Statisztika könyvek

Abari Kálmán

2021-08-30

# 1 Bevezetés

Jelen jegyzet célja a statisztika tanulmányok segítése. Számos segítő könyv, jegyzet, tutoriál és blog érhető el akár szabadon is az interneten, de a kereskedelmi forgalomban lévő anyagokhoz is számos ingyenesen hozzáférhető adatbázis vagy egyéb kiegészítés létezik. Ezeket a tartalmakat dolgozzuk fel, és tesszük elérhetővé az érdeklődők számára. Könyvünkben megkülönböztetett figyelmet szentelünk a pszichológia témakörének, de más tudományágak anyagai megjelenhetnek itt.

Jelen kötet egy együttműködés keretében jött létre, így számos szerzője van:

* Abari Kálmán

# 2 Ingyenes könyvek

* ONLINESTAT - [A könyv elérése](http://onlinestatbook.com/)  
  Online Statistics Education: A Multimedia Course of Study (<http://onlinestatbook.com/>). Project Leader: David M. Lane, Rice University.
* JASPDATA - [A könyv elérése](https://jasp-stats.org/wp-content/uploads/2020/05/The_JASP_Data_Library_1st_Edition.pdf)  
  Wagenmakers, E.-J., Kucharský, Š., & the JASP Team. (2020). The JASP Data Library (1st ed.). <https://doi.org/10.6084/m9.figshare.9980744>

# 3 Adatbázis példák a könyvekből

## 3.1 ONLINESTAT

**Hivatkozás a könyvre:**

Online Statistics Education: A Multimedia Course of Study (<http://onlinestatbook.com/>). Project Leader: David M. Lane, Rice University.

[A könyv elérése](http://onlinestatbook.com/)

### 3.1.1 A leniency adatbázis

#### Leírás

A mosoly növeli az engedékenységet? A különböző típusú mosolyok eltérően hatékonyak? Mosolyogjunk ha bajban vagyunk?

Bizonyítékok vannak arra, hogy a mosolygás enyhítheti az esetleges jogsértések megítélését. Ez a “mosoly-engedékenység” elnevezésű jelenség volt Marianne LaFrance és Marvin Hecht 1995-ös tanulmányának középpontjában.

A mosoly segít barátokat nyerni és befolyásolni az embereket. A mosolygás hatásaival kapcsolatos kutatások ezt alátámasztották, és kimutatták, hogy a mosolygó embert kellemesebbnek, vonzóbbnak, őszintébbnek, társaságkedvelőbbnek és hozzáértőbbnek ítélik meg, mint a nem mosolygó embert.

A vizsgálatban egy iskolai szabálytalanságot kellett megítélni a vizsgálati személyeknek, melynek leírásához egy fényképet is csatoltak az elkövetőről. A fényképen szereplő személy mosolya 4 kategória egyikébe esik: hamis mosoly, érzelmes mosoly, kényszerült mosoly, semleges “mosoly.” Az utóbbi neutrális arckifejezés kontrollként került a vizsgálatba. Minden vizsgálati személy pontosan egy feltételben szerepelt. Ez korlátozhatja az eredmények általánosíthatóságát.

#### Változók

* smile (faktor) - a fényképen szereplő személy mosolya. Lehetséges értékei:
  + 1 - hamis (false smile)
  + 2 - érzelmes (felt smile)
  + 3 - kényszerült (miserable smile)
  + 4 - semleges (neutral control)
* leniency (numerikus) - annak a mértéke, hogy mennyire voltak engedékenyek az ítéletek.

#### Hivatkozás

* LaFrance, M., & Hecht, M. A. (1995) Why smiles generate leniency. Personality and Social Psychology Bulletin, 21, 207-214.
* [Smiles and Leniency](https://onlinestatbook.com/2/case_studies/leniency.html)
* [One-Factor ANOVA (Between Subjects)](https://onlinestatbook.com/2/analysis_of_variance/one-way.html)

#### Kapcsolódó R sorok

# Beolvasás  
leniency <- read.csv(file = "adat/onlinestat/leniency.csv",  
 sep = ",",   
 dec = ".",  
 header = T,   
 quote = "",   
 comment.char = "",   
 fileEncoding = "UTF-8"  
 )  
# Típuskonverzió  
leniency$smile <- factor(leniency$smile,   
 levels = c("1", "2", "3", "4"),  
 labels = c("hamis", "érzelmes", "kényszerült", "semleges"))  
  
# Leíró statisztikai mutatók  
library(DescTools)  
Desc(formula = leniency~smile, data = leniency, plotit = F)  
#> ----------------------------------------------------------   
#> leniency ~ smile (leniency)  
#>   
#> Summary:   
#> n pairs: 136, valid: 136 (100.0%), missings: 0 (0.0%), groups: 4  
#>   
#>   
#> hamis érzelmes kényszerült semleges  
#> mean 5.368 4.912 4.912 4.118  
#> median 5.500 4.750 4.750 4.000  
#> sd 1.827 1.681 1.454 1.523  
#> IQR 2.875 2.375 1.500 1.875  
#> n 34 34 34 34  
#> np 25.000% 25.000% 25.000% 25.000%  
#> NAs 0 0 0 0  
#> 0s 0 0 0 0  
#>   
#> Kruskal-Wallis rank sum test:  
#> Kruskal-Wallis chi-squared = 9.1747, df = 3, p-value = 0.02706

# Leíró statisztikai mutatók  
library(summarytools)  
st\_options("headings", FALSE)  
stby(data = leniency$leniency, INDICES = leniency$smile,   
 FUN = descr,  
 stats = c("n.valid", "mean", "med", "sd", "iqr"),  
 transpose = FALSE,  
 style="rmarkdown",  
 caption="Leíró statisztika")

Leíró statisztika

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | hamis | érzelmes | kényszerült | semleges |
| N.Valid | 34.00 | 34.00 | 34.00 | 34.00 |
| Mean | 5.37 | 4.91 | 4.91 | 4.12 |
| Median | 5.50 | 4.75 | 4.75 | 4.00 |
| Std.Dev | 1.83 | 1.68 | 1.45 | 1.52 |
| IQR | 2.88 | 2.38 | 1.50 | 1.88 |

# Ábra - dobozdiagram  
library(ggplot2)  
ggplot(data = leniency, mapping = aes(x=smile, y=leniency, fill=smile)) +   
 geom\_violin(trim = FALSE) +   
 geom\_boxplot(alpha=0) +   
 geom\_jitter(height = 0, width = 0.1) +   
 theme(legend.position = "none")

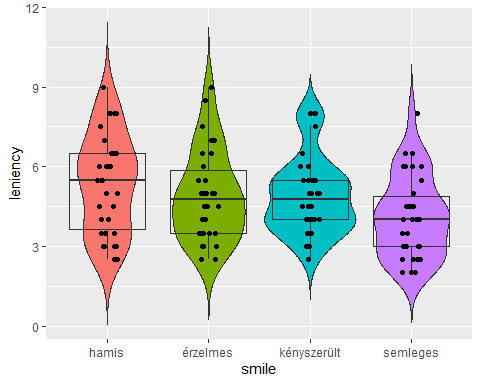


Figure 3.1: Dobozdiagram

# Hipotézisvizsgálat  
leniency\_model <- aov(leniency~smile, data = leniency)  
summary(leniency\_model)  
#> Df Sum Sq Mean Sq F value Pr(>F)   
#> smile 3 27.5 9.178 3.465 0.0182 \*  
#> Residuals 132 349.7 2.649   
#> ---  
#> Signif. codes:   
#> 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
TukeyHSD(leniency\_model)  
#> Tukey multiple comparisons of means  
#> 95% family-wise confidence level  
#>   
#> Fit: aov(formula = leniency ~ smile, data = leniency)  
#>   
#> $smile  
#> diff lwr upr  
#> érzelmes-hamis -0.4558824 -1.483012 0.5712478  
#> kényszerült-hamis -0.4558824 -1.483012 0.5712478  
#> semleges-hamis -1.2500000 -2.277130 -0.2228699  
#> kényszerült-érzelmes 0.0000000 -1.027130 1.0271301  
#> semleges-érzelmes -0.7941176 -1.821248 0.2330125  
#> semleges-kényszerült -0.7941176 -1.821248 0.2330125  
#> p adj  
#> érzelmes-hamis 0.6562329  
#> kényszerült-hamis 0.6562329  
#> semleges-hamis 0.0102192  
#> kényszerült-érzelmes 1.0000000  
#> semleges-érzelmes 0.1888804  
#> semleges-kényszerült 0.1888804  
library(DescTools)  
DunnettTest(leniency~smile, data = leniency,   
 control = "semleges")  
#>   
#> Dunnett's test for comparing several treatments with a control :   
#> 95% family-wise confidence level  
#>   
#> $semleges  
#> diff lwr.ci upr.ci pval  
#> hamis-semleges 1.2500000 0.3119186 2.188081 0.0055  
#> érzelmes-semleges 0.7941176 -0.1439637 1.732199 0.1162  
#> kényszerült-semleges 0.7941176 -0.1439637 1.732199 0.1164  
#>   
#> hamis-semleges \*\*   
#> érzelmes-semleges   
#> kényszerült-semleges   
#>   
#> ---  
#> Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## 3.2 JASPDATA

**Hivatkozás a könyvre:**

Wagenmakers, E.-J., Kucharský, Š., & the JASP Team. (2020). The JASP Data Library (1st ed.). <https://doi.org/10.6084/m9.figshare.9980744>

[A könyv elérése](https://jasp-stats.org/wp-content/uploads/2020/05/The_JASP_Data_Library_1st_Edition.pdf)

### 3.2.1 A sleep adatbázis

#### Leírás

Ez a híres “Sleep” (Alvás) adatmátrix, amelyet Student és Fisher is használt cikkeiben. A vizsgálatban 10 személy szerepelt, akik mindegyike 2 altatót is kapott két különböző időpontban. Az adatbázisban azokat a plusz órákat rögzítettük, amelyeket az egyes altatóknak köszönhetünk.

#### Változók

* extra (numerikus) - Az extra órák száma az adott altató hatására a placebó altatóhoz képest.
* group (faktor) - Az altató típusa. Lehetséges értéke: 1 és 2.
* ID (faktor) - A személy azonosítója.

Az adatbázisban a páros minta hosszú formában van eltárolva.

#### Hivatkozás

* Cushny, A. R and Peebles, A. R. (1905). The action of optical isomers. The Journal of Physiology, 32: 501-510.
* [A könyvben a 21–27. oldalak](https://jasp-stats.org/wp-content/uploads/2020/05/The_JASP_Data_Library_1st_Edition.pdf)

#### Kapcsolódó R sorok

# Beolvasás   
sleep <- read.table(file = "adat/jaspdata/sleep.csv",   
 sep = ",",   
 dec = ".",  
 header = T,   
 quote = "",   
 comment.char = "",   
 fileEncoding = "UTF-8"  
 )  
# Típuskonverzió  
sleep$group <- factor(sleep$group)  
sleep$ID <- factor(sleep$ID)  
str(sleep)  
#> 'data.frame': 20 obs. of 3 variables:  
#> $ extra: num 0.7 -1.6 -0.2 -1.2 -0.1 3.4 3.7 0.8 0 2 ...  
#> $ group: Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...  
#> $ ID : Factor w/ 10 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 10 ...  
  
# Leíró statisztikai mutatók  
library(DescTools)  
Desc(formula = extra~group, data = sleep, plotit = F)  
#> ----------------------------------------------------------   
#> extra ~ group (sleep)  
#>   
#> Summary:   
#> n pairs: 20, valid: 20 (100.0%), missings: 0 (0.0%), groups: 2  
#>   
#>   
#> 1 2  
#> mean 0.750 2.330  
#> median 0.350 1.750  
#> sd 1.789 2.002  
#> IQR 1.875 3.275  
#> n 10 10  
#> np 50.000% 50.000%  
#> NAs 0 0  
#> 0s 1 0  
#>   
#> Kruskal-Wallis rank sum test:  
#> Kruskal-Wallis chi-squared = 3.4378, df = 1, p-value = 0.06372

# Leíró statisztikai mutatók  
library(psych)  
temp <- describeBy(x = extra~group, data = sleep, mat=T, fast=T, digits = 2)  
knitr::kable(temp, caption = "Leíró statisztikai mutatók")

Table 3.1: Leíró statisztikai mutatók

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | item | group1 | vars | n | mean | sd | min | max | range | se |
| X11 | 1 | 1 | 1 | 10 | 0.75 | 1.79 | -1.6 | 3.7 | 5.3 | 0.57 |
| X12 | 2 | 2 | 1 | 10 | 2.33 | 2.00 | -0.1 | 5.5 | 5.6 | 0.63 |

library(flextable)  
library(magrittr)  
temp %>% flextable() %>% autofit()

| item | group1 | vars | n | mean | sd | min | max | range | se |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 10 | 0.75 | 1.79 | -1.6 | 3.7 | 5.3 | 0.57 |
| 2 | 2 | 1 | 10 | 2.33 | 2.00 | -0.1 | 5.5 | 5.6 | 0.63 |

# Leíró statisztikai mutatók  
library(summarytools)  
st\_options("headings", FALSE)  
with(sleep,   
 stby(data = extra,   
 INDICES = group,   
 FUN = descr,  
 stats = c("n.valid", "mean", "sd",   
 "min", "med", "max"),  
 transpose = FALSE,  
 style="rmarkdown",  
 caption="Leíró statisztika"  
 ))

Leíró statisztika

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| N.Valid | 10.00 | 10.00 |
| Mean | 0.75 | 2.33 |
| Std.Dev | 1.79 | 2.00 |
| Min | -1.60 | -0.10 |
| Median | 0.35 | 1.75 |
| Max | 3.70 | 5.50 |

# Ábra  
library(ggplot2)  
ggplot(data = sleep, mapping = aes(x=group, y=extra, fill=group)) +   
 geom\_violin(trim = FALSE) +   
 geom\_boxplot(alpha=0) +   
 geom\_jitter(height = 0, width = 0.1) +   
 theme(legend.position = "none")

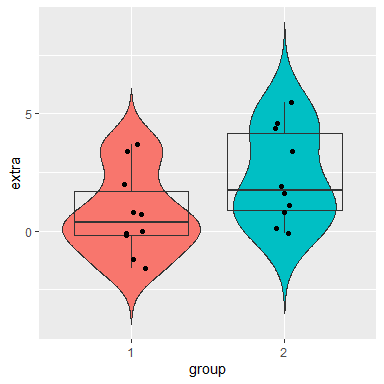


Figure 3.2: Dobozdiagram

# Ábra  
library(ggplot2)  
ggplot(data = sleep, mapping = aes(x=group, y=extra, fill=group)) +   
 geom\_boxplot() +   
 geom\_line(aes(group=ID), alpha=0.7) +   
 geom\_point() +   
 theme(legend.position = "none")

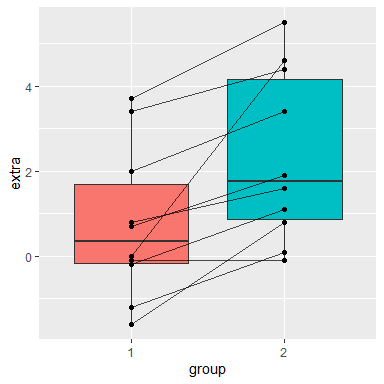


Figure 3.3: Dobozdiagram

# Hipotézisvizsgálat  
library(lsr)  
pairedSamplesTTest(formula = extra~group, data = sleep, id = "ID")  
#>   
#> Paired samples t-test   
#>   
#> Outcome variable: extra   
#> Grouping variable: group   
#> ID variable: ID   
#>   
#> Descriptive statistics:   
#> 1 2 difference  
#> mean 0.750 2.330 -1.580  
#> std dev. 1.789 2.002 1.230  
#>   
#> Hypotheses:   
#> null: population means equal for both measurements  
#> alternative: different population means for each measurement  
#>   
#> Test results:   
#> t-statistic: -4.062   
#> degrees of freedom: 9   
#> p-value: 0.003   
#>   
#> Other information:   
#> two-sided 95% confidence interval: [-2.46, -0.7]   
#> estimated effect size (Cohen's d): 1.285