

Korrelációs számítás



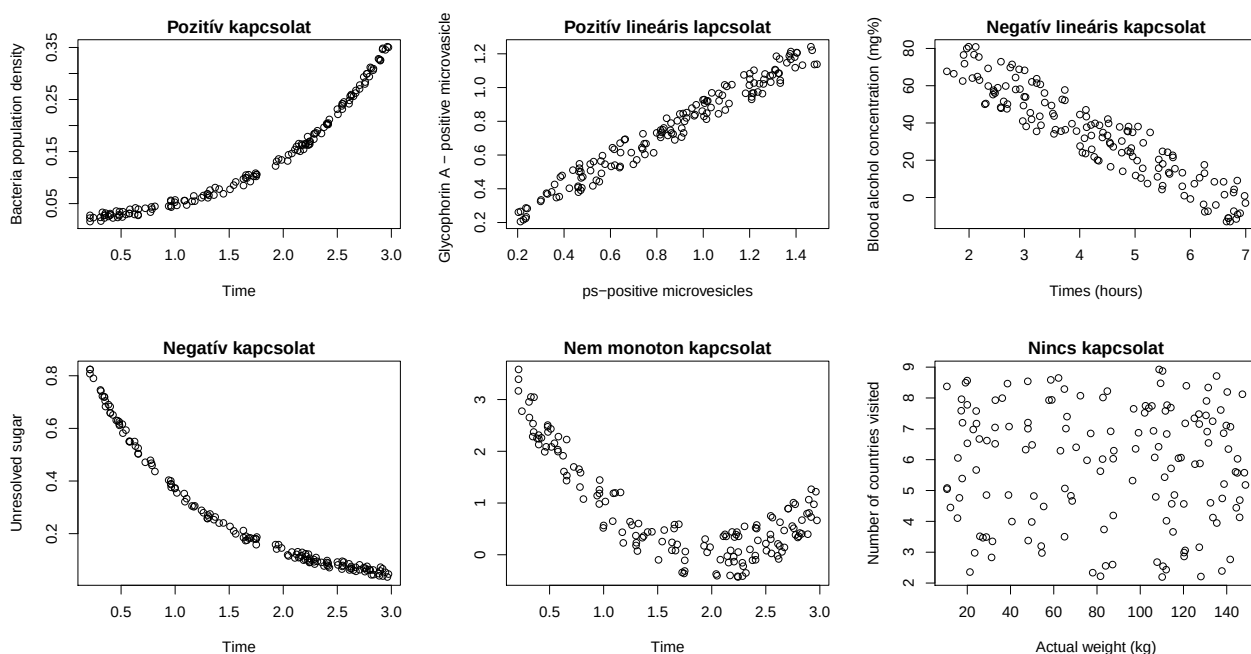
Szűcs Mónika, Griechisch Erika, Rárosi Ferenc
SZTE ÁOK-TTIK Orvosi Fizikai és Orvosi Informatikai Intézet

Utoljára frissítve: 2020. június 4.



Élettudományokkal foglalkozó kutatások során gyakran mérnek több folytonos változót ugyanazokon a mintavételi egységeken (pl.: *testtömeg, vérnyomás, vércukorszint, testmagasság*). Ilyen esetekben felmerül a kérdés, hogy van-e valamilyen kapcsolat a két változó között. A korrelációs számítás két folytonos változó közötti szimmetrikus kapcsolat vizsgálatával foglalkozik.

Lehetséges kapcsolatok két folytonos változó között



Jelen fejezetben csupán lineáris, illetve linearizálható kapcsolatokat vizsgálunk.

1. Pearson-féle korrelációs együttható

A **Pearson-féle korrelációs együtthatót** két **folytonos** változó közötti **lineáris kapcsolat** erősségének mérésre használjuk. A mintából számolt korrelációs együtthatót r -rel jelöljük.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

ahol x_i, y_i : a változók értékei; $\bar{x}, \bar{y}$: mintaátlagok; n : mintaelemszám

A populáció ismeretlen korrelációs együtthatóját ρ -val jelöljük.

SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE

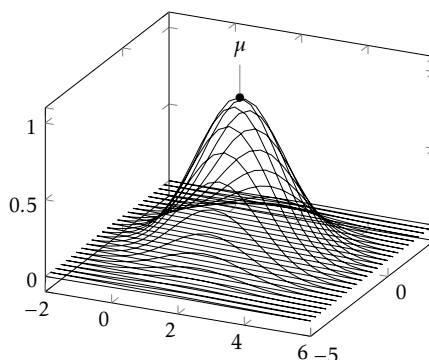
A korrelációs együttható tulajdonságai

- mértékegység nélküli
- $-1 \leq \rho \leq 1$
- ha a két változó független, akkor $\rho = 0$
- $\rho = 0$ esetén nem feltétlenül független a kettő változó. $\rho = 0$ esetén lehetséges, hogy a kettő változó független, de az is lehetséges, hogy van közöttük kapcsolat, de az nem lineáris (nem egyenes) jellegű.
- $\rho \approx \pm 1$ erős pozitív/negatív, lineáris kapcsolatot jelent
- $\rho \approx 0$ gyenge lineáris kapcsolatot jelent

2. Hipotézisvizsgálat a korrelációs együtthatóra

Célja megmutatni, hogy a két változó közötti lineáris összefüggés valódi, vagy csak a véletlen műve, azaz a populációra vonatkozó korrelációs együttható (ρ) eltér-e 0-tól.

Feltétele, hogy a minta kétdimenziós normális eloszlásból származik.



Hipotézisek:

- $H_0 : \rho = 0$, nincs lineáris kapcsolat a két változó között, a változók korrelálatlanok.
- $H_1 : \rho \neq 0$, van lineáris kapcsolat a két változó között.

Számolás és döntés:

I. Próbatasztika

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

Ha a null-hipotézis igaz, a próbatasztika $n - 2$ szabadságfokú t -eloszlást követ.

Döntési szabály:

- Ha $|t| < t_\alpha$, $t \in (-t_\alpha, t_\alpha)$, H_0 -t elfogadjuk, a korreláció nem szignifikáns α szinten, nincs elegendő okunk a két változó közötti lineáris kapcsolat feltételezésére.
- Ha $|t| > t_\alpha$, $t \notin (-t_\alpha, t_\alpha)$, H_0 -t elvetjük, a különbség szignifikáns α szinten, van elegendő okunk a két változó közötti lineáris kapcsolat kimutatására.

SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE

II. p -érték

- Ha $p > \alpha$, H_0 -t elfogadjuk, a korreláció nem szignifikáns α szinten, nincs elegendő okunk a két változó közötti lineáris kapcsolat feltételezésére.
- Ha $p < \alpha$, H_0 , H_0 -t elvetjük, a különbség szignifikáns α szinten, van elegendő okunk a két változó közötti lineáris kapcsolat kimutatására.

Számolás R-rel:

```
> cor(magassag,tomeg)
[1] 0.6726686
> cor.test(magassag,tomeg)
```

Pearson's product-moment correlation

```
data: magassag and tomeg
t = 7.3292, df = 65, p-value = 4.533e-10
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.5158040 0.7858922
sample estimates:
cor
0.6726686
```

Felhasznált irodalom

- Reiczigel Jenő, Harnos Andrea, Solymosi Norbert: Biostatisztika nem statisztikusoknak, Pars Kft. (2014)
- Reiczigel Jenő: Válogatott fejezetek a biostatistikából, SZIE ÁOTK (2005)
<http://www2.univet.hu/users/jreiczig/valfej/val-fej-jegyzet-2005-02-05.pdf> (2019.05.21.)

Jelen tananyag a Szegedi Tudományegyetemen készült az Európai Unió támogatásával.
Projekt azonosító: EFOP-3.4.3-16-2016-00014