnostnih porazdelitev $\overline{Y}, S^2$ in drugih cenilk v primeru majhnih vzorcev. V opravičilo povejmo, da se postopek formalno razlikuje samo po tem, da namesto vrednosti z_α za normalno porazdelitev uporabimo ustrezne vrednosti drugih verjetnostnih porazdelitev.

Vzemimo za začetek spet znani primer, ko za statistično spremenljivko Y ne poznamo aritmetične sredine, varianca pa je znana in enaka σ^2. Tudi če osnovna spremenljivka Y ni porazdeljena normalno, je pri velikem vzorcu $\overline{Y}$, to je vzorčno povprečje, porazdeljeno skoraj normalno. Preizkušali bomo domnevo

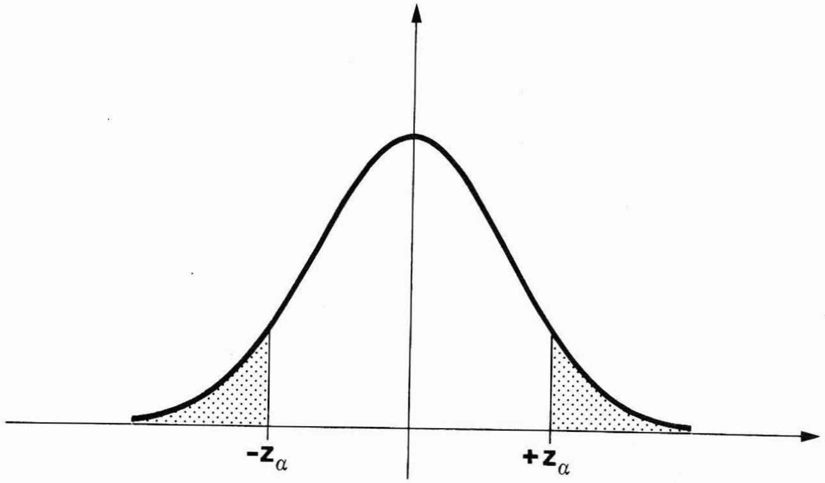
$$H_0 : \mu_y = a$$

Kot vemo iz uvodnega razdelka tega poglavja, je $\overline{Y}$ porazdeljen po zakonu $N(a, \sigma/\sqrt{n})$, če je le hipoteza H_0 pravilna. Slučajna spremenljivka

$$Z = \frac{\overline{Y} - a}{\sigma}\sqrt{n} \tag{1}$$

je v primeru, da je hipoteza $H_0 : \mu_y = a$ pravilna, porazdeljena približno standardizirano normalno, $Z \sim N(0, 1)$. Pri dani stopnji značilnosti zavzame cenilka $\overline{Y}$ določeno vrednost izven intervala zaupanja, ki ustreza temu α, natanko takrat, ko slučajna spremenljivka Z zavzame neko vrednost z, ki po absolutni vrednosti presega kritično vrednost z_α. Tako imamo čisto konkretno testno pravilo:

Ko preverjamo domnevo, da je $\mu_y = a$, iz vzorčnih podatkov izračunamo vrednost cenilke $\overline{Y}$ in nato po (1) še pripadajočo vrednost slučajne spremenljivke Z. Če je $|z| > z_\alpha$ pri dani vrednosti α, rečemo, da so **razlike značilne na stopnji** α in **hipotezo zavrnemo**; v nasprotnem primeru razlike na tej stopnji niso značilne in hipoteze ne zavrnemo.



Slika 5.3: Kritično območje dvostranskega preizkusa

Če določeno domnevo opisane vrste zavrnemo na stopnji značilnosti 0·05, to pomeni, da je za izračunano vrednost (1) veljalo $|z| > 1{\cdot}96$. Ker pri tem $|z|$ ni nujno večja tudi od kritične vrednosti $z_{0{\cdot}01} = 2{\cdot}58$, se seveda lahko zgodi, da hipotezo zavrnemo na stopnji značilnosti 0·05, za stopnjo 0·01 pa moramo reči, da razlike niso značilne in na tej stopnji hipoteze ne moremo zavrniti.

ZGLED 5.3.4:

V trgovini v povprečju prodajo posameznemu kupcu za 2 300 SIT blaga, standardni odklon pa znaša pri tem 700 SIT. V tednu pred novim letom so za vzorec 150 kupcev ugotovili, da so v povprečju plačali po 2 545 SIT. Preizkusi domnevo, da novoletni prazniki značilno vplivajo na velikost zneska, ki ga posamezen kupec potroši za nakup!

Domnevo, da se je zaradi praznikov velikost nakupov povečala, bomo preizkusili tako, da bomo preverili ničelno domnevo, da je velikost ostala nespremenjena, torej $H_0 : \mu_y = 2\,300$. Pri tem bomo predpostavili, da se variabilnost proučevane spremenljivke ni spremenila, zato bomo računali z naslednjimi vrednostmi:

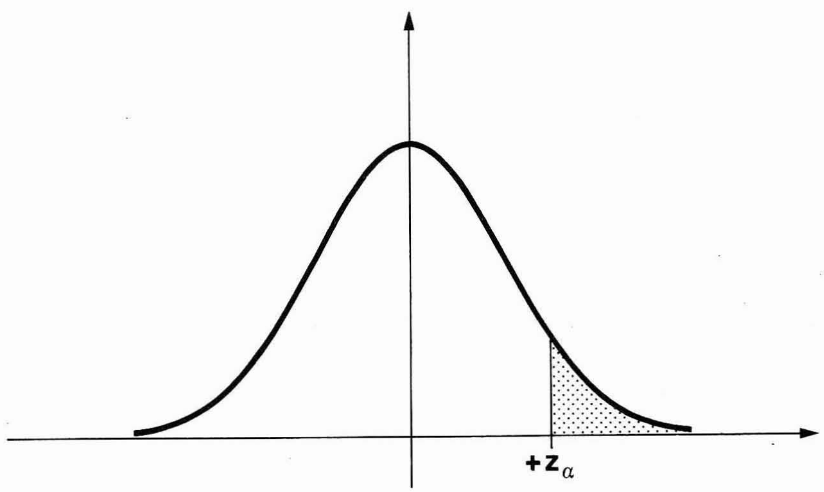
$$\bar{y} = 2\,545, \quad a = 2\,300\,(= \mu_y), \quad \sigma = 700, \quad n = 150$$

Pri teh podatkih dobimo po obrazcu (1)

$$z = \frac{2\,545 - 2\,300}{700} \cdot \sqrt{150} \approx 4{\cdot}29$$

Ta vrednost je večja tudi od kritične vrednosti $z_{0\cdot 001} = 3\cdot 29$, zato moramo hipotezo $H_0 : \mu_y = 2\,300$ zavrniti. Tveganje, da smo pri tem napravili napako I. vrste, je manjše od 0·001, torej zanemarljivo. Zato sklepamo, da novoletni prazniki značilno vplivajo na velikost posameznih nakupov. (Kdo bi si bil mislil!)

Pri dosedanjem delu smo hipotezo $H_0 : \mu_y = a$ zavrnili, če je z obrazcem (1) izračunana vrednost z po absolutni vrednosti presegala kritično vrednost z_α; kritično območje je zato ležalo na obeh straneh ustreznega intervala zaupanja. V takih primerih govorimo o dvostranskih preizkusih oziroma **dvostranskih testih**. Za razliko od njih poznamo **enostranske teste**, pri katerih so za zavrnitev hipoteze $H_0 : \mu_y = a$ ugodne samo vrednosti z, ki so ali manjše od $-z_\alpha$ (če je alternativna hipoteza $H_1 : \mu_y < a$), ali pa samo vrednosti z, ki so večje od z_α (če za alternativno hipotezo vzamemo $H_1 : \mu_y > a$). Za enostranski test se odločamo v primerih, ko je iz vsebine problema jasno, da lahko aritmetična sredina statistične spremenljivke odstopa od predpostavljene (hipotetične) vrednosti samo v eno smer.



Slika 5.4: Kritično območje enostranskega preizkusa

Pri enostranskih preizkusih so kritične vrednosti (pri dani stopnji značilnosti) drugačne kot pri ustreznem dvostranskem testu; če smo prej dobili iz zahteve

$$P(|Y| \le z_{0\cdot 05}) = P(-z_{0\cdot 05} \le Z \le z_{0\cdot 05}) = 2\,\Phi(z_{0\cdot 05}) = 0\cdot 95$$

kritično vrednost $z_{0\cdot 05} = 1\cdot 96$, imamo pri isti stopnji značilnosti za kritično vrednost zahtevo

$$P(Z \ge -z'_{0\cdot 05}) = 0\cdot 95$$

torej je

$$\Phi(z'_{0\cdot 05}) = 0\cdot 45 \Longrightarrow z'_{0\cdot 05} = 1\cdot 645$$

S črtico smo označili, da gre za enostranski preizkus, da bi ločili tudi po simbolu od kritične vrednosti $z_{0\cdot 05} = 1\cdot 96$ za dvostranski preizkus. Za podrobnosti o enostranskih preizkusih bo bralec moral poseči po zahtevnejši literaturi.

Z velikimi vzorci se v praksi dostikrat lotevamo problemov, pri katerih v bistvu ne gre za normalno porazdeljene slučajne spremenljivke, vendar pa njihovi porazdelitveni zakoni z naraščanjem vzorca postajajo podobni normalni porazdelitvi. Med najbolj zanimive sodi binomska porazdelitev, ki jo že dobro poznamo. Kako si pomagamo s pridobljenim znanjem o preizkušanju domnev v zvezi z aritmetično sredino normalne porazdelitve, bomo videli iz naslednjega primera.

ZGLED 5.3.5:

V skladu z Mendelovimi genetskimi zakoni bi moralo biti pri križanju dveh vrst graha v naslednji generaciji razmerje med rumenimi in zelenimi zrni 3 : 1. V slučajnem vzorcu, izbranem iz te generacije, smo našteli 176 rumenih in 48 zelenih zrn. Presodi, ali se vzorčni podatki ujemajo s teoretičnimi napovedmi!

Problem lahko opišemo kot preizkus domneve $H_0 : p = 0\cdot 75$, pri čemer p pomeni verjetnost, da je na slepo izbrano zrno iz proučevane generacije rumeno. Statistiki bi rekli še drugače, namreč, da preverjajo, ali je strukturni delež rumenega graha (v celotni populaciji) res 75%.

Iz verjetnostnega računa vemo, da je frekvenca dogodka, da je izbrano zrno rumeno, porazdeljena po binomskem zakonu $b(n,p)$. To velja natačno sicer samo za vzorce z vračanjem enot, vendar je pri tako veliki populaciji (grahovih zrn) razlika zanemarljiva. Pri velikem n lahko to porazdelitev aproksimiramo z normalno porazdelitvijo, ki ima enako matematično upanje $E(Y) = np$ in enako disperzijo $D(Y) = npq = np(1-p)$; pri tem Y pomeni frekvenco dogodka. Standardizirana slučajna spremenljivka

$$Z = \frac{Y - np}{\sqrt{np\,(1-p)}} \qquad (2)$$

je potem porazdeljena standardizirano normalno, če je le hipoteza, ki jo preizkušamo, pravilna. Vrednosti slučajne spremenljivke Z, ki močno odstopajo od njenega matematičnega upanja $E(Z) = 0$, so malo verjetne in nas navajajo na sklep, da preizkušana domneva ni pravilna. Položaj je torej natanko tak, kot ga poznamo od prej: iz vzorčnih podatkov bomo izračunali pripadajočo vrednost spremenljivke Z in hipotezo H_0 zavrnili, če bo dobljeno število po absolutni vrednosti presegalo kritično vrednost, ki ustreza določeni stopnji značilnosti. Pri nas je frekvenca dogodka enaka 176, velikost vzorca $n = 176 + 48 = 224$ in $p = 0\cdot 75$. Tako dobimo za vrednost cenilke Z

$$z = \frac{176 - 224 \cdot 0{\cdot}75}{\sqrt{224 \cdot 0{\cdot}75 \cdot 0{\cdot}25}} \approx 1{\cdot}23$$

Ker je ta vrednost manjša od $z_{0{\cdot}05} = 1{\cdot}96$, razlike med hipotetično ($n \cdot p = 168$) in dejansko frekvenco (176) niso značilne, hipoteze $H_0 : \ p = 0{\cdot}75$ ne bomo zavrnili.

ZGLED 5.3.6:

Pri 180 metih igralne kocke je 35-krat padla šestica. Preizkusi domnevo, da verjetnost za padec šestice ni $\frac{1}{6}$!

Problema se bomo spet lotili posredno, tako da bomo preizkusili domnevo $H_0 : \ p = \frac{1}{6}$. Ker je v skladu z (2)

$$z = \frac{35 - 180 \cdot \frac{1}{6}}{\sqrt{180 \cdot \frac{1}{6} \cdot \frac{5}{6}}} = 1$$

hipoteze H_0 ne moremo zavrniti, razlika med hipotetično frekvenco $np = 180 \cdot \frac{1}{6} = 30$ in eksperimentalno frekvenco 35 ni značilna. Ob tem velja opozoriti, da takšna ugotovitev ne pomeni, da trdimo, da je verjetnost za padec šestice natanko $\frac{1}{6}$, le vzorčni podatki nam ne dajejo možnosti, da bi tako domnevo zavrnili s primerno majhnim tveganjem.

Toliko naj zadošča za uvod v preizkušanje domnev. Bralec bo pri nadaljevanju svoje izobraževalne poti z veliko verjetnostjo prišel v položaj, ko se bo treba tega dela matematične statistike lotiti bolj zares. Upamo, da mu je učbenik pomagal vsaj odškrniti vrata na to zanimivo področje...

V A J E

1. Statistična spremenljivka Y je porazdeljena normalno, s standardnim odklonom $\sigma = 16$ in neznano aritmetično sredino. Za slučajen vzorec velikosti $n = 100$ smo dobili $\bar{y} = 82$. Preizkusi domnevo, da je aritmetična sredina enaka 86.

2. Dolgoletne izkušnje kažejo, da je povprečen rezultat sprejemnega izpita na neki šoli 64 točk, standardni odklon znaša 8 točk. Na sprejemnem izpitu je skupina 55 kandidatov iz Ljubljane dosegla v povprečju 68 točk. Ali lahko sklepamo, da so razlike med njihovim znanjem in povprečnim znanjem kandidatov za vpis na to šolo značilne? (Predpostavili bomo, da gre za slučajen vzorec.)

3. Za slučajen vzorec stotih žarnic nekega proizvajalca so ugotovili povprečno življenjsko dobo 1 570 ur, standardni odklon je bil ocenjen na 120 ur. Preizkusi domnevo $H_0 : \ \mu_y = 1\,600$ pri stopnjah značilnosti 0·05 in 0·01.

4. Vrvi nekega proizvajalca prenesejo v povprečju obremenitev 18 000 N, $\sigma = 1\,000$ N. Proizvajalec uvaja novo tehnologijo, ki naj bi povečala natezno trdnost vrvi. Za slučajen vzorec 50 vrvi so ugotovili, da so v povprečju zdržale obremenitev 18 500 N. Presodi, ali se je kakovost vrvi res spremenila.

5. Za določeno generacijo učencev je bilo ugotovljeno, da so pri preizkusu znanja iz tujega jezika v povprečju dosegli 74·5% vseh možnih točk, pri čemer je bil standardni odklon spremenljivke 8·0%. 200 učencev neke šole pa je v povprečju doseglo 75·9% vseh možnih točk. Ugotovi, ali gre za značilne razlike. (Za vse tri običajne stopnje značilnosti).

6. Z vzorcem 50 vojakov so želeli ugotoviti povprečno maso vojakov v garnizonu. Na vzorcu so bile ugotovljene naslednje mase (v kg): 59, 58, 65, 58, 69, 64, 61, 62, 66, 80, 75, 78, 61, 67, 45, 67, 51, 76, 73, 79, 56, 71, 72, 64, 51, 66, 54, 60, 69, 59, 72, 68, 68, 54, 66, 57, 58, 75, 60, 70, 79, 65, 53, 67, 60, 80, 73, 56, 74, 65.

   Pri stopnji značilnosti 0·05 preizkusi domnevo, da je aritmetična sredina mas v celotni populaciji enaka 66·5 kg. (Predpostavljamo normalno porazdelitev spremenljivke, populacija pa je dovolj velika, da nam ni treba upoštevati popravnega faktorja za vzorce brez vračanja enot.)

7. V posodi imamo po velikosti enake kroglice, od njih so nekatere bele, druge črne. Radi bi preizkusili domnevo, da je obojih enako. V ta namen izberemo slučajen vzorec $n = 64$ kroglic, z vračanjem elementov. Testno pravilo postavimo takole: če bo v vzorcu od 28 do 36 belih kroglic (vključno), bomo domnevo, da je obojih kroglic enako, sprejeli; če bo frekvenca belih kroglic manjša od 28 ali večja od 36, pa bomo to hipotezo zavrnili. Oceni, kolikšna je verjetnost za napako I. vrste.

8. Pri stotih metih dveh igralnih kock je bila triindvajsetkrat vsota pik enaka sedem. Na stopnji značilnosti 0·05 presodi o domnevi, da sta igralni kocki pošteni, da je torej porazdelitev števila pik na posamezni kocki enakomerna. Oblikuj pravilo za odločanje, pri katerih frekvencah opisanega dogodka ("vsota pik na obeh kockah je 7") v 100 ponovitvah poskusa moramo pri 5% tveganju posumiti v poštenost kock.

9. Dolga leta je bilo med pozitivnimi izpitnimi ocenami približno 10% odličnih. Lansko leto pa jih je med 300 študenti, ki so jih naključno izbrali izmed vseh študentov, ki so pozitivno opravili izpit, kar 40 dobilo odlično oceno. Pri stopnjah značilnosti 0·05 in 0·01 presodi, ali je rezultat značilno drugačen od dolgoletnega povprečja.

10. Sestavi računalniški program, ki bo iz podatkov o vrednostih spremenljivke, ki so zapisani v vektorju $Y(I)$, $I = 1,\ 2,\ \ldots,\ N$, izbral v slučajen vzorec predpisani delež populacije
- z vračanjem enot;
- brez vračanja enot;

nato pa izračunal vrednosti statistik $\overline{Y}$ in S^2 na tem vzorcu in pri poljubni stopnji značilnosti (vhodni podatek) preizkusil ničelno domnevo $H_0 : \mu_y = a$, pri čemer naj ima uporabnik možnost preizkusa za različne vrednosti parametra a.

11. △ △ Če si matematično - računalniški navdušenec, poskusi napisati še program, ki naj generira

a) izbrano število vrednosti spremenljivke, ki bo normalno porazdeljena

b) z matematičnim upanjem in standardnim odklonom, ki ju določi uporabnik programa.

Med primerne lahko štejemo vsaj dve poti:

a) preko binomske porazdelitve, tako da simuliramo - na primer - frekvenco grba v dovolj velikem številu metov kovanca in dobljeno porazdelitev ustrezno prilagodimo, torej standardiziramo na $N(0,1)$ in nato spet raztegnemo in premaknemo na $N(a,\sigma)$;

b) z neposredno uporabo generatorjev slučajnih števil, ki jih imaš na voljo v različnih programskih orodjih (funkcije RND ipd.).

V drugem primeru bo najbrž koristen naslednji recept, ki - na osnovi centralnega limitnega izreka - pove, kako iz enakomerno porazdeljenih vrednosti (RND) narediš standardizirano normalno porazdelitev:

Naj bodo $Y_1, Y_2, \ldots, Y_n$ neodvisne slučajne spremenljivke, porazdeljene enakomerno na intervalu (0,1). Potem je slučajna spremenljivka

$$Z = \frac{\left(\sum_{k=1}^{n} Y_k\right) - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$$

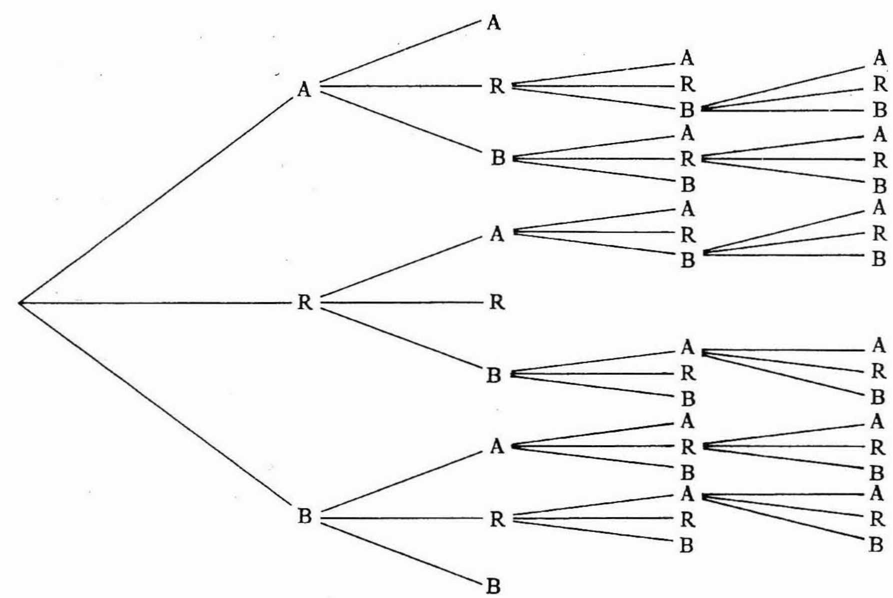
pri dovolj velikem n porazdeljena približno standardizirano normalno.

6. REŠITVE

Kakorkoli se že človek trudi, je verjetnost, da je vsaj ena rešitev napačna, vedno večja od verjetnosti nasprotnega dogodka, po Murphyju pa gre sploh za gotov dogodek. Zato avtor kakopak ne obljublja cekinov...

2.1 Osnovni kombinatorični prijemi (str. 7, str. 12)

1. 192 **2.** a) $2\cdot 2\cdot 4\cdot 3\cdot 3\cdot 4 = 576$ b) 288 c) 144
3. 64 **4.** 625 **5.** 16 **7.** 42 **8.** a) 25 b) 20
9. $25\cdot 24 = 600$ **10.** 900 **11.** 15 600 **12.** 33 (slika 6.1)

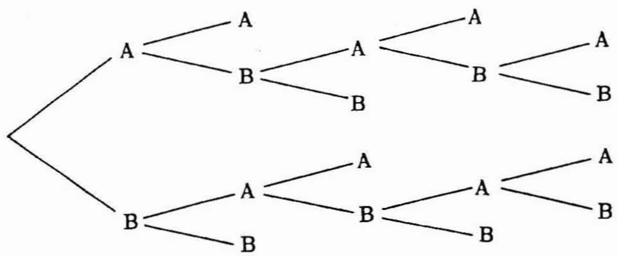


Slika 6.1: Potek šahovskega dvoboja

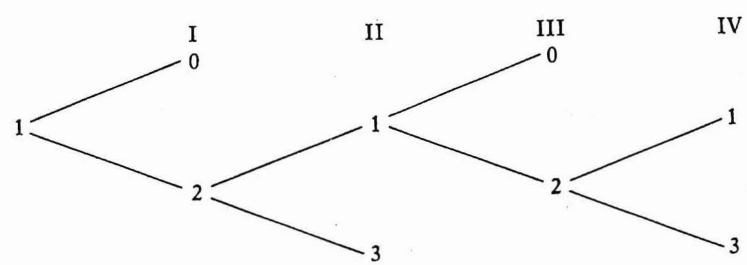
13. A - Andrej, B - Bojan, 10 načinov (slika 6.2). **14.** Glej sliko 6.3!
15. $2+2\cdot 2+2\cdot 2\cdot 2+2\cdot 2\cdot 2\cdot 2 = 30$
Kombinatorično drevo prikazuje, kako nastanejo Morsejevi znaki, ki so sestavljeni iz treh osnovnih znakov, torej pik ali črtic (slika 6.4).

16. glej sliko 6.5!

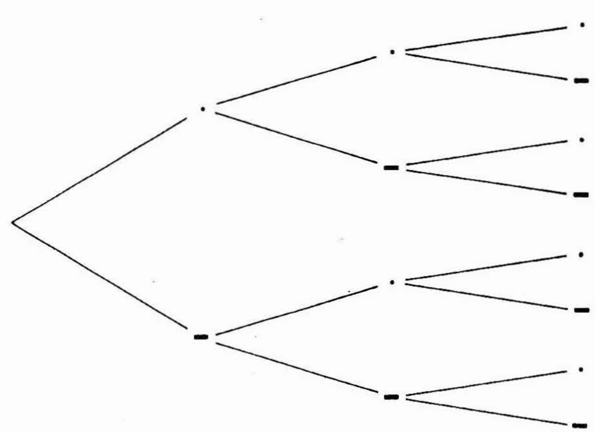
17. AMA, AMAI, AMAN, AMANIJA, AMENONA, AMIN, AMONIT, ANA, ANAM, ANEMONA, ANIMA, ANONIMEN



Slika 6.2: Potek teniškega dvoboja



Slika 6.3: Možni poteki stav na avtomatu



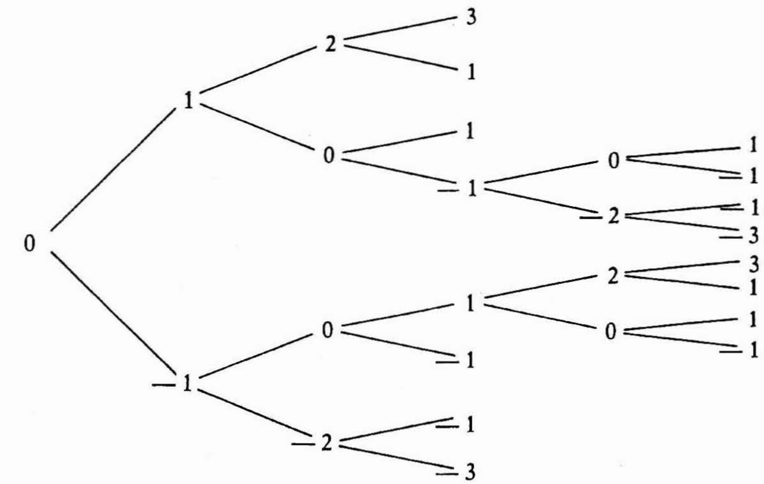
Slika 6.4: Morsejevi znaki

2.2 Permutacije (str. 14. str. 20)

1. abc, acb, bac, bca, cab, cba

2. Primer: $P:\ (a\,b\,c) \mapsto (b\,c\,a)$

3. 1234, 1324, 2134, 2314, 3124, 3214



Slika 6.5: Gibanje po premici

4. $10! = 3\,628\,800$ **5.** $9! = 362\,880$, VSONLJIEA

6. $6! = 720$ **7.** 120

8. Pot lahko začnemo v katerikoli točki (8 možnosti), nadaljujemo v katerokoli od preostalih točk (7 možnosti),..., kar nam da po osnovnem izreku kombinatorike $8 \cdot 7 \cdot \ldots \cdot 2 \cdot 1 = 8! = 40320$ usmerjenih poti; če smeri ne razlikujemo (kar pomeni, da na primer razporedov 12345678 in 87654321 med seboj ne razlikujemo), je takih poti dvakrat manj, to je 20160.

9. $6! = 720$ urnikov; $720 \cdot 15 = 10\,800$ sekund ali tri ure.

10. Eden od načinov računanja: od števila vseh razporedb petih elementov odštejemo tistih $4!$, pri katerih je dekle na začetku vrste, in tistih $4!$, pri katerih stoji na koncu vrste; tako imamo $5! - 2 \cdot 4! = 72$ razporedb, ki ustrezajo pogojem.

11. Z govornikom A na prvem mestu imamo 6 ugodnih vrstnih redov; če je A na drugem mestu, imamo dve možnosti za zasedbo prvega mesta (C ali D), tretje in četrto mesto pa lahko zasedemo na $2! = 2$ načina, po osnovnem izreku kombinatorike je torej ugodnih vrstnih redov z A na drugem mestu $2 \cdot 2! = 4$; če je A na tretjem mestu, mora biti B nujno na četrtem, vrstna reda sta dva (permutiramo C in D na prvih dveh mestih). Skupno imamo torej 12 dovoljenih razporedov. Rezultat je pričakovan: v vseh permutacijah je A natanko tolikokrat pred B, kolikorkrat je B pred A, dovoljenih je polovica od skupaj 24 permutacij.

12. **a)** 24 **b)** 72 **c)** $5! - 3! = 114$

13. **a)** $5! = 120$ **b)** $4 \cdot 4! = 96$ ali $5! - 4! = 96$ **c)** $4! = 24$

14. a) $4! \cdot (5! \cdot 4! \cdot 3! \cdot 3!) = 2\,488\,320$
b) $15! = 1\,307\,674\,368\,000 \approx 1{\cdot}308 \cdot 10^{12}$
c) $10! \cdot 11 \cdot 5! = 4\,790\,016\,000$

15. a) $30\,240$ b) $7/8$ c) $(1 - n/(n^2 + 2n)$
16. a) $27!/25!$ b) $7!/10!$ c) $8!/6!$ č) $(n+3)!/(n-4)!$
17. $151\,200$ **18.** $7!/(3!2!2!) = 210$ **19.** $10!/(6!4!) = 210$
20. Treba je ugotoviti, kdaj je vrednost $n!/(k!(n-k)!)$ maksimalna. Ker števec ni odvisen od vrednosti k, se bo to zgodilo natanko takrat, ko bo imenovalec tega ulomka minimalen. Zaradi simetrije ni omejitev splošnosti, če predpostavimo, da je $k \leq n - k$. Če k zmanjšamo za 1, smo produkt $k!$ zamenjali s $(k-1)!$, k produktu $(n-k)!$ pa pridobili še en faktor $n - k + 1$. Povečanje imenovalca, ki pri tem nastane, je še najmanjše, kadar sta si "izgubljeni" k in "pridobljeni" $n - k + 1$ zelo blizu, to pa je takrat, ko je $k = n - k = n/2$ (pri sodem n) oziroma $k = (n-1)/2$ pri lihi vrednosti n. Za vajo ponazori opisano razmišljanje pri $n = 10$ in $n = 11$!
21. $5!/(2!3!) + 5!/(2!3!) = 20$ **22.** 60 **23.** $15!/(5!5!5!) = 756\,756$
24. $(3n)!/(n!n!n!)$ **25.** $10!/(5!3!2!) = 2520$
26. $(k+m+n)!/(k!m!n!)$ **27.** a) 12 b) 6 c) 6
28. $10!/(3!3!4!) = 4200$ **29.** $\approx 16{\cdot}7\%$

2.3 Variacije (str. 23, str. 25)

1. a) AB, AC, AD, ..., DB, DC; 12
b) AA, AB, AC, ..., DB, DC, DD; 16
2. 210, 5040, 360 **3.** 343; 2187
4. a) 5 b) 16 c) 8 d) 2 **5.** 60, 125
6. $3 \cdot 4 \cdot 3 = 36$, $3 \cdot 5 \cdot 5 = 75$ **7.** $90\,000$
8. $4^r = 4096 \Rightarrow r = 6$ **9.** $13^4 = 28\,561$
10. a) 720 b) 1000 **11.** a) $32^4 = 1\,048\,576$ b) $32 \cdot 31 \cdot 30 \cdot 29 = 863\,040$
12. a) $n \cdot (n-1) \cdot (n-2)$ b) $n \cdot n \cdot n = n^3$
13. 48, 1249 **14.** 24, 64 **15.** $151\,200$
16. $3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 243$ **17.** Glej zglede v besedilu razdelka!

2.4 Kombinacije (str. 27, str. 35)

1. AB, AC, AČ, AD, BC, BČ, BD, CČ, CD, ČD;
AA, AB, AC, AČ, AD, BB, BC, BČ, BD, CC, CČ, CD, ČČ, ČD, DD

2. ABC, ABČ, ABD, ACČ, ACD, AČD, BCČ, BCD, BČD, CČD
3. a) 10 b) 35 c) 35 č) 4950 d) $1\,192\,052\,400$
4. a) $\binom{n}{n-2} = \binom{n}{2} = n(n-1)/2$ b) $(n+5)(n+4)/2$
c) 35 č) n/k d) $\binom{n}{k}$
5. a) $n = 3$ b) $n = 2$ c) $n = 2$
6. $5\,006\,386$ **7.** 252 **8.** 120 35 **9.** 10; 6
10. $\binom{10}{4} - \binom{9}{3} = 126$ **11.** 210 **12.** $\binom{n}{2}$
13. Ločimo dve možnosti, $m = n - 1$ in $m < n - 1$ v prvem primeru je (skica!) premic $1 + (m-1) = m = n - 1$, v drugem primeru pa imamo premic $1 + m \cdot (n-m) + \binom{n-m}{2}$.
14. a) $\binom{6}{2} = 15$ b) $\binom{6}{3} = 20$ c) $\binom{5}{2} = 10$ d) 4
15. $\binom{n}{3} = n \Rightarrow n = 4$ **16.** $\binom{10}{4} - \binom{2}{2}\binom{8}{2} = 182$
17. 0000, 0001, 0011, 0111, 1111
18. a) $\binom{6+4-1}{4} = \binom{9}{4} = 126$ b) $\binom{9}{6} = 84$; c) $\binom{16}{5} = 4368$ č) $\binom{11}{3} = 165$
19. $\binom{n+2-1}{2} = 276 \Rightarrow n = 23$
20. $\binom{10+4-1}{4} = \binom{13}{4} = 715$ (všteto je število 0000!)
21. $\binom{8+12-1}{12} = 50\,388$ **22.** $\binom{10+3-1}{3} = 220$
23. a) $\binom{12+4-1}{4} = 1\,365$ b) $\binom{12+0-1}{0} + \binom{12+1-1}{1} + \ldots \binom{12+4-1}{4} = 1\,820$
24. a) $\binom{12}{4}\binom{7}{3} = 17\,325$ b) 784 c) $1\,200$
25. $\binom{15}{2}\binom{12}{1} = 1\,260$
26. 350 **27.** 180 **28.** 252 13923
29. Izborov s predpisanim deležem samoglasnikov je $\binom{5}{2}\binom{20}{3} = 11\,400$, vsako tako besedo s petimi različnimi črkami pa lahko še permutiramo na $5! = 120$ načinov, zato je vseh besed opisane vrste $11\,400 \cdot 120 = 1\,368\,800$, kar je - razumljivo - znatno manj kot vseh besed s petimi različnimi črkami, ki jih je $25 \cdot 24 \cdot 23 \cdot 22 \cdot 21 = 6\,375\,600$.
30. $\binom{5}{2}\binom{19}{3} \cdot 120 = 205\,200$ **31.** $\binom{19}{4} = 3876$
32. a) $\binom{5}{0}\binom{7}{3} + \binom{5}{1}\binom{7}{2} + \binom{5}{2}\binom{7}{1} + \binom{5}{3}\binom{7}{0} = 220$ b) 4
33. $\binom{5}{1}\binom{20}{2} + \binom{5}{2}\binom{20}{1} + \binom{5}{3}\binom{20}{0} = 1\,160$ ali $\binom{25}{3} - \binom{20}{3} = 1\,160$
34. a) 60 b) 30 c) 42 **35.** a) 91 b) 286 c) 90

2.5 Binomski izrek (str. 38, str. 40)

1. a) $x^5 + 5x^4 + 10x^3 + 10x^2 + 5x + 1$
b) $y^6 - 12y^5 + 60y^4 - 160y^3 + 240y^2 - 192y + 64$
c) $x^4 + 12x^3y + 54x^2y^2 + 108xy^3 + 81y^4$

č) $32x^5 + 80x^4y^2 + 80x^3y^4 + 40x^2y^6 + 10xy^8 + y^{10}$

d) $16x^4 - 32x^3y + 24x^2y^2 - 8xy^3 + y^4$

e) $a^8 - 8a^7\sqrt[3]{b} + 28a^6\sqrt[3]{b^2} - 56a^5b + 70a^4b\sqrt[3]{b} - 56a^3b\sqrt[3]{b^2} +$
$+28a^2b^2 - 8ab^2\sqrt[3]{b} + b^2\sqrt[3]{b^2}$

2. a) $\binom{15}{9}(x^2)^6(-2y^3)^9 = -2\,562\,560\,x^{12}y^{27}$

b) $\binom{8}{5}(x^{1/2})^3(-2x)^5 = -1792\,x^6\sqrt{x}$

3. $-1792\,x^{10}y^9$ **4.** $945\,x^3y^6z^8$ **5.** $x^{20} - 20x^{17} + 180x^{14}$

6. $x^{12} + 24x^9 + 264x^6$ **7.** $(1 - 0{\cdot}006)^4 \approx 0{\cdot}976\,216 \approx 0{\cdot}976$ 64·578

8. $\binom{n}{5} = \binom{n}{12} \Rightarrow n = 17$ **9.** $2^n - 1 = 2^5 - 1 = 31$

10. $2^{10} - \binom{10}{0} - \binom{10}{1} - \binom{10}{2} = 968$. Do enakega rezultata lahko pridemo tudi z neposrednim (bistveno napornejšim) računom.

11. $0 = 0^n = (1-1)^n = \binom{n}{0} - \binom{n}{1} + ... \pm \binom{n}{n}$

2.6 Načelo vključitev in izključitev (str. 41, str. 45)

1. 35% **2.** $m(A \cup B \cup C) = 733 \Rightarrow 999 - 733 = 266$

3. $8! = 40\,320$; $40\,320 - 25\,487 = 14\,833$

4. Nasvet: od ploščine celotnega enakostraničnega trikotnika, ki ga določajo središča krogov, odštejemo vsoto ploščin treh krožnih izsekov; ta je enaka polovici ploščine kroga, zato je končni rezultat
$p = (2R)^2\sqrt{3}/4 - (R^2\pi/2) \approx 0{\cdot}16\,R^2$

3.1 Osnovni statistični pojmi (str. 47, str. 50)

1. Statistična populacija je množica (najvišjih) vrhov v Republiki Sloveniji. Statistične enote so posamezni vrhovi, na primer Triglav, Škrlatica, Veliki Mangart,..., za njih pa ugotavljamo vrednosti treh spremenljivk: a) gorska skupina, v kateri leži vrh; b) pogorje, v katerem je vrh; c) njegova nadmorska višina. Zadnja spremenljivka je številska (numerična), načelno lahko zavzame poljubno pozitivno realno vrednost, ki ne presega višine Triglava, v bistvu pa že pri opredelitvi populacije določimo, kateri je najnižji vrh, ki ga še uvrstimo vanjo (na primer: vsi vsaj 1000 m visoki vrhovi), po drugi strani pa višine običajno zaokrožimo na meter natančno in je zato na prvi pogled zaloga vrednosti te spremenljivke podmnožica v množici celih števil; kljub temu ne smemo pozabiti, da gre za zvezno spremenljivko.

2. Populacijo sestavljajo vse občine, ki so obstajale v Republiki Sloveniji v trenutku popisa. Za razliko od ostalih dveh spremenljivk je število prebivalcev značilna celoštevilska, torej nezvezna spremenljivka.

3. "Zakonsko stanje" je atributivna spremenljivka, osnovni možni vrednosti sta "poročen" in "neporočen", lahko pa pripravimo bogatejšo lestvico, v kateri podmnožico trenutno neporočenih razčlenimo po vzrokih.

4. Razmišljaj o strukturi lastnikov telefonov in strukturi celotne populacije državljanov Slovenije.

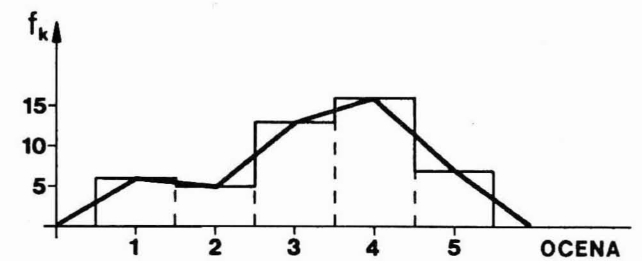
3.2 Urejanje in prikazovanje podatkov (str. 52, str. 60)

1. Porazdelitev ocen

Ocena	f_k
1	6
2	5
3	13
4	16
5	7
Skupaj	47

2. Porazdelitev samoglasnikov

x_k	f_k	f_k^o
a	11	21·57
e	18	35·29
i	9	17·65
o	9	17·65
u	4	7·84
S	51	100·00

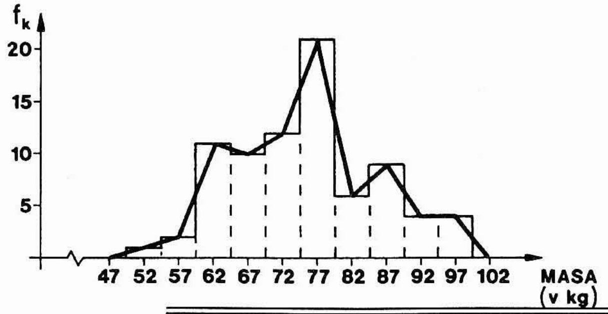


Slika 6.6: Ocene - poligon in histogram

3. Frekvenčna porazdelitev mas

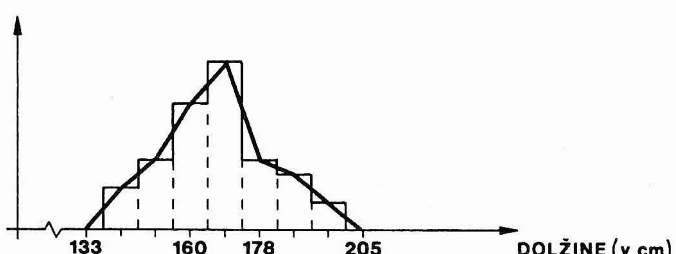
R	Masa (kg)	f_k	f_k^o	F_k	F_k^o
1	50 - 54	1	0·0125	0	0·0000
2	55 - 59	2	0·0255	1	0·0125
3	60 - 64	11	0·1375	3	0·0375
4	65 - 69	10	0·1250	14	0·1750
5	70 - 74	12	0·1500	24	0·3000
6	75 - 79	21	0·2625	36	0·4500
7	80 - 84	6	0·0750	57	0·7125
8	85 - 89	9	0·1125	63	0·7875
9	90 - 94	4	0·0500	72	0·9000
10	95 - 99	4	0·0500	76	0·9500
	Skupaj	80	1·0000	80	1·0000

Slika 6.7: Poligon in histogram za porazdelitev mas

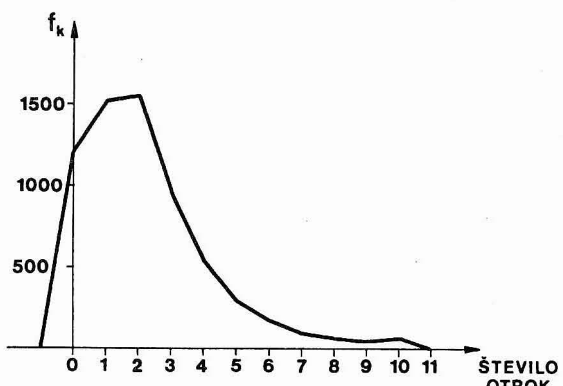


4. Frekvenčna porazdelitev dolžin ostankov (tabela in slika 6.8). Daljših od 155 cm je 80% ostankov.

R	Dolžine (cm)	f_k	f_k^o	F_k	F_k^o
1	138 - 146	3	0'0750	0	0'0000
2	147 - 155	5	0'1250	3	0'0750
3	156 - 164	9	0'2250	8	0'2000
4	165 - 173	12	0'3000	17	0'4250
5	174 - 182	5	0'1250	29	0'7250
6	183 - 191	4	0'1000	34	0'8500
7	192 - 200	2	0'0500	38	0'9500
	Skupaj	40	1'0000	40	1'0000



Slika 6.8: Porazdelitev dolžin ostankov



Slika 6.9: Porazdelitev števila otrok

5. Porazdelitev števila otrok: vsaj enega otroka ima 81'28% družin, več kot dva otroka ima 34'63% in kvečjemu tri otroke premore 80'01% družin.

R	Otrok	f_k	f_k^o	F_k	F_k^o
1	0	1230	0'1872	0	0'0000
2	1	1520	0'2314	1230	0'1872
3	2	1545	0'2351	2750	0'4186
4	3	962	0'1464	4295	0'6537
5	4	537	0'0817	5257	0'8001
6	5	301	0'0458	5794	0'8819
7	6	174	0'0265	6095	0'9277
8	7	108	0'0164	6269	0'9542
9	8	69	0'0105	6377	0'9706
...	...	...	...	...	...
	Skupaj	6570	1'0000	6570	1'0000

6. Porazdelitev telesnih višin. Vsaj 180 cm meri 14 članov društva oziroma 11'67% vseh članov.

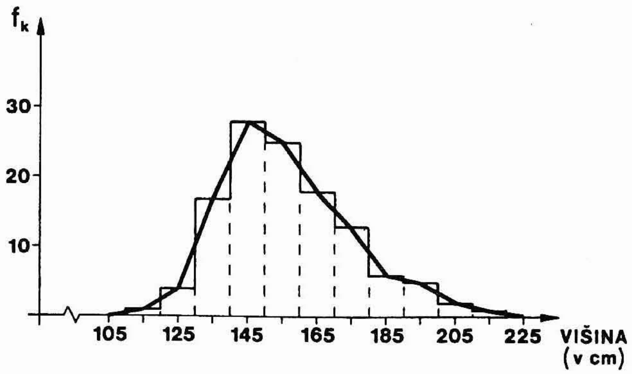
(Slika 6.10 na naslednji strani.)

R	Višine (cm)	f_k	f_k^o	F_k	F_k^o
1	110 - 119	1	0'0083	0	0'0000
2	120 - 129	4	0'0333	1	0'0083
3	130 - 139	17	0'1417	5	0'0416
4	140 - 149	28	0'2333	22	0'1833
5	150 - 159	25	0'2083	50	0'4166
6	160 - 169	18	0'1500	75	0'6250
7	170 - 179	13	0'1083	93	0'7750
8	180 - 189	6	0'0500	106	0'8833
9	190 - 199	5	0'0418	112	0'9333
10	200 - 209	2	0'0167	117	0'9751
11	210 - 219	1	0'0083	119	0'9918
	Skupaj	120	1'0000	120	1'0000

7. Nasvet: za risanje frekvenčnega kolača uporabi program za delo s preglednicami, ki ga običajno uporabljaš!

Cifra	0	1	2	3	4	5	6	7	8	9
f_k	0	3	5	7	9	11	13	15	17	19

8. Eden od možnih načinov je predstavljanje po načelu, naj bo frekvenci posameznega razreda premo sorazmerna ploščina pripadajočega pravokotnika v histogramu.



Slika 6.10: Porazdelitev telesnih višin (naloga 6)

3.3 Srednje vrednosti (str. 62, str. 69)

2. A: 9·625 G: 8·011 H: 6· 136

3. Natanko takrat, ko je $a = b$.

4. 7·0 **5.** 40· 21 d.e.

6. a) 166·80 cm b) 166·975 ≈ 167 cm

7. Ni izrazitega modusa, pred grupiranjem v razrede ima največjo frekvenco dolžina 155 cm (pojavlja se trikrat), osem različnih dolžin pa nastopa po dvakrat.

Mediana iz negrupiranih podatkov: 166·0 cm

Mediana iz grupiranih podatkov: 166·0 cm

8. Družina ima povprečno po 2·24 otroka. (Kaj to pomeni v praksi?)

9. $(1 + 2 + \ldots n)/n = [n(n+1)/2]/n = (n+1)/2$

10. Avtobus mora voziti najmanj 80 km/h.

3.4 Mere variabilnosti (str. 71, str. 74)

1. 4·560 ; 2·135 **2.** $\sigma = 1{\cdot}22$ $K_v = 0{\cdot}\ 37$

3. Variacijski razmik pri negrupiranih podatkih je 57 (cm), standardni odklon, izračunan iz negrupiranih podatkov, pa znaša 12·9 (cm) in je tako precej manjši od standardnega odklona, ki ga dobimo po grupiranju in je 13·7 cm. (Preglej še enkrat negrupirane podatke in način grupiranja ter poskusi ugotoviti, zakaj je prišlo do te razlike!

4. Povprečen čas za prevoz na delo je slabih 38 minut (37·38 minute), vendar je variabilnost časov relativno velika, $\sigma = 31{\cdot}18$ minute, koeficient variacije pa je 0·82, zato iz aritmetične sredine v bistvu zelo malo izvemo o porazdelitvi.

3.5 Indeksna števila (str. 75, str. 80)

2.

LETO	N	B='85	B='91	VER
1985	134	100·0	73·6	–
1986	152	113·4	83·5	113·4
1987	187	139·6	102·7	123·0
1988	171	127·6	94·0	91·4
1989	192	143·3	105·5	112·3
1990	198	147·8	108·8	103·1
1991	182	135·8	100·0	91·9

a) Število nesreč je bilo vsa leta večje kot leta 1985 (bazni indeksi vsi večji od 100).
b) Najhitrejša rast glede na preteklo leto (= največji verižni indeks) je bila v letu 1987.
c) Za približno 6·4%.

3. Število nesreč se je vsako leto - v povprečju - povečalo za približno 5·2%.

4. Povprečni koeficient dinamike je 1·07309465..., prodaja je naraščala (v povprečju) za približno 7·3% letno.

LETO	PR	B='80	B='86	VER
1980	245	100·0	49·1	–
1981	354	144·5	79·9	144·5
1982	453	184·9	90·8	128·0
1983	378	154·3	75·8	83·4
1984	424	173·1	85·0	112·2
1985	488	199·2	97·8	115·1
1986	499	203·7	100·0	102·3
1987	512	209·0	102·6	102·6
1988	476	194·3	95·4	93·0
1989	523	213·5	104·8	109·9
1990	555	226·5	111·2	106·1
1991	511	208·6	102·4	92·1
1992	572	233·5	114·6	111·9
1993	613	250·2	122·8	107·2

5. a) Manjkajoča podatka: 1988 – 476, 1990 – 504
Manjkajoča verižna indeksa: 1985 – 121· 2, 1992 – 111·9

b) Bazni indeksi z osnovo v letu 1980: 100·0, 140· 0, 159·4, 173·8, 195· 0, 236·3, 255·6, 320· 0, 297·5, 326·9, 315· 0, 319·4, 257·5, 387· 5 ;

c) Bazni indeksi z osnovo v letu 1986: 39·1, 54· 8, 62·3, 68·0, 76· 3, 92·4, 100·0, 125· 2, 116·4, 127·9, 123· 2, 124·9, 139·9, 151· 6;

č) Povprečni koeficient dinamike: 1· 1088488... ≈ 1·11.

4.1 Poskusi in dogodki (str. 83, str. 92)

1. $A \subset D, \quad C \subset B, \quad C \subset E$

2. $A \cap A = A, \quad A \cup A = A$

3. $C \subset A, \quad D = A, \quad E = B, \quad E = A', \quad G = D, \quad G = B' \ldots$

5. $B = A_1 \cap A_2' \cap A_3', \qquad C = A_1 \cap A_2 \cap A_3$

$D = (A_1 \cap A_2' \cap A_3') \cup (A_1' \cap A_2 \cap A_3') \cup (A_1' \cap A_2' \cap A_3) \ldots$

6. $C = [(A_1 \cap A_2') \cup (A_1' \cap A_2) \cup (A_1 \cap A_2)] \cap$

$\cap [(B_1 \cap B_2 \cap B_3') \cup (B_1 \cap B_2' \cap B_3) \cup (B_1' \cap B_2 \cap B_3)]$

7. A : 0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111

B : 0000, 1111

C : 0011, 0101, 0110, 1000, 1001, 1010, 1100

D : 0011, 0101, 0110, 1000, 1001, 1010,
1100, 0111, 1011, 1101, 1110, 1111

8. E_{ij} naj bo dogodek, da v prvem poskusu dobimo i- to, v drugem pa j-to kroglo, $(i,\ j = 1,\ 2,\ 3,\ 4)$.

$A = E_{11} \cup E_{13} \cup E_{31} \cup E_{33} \qquad B = E_{12} \cup E_{21} \cup E_{24} \cup E_{42}$

9. Uporabili bomo C za številko in M za grb ter izpuščali znak $\cap$ za produkt dogodkov:

$A = MMC \cup MCM \cup CMM \cup MMM$

$B = MMC \cup MCM \cup CMM \cup MCC \cup CMC \cup CCM \cup CCC$

$C = MMC \cup MCC \cup CMC \cup CCC$

$D = MMC \cup CMM \cup MMM \cup MCC \cup CCM \cup CCC$

10. a) 8 b) 11 c) 11 č) 14 d) 25 e) 0
f) 1 g) 1 h) 7 i) 4 j) 8 k) 11

11. $A = MM \cup RR$, $B = BM \cup BR$, $C = MR$; Da!

12. $(A \cup B)' = A' \cap B'$, zato tretji dogodek ni združljiv s prvim oziroma drugim (da sta ta dva nezdružljiva, je očitno). Da dokažemo, da je vsota teh treh dogodkov res gotov dogodek, v izrazu $A \cup (A' \cap B) \cup (A \cup B)'$ uporabimo distributivnostni zakon (enačba (5) na strani 90).

13. a) $B \subset A$, $C \subset A$ b) $A \subset (B \cap C)$

14. Če ne bo šlo, si pomagaj s sliko 4.3 na strani 111.

15. $A = MCMM \cup CMCMM \cup CMCMCC \cup \ldots$;

$B = CC \cup CMM \cup CMCC \cup CMCMM \cup \ldots; \ldots$

16. $\{M,\ CM,\ CCM,\ CCCM,\ CCCCM,\ CCCCC\}$

4.2 Verjetnost slučajnega dogodka (str. 95, str. 101)

4. $P(A) = 1/6 \quad P(B) = 1/2 \quad P(C) = 1/6$

$P(D) = 1/2 \quad P(E) = 2/3$

5. $P(E_i) = i \cdot p \implies p + 2p + \ldots + 6p = 1 \implies p = 1/21$

$P(A) = 6/21, \quad P(B) = 9/21, \quad P(C) = 3/21,$

$P(D) = 12/21, \quad P(E) = 10/21$

Primerjaj z verjetnostmi za pošteno kocko!

6. $P(A) = 1/4! \approx 0{\cdot}042$

7. $P(A) = 1/4^4 \approx 0{\cdot}004$

8. $P(A) = 1/(4!/2!) \approx 0{\cdot}083$

9. $P(A) = 1/3, \quad P(B) = 1/2, \quad P(C) = (30 + 45 - 15)/90 = 2/3.$

10. $P(A) = 100/5000 = 0{\cdot}02, \quad P(B) = \binom{100}{1}\binom{4900}{1}/\binom{5000}{2} \approx 0{\cdot}039$

11. $P(A) = 10/1140 \approx 0{\cdot}0088, \quad P(B) \approx 0{\cdot}189, \quad P(C) \approx 0{\cdot}399$

12. $P(A) = 0{\cdot}512, \quad P(B) = 0{\cdot}096$

13. $P(A) = 0{\cdot}5, \quad P(B) = 0{\cdot}2$

14. $m = 1,\ n = 6!/(3!2!) = 60 \implies P(A) = 1/60 \approx 0{\cdot}017$

15. $P(A) = 6/36 \approx 0{\cdot}167, \quad P(B) = 9/36 = 0{\cdot}25,$

$P(C) = 15/36 \approx 0{\cdot}417, \quad P(D) = 21/36 \approx 0{\cdot}583,$

$P(E) = 16/36 \approx 0{\cdot}444, \quad P(F) = 6/36 \approx 0{\cdot}167,$

$P(G) = 1/36 \approx 0{\cdot}028, \quad P(H) = 11/36 \approx 0{\cdot}306,$

$P(I) = 35/36 \approx 0{\cdot}972$

16. $P(A) = (6 \cdot 5 \cdot 4 \cdot 3 \cdot 2)/6^6 \approx 0{\cdot}093,$

$P(B) = (12!/(2! \cdot 2! \cdot 2! \cdot 2! \cdot 2! \cdot 2!))/6^{12} \approx 0{\cdot}0034$

17. $P(A) = (2 \cdot 4 \cdot 3)/(5 \cdot 4 \cdot 3) = 0{\cdot}4$

18. $P(A) = 12/40 = 0{\cdot}30, \quad P(B) = 9/40 = 0{\cdot}225, \quad P(C) = 0/40 = 0,$

$P(D) = 7/40 = 0{\cdot}175, \quad P(E) = 4/40 = 0{\cdot}10$

19. $P(A) \approx 0{\cdot}005$

20. $P(B_n) = n!/n^n$; te verjetnosti z naraščanjem n hitro padajo,

$P(B_2) = 0{\cdot}50,\ P(B_4) \approx 0{\cdot}094, \ldots\ P(B_{10}) \approx 0{\cdot}0036, \ldots$

4.3 Lastnosti in računanje verjetnosti (str. 104, str. 108)

3. $P(A) = \binom{4}{2}\binom{46}{2}/\binom{50}{4} \approx 0{\cdot}027$

4. $\binom{N}{q}\binom{M}{k-q}/\binom{M+N}{k}$ (Za katere vrednosti parametrov velja ta obrazec?)

5. $\binom{8}{5}\binom{2}{0}/\binom{10}{5} = 56/252 \approx 0{\cdot}22$

6. $P(A) = 1 - (5/6)(5/6) = 11/36 \approx 0{\cdot}31$

7. $P(A) = 1 - (5/6)^4 \approx 0{\cdot}518, \quad P(B) = 1 - (35/36)^{24} \approx 0{\cdot}491$

8. $P(A) = \binom{4}{3}\binom{28}{0}/\binom{32}{3} = 4/4960 \approx 0{\cdot}0008\,,$
$P(B) = \binom{8}{3}\binom{24}{0}/\binom{32}{3} = 56/4960 \approx 0{\cdot}011\,,$
$P(C) = \binom{8}{2}\binom{24}{1}/\binom{32}{3} = 672/4960 \approx 0{\cdot}135\,,$
$P(D) = \left(\binom{16}{3} + \binom{16}{3}\right)/\binom{32}{3} = 1120/4960 \approx 0{\cdot}226\,,$
$P(E) = 1 - \binom{28}{3}/\binom{32}{3} = 1684/4960 \approx 0{\cdot}0{\cdot}34$

9. $P(A) = \binom{3}{3}\binom{7}{6}/\binom{10}{9} = 7/10 = 0{\cdot}7$

10. Sočasno izbiranje je enakovredno zaporednemu izbiranju posameznih elementov brez vračanja le-teh v populacijo. Po podatkih iz naloge smo izmed $N+M$ že izbrali k izdelkov, tako da je v populaciji še $N+M-k$ izdelkov; ker so bili vsi dobri, je med njimi še $N-k$ dobrih. Kakšna je verjetnost, da bo tudi naslednji dober, najbrž ni težko izračunati.

11. $P(A\cup B\cup C) = P(A)+P(B)+P(C)-$
$P(A\cap B) - P(A\cap C) - P(B\cap C) + P(A\cap B\cap C)$

4.4 Pogojna verjetnost in verjetnost produkta (str. 110, str. 117)

1. $P(A/B) = P(A\cap B)/P(B) = P(N)/P(B) = 0 \neq P(A)$
2. $P(A/B) = 2/5\,,\quad P(B/A) = 2/3$
3. $P(A\cap B) = 1/6\,,\quad P(A\cup B) = 2/3\,,\quad P(A\cap B'\cap C') = 1/9$
4. $P(A) = 0{\cdot}50\,,\quad P(B) = 0{\cdot}375\,,\ldots$
5. $P(A) = (7\cdot6\cdot5\cdot4)/(10\cdot9\cdot8\cdot7) \approx 0{\cdot}17\,,$
$P(B) = (3\cdot2\cdot7\cdot6)/(10\cdot9\cdot8\cdot7) \approx 0{\cdot}09\,,$
6. $P(A) \approx 0{\cdot}025\,,\quad P(B) \approx 0{\cdot}002\,,$
7. $P(A/B) = P(A\cap B)/P(B) = (1/36)/(5/36) = 1/5 = 0{\cdot}2$
9. $P(A) = 4/36 + (32/36)(6/36) = 56/216 \approx 0{\cdot}259$
10. $P(A) = 1-(1-0{\cdot}08)^{50} \approx 0{\cdot}985$
11. $1-(1-p)^5 > 0{\cdot}99 \Longrightarrow p \geq 0{\cdot}602$
12. $P(A) = (1/2) + (1/4)(1/2) + (1/4)^2(1/2) + \ldots$
13. $P(A) = (1/6) + (25/36)(1/6) + (25/36)^2(1/6) + \ldots$
14. Glej tudi zgled 4.4.6!

	I. (B)	II. (B')	
1. tovarna (A)	0·36	0·24	0·60
2. tovarna (A')	0·32	0·08	0·40
	0·68	0·32	1·00

15. $P(A) = 0{\cdot}20 + 0{\cdot}10 - 0{\cdot}05 = 0{\cdot}25$
$P(B) = P(M/J) = P(M\cap J)/P(J) = 0{\cdot}05/0{\cdot}10 = 0{\cdot}50$
$P(C) = P(M'/J') = P(M'\cap J')/P(J') = 0{\cdot}75/0{\cdot}90 \approx 0{\cdot}83$

16. $(\frac{1}{6}\cdot\frac{3}{4}\cdot\frac{2}{3})/[\frac{1}{6}\cdot\frac{3}{4}\cdot\frac{2}{3} + \frac{5}{6}\cdot\frac{1}{4}\cdot\frac{2}{3} + \frac{5}{6}\cdot\frac{3}{4}\cdot\frac{1}{3}]$

17. a) $P(Z_{25} : P(Z_{50} = 2:1 \Longrightarrow P(Z_{50}) = 0{\cdot}30 \Longrightarrow$
$\Longrightarrow P(A) = 0{\cdot}60 + 0{\cdot}40\cdot0{\cdot}30 = 0{\cdot}72$
b) $P(Z_{25} : P(Z_{50} = 4:1 \Longrightarrow P(Z_{50}) = 0{\cdot}15 \Longrightarrow$
$\Longrightarrow P(A) = 0{\cdot}60 + 0{\cdot}40\cdot0{\cdot}15 = 0{\cdot}66$

18. $P(A) = \frac{4}{31}\cdot\frac{1}{2} \approx 0{\cdot}065$
$P(B) = 1\cdot\frac{1}{2} + \frac{3}{31}\cdot\frac{1}{2} \approx 0{\cdot}55$

19. $P(A) = 0{\cdot}0158\ldots \approx 0{\cdot}016$ (Pomagaj si z zgledom 4.4.6!)

4.5 Dvofazni poskusi in Bayesov obrazec (str. 120, str. 124)

1. $109/180 \approx 0{\cdot}606$
2. a) $49/120 \approx 0{\cdot}408$ b) $\approx 0{\cdot}102$
3. $P(A) \approx 0{\cdot}685$ **4.** $P(A) \approx 0{\cdot}527$ **5.** $P(A) \approx 0{\cdot}560$
6. a) $P(A) \approx 0{\cdot}817\,,$ b) $P(A) \approx 0{\cdot}232$
7. a) $P(A) \approx 0{\cdot}72\,,$ b) $P(A) \approx 0{\cdot}05$
8. a) $P(A) \approx 0{\cdot}27\,,$ b)$P(H_1/A) \approx 0{\cdot}689$
9. $P(H_1/A) = 2/3 \approx 0{\cdot}67$
10. $P(H_1/A) = 7/12 \approx 0{\cdot}0{\cdot}58$
11. $P(H_1/A) = 5/8 = 0{\cdot}625$
12. $P(H_1/A) = 0{\cdot}25$

4.6 Zaporedja poskusov, Bernoullijev obrazec (str. 126, str. 134)

1. Še vedno $1/2$.
2. a) $45/1024 \approx 0{\cdot}044$ b) $56/1024 \approx 0{\cdot}055$ c) $1013/1024 \approx 0{\cdot}989$
3. $P(A) = 28/2^8 \approx 0{\cdot}109$
4. $P(A) = 35/128 \approx 0{\cdot}273\,,\quad P(B) = 84/512 \approx 0{\cdot}164$
5. $P(A) \approx 2{\cdot}6\cdot10^{-6}\,,\quad P(B) = 1-P(A)\,,\quad P(C) = 1-0{\cdot}96^4 \approx 0{\cdot}151$
6. $P(A) = \binom{80}{4}\,0{\cdot}06^4\,0{\cdot}94^{76} \approx 0{\cdot}186$
7. $n \geq 4$
8. $P(A) = \binom{10}{2}\,(6/36)^2\,(30/36)^8 \approx 0{\cdot}291$
9. a) $P(A) \approx 0{\cdot}358$ b) $P(B) \approx 0{\cdot}016$
10. $P(A) = \binom{5}{3}\,0{\cdot}518^3\,0{\cdot}482^2 \approx 0{\cdot}323$
$P(B) = 1-0{\cdot}485^5 \approx 0{\cdot}974\,,\quad P(C) = 0{\cdot}518$

11. $1 - 0'999^{50} \approx 0'04879$, približno 5% reklamacij.

12. a) 15 b) 17 in 18

13. $\binom{9}{6} 0'45^6\, 0'55^4 \approx 0'064$

14. $\frac{6!}{2!2!2!}\left(\frac{4}{10}\right)^2\left(\frac{4}{10}\right)^2\left(\frac{2}{10}\right)^2 \approx 0'092$

4.7 Slučajne spremenljivke (str. 136, str. 149)

1. Vsota pik pri metu dveh poštenih igralnih kock

$$X : \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{6} & \frac{1}{6} \end{pmatrix}$$

2. Večje od števil pik pri metu dveh kock

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{36} & \frac{3}{36} & \frac{5}{36} & \frac{7}{36} & \frac{9}{36} & \frac{11}{36} \end{pmatrix}$$

3. Enakomerna porazdelitev na $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{pmatrix}$$

4. Frekvenca številke v dveh (treh) metih kovanca

$$X : \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix} \qquad X : \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{pmatrix}$$

5. Frekvenca bele krogle (vračanje)

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{1296}{10000} & \frac{3456}{10000} & \frac{3456}{10000} & \frac{1536}{10000} & \frac{256}{10000} \end{pmatrix}, \quad P(X \geq 1) = 0'8704$$

6. Frekvenca bele krogle (brez vračanja)

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{15}{210} & \frac{80}{210} & \frac{90}{210} & \frac{24}{210} & \frac{1}{210} \end{pmatrix}, \quad P(X \geq 1) \approx 0'93$$

7. Število porabljenih nabojev (pozor na $P(X = 4)$!)

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0'800 & 0'160 & 0'032 & 0'008 \end{pmatrix}$$

8. Število pregledanih naprav

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0'1 & 0'09 & 0'081 & 0'0729 & 0'6561 \end{pmatrix}$$

9. Računanje verjetnosti iz verjetnostne sheme

$$X : \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ \frac{1}{10} & \frac{2}{10} & \frac{2}{10} & \frac{4}{10} & \frac{1}{10} \end{pmatrix}$$

a) 5/10 b) 7/10 c) 3/10 č) 5/10 d) 6/10 e) 2/10

10. $P(X = 1) = 0'4$,
$P(X = 2) = 0'6 \cdot 0'6 = 0'36$,
$P(X = 3) = 0'6 \cdot 0'4 \cdot 0'4 \ldots$

Naloga	$E(X)$	$D(X)$	$\sigma(X)$
11.	3'50	2'92	1'71
12.	7'00	5'83	2'42
13.	1'50	0'75	0'87
14.	3000	2500	50
15.	1'94	1'43	1'20

Naloga	$E(X)$	$D(X)$	$\sigma(X)$
16.	-0'25	7'69	2'77
17.	1'37	0'40	0'63
18.	240	96	9'80
19.	0'80	0'36	0'60
20.	0'60	0'54	0'73

21. $P(|X - E(X)| < 100) \geq 1 - \frac{D(X)}{\varepsilon^2} = 1 - \frac{2500}{10000} = 0'75$

22. $E(X) = 160$, $D(X) = 96$, vsaj $0'76$

23. $P(|X - 5000| > 1000) < 0'0025$

4.8 Normalna porazdelitev (str. 152, str. 162)

1. Iskane verjetnosti so (glej tabelo):
a) 0'2119 b) 0'9332 c) 0' 8413 č) 0'6915
d) 0'3446 e) 0'1752 f) 0' 7745

2. Iskane verjetnosti so (glej tabelo):
a) 0'1587 b) 0'9332 c) 0' 9772
č) 0'8185 d) 0'4938 e) 0' 7745

3. 0'1587

4. $P(\text{sprejme}) = 0'8185 \Longrightarrow$ kupec zavrne več kot 18% plošč!

5. $P(X > 5100) = \frac{1}{2} - \Phi(2) \approx 0{\cdot}023$ 6. $P(X \leq 80) \approx 0{\cdot}023$

4.9 Zakon velikih števil (str. 163, str. 165)

1. $0{\cdot}778$, **2.** $> 0{\cdot}609$, **3.** $n \geq 129$

5.1 Vzorčenje (str. 169, str. 178)

1. $\mu_y = 5 = \mu_{\bar{y}}$, $\sigma_{\bar{y}}^2 = 4{\cdot}\overline{6}$, $E(S^2) = 4{\cdot}\overline{6}$, (Glej zgled 5.1.1!)

2. $E(S^2) = 5{\cdot}6 \Longrightarrow \frac{6}{6-1}E(S^2) = \sigma_y^2$ **3.** $\sigma_{\bar{y}} \approx 0{\cdot}93$

4. $\mu_y = 68{\cdot}0$, $\sigma_{\bar{y}}^2 = 0{\cdot}36$ oziroma $\sigma_{\bar{y}}^2 = 0{\cdot}33$

5. $\Phi(\frac{68{\cdot}3-68{\cdot}0}{0{\cdot}6}) - \Phi(\frac{66{\cdot}8-68{\cdot}0}{0{\cdot}6}) = \Phi(0{\cdot}5) + \Phi(2) \approx 0{\cdot}669$

6. a) $\frac{1}{2} - \Phi(\frac{69{\cdot}2-68{\cdot}0}{0{\cdot}6}) \approx 0{\cdot}023 = 2{\cdot}3\%$ b) 50%

7. 5177, 2746, 4042, 3312, *9044*, 4662, 1628, *9893*, 5820, 4186, 1964, 0870. Skupin v poševnem tisku ne uporabimo. (Zakaj?)

8. Ocena: 169·0, 8·77, dejansko: 165·12, 8·37. **10.** a) 7· 2 b) 8·4

5.2 Ocenjevanje parametrov (str. 181, str. 188)

3. Dogodek ni nemogoč, je pa malo verjeten. Utemelji!

4. $\alpha = 0{\cdot}05$: $[164{\cdot}3\,, 173{\cdot}7]$ že "pokrije" pravo vrednost 165· 12.

6. $20°40'32'' \pm 20''$ **7.** 11· 16 oziroma 4·48 minute

8. $[4{\cdot}412\,, 5{\cdot}588]$ **9.** $[6{\cdot}\,85\,, 7{\cdot}67]$ **10.** $[190\,800\,, 201\,200]$

11. $[134{\cdot}4\,, 145{\cdot}6]$ **12.** a) $n \geq 139$ b) $n \geq 73$ 13. $n \geq 385$

5.3 Preizkušanje domnev (str. 191, str. 198)

1. $z = -2{\cdot}50$, H_0 zavrnemo na stopnji 0·05.

2. Razlike značilne celo na stopnji 0·001.

3. $z = -2{\cdot}50$, H_0 zavrnemo na stopnji 0· 05, pri 0·01 pa ne.

4. Pri enostranskem testu lahko hipotezo $H_0 : \mu_y = 1600$ zavrnemo celo pri stopnji značilnosti 0·01.

5. Značilna razlika na stopnji 0·05.

6. Hipoteze $H_0 : \mu_y = 66{\cdot}5$ ne moremo zavrniti. **7.** 0·26

8. $z \approx 1{\cdot}70$, hipoteze, da je kocka poštena, ne moremo zavrniti.

9. Razlike značilne na stopnji 0·05, ne pa na stopnji 0·01.

LITERATURA

To je samo skromen spisek del, iz katerih sem največ pobral. Za dela podpisanega avtorja kakopak dodatno velja tranzitivnost.

AVSEC F., COKAN A., PUCELJ I.: Kombinatorika, verjetnostni račun in statistika, zbirka nalog za srednje šole. Ljubljana: Državna založba Slovenije, 1986.

BLAZNIK A., COKAN A., PAVLIČ G.: Matematični priročnik za srednje šole. Ljubljana: Državna založba Slovenije, 1990.

ČIBEJ J. A.: Matematika. Kombinatorika. Ljubljana: Državna založba Slovenije, 1984. (6. natis Ljubljana: Državna založba Slovenije, 1991.)

ČIBEJ J. A.: Matematika. Verjetnostni račun in statistika. Ljubljana: Državna založba Slovenije, 1984.

ČIBEJ J. A.: Statistika. Učbeniško gradivo za naravoslovno - matematično usmeritev. Ljubljana: Zavod SR Slovenije za šolstvo, 1985.

HOEL P. G.: Elementary Statistics. New York: John Wiley & Sons, 1976.

INDIHAR S., KAVKLER I.: Matematika I, I. del. Maribor: Visoka ekonomsko - komercialna šola, 1987.

JAMNIK R.: Verjetnostni račun. Ljubljana: Mladinska knjiga, 1971.

KOŠMELJ B.: Statistika. (Srednje izobraževanje.) 3. natis. Ljubljana: Državna založba Slovenije, 1987.

PITMAN J., Probability. New York: Springer-Verlag, 1993.

RENDER B., STAR R. M.: Quantitative Analysis for Management. Needham Heights: Allyn and Bacon, 1988.

ROWNTREE D.: Statistics without Tears. A Primer for Non-mathematicians. London: Penguin Group, 1981.

SOMMER E., SOMMER D., HÖFFLIN H.: Mathematik für Wirtschaftsgymnasien - Analysis. Bad Homburg vor der Höhe: Verlag Dr. Max Gehlen, 1991.

VADNAL A.: Elementarni uvod v verjetnostni račun. Ljubljana: Državna založba Slovenije, 1972b.

Jože Andrej Čibej
MATEMATIKA
Kombinatorika. Verjetnostni račun. Statistika

Uredila *Milena Strnad*

Opremil *Vili Vrhovec*
Likovno-grafično uredila *Irena Petrič*
Tehnične risbe narisal *Darko Simeršek*

Izdala in založila *DZS, d.d.*
Za DZS *Bojan Petan*
Za Divizijo založništev *Andrej Založnik*
Glavna urednica *Tanja Železnik*

Prelom in priprava za tisk *J. A. Čibej*
Natisnila *Grafika Soča, d. d.*
Ljubljana 2004
Četrta prenovljena izdaja
Četrti natis