

1. feladat. Folytonos valószínűségi változó szimulálása

Legyen (X, Y) a következő eloszlású kétdimenziós valószínűségi változó:

$$X \sim \text{Pareto}(x_{\min}, \alpha) \quad \text{és} \quad Y = \sin(X + E), \quad \text{ahol} \quad E \sim \text{Egyenl}(0, 1).$$

- Legyen x_{\min} a Neptun-kódod¹ első betűjének, míg α az utolsó betűjének angol ABC-beli sorszáma.
- Szimulálj az (X, Y) párból 1000 független példányt!
- A szimulált minta alapján számold ki az X és Y közti Pearson korreláció és a rangkorreláció értékét is! Adj rövid szöveges leírást arról, amit tapasztalsz!
- Ábrázold a kapott pontpárokat pontfelhő ábrán!

Segítség: a Pareto-eloszlás sűrűségfüggvénye:

$$f(x) = \alpha \frac{x_{\min}^\alpha}{x^{\alpha+1}}, \quad \text{ha} \quad x \in [x_{\min}, \infty).$$

2. feladat. Normalitásvizsgálat

Legyen $\Theta = \text{"a Neptun-kódodban}^2 \text{ szereplő számok összege"} \pmod{49}$.

- Végezd el az előadáson és a laboron is tanult heurisztikus normalitásvizsgálatot a mellékelt *crimes.xlsx* adatfájl Θ sorszámának megfelelő oszlopán³ úgy, hogy 10 osztópontot használj!
- Az adódó pontokra rajzold ki a négyzetes értelemben legjobban illeszkedő egyenest!
- A kapott egyenes milyen becslést szolgáltat az oszlophoz tartozó változó várható értékére és szórására? Számold ki!

3. feladat. Hipotézisvizsgálat

Ismét a mellékelt *crimes.xlsx* adatfájl Θ oszlopán/változóján dolgozunk. U/Z-próbával vizsgáld meg azt a hipotézist, hogy ennek a változónak a várható értéke a fájl második lapján (az adott sorszámhoz tartozó) érték-e! Dolgozz 0.03 elsőfajú hibával. A megoldáshoz írd az egyik cellába rövid szöveges következtetést is!

Segítség: Nagy minta esetén az U-próba a háttérváltozó normalitása és a szórás ismeretének feltételei nélkül is használható. Annyit kell csak tenni, hogy az ismert szórás helyébe a korrigált empirikus szórást helyettesítsük (lásd lenti képlet), ezután mindent a megszokott módon kell csinálni (beleértve azt is, hogy a standard normális eloszlás megfelelő kvantilisét használjuk az elfogadási intervallum felvételéhez):

$$U = \frac{\bar{X} - \mu_0}{\frac{s^*}{\sqrt{n}}},$$

ahol μ_0 a H_0 hipotézisbeli érték a várható értékre.

4. feladat. Diszkrét valószínűségi változó szimulálása

Szimulálj 500 elemű mintát a

$$p(A) = 0.05, \quad p(B) = 0.45, \quad p(C) = 0.1, \quad p(D) = 0.15, \quad p(F) = 0.25$$

súlyfüggvényű diszkrét ordinális háttérváltozóból! (Amerikai osztályzatok, így nincs E .)

- Számold ki az alsó, felső kvartiliseket és mediánt!
- Ha magyar osztályzatok szerint menne a kódolás, átlagot is tudnánk számolni. Tedd ezt meg! Mit tapasztalunk, ha az átlagot és a mediánt összehasonlítjuk? Az eloszlások melyik tanult mértékével/statisztikájával magyarázható ez? Számoljuk ennek értékét is ki!

¹Ha a két betű azonos vagy eleve csak egy betű van, az nem probléma.

²Ha nem tartalmaz számot, legyen ez az összeg 0.

³A sorszámokat az adatfájl második lapja tartalmazza, a szükséges adatot másold át a megoldásfájlodba.