Vzorčne porazdelitve

Statistika

Cilji

Razumeti in uporabiti:

- naključne vzorce
- Centralni limitni izrek (CLI)
- vzorčne porazdelitve

Normalna porazdelitev

- L. 1738 je Abraham De Moivre (1667-1754) objavil aproksimacijo binomske porazdelitve,
  ki je normalna krivulja.
- L.1809 je Karl Frederic Gauss (1777-1855) raziskoval matematično ozadje planetarnih orbit, ko je izpeljal normalo porazdelitveno funkcijo.

Normalna porazdelitev

zaloga vrednosti so vsa realna števila

verjetnostna gostota:

$$p(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{\left[-(x-\mu)^2/(2\sigma^2)\right]}, \qquad -\infty<x<\infty$$

povprečje: μ

varianca: σ^2

rodovna funkcija momentov: $e^{\left[\mu t+\sigma^2\left(t^2/2\right)\right]}$

Standardna normalna porazdelitev (Z)

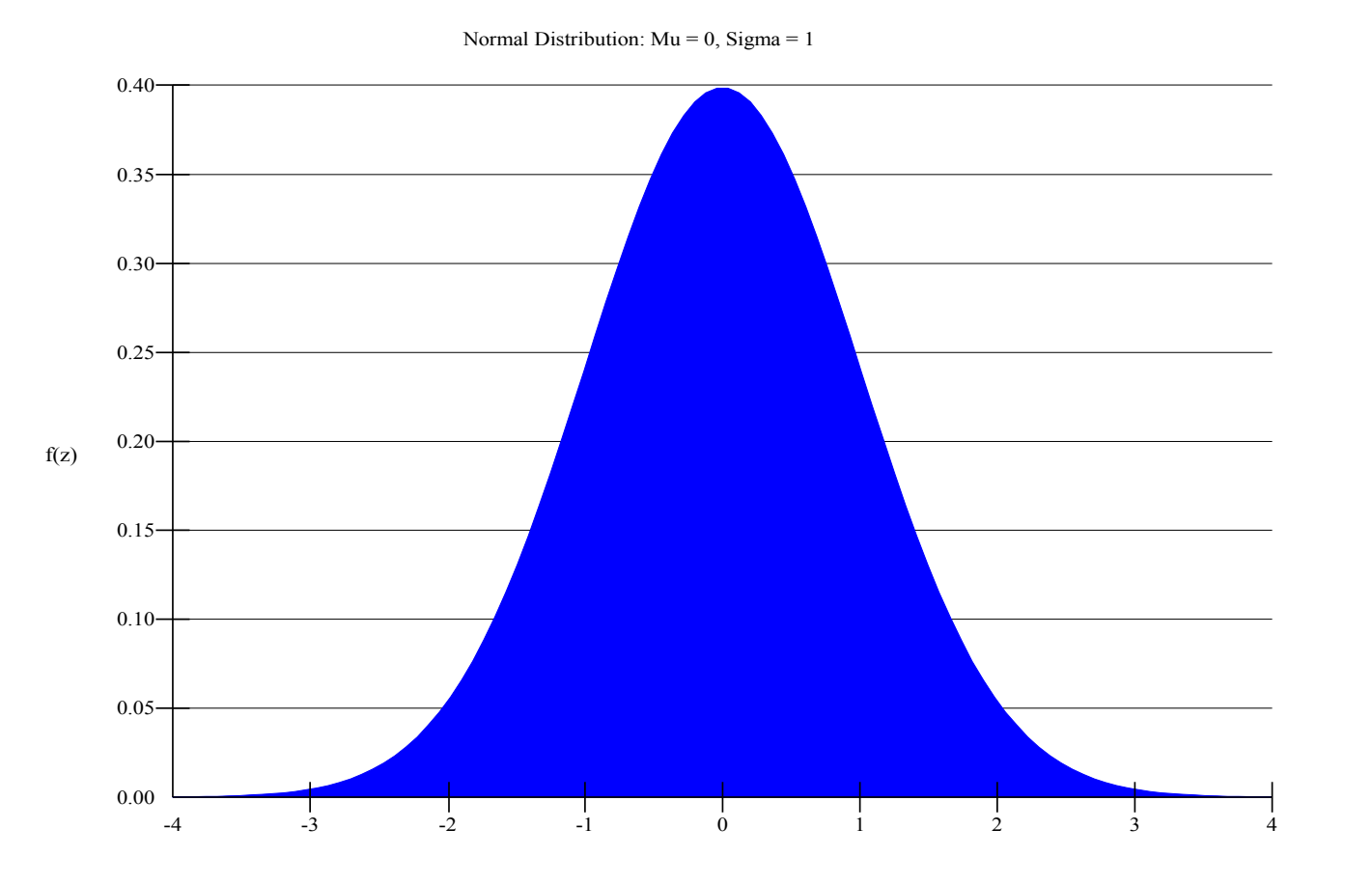
$$p(z;\mu,\sigma)=\frac{1}{\sqrt{2\pi}}e^{-\left(z^2/2\right)},\quad -\infty<z<\infty$$

povprecje: 0

varianca: 1

rodovna funkcija momentov: $e^{t^2/2}$

Standardna normalna porazdelitev (Z)



Naključni vzorec

Vsakega člana populacije izberemo z enako verjetnostjo

- neskončna populacija
- vzorčenje z vračanjem
- velikost vzorca je manj kot 5% celotne populacije

Naključni vzorec

Slučajne spremenljivke, ki sestavljajo naključni vzorec so:

- neodvisne
- enako porazdeljene

  - $E(X_1)=E(X_2)=\ldots=E(X_n)=E(X) = \mu$
  - $V(X_1)=V(X_2)=\ldots=V(X_n)=V(X) = \sigma^2$

Matematično upanje – pričakovana vrednost – enostavno povprečje

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n} E(X_1 + X_2 + \cdots + X_n)$$

$$¿ \frac{1}{n}\left[E(X_1) + E(X_2) + \cdots + E(X_n)\right]$$

$$¿ \frac{1}{n}\left[n\, E(X)\right] = E(X) = \mu$$

Varianca enostavnega povprečja

$$V(\bar{X}) = V\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$

$$¿\ \frac{1}{n^2} V(X_1 + X_2 + \cdots + X_n)$$

$$¿\ \frac{1}{n^2}\left[V(X_1) + V(X_2) + \cdots + V(X_n)\right]$$

$$¿\ \frac{1}{n^2}\left[n V(X)\right] = \frac{1}{n} V(X) = \frac{\sigma^2}{n}$$

Centralni limitni izrek (CLI)

L. 1810 je Pierre Laplace (1749-1827) študiral anomalije orbit Jupitera in Saturna, ko je izpeljal razširitev De Moivrovega limitnega izreka, “Vsaka vsota ali povprečje, če je število členov dovolj veliko, približno normalno porazdeljena.”

Centralni limitni izrek

Če imamo dovolj *velik slučajni vzorec* iz porazdelitve s končnim povprečjem in varianco, potem je porazdelitev ***enostavnega povprečja*** vsaj približno normalna.

Dokaz centralnega limitnega izreka

Naj bo $Z_i = \frac{X_i - \mu}{\sigma} \Rightarrow M_Z(t) = 1 + \frac{t^2}{2} + \frac{t^3}{3!} E\left(Z_i^3\right) + \cdots$

Za $Y_n = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i$

$$\Rightarrow M_n(t) = \left[M_Z\left(\frac{t}{\sqrt{n}}\right) \right]^n = \left(1 + \frac{t^2}{2\mathrm{n}} + \frac{t^3}{3!\, n^{3/2}} k + \cdots \right)^n$$

kjer je $k = E\left(Z_i^3\right)$.

Dokaz centralnega limitnega izreka

$$\log M_n(t) = n \log\left(1 + \frac{t^2}{2\mathrm{n}} + \frac{t^3}{3!\,n^{3/2}} k + \cdots\right)$$

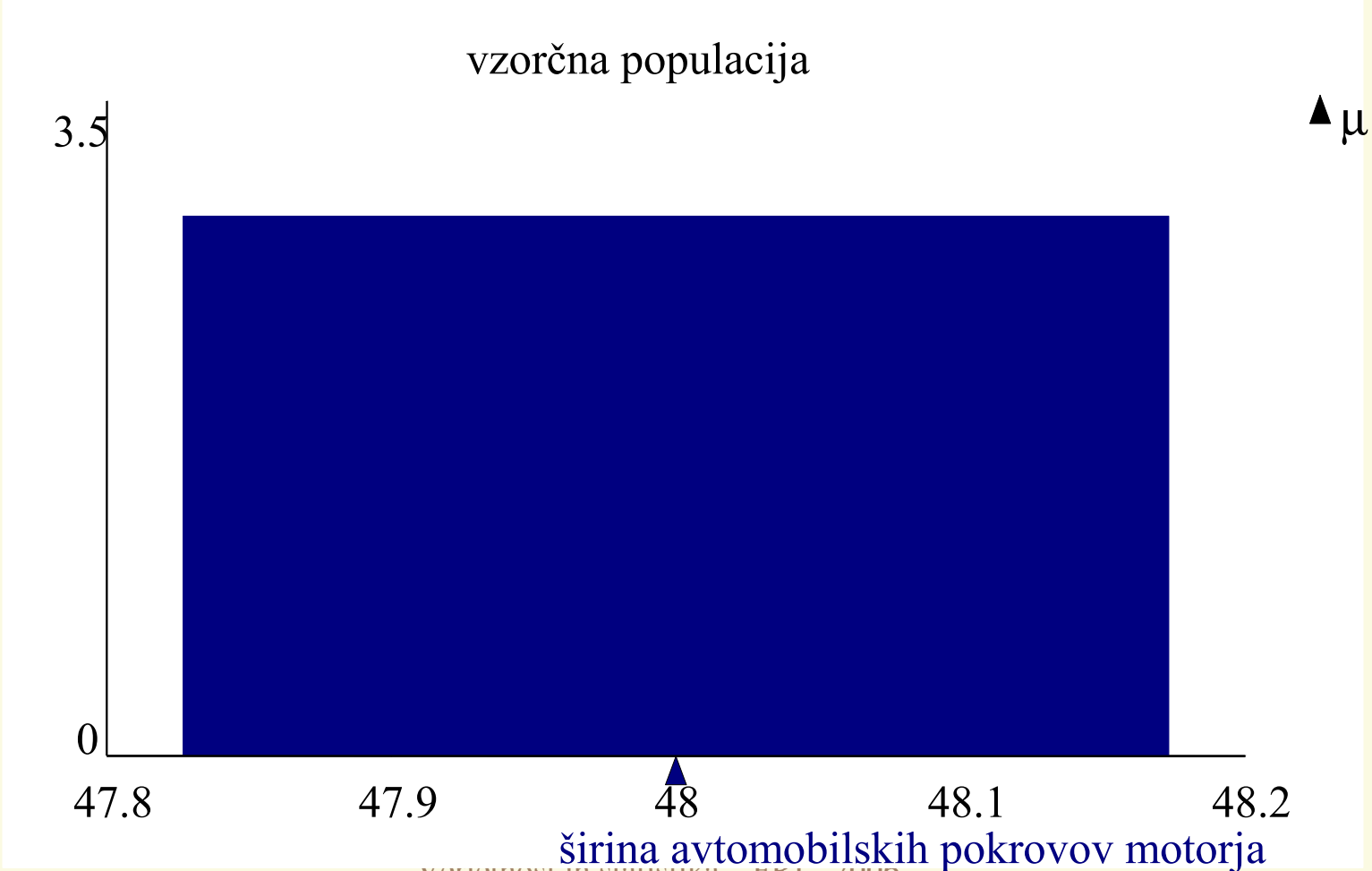
$$\text{Naj bo} \quad x = \left(\frac{t^2}{2\mathrm{n}} + \frac{t^{3\mathrm{k}}}{3!\,n^{3/2}} + \cdots\right)$$

$$\Rightarrow \log M_n(t) = n \log(1 + x) = n\left(x - \frac{x^2}{2} + \cdots\right)$$

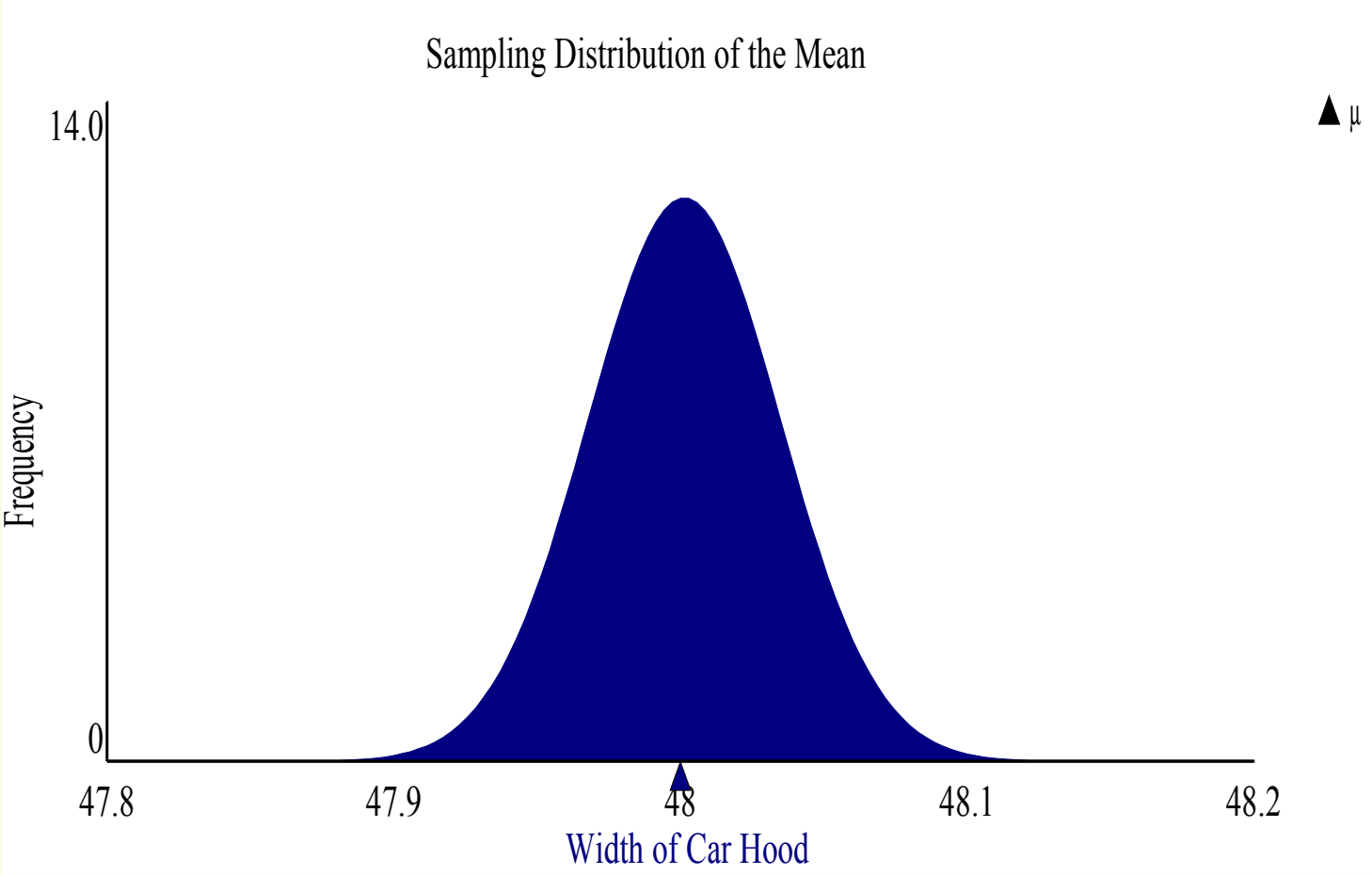
$$¿\, n\left[\left(\frac{t^2}{2\mathrm{n}} + \frac{t^{3\mathrm{k}}}{3!\,n^{3/2}} + \cdots\right) - \frac{1}{2}\left(\frac{t^2}{2\mathrm{n}} + \frac{t^{3\mathrm{k}}}{3!\,n^{3/2}} + \cdots\right)^2 + \cdots\right]$$

$$\Rightarrow \lim_{n \to \infty} \log M_n(t) = \frac{t^2}{2} \Rightarrow \lim_{n \to \infty} M_n(t) = e^{t^2/2}$$

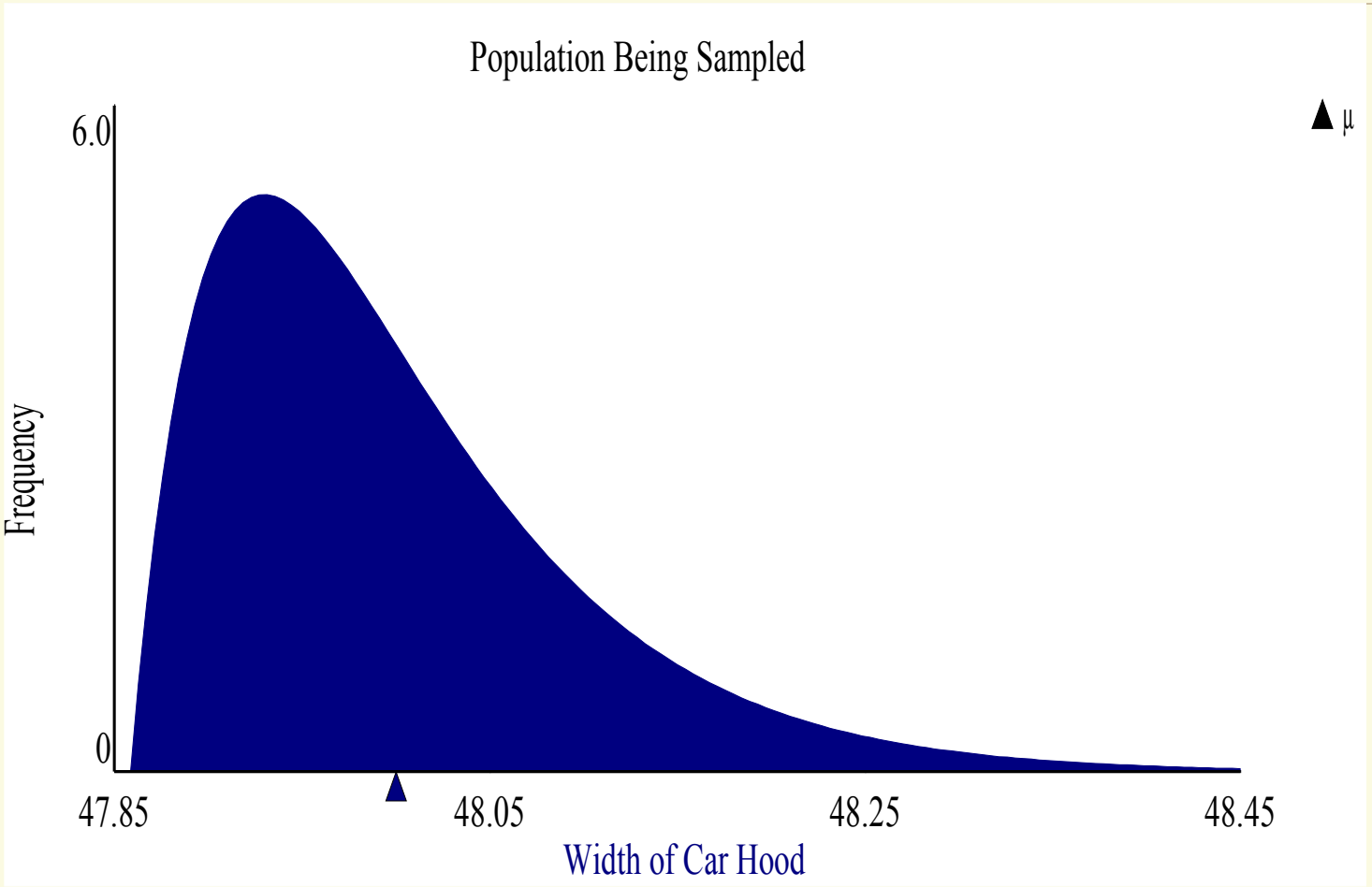
Centralni limitni izrek



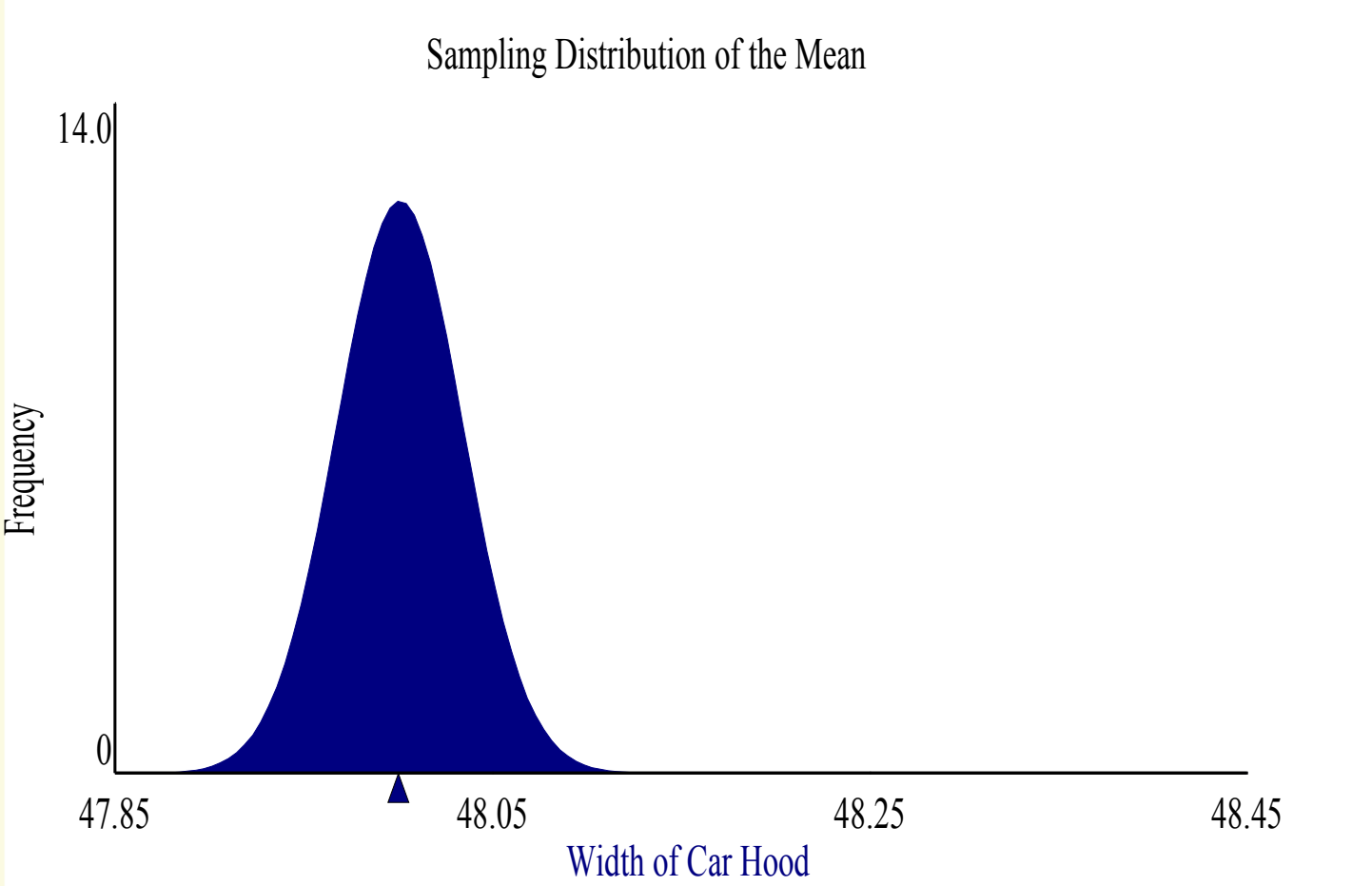
Centralni limitni izrek



Centralni limitni izrek



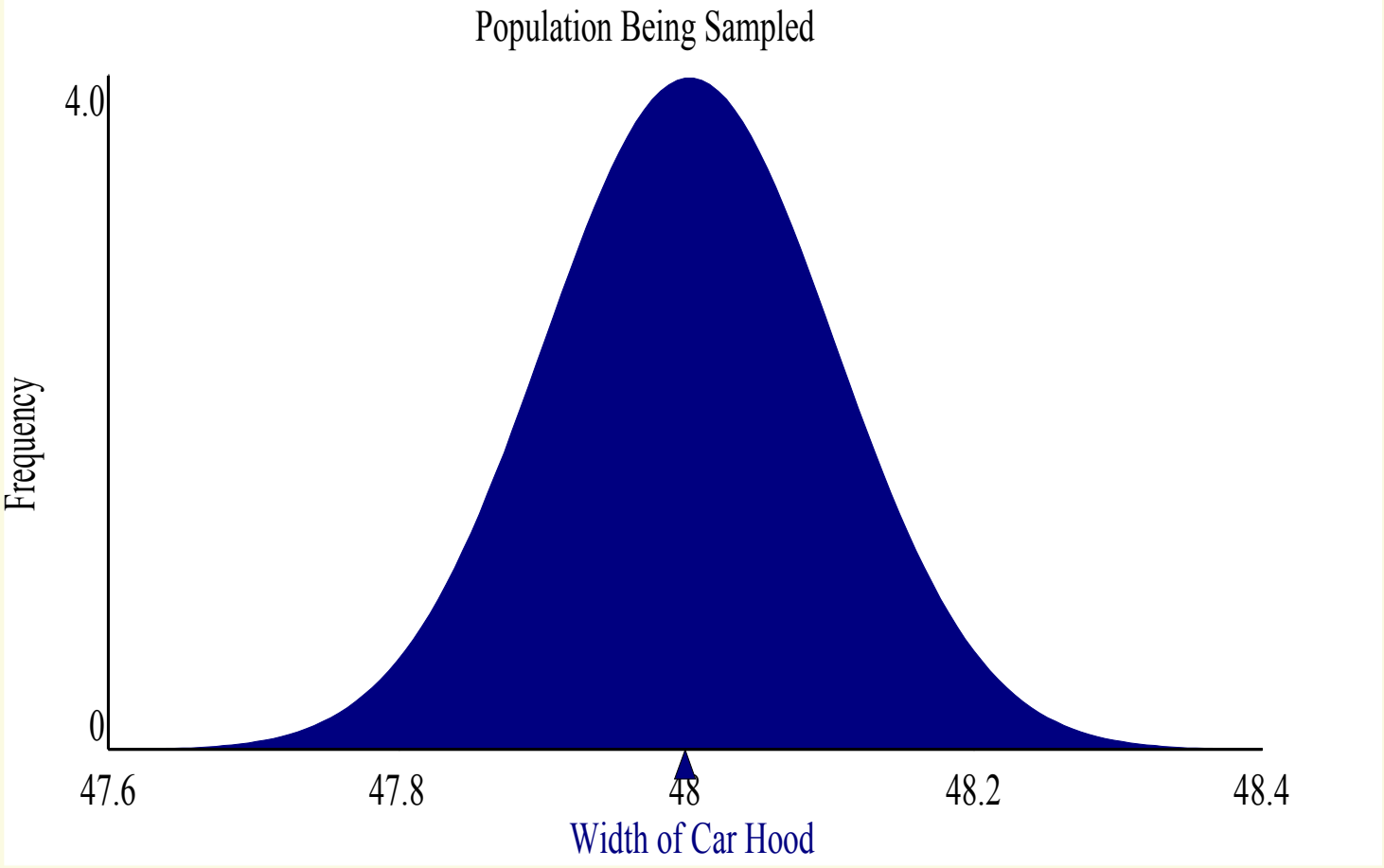
Centralni limitni izrek



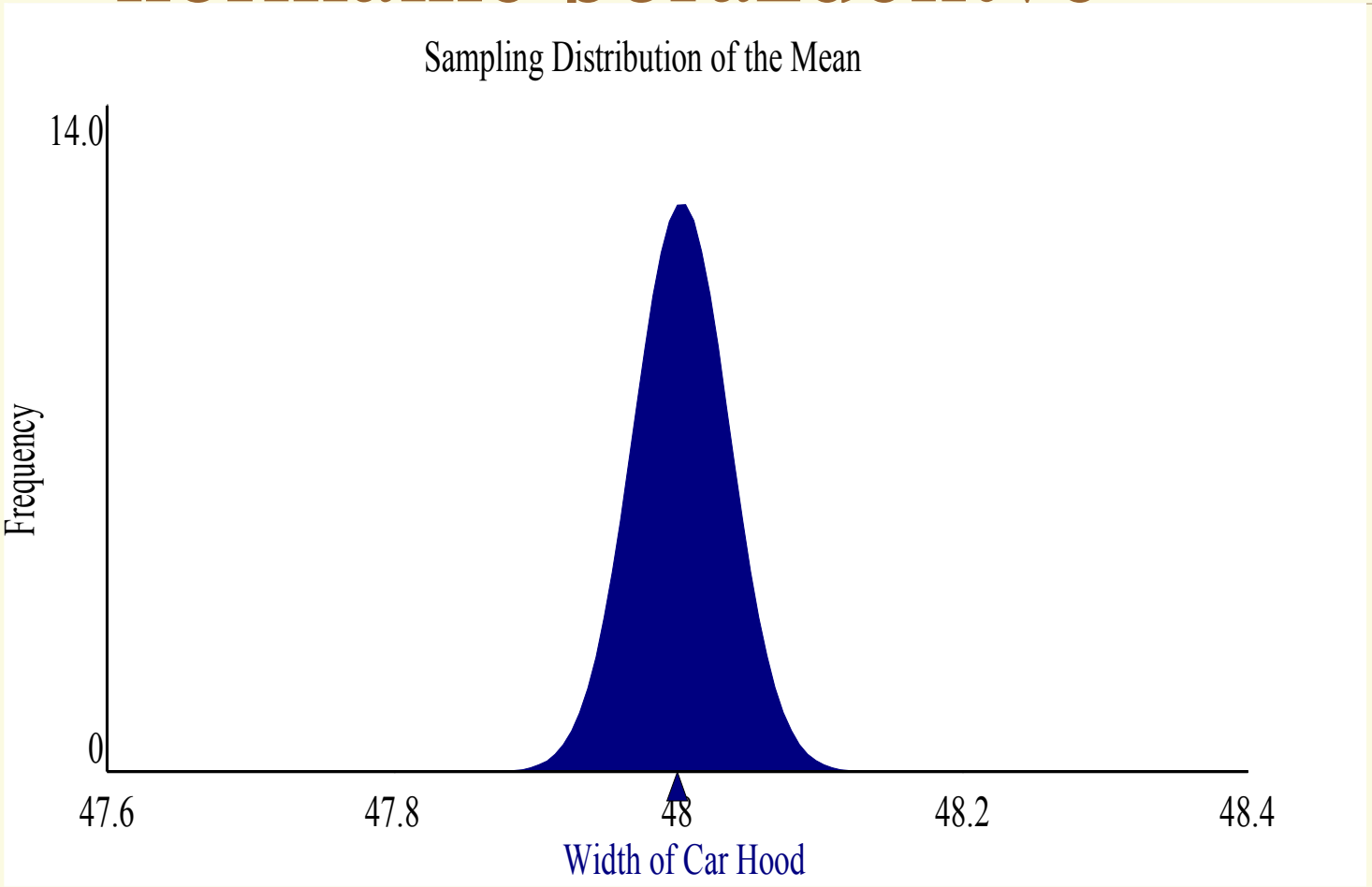
Reprodukcijska lastnost normalne porazdelitve

Vsaka linearna kombinacija <u>neodvisne</u> in <u>normalno</u> porazdeljene slučajne spremenljivke je tudi sama normalno porazdeljena.

Reprodukcijska lastnost normalne porazdelitve



Reprodukcijska lastnost normalne porazdelitve

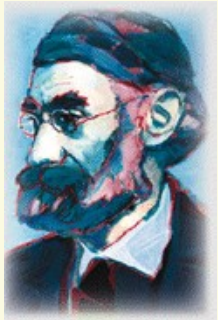


Porazdelitev enostavnega povprečja

Če lahko uporabimo bodisi CLI ali reprodukcijsko lastnost normalne distribucije, potem je porazdelitev enostavnega povprečja normalna s parametri:

- povprečje: μ
- varianca: σ^2/n

Hi-kvadrat porazdelitev (χ^2)



- L. 1863 je Ernst Abbe (1840-1905), nemški fizik, prvi izpeljal hi-kvadrat porazelitev, ko je preučeval porazdelitev vsote kvadratov napak.
- L. 1878 je Ludwig Boltzmann izpeljal hi-kvadrat porazdelitev z 2-ma in 3-emi prostostnimi stopnjami, ko je študiral kinetično energijo molekul.
- Karl Pearson (1857-1937) je demonstriral uporabnost hi-kvadrat porazdelitve statistikom.

Hi-kvadrat porazdelitev(χ^2)

- je član Gama družine
- parametri
  - število prostostnih stopenj (ν)
- povprečje
  - število prostostnih stopenj (ν)
- varianca
  - 2ν

Hi-kvadrat (χ^2)

Število prostostnih stopenj za vsoto dveh neodvisnih slučajnih spremenljivk,

ki sta porazdeljeni po hi-kvadrat,

je enaka vsoti prostostnih stopenj,

ki ustrezata vsaki od hi-kvadrat porazdelitev: $(\nu_1 + \nu_2)$.

Hi-kvadrat (χ^2)

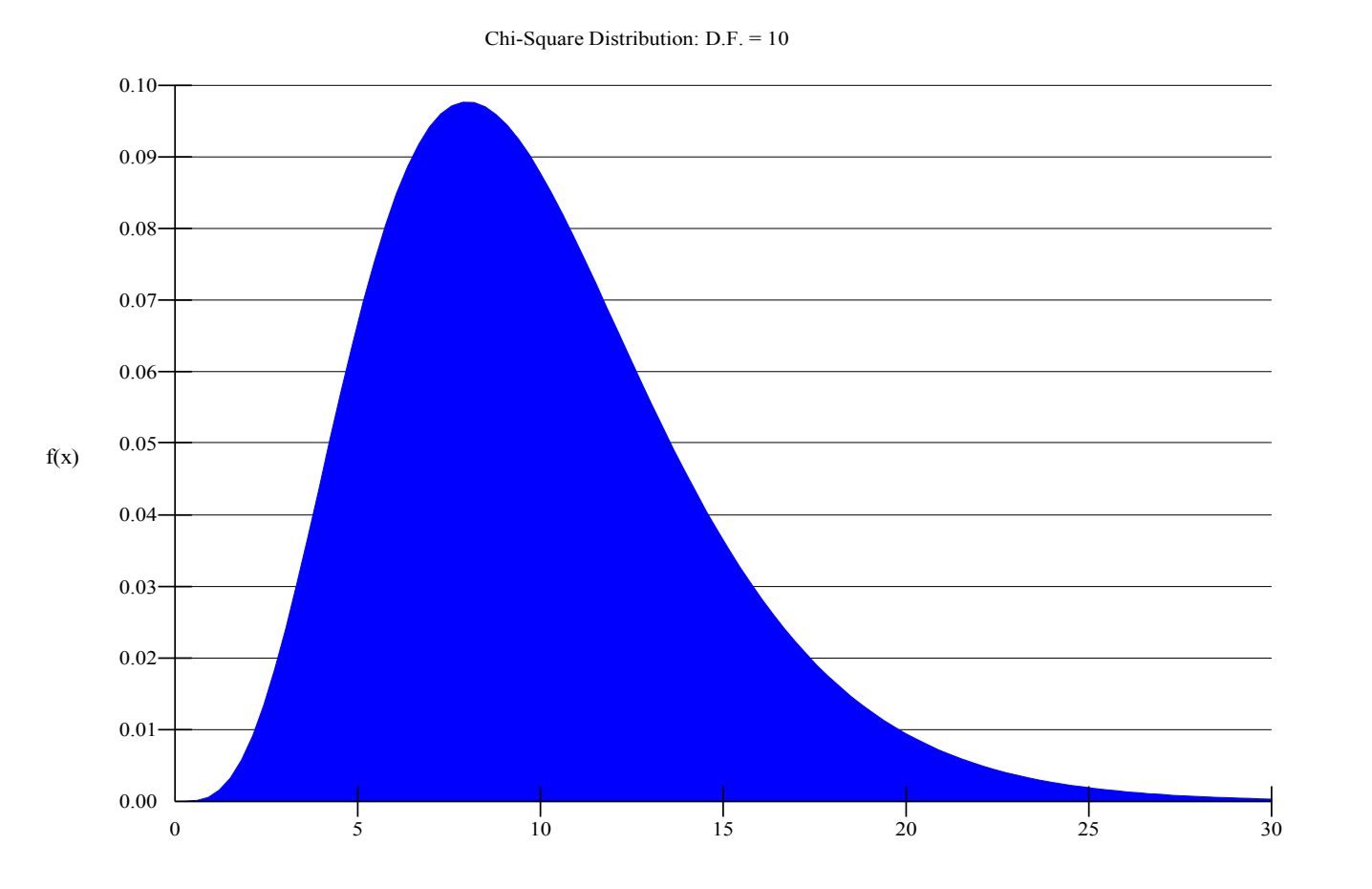
$$\chi_n^2 = \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left[\frac{(X_i - \mu)}{\sigma} \right]^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2}$$

¿ $$\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} \Rightarrow \chi_{n-1}^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2}$$

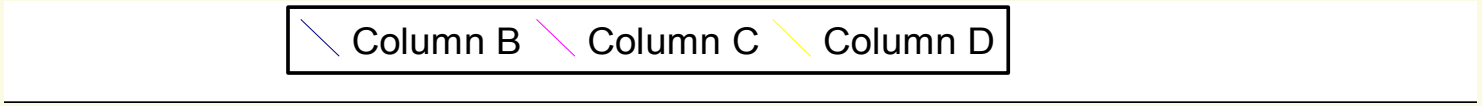
$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} \Rightarrow \sum_{i=1}^{n} (X_i - \bar{X})^2 = S^2 (n-1)$$

$$\chi_{n-1}^2 = \frac{S^2 (n-1)}{\sigma^2}$$

Hi-kvadrat porazdelitev (χ^2)



Hi-kvadrat (χ^2)



Studentova *T*-porazdelitev

- L. 1908 je William Gosset, pivovar pri Guinessu, objavil (pod psevdonimom Student) popravek za majhne velikosti vzorcev, ki ga je razvil medtem, ko je študiral, kako kvaliteta kvasovk vpliva na kvaliteto proizvedenega piva.
- Kasneje je Harold Hotelling izpeljal Studentovo *t*-porazdelitev, in jo poimenoval v čast Williamu Gossetu.

Studentova *T*-porazdelitev

- Kvocient dveh neodvisnih slučajnih spremenljivk.
  - standardna normalna (Z)
  - hi-kvadrat (χ^2)

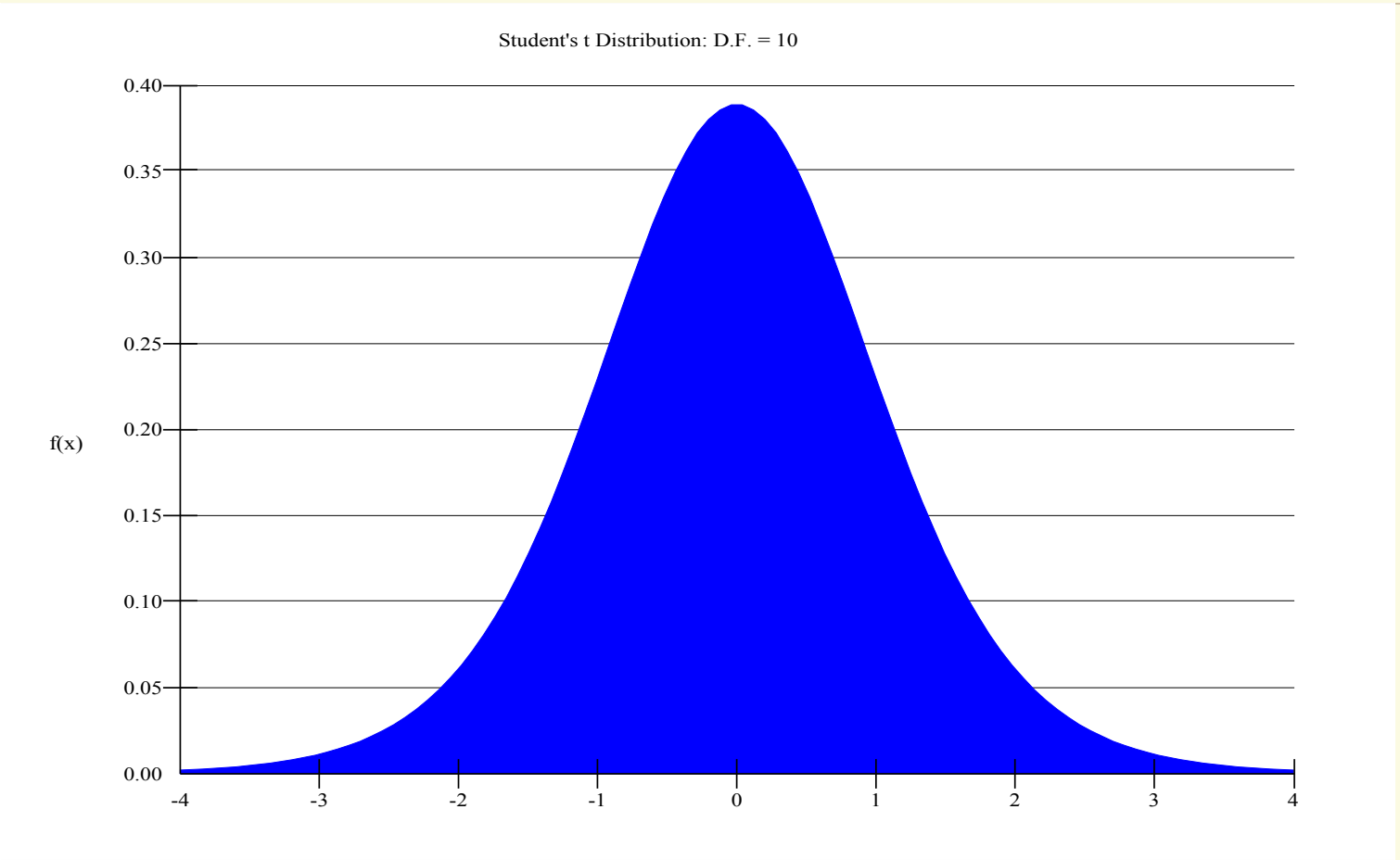
$$T_\nu = \frac{Z}{\chi^2/\nu}$$

- parameter
  - število prostostnih stopenj (ν)

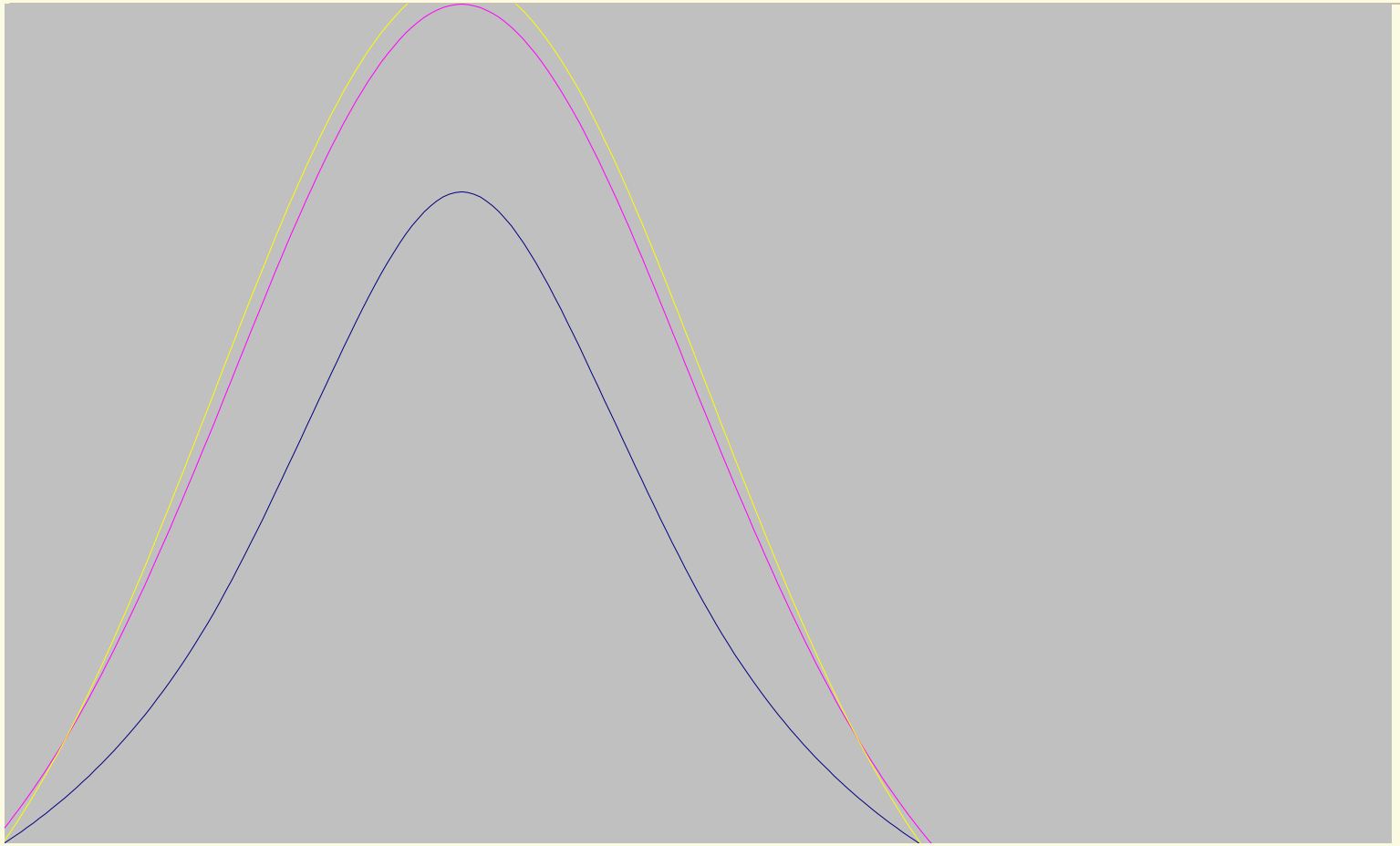
Studentova *T*-porazdelitev

- povprečje
  - nedefinirano za $\nu = 1$ (Cauchyjeva porazdelitev)
  - sicer $\mu = 0$
- varianca $\dfrac{v}{v-2}$ za $v > 2$

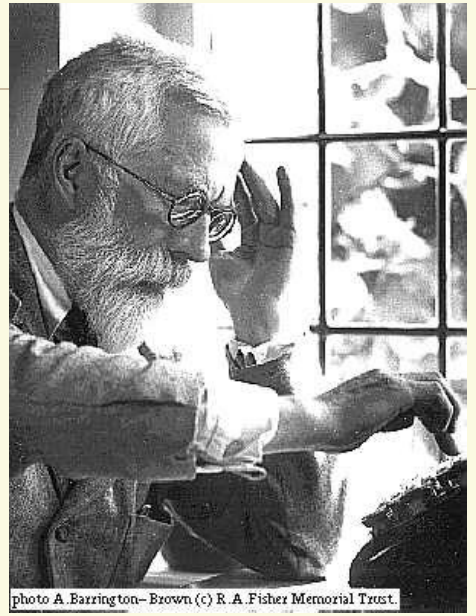
Studentova *T*-porazdelitev



Studentova *T*-porazdelitev



F-porazdelitev



Sir Ronald Fisher (1890-1962)
je razvil *F*-porazdelitev,
ko je preučeval razlike v donosu letin.

F-porazdelitev

- Kvocient dveh neodvisnih hi-kvadrat slučajnih spremenljivk
- Parameteri
  - število prostostnih stopenj, ki ustrezajo števcu (ν_1),
  - število prostostnih stopenj, ki ustrezajo imenovalcu (ν_2).

F-porazdelitev

- povprečje

$$\frac{\nu_2}{\nu_2 - 2}$$

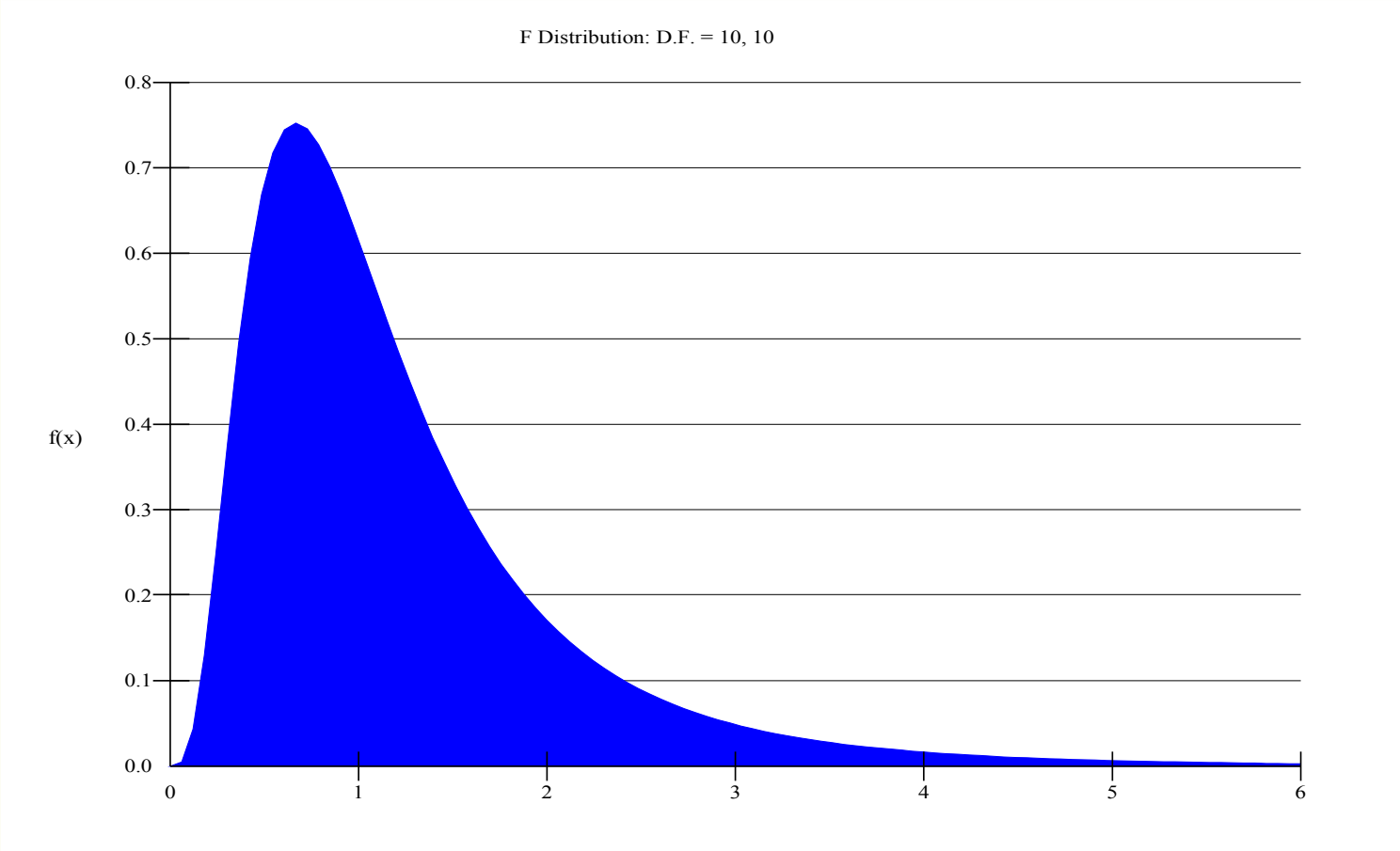
za $\nu_2 > 2$

- varianca

$$\frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$$

za $\nu_2 > 4$

F-porazdelitev



F-porazdelitev

$$F_{\nu_1,\nu_2} = \frac{\chi_1^2 / \nu_1}{\chi_2^2 / \nu_2}$$

$$F_{1-\alpha,\nu_1,\nu_2} = 1 / F_{\alpha,\nu_2,\nu_1}$$

Ponovitev

- Kaj je slučajni vzorec?
- Kaj je centralni limitni izrek?
- Izračunajte verjetnosti z uporabo:
  - standardne normalne porazdelitve (Z),
  - hi-kvadrat porazdelitev (χ^2),
  - T-porazdelitev,
  - F-porazdelitev.