

Leíró és matematikai statisztika gyakorlat

2018/2019 II. félév

beadandó feladat

A statisztikai elemzés dokumentációját LaTeX-ben vagy Word-ben kell megírni, az elemzésekhez szükséges programot pedig az R statisztikai programcsomag segítségével kell elkészíteni. A megoldásokat elegendő email-ben elküldeni, a szöveges részt .pdf, a forráskódot pedig .txt, illetve a felhasznált adatokat .RData formátumban.

Beadási határidő: 2019. május 12.

Az elemzés vázlatos felépítése:

1. A vizsgált adatok rövid bemutatása.
2. A elemzéshez használt adatsorok áttekintése, egyszerű ábrák készítése. Esetleges adathibák kiszűrése.
3. Az adatsorok egyedi elemzése: alapstatisztikák kiszámítása, szöveges értelmezése; grafikonok készítése (hisztogram, boxplot).
4. Normalitásvizsgálat (Q-Q plot, Shapiro–Wilk-próba).
5. Hipotézisvizsgálat.

Formai követelmények: a beadandó legyen összefüggő, tömör, lényegre törő, mintha egy néhány oldalas jelentés lenne. A grafikonoknak legyen címe es tengelyfelirataik (ahol értelmes). A táblázatok legyenek áttekinthetőek és szintén legyen címük. Az R kód legyen áttekinthető, rendezett, és legyen benne rövid magyarázat.

Néhány adatbázis, ahol érdemes adatokat gyűjteni a beadandóhoz:

- EU-statisztikák:
<https://ec.europa.eu/eurostat/data/browse-statistics-by-theme>
- Kaggle:
<https://www.kaggle.com/datasets>
- Földtani adatok:
<https://data.usgs.gov/datacatalog/>
<https://www.bgs.ac.uk/opengeoscience/>
- R adatbázisok:
<https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.html>