

Változók jellemzése (a minta eloszlása)



Szűcs Mónika, Griechisch Erika, Rárosi Ferenc
SZTE ÁOK-TTIK Orvosi Fizikai és Orvosi Informatikai Intézet

Utoljára frissítve: 2018. augusztus 31.



1. A leckében előforduló fogalmak, jelölések

Megfigyelési, mintavételi egység: A vizsgálat alanya vagy tárgya. A mintavételi egységeken végezzük el a vizsgálatokat, méréseket, megfigyeléseket. Orvosi gyakorlatban a leggyakoribb mintavételi egység egy ember, egy szerv, egy vérminta, más vizsgálatok esetén gyakori példa egy család, egy település, egy erdő, egy állat, egy faj, egy iskolai osztály.

Populáció: Az összes lehetséges mintavételi egységet tartalmazó halmaz, amiről szeretnénk információt nyerni a vizsgálat, kísérlet során. A populáció nagyon ritkán egy jól meghatározott véges halmaz, mint például egy ritka trópusi betegségben szenvedő 200 beteg (rajtuk kívül senki mást nem fertőzött meg ez a betegség, és már nem is fog). Gyakoribb azonban hogy a populáció nagysága végtelen, nem lehet az összes múlt, jelen és jövőbeli egyedet beazonosítani. *Biostatistikában, és ebben a leckében is populáció alatt nem magukat a mintavételi egységek sokaságát, hanem az azokon mért értékeket értjük.*

Minta: A populáció egy részhalmaza, azon mintavételi egységek (egyedek) melyeken a vizsgálat ténylegesen megtörtént. *Biostatistikában, és ebben a leckében is minta alatt nem magukat a megvizsgált mintavételi egységeket, hanem az azokon mért értékeket értjük.*

Adatbázis (adatmátrix): A minta adatait tartalmazza táblázatos formában. Minden sora egy-egy megfigyelési egységnek felel meg, oszlopaiban pedig a különböző mérések, megfigyelések szerepelnek. (Leggyakrabban egy megfigyelési egységhez NEM tartozhat két vagy több sor; különböző mérések, megfigyelések különböző oszlopokban kell hogy szerepeljenek.)

Változó: Az adatbázis oszlopait változóknak nevezzük.

2. Változók típusai

A változó típusa meghatározza a változóval végezhető műveleteket.

Ebben a leckében a véletlen változók **két nagy csoportjával** foglalkozunk: diszkrét és folytonos.

2.1. Diszkrét változó eloszlása

Azt mondjuk, hogy X *diszkrét változó* ha a lehetséges értékeinek száma véges, általában néhány kategória, pl. *nem, vércsoport, korcsoport*.

Egy *diszkrét változó eloszlása* megadja, a lehetséges értékeket és azok gyakoriságát.

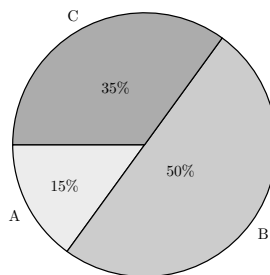
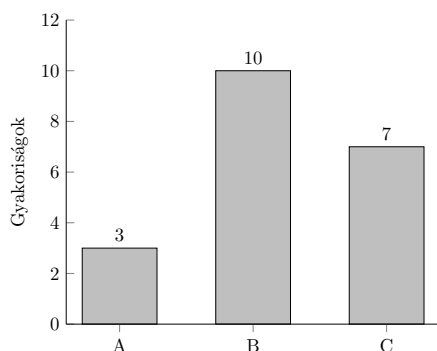
2.2. Diszkrét változó eloszlásának jellemzése

- számokkal

(abszolút) gyakorisági táblázat: a gyakoriság megadja, hogy a változó hányszor veszi fel az adott értéket.

relatív gyakorisági táblázat: az adott kimenetel gyakorisága osztva a teljes esetszámmal.

- ábrákkal
 - relatív gyakorisági/gyakorisági oszlopdiagram
 - relatív gyakorisági/gyakorisági kördiagram



2.3. Folytonos változó eloszlása

Azt mondjuk, hogy X folytonos változó, ha végtelen sok értéket vehet fel egy adott intervallumból, pl. koncentráció, hőmérséklet, vérnyomás, életkor.

Egy folytonos változó eloszlása megadja hogy melyek a lehetséges értékek, és ezek milyen gyakran esnek bizonyos intervallumokba.

2.4. Folytonos változók eloszlásának jellemzése

- Az eloszlás közepének mérőszámai

átlag számtani közép
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

medián az a szám, amelynél az adatok fele kisebb, amely tehát a rendezett adatsor közepe

Kiszámítása: rendezett mintában...

- páratlan mintaelemszám esetén a középső elem
- páros mintaelemszám esetén a középső két elem átlaga

módusz a leggyakrabban előforduló érték(ek)

- Az eloszlás szóródásának mérőszámai

terjedelem a legnagyobb (maximum) és a legkisebb (minimum) érték közötti különbség.

interkvartilis terjedelelem a felső és az alsó kvartilis különbsége. Ez az intervallum tartalmazza az adatok középső 50%-át. $IQR = Q_3 - Q_1$

- alsó kvartilis** (Q_1) az az érték, amelynél az adatok 25%-a kisebb
- felső kvartilis** (Q_3) az az érték, amelynél az adatok 75%-a kisebb.

SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE

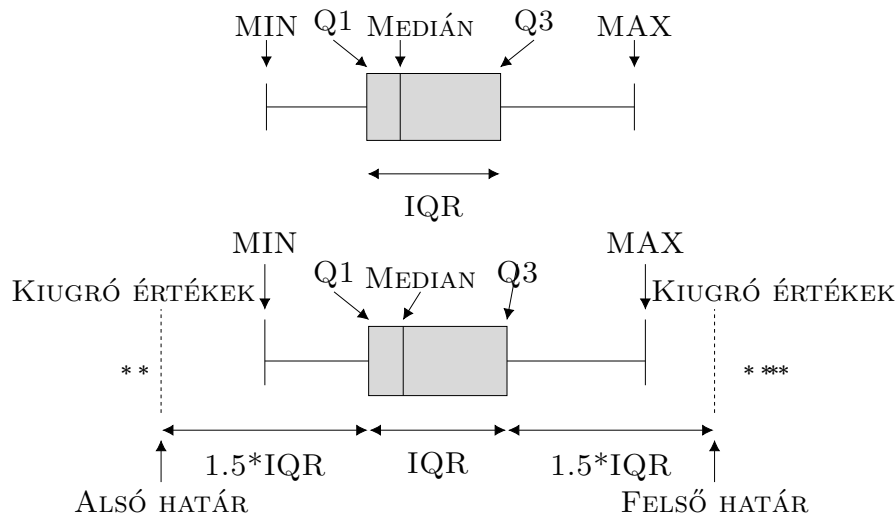
variância $\text{var} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$

szórás (standard deviáció) megadja a minta-adatok szóródását a mintaátlag körül.

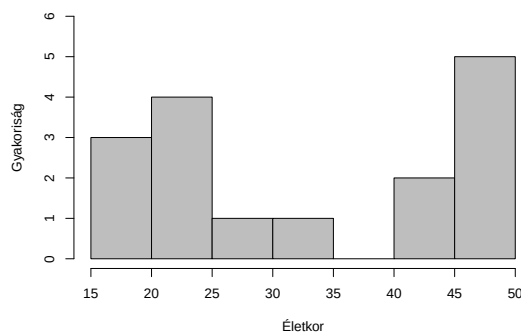
$$\text{sd} = \sqrt{\text{var}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

- ábrák

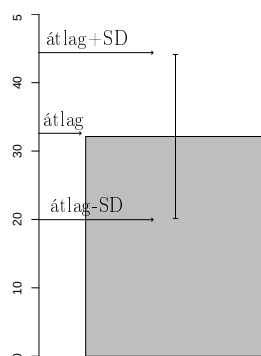
box-diagram (doboz-ábra)



hisztogram



átlag-szórás diagram



SZÉCHENYI 2020



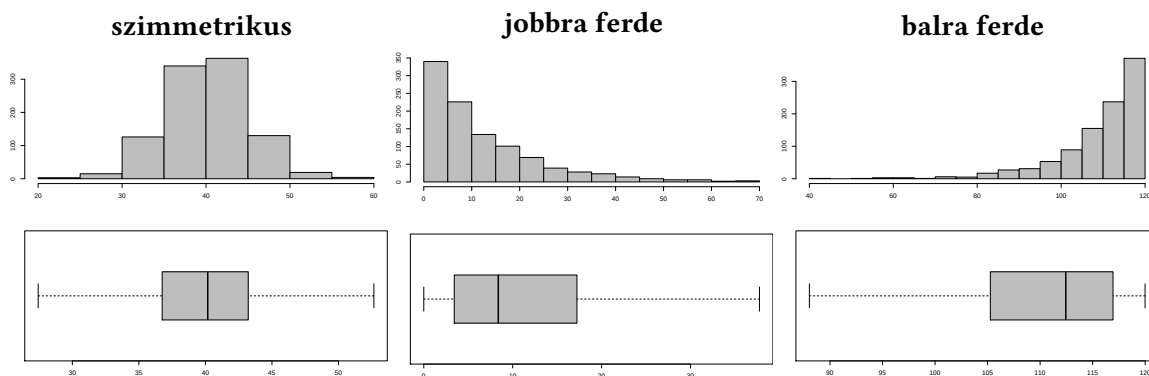
MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE

- Az eloszlás alakja



2.5. Az adatok transzformálása

Összeadás, kivonás Az adatokhoz ugyanazt a számot hozzáadva (kivonva) az eloszlás közepét jellemző mérőszámok a hozzáadott (kivont számmal) növekednek (csökkennek).

Az adatokhoz ugyanazt a számot hozzáadva (kivonva) az eloszlás szóródását jellemző mérőszámok nem változnak.

Szorzás, osztás Az adatokat ugyanazzal a számmal szorozva (osztva) a közép mérőszámai az adott számmal szorzódnak (osztódnak).

Az adatokat ugyanazzal a számmal szorozva (osztva) a szóródás mérőszámai a variancia kivételével az adott szám abszolút értékével szorzódnak (osztódnak), a variancia az adott szám négyzetével szorzódik (osztódik).

Felhasznált irodalom

- Reiczigél Jenő, Harnos Andrea, Solymosi Norbert: Biostatisztika nem statisztikusoknak, Pars Kft. (2014)
- Reiczigél Jenő: Válogatott fejezetek a biostatistikából, SZIE ÁOTK, (2005), <https://docplayer.hu/279567-Reiczigél-jenő-valogatott-fejezetek-a-biostatistikabol.html> (hozzáférés: 2018.08 31.)