1. Feladat

December 4, 2024

0 Előkészületek

0.1 Szükséges könyvtárak importálása

[21]:
```
import pandas as pd
import numpy as np
from scipy import stats
```

0.2 Adatok beolvasása

[22]:
```
# Kategóriák definiálása
kategoriak = ['szuperhos', 'anti-hos', 'mellekszereplo', 'fogonosz',␣
↪'allatsegito']

# Adatok beolvasása string-ként
with open('data/bead1.csv', 'r') as file:
    lines = file.readlines()

# Az első sor elhagyása (mivel az a kategóriákat tartalmazza)
# Az értékek átalakítása soronként listává
data = [list(map(float, line.strip().strip('"').split(','))) for line in lines[1:
↪]]

# DataFrame létrehozása
df = pd.DataFrame(data, columns=kategoriak)


# Adatok átalakítása long formátumba
df_long = df.melt(var_name='Kategória', value_name='Gumimaci pontszám')
# Alapvető statisztikai jellemzők
print("Alapvető statisztikai jellemzők kategóriánként:")
print(df_long.groupby('Kategória')['Gumimaci pontszám'].describe())
```

```
Alapvető statisztikai jellemzők kategóriánként:
                count    mean       std   min     25%    50%     75%    max
Kategória
allatsegito      50.0  5.4096  3.130664  1.37  2.8850  4.335  9.6700  10.00
anti-hos         50.0  2.0552  1.655914  0.14  0.8300  1.545  3.0850   6.71
```

fogonosz	50.0	5.4766	2.125727	1.30	3.7325	5.690	7.1775	9.21
mellekszereplo	50.0	2.9046	1.635708	0.17	1.5675	2.735	4.0050	7.24
szuperhos	50.0	4.4258	2.879298	0.52	1.8700	3.770	6.0975	9.90

1 Hipotézisvizsgálat

1.1 Próba meghatározása

Az adatok eloszlásáról nem tudunk semmit, csak hogy számok és a nagyságuk sorrendje számít, így ordinális változóknak tekintjük a gumimaci pontszámokat. A szereplők egymástól függetlenek és 5 mintánk van, így Kruskal-Wallis próbát hajtunk végre.

1.2 Kruskal-Wallis próba

1.2.1 Hipotézisek megfogalmazása

Hipotézispár:

H_0: A kategóriák pontszámainak eloszlása azonos
H_1: Van két olyan kategória, amelyek pontszámainak eloszlása különbözik
Szignifikanciaszint: $\varepsilon = 0.05$

1.2.2 Próbastatisztika számítása

```
[23]: h_stat, p_value = stats.kruskal(*[group['Gumimaci pontszám'].values
                                       for name, group in df_long.
       ↪groupby('Kategória')])
      print("Kruskal-Wallis teszt eredménye:")
      print(f"H-statisztika = {h_stat:.4f}")
```

```
Kruskal-Wallis teszt eredménye:
H-statisztika = 68.1814
```

1.2.3 Döntés a kritikus érték alapján

Paraméterek:

Kategóriák száma (k) = 5
Szabadságfok (df) = k-1 = 4
Szignifikanciaszint (ε) = 0.05
H-statisztika = 68.1814
χ^2(0.05,4) kritikus érték (táblázat alapján) = 9.49

Döntési szabály:

Ha H > $\chi^2(\varepsilon,df)$ → elvetjük H0-t
Ha H ≤ $\chi^2(\varepsilon,df)$ → nem vetjük el H0-t

Összehasonlítás:

$68.1814 > 9.49$
A H-statisztika értéke nagyobb, mint a kritikus érték

1.2.4 Következtetés:

A H-statisztika meghaladja a kritikus értéket, ezért $\varepsilon = 0.05$ szignifikanciaszinten elvetjük a nullhipotézist.
Azaz statisztikailag kimutatható, hogy van különbség a kategóriák gumimaci pontszámai között.

2 Post-hoc tesztek

Mivel szignifikáns eltérést találtunk, ezért páronként meg kell vizsgálnunk a kategóriákat. A változóink ordinálisak, páronként végezzük a teszteket (tehát minden teszt esetén 2 mintát vetünk össze), a mintáink nem összefüggők.
Páronként 2 független mintás ordinális próbát, azaz Mann-Whitney próbát hajtunk végre.

2.1 Mann-Whitney Z teszt páronként

[24]:
```
kategoriak = df_long['Kategória'].unique()
alpha = 0.05 # szignifikanciaszint

# Kritikus érték (kétoldali próba) normális eloszlás táblázatból
z_critical = 1.96  # z0.975 = 1.96

print(f"\nPáronkénti Mann-Whitney Z teszt eredményei:")
print(f"Kritikus érték (z{1-alpha/2:.3f}): {z_critical}")
print("-" * 50)

results = []
for i in range(len(kategoriak)):
    for j in range(i+1, len(kategoriak)):
        x = df_long[df_long['Kategória'] == kategoriak[i]]['Gumimaci pontszám'].
↪values
        y = df_long[df_long['Kategória'] == kategoriak[j]]['Gumimaci pontszám'].
↪values

        # Mann-Whitney teszt
        stat, p_value = stats.mannwhitneyu(x, y, alternative='two-sided')

        # Z-érték kiszámítása a p-értékből
        z_stat = stats.norm.ppf(1 - p_value/2)

        results.append({
            'Kategória 1': kategoriak[i],
            'Kategória 2': kategoriak[j],
            'Z-érték': z_stat,
```

```
            '|Z|': abs(z_stat),
            'Szignifikáns': abs(z_stat) > z_critical
        })
        print(f"{kategoriak[i]} vs {kategoriak[j]}: |Z| = {abs(z_stat):.4f} {'*'␣
↪if abs(z_stat) > z_critical else ''}")

# Eredmények DataFrame-be rendezése és megjelenítése
results_df = pd.DataFrame(results)
print("\nÖsszes páronkénti összehasonlítás eredménye:")
print(results_df)

# Szignifikáns különbségek kiírása
print("\nSzignifikáns különbségek:")
sig_pairs = results_df[results_df['Szignifikáns']].apply(
    lambda x: f"{x['Kategória 1']} vs {x['Kategória 2']} (|Z| = {x['|Z|']:.
↪4f})", axis=1
)
for pair in sig_pairs:
    print(f"{pair}")

# Nem szignifikáns különbségek kiírása
print("\nNem szignifikáns különbségek:")
nonsig_pairs = results_df[~results_df['Szignifikáns']].apply(
    lambda x: f"{x['Kategória 1']} vs {x['Kategória 2']} (|Z| = {x['|Z|']:.
↪4f})", axis=1
)
for pair in nonsig_pairs:
    print(f"{pair}")
```

```
Páronkénti Mann-Whitney Z teszt eredményei:
Kritikus érték (z0.975): 1.96
--------------------------------------------------
szuperhos vs anti-hos: |Z| = 4.4880 *
szuperhos vs mellekszereplo: |Z| = 2.4611 *
szuperhos vs fogonosz: |Z| = 2.2543 *
szuperhos vs allatsegito: |Z| = 1.8492
anti-hos vs mellekszereplo: |Z| = 2.7404 *
anti-hos vs fogonosz: |Z| = 6.7423 *
anti-hos vs allatsegito: |Z| = 5.8649 *
mellekszereplo vs fogonosz: |Z| = 5.5497 *
mellekszereplo vs allatsegito: |Z| = 4.1192 *
fogonosz vs allatsegito: |Z| = 0.7314

Összes páronkénti összehasonlítás eredménye:
      Kategória 1     Kategória 2   Z-érték       |Z|  Szignifikáns
0       szuperhos        anti-hos  4.487958  4.487958          True
```

```
1       szuperhos  mellekszereplo  2.461145  2.461145          True
2       szuperhos        fogonosz  2.254313  2.254313          True
3       szuperhos     allatsegito  1.849159  1.849159         False
4        anti-hos  mellekszereplo  2.740367  2.740367          True
5        anti-hos        fogonosz  6.742276  6.742276          True
6        anti-hos     allatsegito  5.864852  5.864852          True
7  mellekszereplo        fogonosz  5.549675  5.549675          True
8  mellekszereplo     allatsegito  4.119246  4.119246          True
9        fogonosz     allatsegito  0.731384  0.731384         False

Szignifikáns különbségek:
szuperhos vs anti-hos (|Z| = 4.4880)
szuperhos vs mellekszereplo (|Z| = 2.4611)
szuperhos vs fogonosz (|Z| = 2.2543)
anti-hos vs mellekszereplo (|Z| = 2.7404)
anti-hos vs fogonosz (|Z| = 6.7423)
anti-hos vs allatsegito (|Z| = 5.8649)
mellekszereplo vs fogonosz (|Z| = 5.5497)
mellekszereplo vs allatsegito (|Z| = 4.1192)

Nem szignifikáns különbségek:
szuperhos vs allatsegito (|Z| = 1.8492)
fogonosz vs allatsegito (|Z| = 0.7314)
```

Látható tehát, hogy a legtöbb kategória között szignifikáns különbség van az eloszlásuk tekintetében. Egyedül a szuperhos-allatsegito és fogonosz-allatsegito párosok eloszlásában nincs szignifikáns különbség.