**TÁRGYTEMATIKA**

**A tárgy neve: Adatelemzés és statisztikai következtetés**

**A tárgy kódja: PSZB17-210**

**Tárgyfelelős: Dr. Nagy Tamás**

**Kurzus-specifikus információk**

|  |
| --- |
| **A kurzus alapadatai** |

**A kurzus specifikus neve:** Adatelemzés és statisztikai következtetés - gyakorlat

**A kurzus kódja:** PSZB17-210-6

**A kurzus időpontja és helyszíne:** Szerda 15:00-17:30 IZU 422

**Az oktató neve:** Dr. Kekecs Zoltán

**Az oktató tanszéke:** Affektív Pszichológia Tanszék, PPK

**Az oktató email címe:** [kekecs.zoltan@ppk.elte.hu](mailto:kekecs.zoltan@ppk.elte.hu)

**Fogadóóra időpontja és helyszíne:** Hétfő, 10:00 – 11:00 IZU 123/125 vagy Teams-en (előzetes egyeztetés szükséges email-ben)

**Tanszéki demonstrátorok:** Palágyi Árpád, Beregrai Vivien

|  |
| --- |
| **Részletes tematika** |

1. **óra  - Február 14.**

Óramegbeszélés, bevezetés az R-be

Gyakorláshoz ajánlott: DataCamp: [Introduction to R (Linkek egy külső oldalra)](https://www.datacamp.com/courses/free-introduction-to-r) (Chapter 1-2: Basics, Vectors)

1. **óra - Február 21.**

Adatmenedzsment az R-ben

Gyakorláshoz ajánlott: DataCamp: [Introduction to tidyverse (Linkek egy külső oldalra)](https://www.datacamp.com/courses/introduction-to-the-tidyverse) (Chapter 1 - Data Wrangling, Chapter 3 - Grouping and summarizing)

Hasznos anyag: [Tidyverse Cheet Sheet (Linkek egy külső oldalra)](https://datacamp-community-prod.s3.amazonaws.com/e63a8f6b-2aa3-4006-89e0-badc294b179c)

1. **óra – Február 28.**

Adatvizualizáció (eloszlásfüggvények, hisztogram, pont-, oszlop-, vonal- és dobozdiagramok), eloszlások, kvantilisek, adatok összesítése, gyakoriság és százalék, középértékek és szóródás,

Gyakorláshoz ajánlott: DataCamp: [Introduction to the tidyverse (Linkek egy külső oldalra)](https://www.datacamp.com/courses/introduction-to-the-tidyverse) (Chapter 2 - Data visualization, Chapter 4 - Types of visualization)

1. **óra – Március 6.**

Feltáró adatelemzés, adattisztítás, egyváltozós és többváltozós adatvizualizáció. Faktorok.

Gyakorláshoz ajánlott: DataCamp: [Exploratory data analysis (Linkek egy külső oldalra)](https://www.datacamp.com/courses/exploratory-data-analysis) in R (Chapter 1 - Exploring Categorical Data, Chapter 2 - Exploring Numerical Data)

1. **óra – Március 13.**

Statisztikai inferencia, p-érték, konfidencia intervallum, két változó együttjárása; t-teszt, Khi-négyzet próba, Pearson- és Spearman-féle korrelációk, ezen kapcsolatok vizualizációja

Gyakorláshoz ajánlott: DataCamp: [Inference for Numerical Data (Linkek egy külső oldalra)](https://www.datacamp.com/courses/inference-for-numerical-data) in R (Chapter 1, 2, 3);

1. **óra – Március 20.** - **Első ZH**

**Április 4. – Pót ZH** **(8:30-11:00)** – IZU121 – (Aki pót ZH-zni akar, hoznia kell saját gépet amin fut az R és tud csatlakozni az ELTE wifihez!)

1. **óra – Április 3.**

Lineáris regresszió; regressziós egyenlet; reziduális hiba; modell hatékonyság

Gyakorláshoz ajánlott: DataCamp: [Introduction to regression in R, (Chapters 1, 2)](https://app.datacamp.com/learn/courses/introduction-to-regression-in-r)

1. **óra – Április 10.**

Speciális prediktorok és azok értelmezése: kategorikus változók (átlagok összehasonlítása), interakciók, nem-lineáris összefüggések modellezése

Gyakorláshoz ajánlott: DataCamp: Intermediate [Statistical Modeling in R (Part 2) (Linkek egy külső oldalra)](https://www.datacamp.com/courses/statistical-modeling-in-r-part-2) (Chapter 1)

1. **óra – Április 17.**

Ismételt mérések összehasonlítása

Gyakorláshoz ajánlott: DataCamp: [Hierarchical and Mixed Effects Models (Linkek egy külső oldalra)](https://www.datacamp.com/courses/hierarchical-and-mixed-effects-models) in R (Chapter 1, 2, 4)

1. **óra – Április 24.**

Nem folytonos kimeneti változók: Klasszifikáció, Generalizált Lineáris regresszió, Binomiális logisztikus regresszió

Gyakorláshoz ajánlott: DataCamp: [Introduction to regression in R, (Chapters 4)](https://app.datacamp.com/learn/courses/introduction-to-regression-in-r), [Intermediate Regression in R](https://app.datacamp.com/learn/courses/intermediate-regression-in-r) (Chapter 4).

1. **óra – Május 1. (Nemzeti ünnep, nincs óra, az anyag opcionális, önállóan feldolgozható a Canvas-on lévő anyagok alapján, az ehhez az órához tartozó házi feladatok és gyakorló feladatok is opcionálisak)**

Egyéb statisztikai eljárások a pszichológiában mint főkomponens-elemzés (PCA), feltáró faktoranalízis (EFA),

Gyakorláshoz ajánlott: DataCamp: [Dimensionality Reduction in R (Linkek egy külső oldalra)](https://www.datacamp.com/courses/dimensionality-reduction-in-r) (Chapters 1, 2, 3, 4)

1. **óra – Május 8.**

Model diagnosztika: kiugró értékek, Cook-távolság, hiányzó értékek kezelése; A lineáris modellek előfeltevései, azok sérülésének diagnosztikája és kezelése: reziduálisok normalitása, homoszkedaszticitás, multikollinearitás, linearitás

Gyakorláshoz ajánlott: DataCamp: [Inference for Linear Regression (Linkek egy külső oldalra)](https://www.datacamp.com/courses/inference-for-linear-regression) (Chapter 4 - Technical Conditions in linear regression).

1. **óra – Május 15. - Második ZH**

**Május 21. 13:00-15:30 (IZU121) – Pót ZH** - Aki pót ZH-zni akar, hoznia kell saját gépet amin fut az R és tud csatlakozni az ELTE wifihez!)

|  |
| --- |
| **Kurzussal kapcsolatos egyéb specifikus információk** |

**A kurzus célja**

Az órákon a hallgatók az előadáson tanult eljárásokat tanulják meg a gyakorlatba ültetni. A kurzus elvégzésével a hallgatók képet kapnak a pszichológiai témájú tudományos kutatásokban, és egyéb gyakorlati területeken alkalmazott statisztikai eljárásokról, és képessé válnak arra, hogy a megismert elemzési és ábrázolási eljárásokat maguk is elvégezzék.

**A kurzus teljesítésének feltételei**

A gyakorlati jegy megszerzésének feltétele (a teljesítés feltételeit a kurzus jegyzőkönyvben rögzítjük)

1. Aktív órai részvétel
2. A házi feladatok rendszeres elkészítése
3. A két zárthelyi vizsga legalább elégséges (60%) teljesítése

Az **aktív órai részvétel** a gyakorlófeladatok feltöltésével igazolható az óra napján éjfélig. Amennyiben a gyakorlófeladatok közül háromnál több hiányzik, a kurzus elvégzése sikertelen.

Az órák végén adott **házi feladatok** *megfelelt/nem felelt meg* értékelést kapnak. Az aktuális órához tartozó házi feladatot mindig az adott héten vasárnap éjfélig kell beküldeni. Amennyiben a házi feladatok közül háromnál több nem megfelelt értékelést kap (vagyis nincs beküldve, vagy hiányosan van beküldve), a kurzus elvégzése sikertelen.

A félév során két **zárthelyi vizsgát**(ZH-t) kell majd elvégezni. Az ezekre a ZH-kra kapott értékelés átlaga adja majd a félév végi értékelést. A gyakorlat elvégzéséhez a félév végi értékelésnek legalább 60%-nak kell lennie. Ez az értékelés beszámít majd a kurzusra kapott jegybe mint „kollokvium jegy”.  
A ZH nyílt könyv (open books) jellegű, használhatóak jegyzetek és internetes anyagok. De a ZH önálló munka! A ZH felügyelőjével folytatott kommunikáción kívül mindenfajta más interakció tiltott a ZH során, ideértve a kurzus hallgatóival való interakciót.

**DataCamp és bónuszpontok**

Minden órához tartoznak DataCamp kurzusok, melyek az adott óra anyagával átfednek (lásd a részletes tematikában). Ezeknek a DataCamp kurzusoknak az elvégzése nem kötelező, de ajánlott, mert nagyon hasznos gyakorlati tapasztalatot nyújtanak, és a tudás megerősítését szolgálják. A félév végén a DataCamp-en szerzett minden 1000XP után 1 bónusz százalékpontot adunk a végső gyakorlati értékelésben, de csak azoknak, akik már elértek legalább 60%-os értékelést bónuszpontok nélkül (maximum 20 bónusz százalékpont szerezhető így).

DataCamp-en a **második ZH napján éjfélig** összegyűjtött XP-ket a számítjuk be.

**Csalás, másolás, plágium**

Ha kiderül, hogy valaki csalt, például kommunikált a ZH során, vagy mások anyagának lemásolásával oldott meg valamely beadandó feladatot vagy ZH feladatot, az érintett diákok nem kaphatnak jegyet a kurzusra az adott félévben! Ez vonatkozik a kommunikációban résztvevő mindkét félre, és a másolóra csakúgy, mint arra, akiről másoltak. Ezen felül a csalással és plágiummal kapcsolatos ügyek bejelentésre kerülnek az egyetemi hatóságok felé is.

**Ajánlott irodalom**

Poldrack, R. A. (2018) *Statistical Thinking for the 21st Century.* URL: [http://statsthinking21.org/ (Linkek egy külső oldalra)](http://statsthinking21.org/)

Reiczigel, J., Harmos, A. & Solymosi, N. (2010). Biostatisztika nem statisztikusoknak. Pars Kft., Nagykovácsi.

Vargha, A. (2015). *Matematikai Statisztika – Pszichológiai, biológiai és nyelvészeti alkalmazásokkal.* Budapest: Pólya Kiadó

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R.*Sage, London.

**Általános információk a kurzusról**

Az alábbi információ az Adatelemzés kurzusra általánosságban vonatkozik, mind az előadásról és a gyakorlatról egyaránt vannak benne információk.

|  |
| --- |
| **Oktatás célja** |

**A tárgy képzési célja:**

A kurzus célja az adatelemzés statisztikai módszereinek készség szintű gyakorlati alkalmazásához és az eredmények értelmezéséhez szükséges alapfogalmak, elméleti statisztikai alapok és elemzési módok elsajátítása.

**Tanulási eredmények, kompetenciák**

tudás:

* Ismeri a statisztika alapfogalmait, szakkifejezéseit, egyszerű (egyváltozós) elemzési módjait
* Ismeri az egyszerű (egyváltozós) statisztikai elemzések alkalmazásának feltételeit, az eredmények értelemzésének lehetőségeit
* Ismeri a statisztikai elemzések közlésének formai és tartalmi követelményeit (APA formátum)
* Tisztában van a statisztikai elemzések korlátaival

attitűd:

* Érzékenységet és érdeklődést mutat a pszichológiai jelenségek és problémák tudományos vizsgálata iránt
* Ismeretei alkalmazása során rugalmassággal és kreativitással rendelkezik.

képesség:

* Képes leíró statisztikai elemzéseket elvégezni
* Képes pszichológiai jelenségekre vonatkozóan statisztikai hipotézist megfogalmazni, és a hipotézist a megfelelő statisztikai módszerekkel ellenőrizni
* Képes statisztikai elemzéseket önállóan futtatni és az eredményeket megfelelően értelmezni statisztikai szoftverek segítségével (pl. SPSS, JASP, PSPP, CogStat, ROPstat)

|  |
| --- |
| **Tantárgy tartalma** |

**Fő tartalmi, tematikai egységek**

Statisztikai alapfogalmak: változók és fajtáik, minta és populáció, eloszlások és jellemzőik, leíró és következtetési statisztika, hipotézis ellenőrzés, szignifikancia, hatásméret,

Csoportok és mérések összehasonlítására alkalmas paraméteres és nem paraméteres statisztikai próbák: egymintás-, összetartozó mintás- és kétmintás t próba; varianciaanalízis; Mann-Whitney próba, Wilcoxon próba, Előjel Próba, Kruskal-Wallis próba, Friedman próba

Diszkrét változók kapcsolatának vizsgálata**:** Khí-négyzet próba, McNemar próba

Kvantitatív változók kapcsolatának korrelációs és lineáris regressziós vizsgálata

**Tervezett tanulási tevékenységek, tanítási módszerek**

előadás, vezetett egyéni gyakorlatok, önálló egyéni gyakorlatok

|  |
| --- |
| **Számonkérési és értékelési rendszere** |

**Követelmények és az értékelés módja, szempontjai:**

követelmények

* aktív részvétel az órán
* az órákon tárgyalt elméleti anyag ismerete
* a gyakorlati feladatok, házi feladatok teljesítése

az értékelés módja:

Írásbeli vizsga a kurzus elméleti részéből (az előadásokon elhangzott anyag valamint a kötelező irodalom). Kollokvium érdemjegye ötfokú osztályzat.

Gyakorlati jegy – ötfokú osztályzat a gyakorlati feladatok, házi feladatok, és a zárthelyi dolgozatok teljesítése alapján

**Végső jegy**: ötfokú osztályzat. A kollokvium jegy (50%) és a gyakorlati jegy (50%) átlaga a kerekítés matematikai szabályainak megfelelően. Egyik részjegy sem lehet elégtelen.

az értékelés szempontjai:

* A statisztika alapfogalmak és elemzési módok elméleti szintű ismeretének mennyisége és minősége
* A statisztikai adatelemzés készség szintű gyakorlati alkalmazásának és az eredmények értelmezésének minősége

|  |
| --- |
| **Irodalom** |

**Kötelező irodalom**

* Vargha A. (2007) *Matematikai statisztika. Pszichológiai, nyelvészeti és biológiai alkalmazásokkal.* Budapest: Pólya Kiadó.