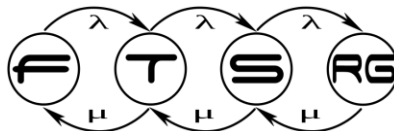


Kísérlettervezés alapfogalmak

**Budapest University of Technology and Economics
Fault Tolerant Systems Research Group**



Kísérlettervezés

- Cél: a modell paraméterezése a valóság alapján
 - Vagy absztrakt modell a konkrét modell alapján
- Információt **kísérletek** révén szerzünk
 - Pontosan mit szeretnénk tudni?
 - Ehhez milyen megfigyelést, hányszor kell elvégezni?
 - A kapott eredményekből mire lehet következtetni?
- (statisztikai) **kísérlettervezés** (Design of Experiment, DOE)
 - hatékony eljárás a kísérletek tervezésére és elemzésére
 - valós és objektív konklúziók levonásához
- Kísérletterv: még a kísérlet elvégzése előtt

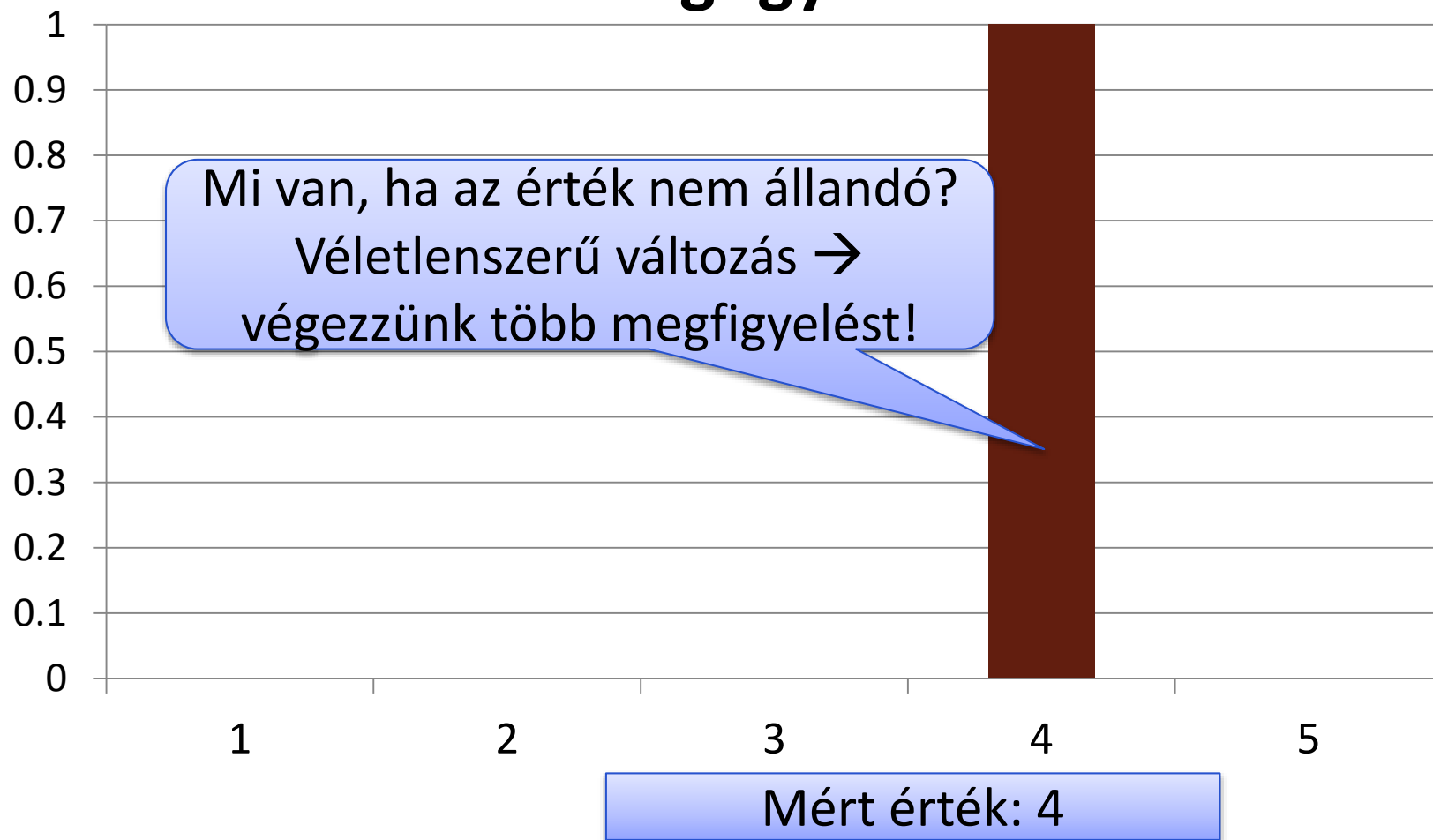
Kísérlettervezés

- Mire jó?
 - Alternatívák közötti választás
 - Érzékeny paraméterek, kulcsfaktorok
 - Megfelelő célérték, változékonyság csökkentése
 - Robosztussá tétel
- Fontos:
 - Világos cél, egyértelmű eredmények
 - Kis méret, alacsony költség
 - Valós viszonyok

Számszerű jellemzők mérése

- Mérjük meg egy jellemző értékét! (1..5)

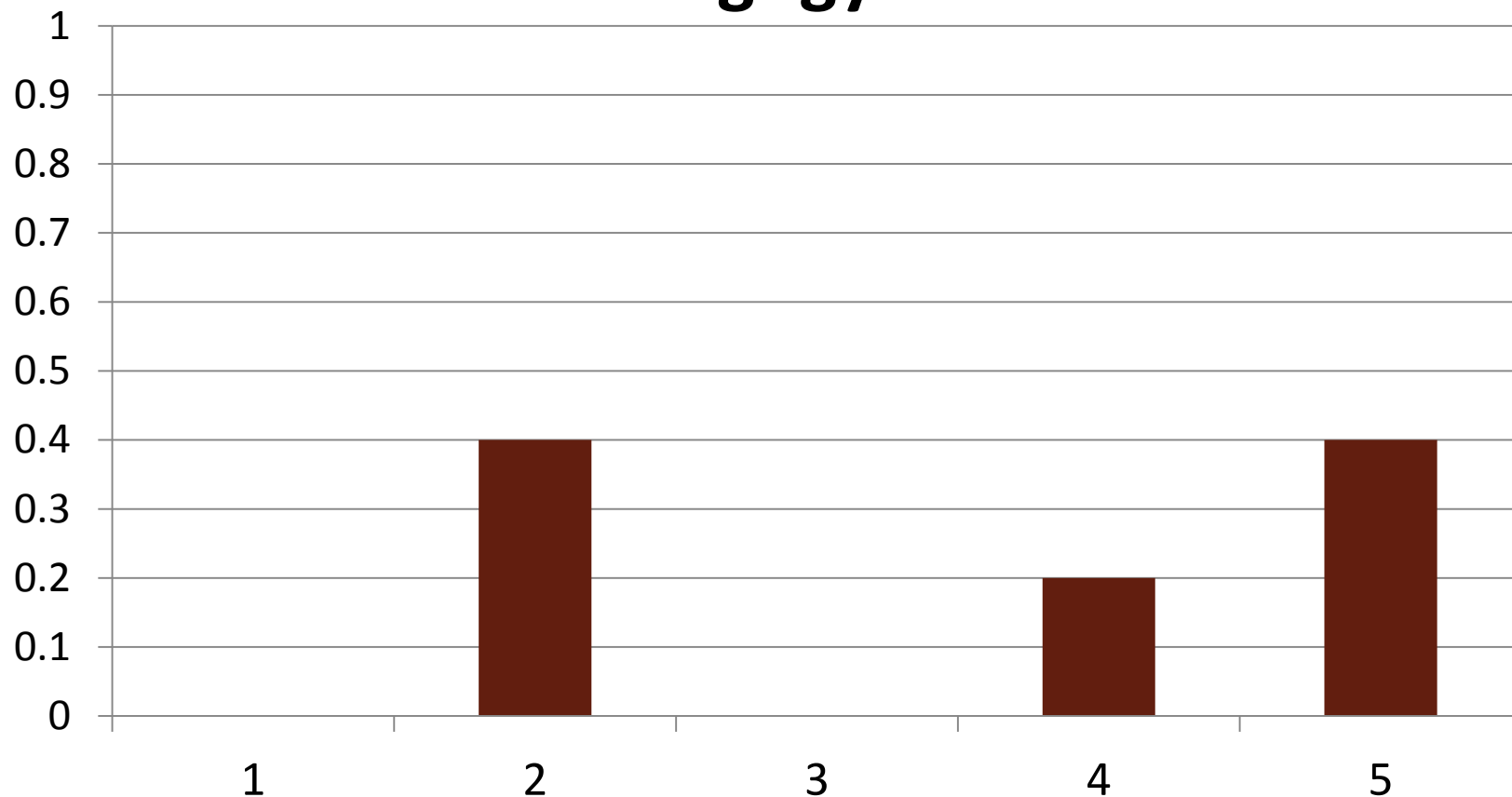
1 megfigyelés



Számszerű jellemzők mérése

- Mérjük meg egy jellemző értékét! (1..5)

5 megfigyelés

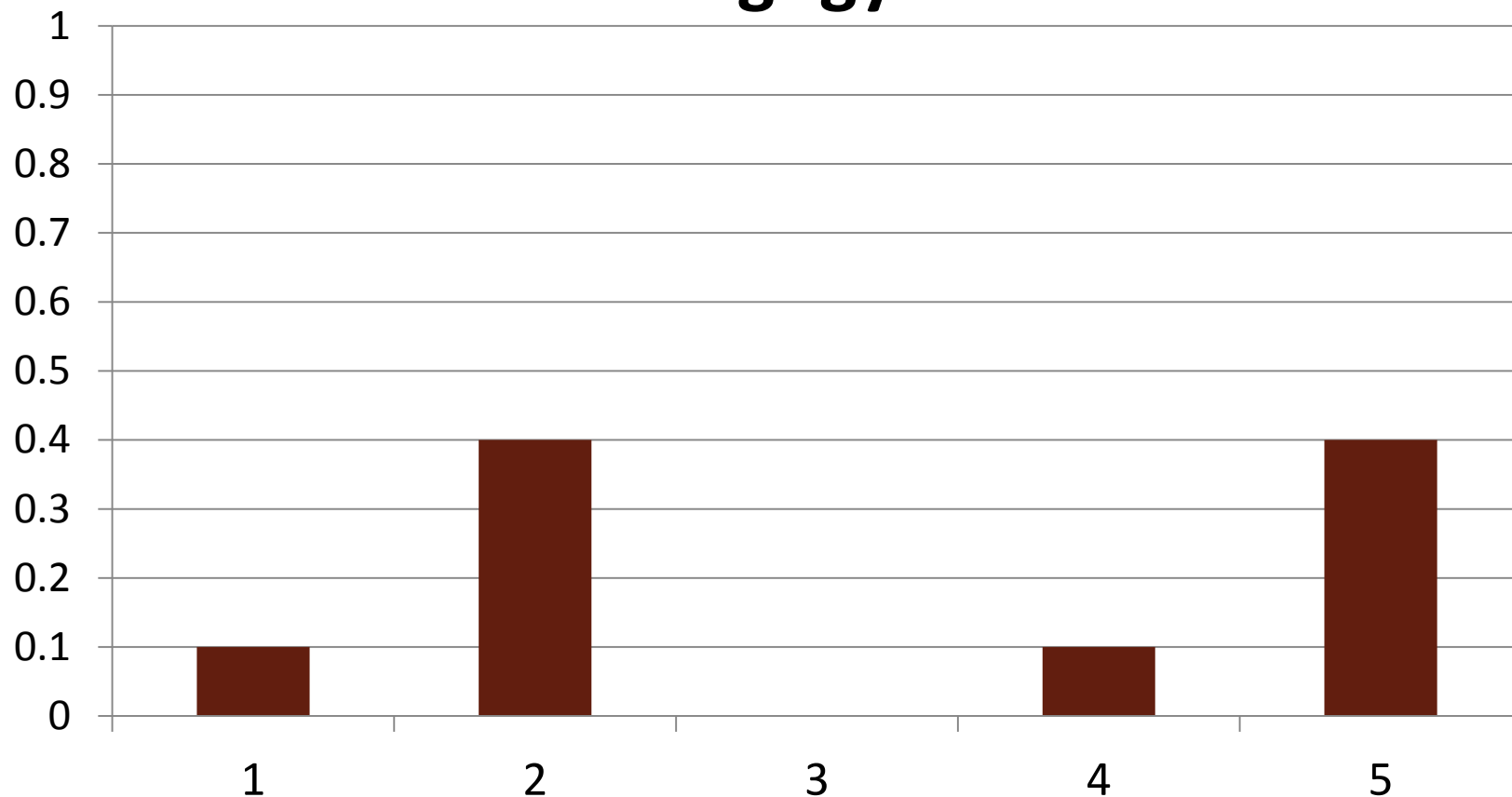


Számított átlag: 3,6

Számszerű jellemzők mérése

- Mérjük meg egy jellemző értékét! (1..5)

10 megfigyelés

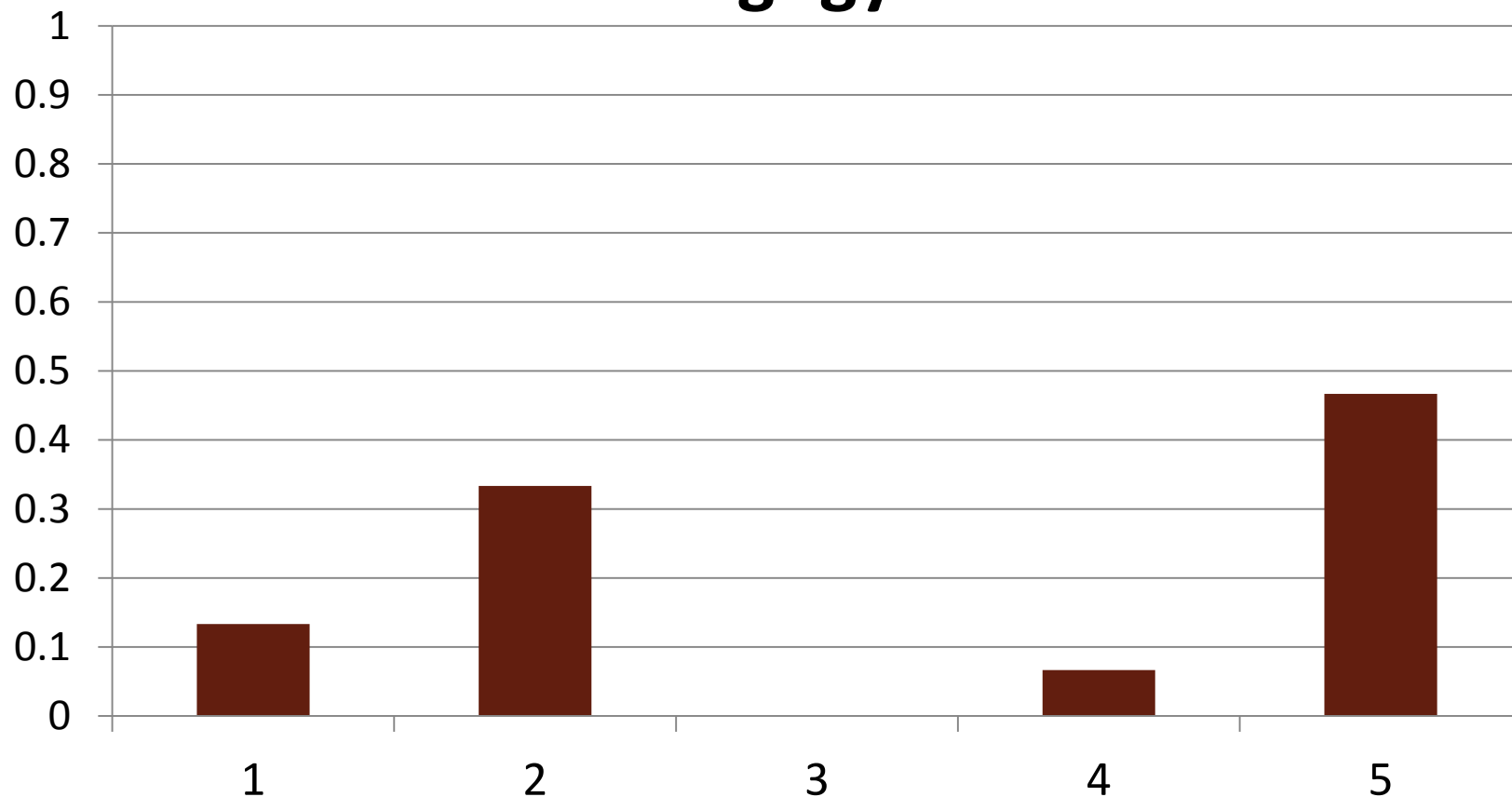


Számított átlag: 3,3

Számszerű jellemzők mérése

- Mérjük meg egy jellemző értékét! (1..5)

15 megfigyelés

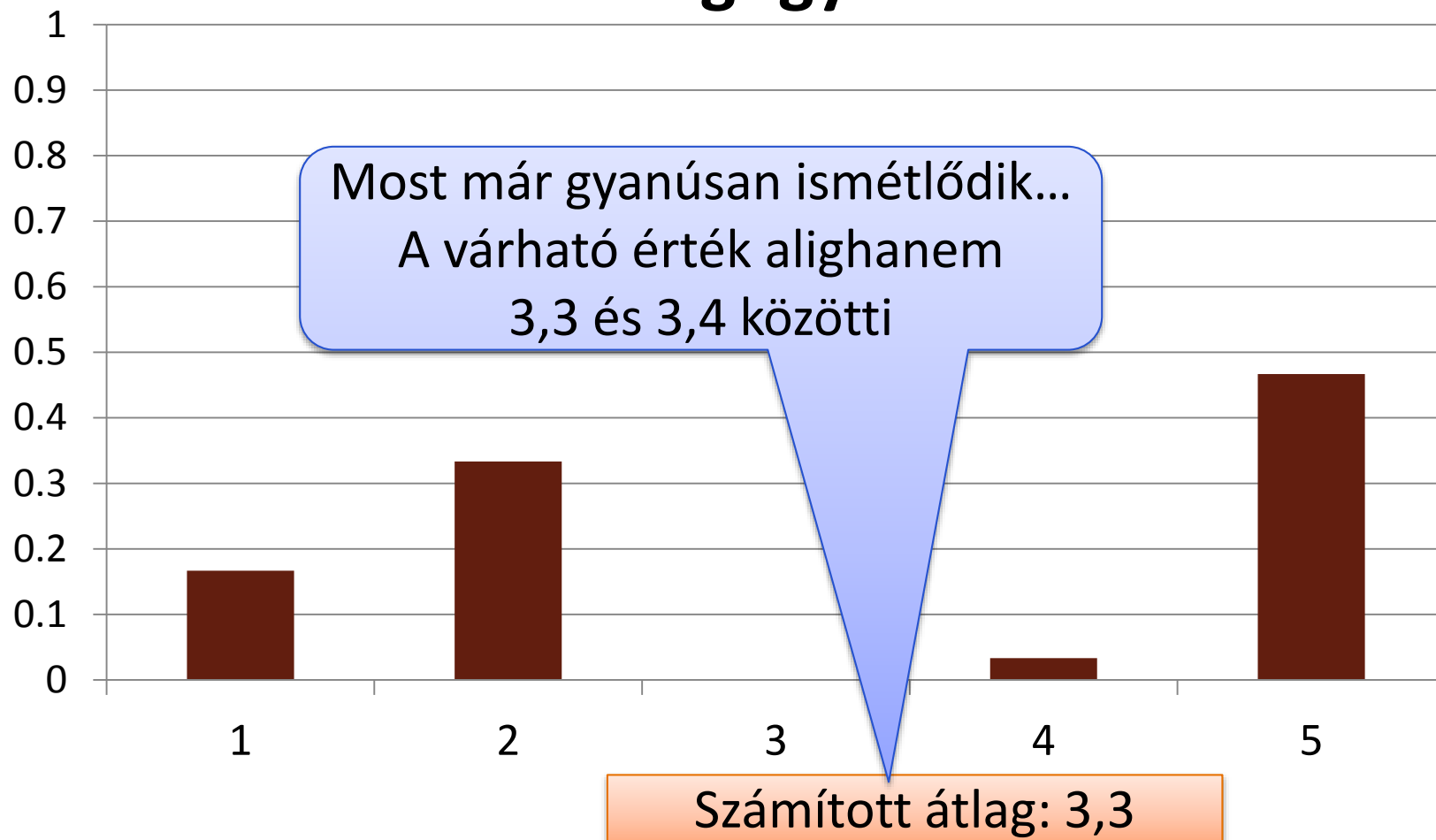


Számított átlag: 3,4

Számszerű jellemzők mérése

- Mérjük meg egy jellemző értékét! (1..5)

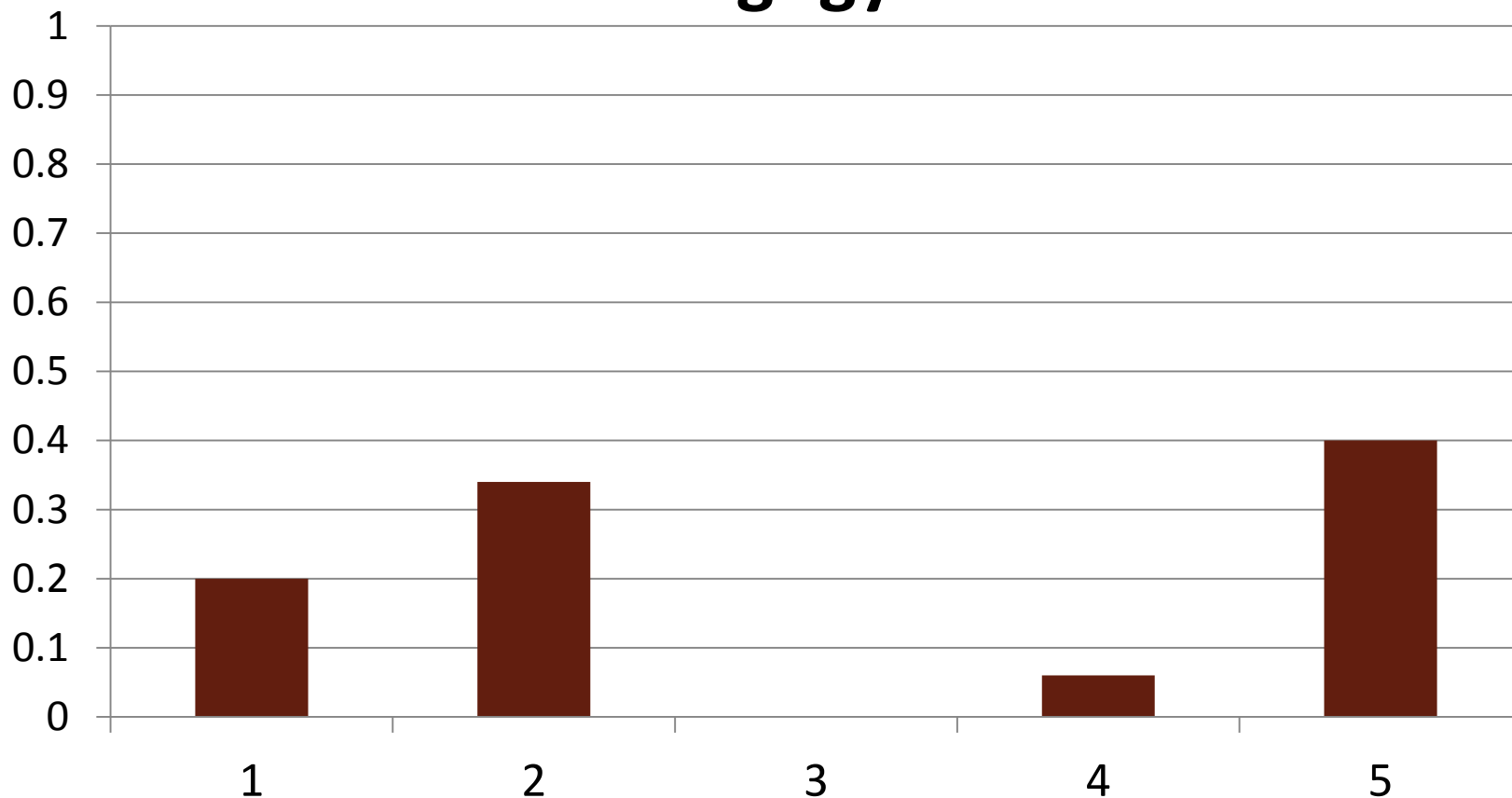
30 megfigyelés



Számszerű jellemzők mérése

- Mérjük meg egy jellemző értékét! (1..5)

50 megfigyelés



Számított átlag: 3,12

Számszerű jellemzők mérése

- Mérjük meg egy jellemző értékét! (1..5)

100 megfigyelés



Számított átlag: 2,88

Ismétlés: tapasztalati átlag, szórás

- Valószínűségi változó: E (vizsgálandó jelenség)
 - Várható érték: $\mu = E(X)$ átlagos viselkedés
 - Szórás: $\sigma = \sqrt{E(X - \mu)^2}$ (eltérések mértéke)
- Mintavétel: $x_1, x_2, \dots, x_t$ (mérések, megfigyelések)
 - Tapasztalati átlag: $\bar{x} = \frac{x_1 + x_2 + \dots + x_t}{t}$
 - Szórásra **nem jó**: $\sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_t - \bar{x})^2}{t}}$
 - Korrigált tapasztalati szórás: $\sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_t - \bar{x})^2}{t-1}}$

Tapasztalati átlag

- Módszer: számtani átlag képzése
 - Ismételt megfigyelések
 - Egymástól függetlenül
 - Azonos feltételek mellett
- Kérdések
 - Hány megfigyelést kell végezni?
 - A tapasztalati átlag mennyire jellemzi a valódi várható értéket?
- Először tisztázandó:
 - *A tapasztalati átlag eloszlása*

Tapasztalati átlag eloszlása

- Kísérlet = megfigyelések sorozata
- Megfigyelések sorozatának **tapasztalati átlaga**:
 - Egy jellemzőt t db független megfigyeléssel mérve,
 - majd a mért értékeket átlagolva kapott eredmény
- **Centrális határeloszlás tételéből** következik:
 - Tetszőleges eloszlású jellemző
 - (de legyen *véges* m várható értékű és s szórású)
 - tapasztalati átlaga $t \rightarrow \infty$ esetén közelítőleg
 - **normális eloszlású**,
 - $\mu = m$ várható értékkel és $\sigma = s/\sqrt{t}$ szórással

Tapasztalati átlag eloszlása

- Kísérlet = megfigyelések sorozata
- Megfigyelések sorozatának **tapasztalati átlaga**:
 - Egy jellemzőt t db független megfigyeléssel mérve,
Ökölszabály:
ismert szórásnál $t > 30$,
ismertetlen szórásnál $t > 100$ után kezd elfogadható lenni a közelítés
 - (de legyen *véges* m várható értékű és s szórású)
 - tapasztalati átlaga $t \rightarrow \infty$ esetén közelítőleg
 - **normális eloszlású**,
 - $\mu = m$ várható értékkel és $\sigma = s/\sqrt{t}$ szórással

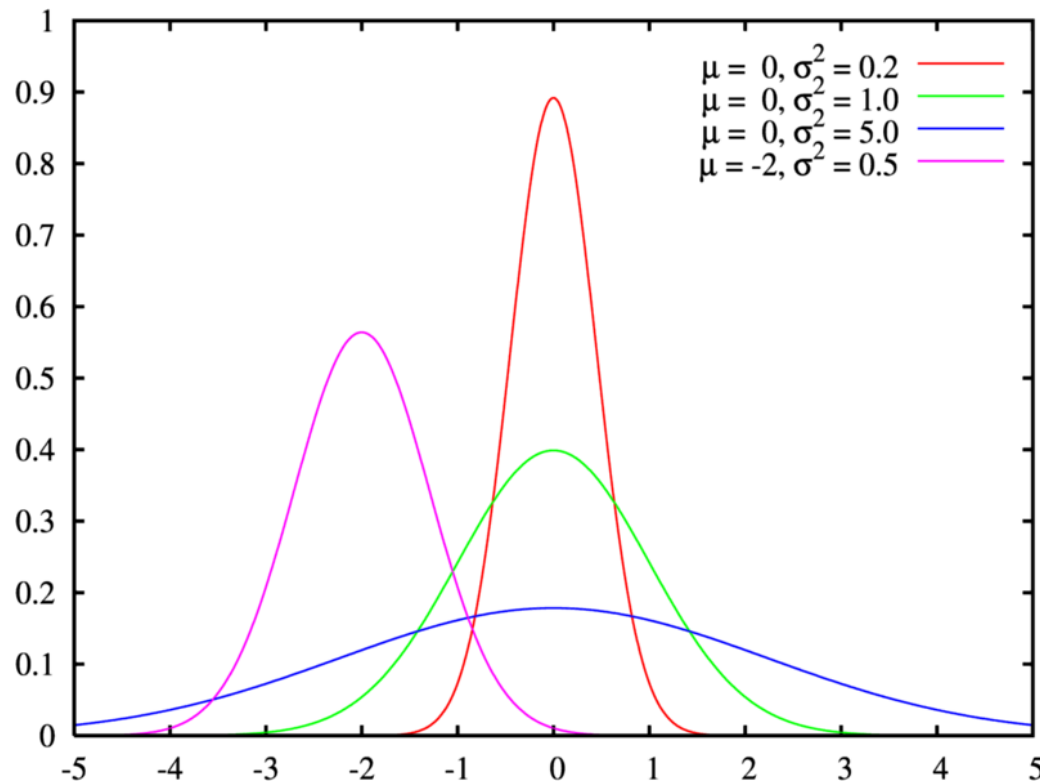
A normális (Gauss) eloszlás

- Valószínűsűrűség-függvénye: (nem kérdezzük)

$$f(\textcolor{red}{x}) = \frac{1}{\textcolor{blue}{\sigma}\sqrt{2\pi}} e^{-\frac{(\textcolor{red}{x}-\textcolor{violet}{\mu})^2}{2\textcolor{blue}{\sigma}^2}}$$

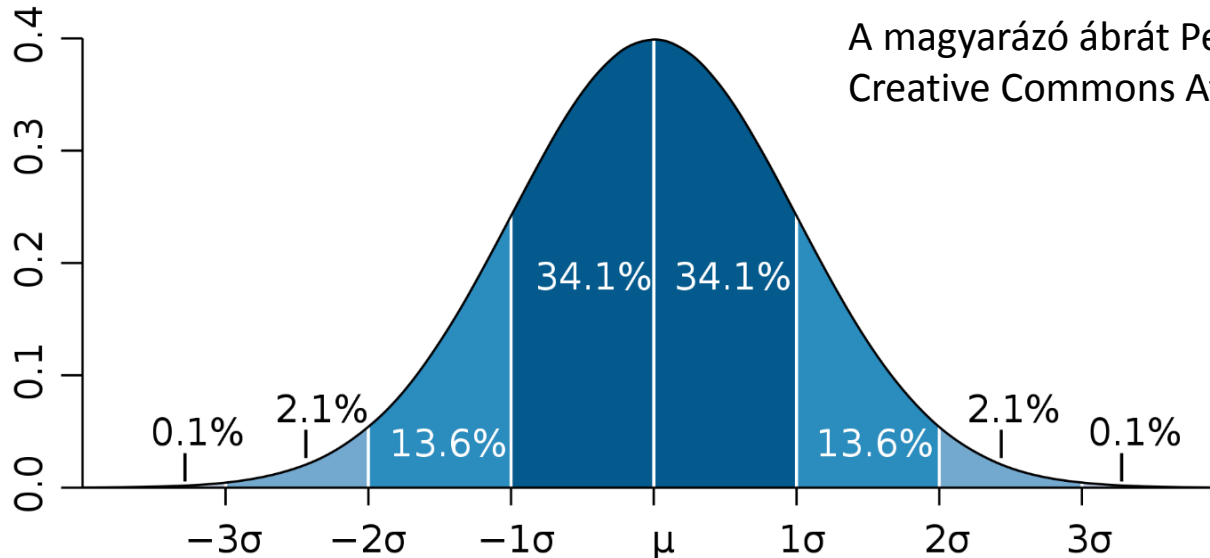
- Paraméterek

- Várható értéke μ
 - $\rightarrow m$ a mi esetünkben
- Szórása σ
 - $\rightarrow s/\sqrt{t}$ esetünkben



A normális (Gauss) eloszlás

- A várható érték körül koncentrálódnak



A magyarázó ábrát Petter Strandmark tette közzé
Creative Commons Attribution 2.5 Generic licenc alatt

- A normális eloszlású változó...
 - az esetek **68%**-ában legfeljebb **1 σ** messze kerül μ -től
 - az esetek **95%**-ában legfeljebb **2 σ** messze kerül μ -től
 - az esetek **99,7%**-ában legfeljebb **3 σ** messze kerül μ -től
 - ...

Konfidenciaintervallumok

- Ha tehát a tetszőleges eloszlású, s szórással vizsgált jellemzőről t db (>30) megfigyelést végzünk
- A tapasztalati átlagáról...
 - **68%** biztonsággal kijelenthető, hogy legfeljebb $s/\sqrt{t}$ pontatlansággal becsli m értékét
 - **95%** biztonsággal $2s/\sqrt{t}$ sugarú intervallumba esik
 - **99,7%** biztonsággal $3s/\sqrt{t}$ sugarú intervallumba esik
- És t növelésével gyökösen szűkül az intervallum

Konfidenciaintervallumok

- Ha tehát a tetszőleges eloszlású, s szórású vizsgált jellemzőről t db (>30) megfigyelés végzünk
- A **Konfidenciaszint** garáról...
 - **68%** biztonsággal kijelenthetjük, hogy a vizsgált jellemző a legfeljebb $s/\sqrt{t}$ pontatlansággal a **Konfidenciaintervallum sugara (félszélessége)** intervallumba esik
 - **95%** biztonsággal $2s/\sqrt{t}$ sugarú intervallumba esik
 - **99,7%** biztonsággal $3s/\sqrt{t}$ sugarú intervallumba esik
- És t növelésével gyökösen szűkül az intervallum

Kísérlettervezés példa

- A várható értékre 30 megfigyelés
 - Tapasztalati átlag: 2,3 s (jó-e ez? kell még mérni?)
 - Tapasztalati szórás: $s = 1,1$ s
- Cél
 - 99,7%-os konfidenciaintervallum 0,6 s széles legyen
- Kísérlettervezés
 - Elvárt sugár (félszélesség) = $3\sigma = 3s/\sqrt{t} < 0,3$ s
 - (ez a σ az átlag szórása, nem az eredeti mért jellemzőé!)
 - Ezért $t = 121$ megfigyelés kell legalább
- Hol a csalás?

Korrekcio

- Többször a tényleges eloszlás paraméterei *a priori* ismeretlenek (különben minek mérnénk?)
- Így nem használható fel a tényleges s szórás
- Csak a tapasztalati szórás használható → Gauss/normális helyett Student t-eloszlás
 - (más konfidenciaintervallumok)
- $t \rightarrow \infty$ esetén Student → normális
- Ökölszabály: $t > 100$ esetén használható a Gauss

Statisztikai próbák (ízelítő)

- Hipotézisek
 - **Nullhipotézis:** a világ X módon működik
 - **Alternatív hipotézis:** a világ nem X módon működik
- Kísérlet eredménye: Z
 - Ha az eltérés X -től nem írható a véletlen számlájára, akkor **szignifikáns** az eredmény
- p -próba (leegyszerűsítve)
 - $p = \mathbf{Pr}\{X \text{ esetén a kísérletnek } Z \text{ lesz az eredménye}\}$
 - Ha $p \ll 1$, akkor a nullhipotézist cáfoltuk
- A nullhipotézis sosem igazolható
 - Vagy szignifikánsan cáfoltuk, vagy nem

Statisztikai próbák (ízelítő)

■ Hipotézisek

- **Nullhipotézis:** a világ X módon működik
- **Alternatív hipotézis:** a világ nem X módon működik

■ Kísérlet a 7. előadás 7. feladatában

Pl.

- X és Y jellemző független valószínűségi változók; a nullhipotézis megnevezésére, vagy a vállalt SLA be van tartva

■ p -próba (leegyszerűsítve)

- $p = \mathbf{Pr}\{X \text{ esetén a kísérletnek } Z \text{ lesz az eredménye}\}$
- Ha $p \ll 1$, akkor a nullhipotézist cáfoltuk

■ A nullhipotézis sosem igazolható

- Vagy szignifikánsan cáfoltuk, vagy nem

Statisztikai próbák (ízelítő)

■ A/B próba

○ ,A' honlapterv esetén

- 300 mért látogatóból 4 vásárolt

○ ,B' honlapterv esetén

- 50 mért látogatóból 2 vásárolt

○ Mondhatjuk-e, hogy az ,B' jobb az ,A'-nál?

- Szignifikancia: $p=25\%$ $\rightarrow$ több mérés kell még
 - (itt most a magasabb p érték jelzi a szignifikánsabb eredményt)
- <http://web-tracking-guide.com/statistical-significance-calculator.html>