

Beadandó Feladatok

1. Feladat (5 pont)

A `bead1.csv` fájl különböző édességek cukortartalmát tartalmazza. A feladat célja annak statisztikai vizsgálata, hogy a különböző édességek cukortartalma szignifikánsan eltér-e egymástól, vagy mindegyik ugyanolyan édes. A fájl az alábbi változókat tartalmazza:

- **Cukortartalom:** az édességek cukortartalma grammal (100g édesség esetén),
 - **Kategória:** az édesség típusa, amely öt szintet (1-5) vehet fel: **csokoládé, gumicukor, keksz, fagylalt, cukorka**.
1. **Hipotézisvizsgálat:** Vizsgáld meg, hogy az $\varepsilon = 0.05$ szignifikanciaszinten kimutatható-e szignifikáns különbség a kategóriák cukortartalma között.
 2. **Post-hoc tesztek:** Amennyiben szignifikáns eltérést találsz, végezz el egy post-hoc elemzést annak megállapítására, hogy mely kategóriák között van statisztikailag szignifikáns különbség. A post-hoc elemzés során páronként hasonlítsd össze a kategóriákat (**csokoládé, gumicukor, keksz, fagylalt, cukorka**), és az eredményeket részletesen kommentáld.

2. Lineáris Regressziós Feladat (20 pont)

A `bead2.csv` fájl édességek jellemzőit tartalmazza. A feladat egy lineáris regressziós modell felállítása, ahol az eredményváltozó (Y) egy édesség népszerűségi pontszáma (1-10 között), a magyarázó változók pedig az édesség kalóriatartalma (X_1 , kcal/100g) és a cukortartalma (X_2 , gr/ 100g). Az alábbi elemzéseket végezd el:

1. Becslések (7 pont)

1. **Az együtthatók pontbecslése (3 pont):** Számítsd ki a regressziós együtthatók pontbecslését és a standardizált együtthatókat. Értelmezd az eredményeket, és írd fel egy lineáris modellt.
2. **Előrejelzés készítése (1 pont):** Készíts előrejelzést az édesség népszerűségi pontszámára, ha az édesség kalóriatartalma $X_1 = 450$ kcal/100g és a cukortartalma $X_2 = 30$ gramm.
3. **Konfidenciaintervallum az együtthatókra (2 pont):** Számítsd ki a 95%-os konfidenciaintervallumokat az együtthatókra, és készíts konfidenciaintervallumot a regressziós egyenesre. Értelmezd az eredményeket.
4. **Előrejelzési intervallum (1 pont):** Számítsd ki a 95%-os előrejelzési intervallumot, ha az édesség kalóriatartalma $X_1 = 450$ kcal/100g és a cukortartalma $X_2 = 30$ gramm.

2. Illeszkedésdiagnosztika (3 pont)

1. **Determinációs együttható (R^2) és korrigált R^2 (3 pont):** Számítsd ki az R^2 és korrigált R^2 értékeket. Magyarázd el ezek jelentőségét az illeszkedés szempontjából, és tárgyald a korrigált R^2 hasznosságát.

3. Modelldiagnosztika (10 pont)

1. **Modelldiagnosztikai tesztek (2 pont):** Teszteld a modell szignifikanciáját 5%-os szinten, és magyarázd el a teszt fontosságát a modell megbízhatósága szempontjából.
2. **Változók szignifikanciájának tesztelése (3 pont):** Teszteld az egyes együtthatók szignifikanciáját 5%-os szinten és értelmezd az eredményeket.
3. **Multikollinearitás vizsgálata (1 pont):** Számítsd ki a Variance Inflation Factor (VIF) értékét, és magyarázd el, fennáll-e multikollinearitás. Tárgyald, miért jelent problémát a multikollinearitás.

4. **Hibatagok vizsgálata (4 pont):** Teszteld a modell hibatagjának tulajdonságait: várható értéke, normalitása, függetlensége és homoszkedaszticitása 5%-os szinten, és becsüld a varianciát.

3. Feladat (15 pont)

A `bead3.csv` fájl egy édességbolti eladások havi adatát tartalmazza 100 kilogrammban.

Feladatok

1. **Determinisztikus modell illesztése (5 pont):** Illessz determinisztikus modellt az adatokra, és számítsd ki az együttthatókat. Vizsgáld meg a maradéktagokra vonatkozó feltételek teljesülését. Készíts előrejelzést a következő hónapokra.
2. **Exponenciális simítás alkalmazása (5 pont):** Alkalmazz exponenciális simítást, és vizsgáld meg, hogy mennyire illeszkedik a modell az adatokhoz. Teszteld az illeszkedést, és készíts előrejelzést a következő hónapokra.
3. **Box-Jenkins modell (5 pont):** Alkalmazz Box-Jenkins (ARIMA vagy SARIMA) modellt, és teszteld az illeszkedést. Vizsgáld meg a modell paramétereit, és készíts előrejelzést a következő hónapokra.