

Statisztika

Politológus képzés

Daróczi Gergely

Politológia Tanszék

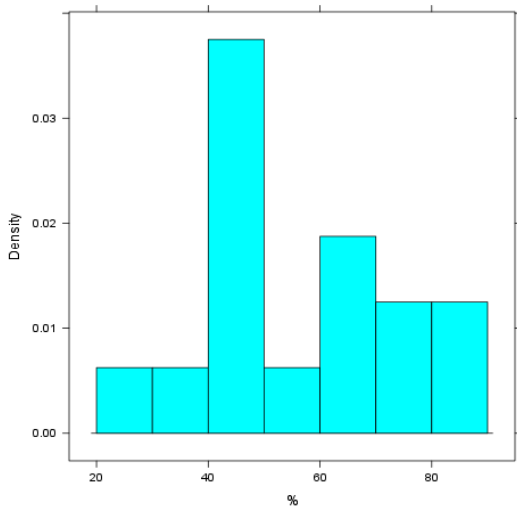
2012. április 17.



PÁZMÁNY PÉTER
KATOLIKUS EGYETEM
Bölcsészettudományi Kar

- 1 Leíró statisztikák
- 2 Középértékek
 - Példa
 - Esettanulmány
- 3 Szóródási mutatók
 - Példa
 - Esettanulmány
- 4 Néhány megjegyzés a grafikonokról
- 5 Számítások

Adatelemzés #1



Átlag: 56.9 %

Medián: 52 %

Módusz: 48 % és 78 %

Szórás: 18,8

Számos különböző középérték (átlag) áll a kutató rendelkezésére - azok előnyeivel és hátrányaival együtt:

- 1 **Számítási átlag:** $\frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$
- 2 **Mértani átlag:** $\sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 x_2 \dots x_n}$
- 3 **Módusz:** leggyakoribb érték
- 4 **Medián:** a statisztikai sokaságot kétfelé vágó érték
- 5 **Midrange:** $\frac{\max x + \min x}{2}$

Mely középérték(ek)kel jellemezné a következőket?

- 1 Egy szemeszterben gyűjtött jegyek.
- 2 A könyvtárban lévő hallgatók száma.
- 3 A legjellemzőbb hajszín a Campuson.
- 4 A budapesti lakosok fizetése.
- 5 Egy sítáborban elköltött eurók száma.

Értelmezze, szükség esetén javítsa a következő állításokat!

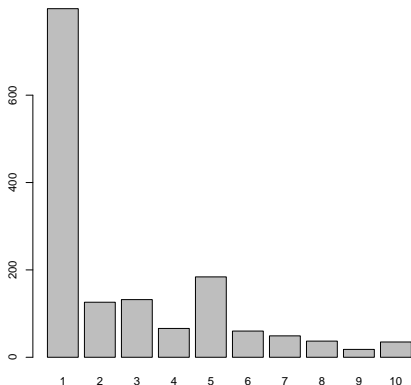
- 1 „1940 és 1948 között az átlagos heti juttatások összege 107 százalékkal emelkedett a United States Steel Corporation-ben.”
- 2 „Az átlagos fizetés \$ 5.000 körül mozgott 1942-ben.”
- 3 „Az autóbalesetben való elhalálozás valószínűsége kétszer nagyobb, mint repülőbalesetet szenvedni (+ statisztikus a repülőn).”
- 4 „Tamás IQ-ja 98, Lindáé pedig 101. Egy újabb példája annak, hogy a nők sokkal intelligensebbek a férfiaknál.”
- 5 „Ezen a héten feleannyit dolgoztam, mint tavaly ilyenkor.”

Milyen középértékekkel jellemezné a következő, European Values Study (Hungary, 2008)-ból származó változókat:

„Kérem mondja meg a következő állításokról, hogy azokat mindig megengedhetőnek tartja-e (10), vagy soha meg nem engedhetőnek tartja (10, vagy valami a kettő között.”

- 1 Állami juttatásokat jogtalanul igénybe venni
- 2 Abortusz
- 3 Válás
- 4 Jegy és bérlet nélkül utazni egy tömegközlekedési eszközön
- 5 Homoszexualitás

„Kérem mondja meg a következő állításokról, hogy azokat mindig megengedhetőnek tartja-e (10), vagy soha meg nem engedhetőnek tartja (10, vagy valami a kettő között).” – **Jegy és bérlet nélkül utazni egy tömegközlekedési eszközön**



Átlag: 2.751

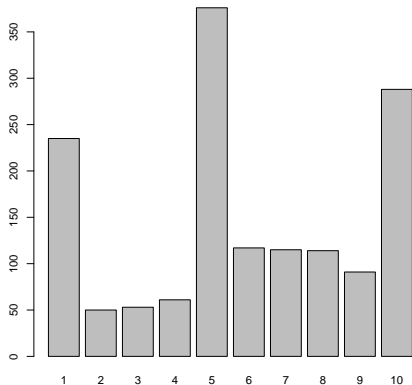
Módusz: 1

Medián: 1

Leíró statisztikák

Esettanulmány

„Kérem mondja meg a következő állításokról, hogy azokat mindig megengedhetőnek tartja-e (10), vagy soha meg nem engedhetőnek tartja (10, vagy valami a kettő között.” – **Válas**



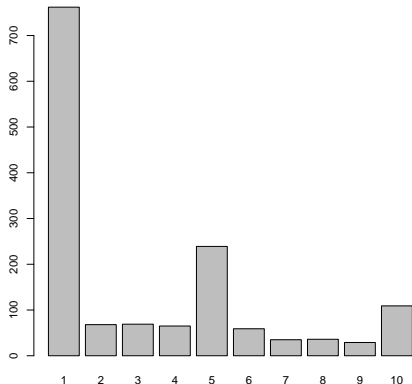
Átlag: 5.824

Módusz: 5

Medián: 5

„Kérem mondja meg a következő állításokról, hogy azokat mindig megengedhetőnek tartja-e (10), vagy soha meg nem engedhetőnek tartja (10, vagy valami a kettő között).” –

Homoszexualitás



Átlag: 3.261

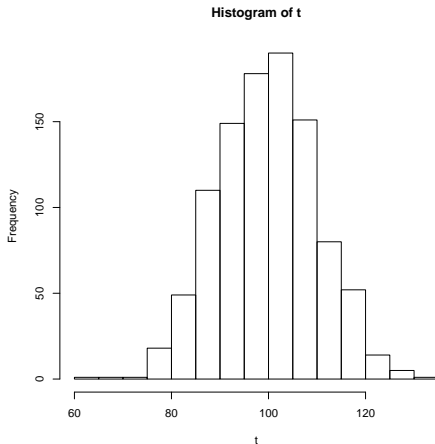
Módusz: 1

Medián: 1

Hallgatók közötti intelligencia-teszt eredménye:



Hallgatók közötti intelligencia-teszt eredménye:

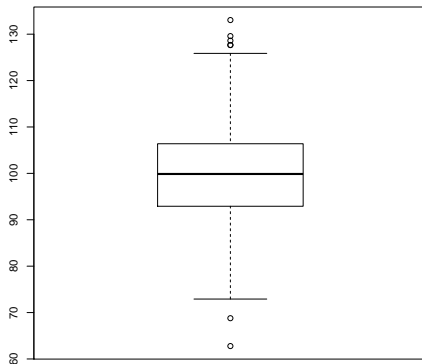


Átlag: 99.6

Módusz: 89.2

Medián: 99.8

Hallgatók közötti intelligencia-teszt eredménye:



Átlag: 99.6

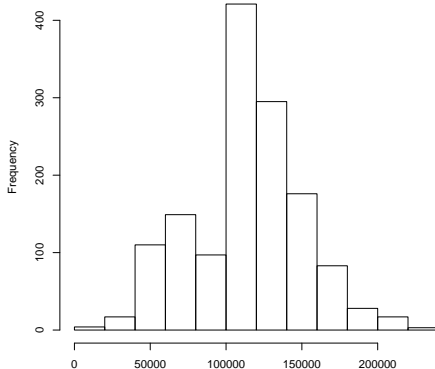
Módusz: 89.2

Medián: 99.8

Magyar munkavállalók fizetése:



Magyar munkavállalók fizetése:

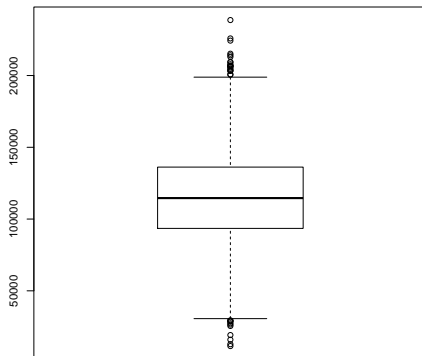


Átlag: 113721

Módusz: 72554

Medián:
114613

Magyar munkavállalók fizetése:



Átlag: 113721

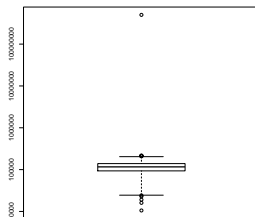
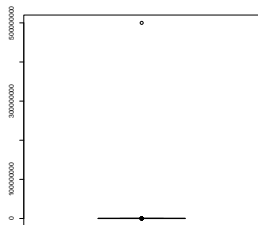
Módusz: 72554

Medián:
114613

Mi történik, ha a mintába egy kiemelkedően gazdag alany is bekerül?



Mi történik, ha a mintába egy kiemelkedően gazdag alany is bekerül?



Átlag: 471150

Módusz: 72554

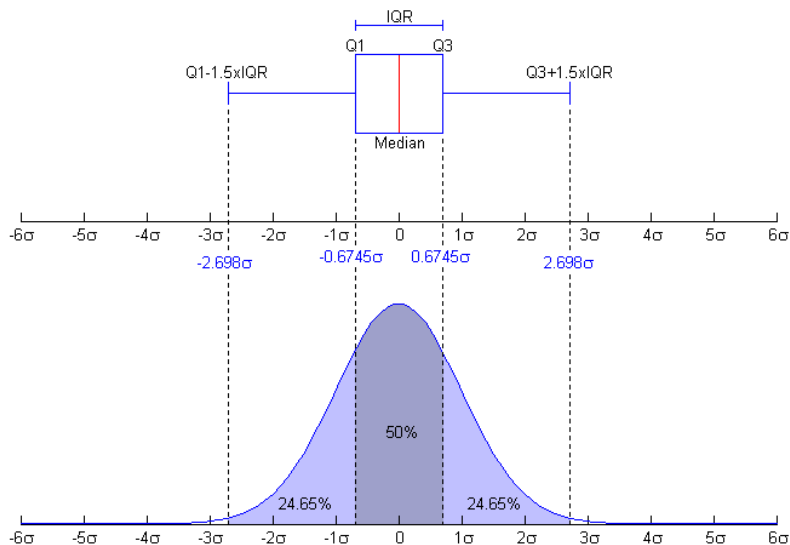
Medián:
116299

Számos különböző szóródási mutató áll a kutató rendelkezésére - azok előnyeivel és hátrányaival együtt. A homogenitás, variabilitás kiragadott mérőeszközei:

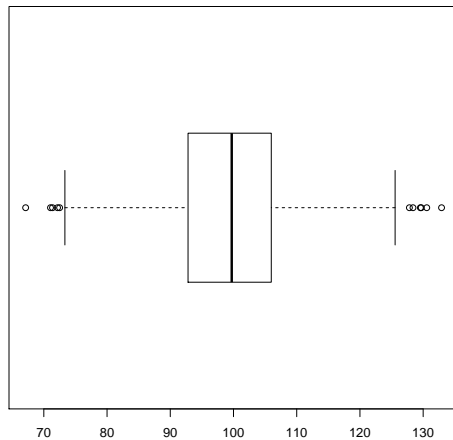
- 1 **Terjedelem:** $\max x - \min x$
- 2 **Empirikus szórás:** $\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{n-1}}$
- 3 **Variancia:** σ^2
- 4 A **kvartilisek** (Q1, Q2, Q3) négy egyenlő részre osztják a rendezett mintát: Q2 felezi a mintát (tehát megegyezik a mediánnal); Q1 a minta alsó felének, Q3 a felső felének közepe.
- 5 **Interkvartilis terjedeleme (IQR):** $Q_3 - Q_1$

Leíró statisztikák

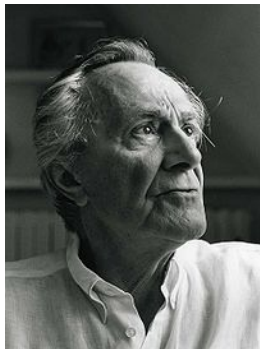
Interkvartilis terjedelem



Hallgatók közötti intelligencia-teszt eredménye:



Lyotard : A posztmodern állapot (1979)



- „vége a nagy elveszéléseknek”
- „anything goes”
- „posztmodern kor és posztmodern kultúra”

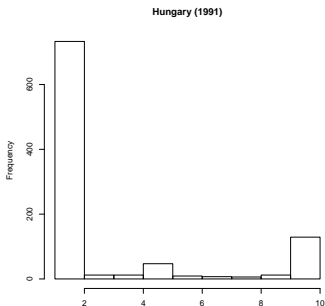
Vajon mi történik a normákkal?

„Kérem mondja meg a következő állításokról, hogy azokat mindig megengedhetőnek tartja-e (10), vagy soha meg nem engedhetőnek tartja (10, vagy valami a kettő között.” –

Homoszexualitás (1982-1991)

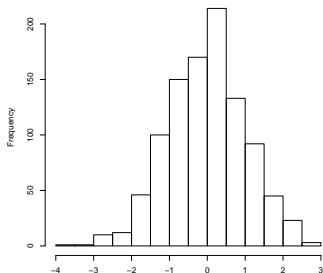


$$\bar{x} = 1.447407; \sigma = 1.419384$$

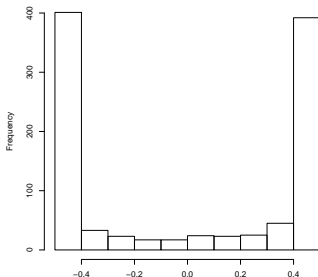


$$\bar{x} = 2.713547; \sigma = 3.230236$$

Mi magyarázza az átlagok alacsony különbségét?



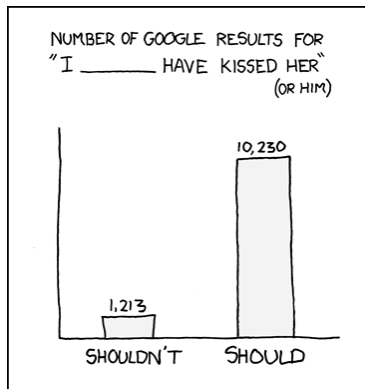
$$\bar{x} = -0.1; \sigma = 1.019$$



$$\bar{x} = 0; \sigma = 0.453$$

Mi a magyarázat arra, hogy az első példában magasabb a szórás?

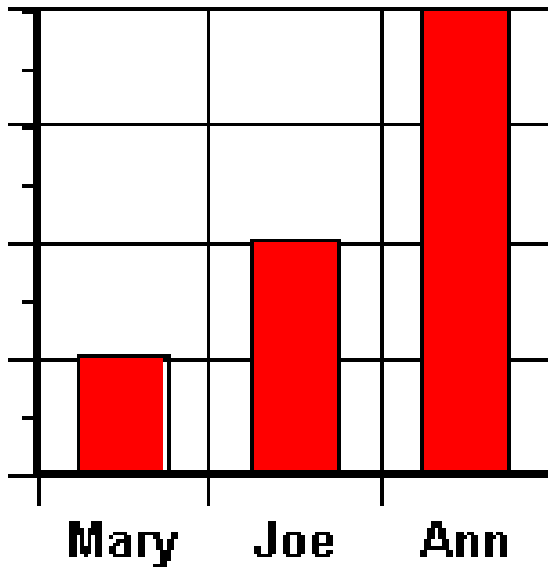
Egy eddig elhallgatott *leíró statisztikai* mutató: **szumma**



Milyen problémát sejtünk a fentebbi ábrához tartozó „kutatás” mögött?

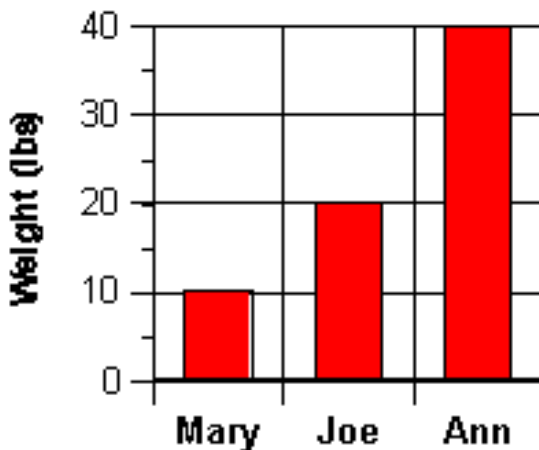
Néhány megjegyzés a grafikonokról

Palacsintás példa



Néhány megjegyzés a grafikonokról

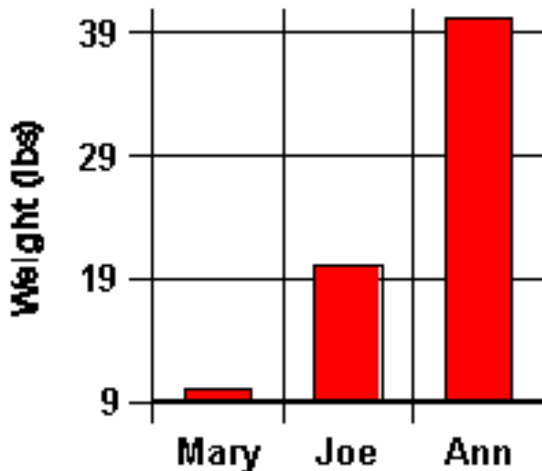
Palacsintás példa



Forrás: <http://faculty.washington.edu/chudler/stat3.html>

Néhány megjegyzés a grafikonokról

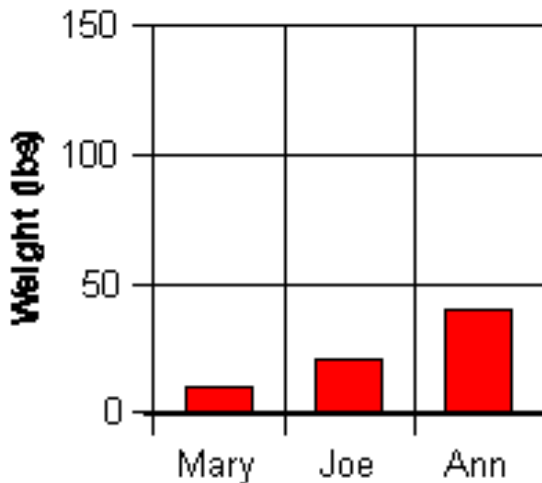
Palacsintás példa



Forrás: <http://faculty.washington.edu/chudler/stat3.html>

Néhány megjegyzés a grafikonokról

Palacsintás példa



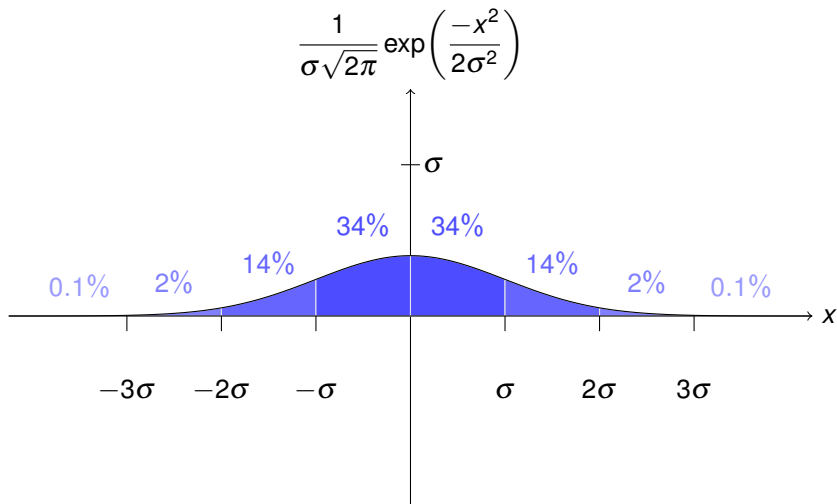
Forrás: <http://faculty.washington.edu/chudler/stat3.html>

Szükséges képletek:

- **számtani átlag:** $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ (becslőfüggvény)
- **korrigált empirikus szórás:** $\sigma = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}}$
- **standard/mintavételi hiba:** $SE = \frac{\sigma}{\sqrt{n}}$
- **konfidencia intervallum:** $\bar{x} \pm SE \cdot z$ ahol $z = 1.96$

Számítások

A várható értékről



standard normális eloszlás: $\bar{x} = 0, \sigma = 1$

Számítások

Egy nagyon egyszerű példa

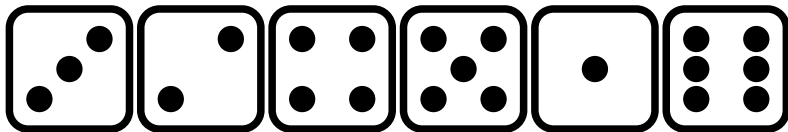
Játékszabály

Egy hatoldalú dobókockával gurítunk:

Ha a dobás páratlan, a játékos a dobott összeget megnyeri.

Amennyiben a dobás páros, úgy ő fizet két dollárt a banknak.

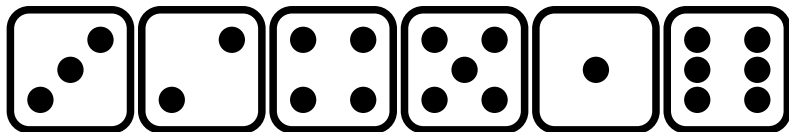
Adatgyűjtés: előzetes számítások helyett belevetjük magunkat a játékba, és a következő értékeket dobjuk.



Folytassuk a játékot?

A mintavételi hiba

Egy egyszerű példa



$$X = \{-2, 2, 4, -2, -2, 6\}$$

$$\bar{x} = \frac{-2+2+4+2+2+6}{6} = \frac{6}{6} = \frac{1}{1} = 1$$

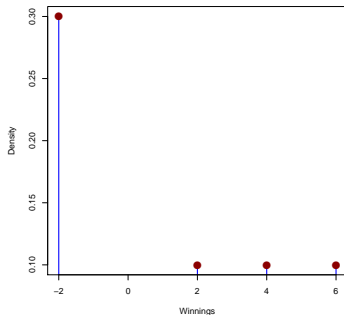
$$\sigma = \sqrt{\frac{(-2-1)^2 + (2-1)^2 + (4-1)^2 + (-2-1)^2 + (-2-1)^2 + (6-1)^2}{5}} =$$
$$= \sqrt{\frac{9+1+9+9+9+25}{5}} = \sqrt{\frac{62}{5}} = \sqrt{12.4} = 3.521363$$

$$SE = \frac{3.521363}{\sqrt{6}} = \frac{3.521363}{2.44949} = 1.437591$$

A várható érték valahol -1.87 és 3.87 között (CI)
található 95%-os döntési szinten.

Hogyan döntenénk?

Felejtkezzünk el az empirikus adatokról egy időre, és vizsgáljuk meg a **valódi** várható értéket!



Mi a probléma az ábrával?

Köszönöm a figyelmet!

Daróczi Gergely

daroczi.gergely@btk.ppke.hu