Baran Sándor

Feladatok a hipotézisvizsgálat témaköréből

mobiDIÁK könyvtár

Baran Sándor

Feladatok a hipotézisvizsgálat témaköréből

mobiDIÁK könyvtár

SOROZATSZERKESZTŐ

Fazekas István

Baran Sándor

Debreceni Egyetem

Feladatok a hipotézisvizsgálat témaköréből

Oktatási segédanyag

mobiDIÁK könyvtár

Debreceni Egyetem

Lektor

Barczy Mátyás
Debreceni Egyetem

mobiDIÁK könyvtár
Debreceni Egyetem
Informatikai Kar
4010 Debrecen, Pf. 12
http://mobidiak.inf.unideb.hu/

A mű „A mobiDIÁK önszervező mobil portál” (IKTA, OMFB-00373/2003)) projekt keretében készült.

Tartalomjegyzék

Bevezetés

Az itt összegyűjtött néhány feladat, melyek jelentős része brit érettségi feladatok fordítása, illetve átdolgozott változata, remélhetőleg segítséget nyújt a leggyakoribb paraméteres és nemparaméteres próbák elsajátításához. A vizsgált paraméteres próbák az egy- illetve kétmintás Z-próba (U-próba), az egy-, két- illetve páros mintás t-próba, az F-próba, valamint a szórásra vonatkozó khi-négyzet próba és az egy szempontú szórásanalízis. A nemparaméteres próbák közül a binomiális, az előjel, a Wilcoxon-féle előjeles rangösszeg próbát és a Mann-Whitney próbát, valamint a khi-négyzet próbával végzett illeszkedés- illetve függetlenségvizsgálatot tekintettük át. A példatár feltételezi a hipotézisvizsgálat elméletének és az alapvető próbáknak az ismeretét, jelölésrendszere pedig igazodik a hazai szakirodalomban általánosan elfogadotthoz (lásd például [2, 3, 5, 6]). A próbák végrehajtásához szükséges táblázatok nagy része megtalálható a hivatkozott jegyzetek és példatárak függelékében, a legteljesebb mű azonban e témában Bolsev és Szmirnov táblázatgyűjteménye [1]. A nemparaméteres próbákkal kapcsolatban magyar nyelven ajánlható még [8], angolul pedig [4, 7]. Ez utóbbi három mű függelékként tartalmazza az itt tárgyalt nemparaméteres próbák táblázatait is. A megoldott feladatoknál a legtöbb próba kritikus értékeit a MATLAB programcsomag segítségével határoztuk meg, a Wilcoxon-féle előjeles rangösszeg próba táblázata pedig megtalálható a Függelékben.

1. Paraméteres próbák

1.1. Z-próba

1.1. Példa Egy teherautórakománnyi félliteres üdítőitalból 10 palackot véletlenszerűen kiválasztva és lemérve azok űrtartalmát az alábbi, milliliterben kifejezett értékeket kaptuk:

$$499,\ 525,\ 498,\ 503,\ 501,\ 497,\ 493,\ 496,\ 500,\ 495.$$

Ismert, hogy a palackokba töltött üdítőital mennyisége normális eloszlású 3 ml szórással. 95%-os döntési szintet használva vizsgálja meg a gyártó azon állítását, hogy a palackokba átlagosan fél liter üdítőitalt töltöttek!

Megoldás.

$$H_0 : \mu = 500;$$
$$H_1 : \mu \neq 500.$$
$$\alpha = 0.05.$$

$n = 10,\ \bar{x} = 500.7,\ \sigma = 3.$

A próbastatisztika:

$$Z = \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n} = \frac{500.7 - 500}{3}\sqrt{10} = 0.7379.$$

Ha H_0 igaz, a próbastatisztika eloszlása standard normális.

A kritikus tartomány: $|Z| \geq Z(0.975) = 1.960$. A kapott érték nem esik bele, így elfogadjuk H_0-t.

Alternatív megoldás

p-érték: $\mathsf{P}(|Z| \geq 0.7379) = 2\mathsf{P}(Z \leq -0.7379) = 2\Phi(-0.7379) = 0.4606 > 0.05$. Elfogadjuk H_0-t. □

1.2. Példa Egy kiterjedt népegészségügyi vizsgálat során megállapították, hogy az egészséges felnőtt populáció esetén a diasztolés (alsó) vérnyomás értékek átlaga 84.8 higanymilliméter, szórása pedig 12.8 higanymilliméter. Az Alsóbezgenyei Atlétikai Klub hat véletlenszerűen kiválasztott versenyzőjénél a klub sportorvosa az alábbi diasztolés értékeket jegyezte fel:

$$79.2,\ 64.6,\ 86.8,\ 73.7,\ 74.9,\ 62.3.$$

a) A sportorvos ezek alapján úgy gondolta, hogy az atléták átlagos diasztolés vérnyomása alacsonyabb, mint 84.8. Feltételezve, hogy az atléták diasztolés vérnyomása normális eloszlást követ, szórása pedig megegyezik a teljes populációra kapott értékkel (12.8 higanymilliméter), döntsön 95%-os szinten, igaza van-e a doktornak!

Az Alsóbezgenyei Sakk Klub versenyzői szintén meglátogatták a fent említett doktort, aki az ő esetükben is feljegyezte hat véletlenszerűen kiválasztott sportoló diasztolés vérnyomás értékét, melyek az alábbiak:

$$84.6,\ 93.2,\ 104.6,\ 106.7,\ 76.3,\ 78.2.$$

b) Hipotéziseit pontosan megfogalmazva döntsön 95%-os szinten, hogy a sakkozók átlagos diasztolés vérnyomása magasabb-e, mint az atlétáké! A sakkozók diasztolés vérnyomásáról szintén feltehetjük, hogy normális eloszlást követ, szórása pedig megegyezik a teljes népesség körében mért értékkel.

Megoldás.

a)

$$H_0 : \mu_x = 84.8;$$
$$H_1 : \mu_x < 84.8. \qquad \text{(egyoldali ellenhipotézis)}$$
$$\alpha = 0.05.$$

$n = 6,\ \bar{x} = 73.5833,\ \sigma_x = 12.8.$

A próbastatisztika:

$$Z = \frac{\bar{x} - \mu_{x,0}}{\sigma_x}\sqrt{n} = \frac{73.5833 - 84.8}{12.8}\sqrt{6} = -2.1465.$$

Ha H_0 igaz, a próbastatisztika eloszlása standard normális.

A kritikus tartomány: $Z \leq Z(0.05) = -1.645$. A kapott érték beleesik, így elvetjük H_0-t.

Alternatív megoldás

p-érték: $\mathsf{P}(Z \leq -2.1465) = \Phi(-2.1465) = 0.0158 < 0.05$. Elvetjük H_0-t.

b)

$$H_0 : \mu_x = \mu_y;$$
$$H_1 : \mu_x < \mu_y. \qquad \text{(egyoldali ellenhipotézis)}$$
$$\alpha = 0.05.$$

$n = m = 6,\ \bar{x} = 73.5833,\ \bar{y} = 90.6,\ \sigma_x = \sigma_y = 12.8.$

A próbastatisztika:

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} = \frac{73.5833 - 90.6}{12.8}\sqrt{3} = -2.3026.$$

Ha H_0 igaz, a próbastatisztika eloszlása standard normális.

A kritikus tartomány: $Z \leq Z(0.05) = -1.645$. A kapott érték beleesik, így elvetjük H_0-t.

Alternatív megoldás

p-érték: $\mathsf{P}(Z \leq -2.3026) = \Phi(-2.3026) < 0.0107 < 0.05$. Elvetjük H_0-t. $\square$

1.3. Példa Az *Ezt idd* teát 200 grammos dobozokban árulják, a csomagológép szórása 4 gramm. A Fogyasztóvédelmi Felügyelőség lemérte öt véletlenszerűen kiválasztott teásdoboz tömegét, melyekre az alábbi grammban kifejezett értékek adódtak:

$$196,\ 202,\ 198,\ 197,\ 190.$$

Hipotéziseit pontosan megfogalmazva és feltételezve, hogy a teásdobozok tömege normális eloszlást követ, döntsön 98%-os szinten, hogy az átlagos töltőtömeg tényleg 200 gramm, avagy kevesebb annál!

1.2. t-próba, F-próba és a szórásra vonatkozó khi-négyzet próba

1.4. Példa Egy gabonaraktárban 60 kg-os kiszerelésben búzát csomagolnak. A havi minőségellenőrzés során azt is meg akarták vizsgálni, hogy a raktárból kikerülő zsákokban tényleg 60 kg búza van-e, ezért lemértek tíz darab véletlenül kiválasztott zsákot. Eredményül a következőket kapták:

$$60.2,\ 63.4,\ 58.8,\ 63.6,\ 64.7,\ 62.5,\ 66.0,\ 59.1,\ 65.1,\ 62.0.$$

Hipotéziseit és az adatokra vonatkozó feltételeit pontosan megfogalmazva döntsön 95%-os szinten, a zsákok átlagos töltőtömege tényleg 60 kg-e!

Megoldás.

$$H_0 : \mu = 60;$$
$$H_1 : \mu \neq 60.$$
$$\alpha = 0.05.$$

Feltételezzük, hogy a zsákok töltőtömege normális eloszlású.

$\bar{x} = 62.54,\ s^{*2} = 6.2938,\ s^* = 2.5087.$

A próbastatisztika:

$$t = \frac{\bar{x} - \mu_0}{s^*}\sqrt{n} = \frac{62.54 - 60}{2.5087}\sqrt{10} = 3.2017.$$

Ha H_0 igaz, a próbastatisztika eloszlása t-eloszlás $\nu = 9$ szabadsági fokkal.

A kritikus tartomány: $|t| \geq t_9(0.975) = 2.262$. A kapott érték beleesik, így elvetjük H_0-t. □

1.5. Példa Egy üzem gyártósorán az egyik szerelési feladatra megadott szintidő 9 perc. Az e ponton dolgozó alkalmazottak már több kérvényben kérték a szintidő felemelését, mivel véleményük szerint az nem elegendő a feladat elvégzésére.

Az üzem vezetősége egy ellenőrt küldött ki, aki 12 véletlenszerűen kiválasztott alkalommal megmérte a feladat elvégzéséhez szükséges időt. Az eredmények az alábbiak:

9.4, 8.8, 9.3, 9.1, 9.4, 8.9, 9.3, 9.2, 9.6, 9.3, 9.3, 9.1.

Hipotéziseit és az adatokra vonatkozó feltételeit pontosan megfogalmazva döntsön 99%-os szinten, igazuk van-e a munkásoknak!

Megoldás.

$$\begin{aligned} H_0 &: \mu = 9; \\ H_1 &: \mu > 9. \qquad \text{(egyoldali ellenhipotézis)} \\ \alpha &= 0.01. \end{aligned}$$

Feltételezzük, hogy a feladat elvégzéséhez szükséges idő normális eloszlású.

$\bar{x} = 9.2250,\ s^{*2} = 0.0493,\ s^* = 0.2221.$

A próbastatisztika:

$$t = \frac{\bar{x} - \mu_0}{s^*}\sqrt{n} = \frac{9.225 - 9}{0.2221}\sqrt{12} = 3.5093.$$

Ha H_0 igaz, a próbastatisztika eloszlása t-eloszlás $\nu = 11$ szabadsági fokkal.

A kritikus tartomány: $t \geq t_{11}(0.99) = 2.718$. A kapott érték beleesik, így elvetjük H_0-t. □

1.6. Példa Az atlétikai világbajnokságon résztvevő kokszföldi csapat néhány versenyzője arra panaszkodott, hogy a leadott doppingtesztjeiket nem megfelelően analizálták és az egyik szernek túlságosan magas koncentrációját mutatták ki, minek következtében a versenybíróság törölte az eredményeiket. A Kokszföldi Atlétikai Szövetség a laboratóriumot tesztelendő nyolc mintát küldött, melyek mindegyikében a kérdéses anyag koncentrációja pontosan 0.500 g/l volt. A laboratórium az alábbi eredményeket szolgáltatta:

$$0.485,\ 0.518,\ 0.460,\ 0.530,\ 0.560,\ 0.550,\ 0.490,\ 0.575.$$

A laboratórium mérési hibáját normálisnak tételezve fel döntsön 95%-os szinten, igazuk van-e az atlétáknak!

1.7. Példa Az Árelhajlásvizsgáló Hivatal összehasonlította két konkurens hipermarket élelmiszerárait. Tíz véletlenszerűen kiválasztott terméket vizsgáltak, melyek árait az alábbi táblázat tartalmazza:

Termék	A	B	C	D	E	F	G	H	I	J
Alfa Hipermarket	464	158	376	112	98	92	38	74	66	38
Beta Hipermarket	432	148	416	104	84	98	36	62	76	34

Az árkülönbségeket normális eloszlásúnak tételezve fel döntsön 95%-os szinten, van-e eltérés a két hipermarket élelmiszereinek árszintje között!

Megoldás. x: Alfa Hipermarket árai; y: Beta Hipermarket árai.

$$H_0 : \mu_x = \mu_y;$$
$$H_1 : \mu_x \neq \mu_y.$$
$$\alpha = 0.05.$$

A különbségek, $d = x - y$: $32,\ 10,\ -40,\ 8,\ 14,\ -6,\ 2,\ 12,\ -10,\ 4.$

Az eredetivel ekvivalens hipotézisek:

$$H_0 : \mu_d = 0;$$
$$H_1 : \mu_d \neq 0.$$

$\bar{d} = 2.6,\ s^{*2} = 357.3778,\ s^* = 18.9044.$

A próbastatisztika:

$$t = \frac{\bar{d} - \mu_0}{s^*}\sqrt{n} = \frac{2.6 - 0}{18.9044}\sqrt{10} = 0.4349.$$

Ha H_0 igaz, a próbastatisztika eloszlása t-eloszlás $\nu = 9$ szabadsági fokkal.

A kritikus tartomány: $|t| \geq t_9(0.975) = 2.262$. A kapott érték nem esik bele, így elfogadjuk H_0-t. □

1.8. Példa A Mindent Tudás Egyeteme másodéves közgazdász hallgatói két zárthelyi dolgozatot írtak statisztikából. Az alábbi táblázat tíz véletlenszerűen kiválasztott hallgató eredményeit tartalmazza:

Hallgató	A	B	C	D	E	F	G	H	I	J
I. dolgozat	57	63	67	82	45	65	53	32	51	27
II. dolgozat	53	62	63	80	46	64	44	28	50	29

A dolgozateredmények eltérését normális eloszlásúnak tételezve fel döntsön 95%-os szinten, van-e különbség a két dolgozat nehézségi foka között!

1.9. Példa A Felsőkutyfalvi Kerékpárüzem kerékrészlegének vezetője arra gyanakszik, hogy az egyik beszállító által készített küllők hosszúsága igencsak változékony. Gyanújának ellenőrzése céljából az adott beszállító termékeiből véletlenszerűen kiválasztott 20 darabot és megmérte azok hosszát. A hossz szórásnégyzetének a minta alapján számolt torzítatlan becslése (azaz a minta korrigált tapasztalati szórásnégyzete) 1.0369 mm^2.

A beszállító állítása szerint a küllők hosszának szórása 0.75 mm.

A küllők hosszát normális eloszlásúnak tételezve fel ellenőrizze, megalapozott-e a részlegvezető gyanúja! Döntsön 95%-os szinten!

Megoldás.

$$H_0 : \sigma = 0.75;$$

$$H_1 : \sigma > 0.75.$$

$$\alpha = 0.05.$$

A próbastatisztika:

$$\mathcal{X}^2 = \frac{(n-1)s^{*2}}{\sigma_0^2} = \frac{19 \times 1.0369}{0.75^2} = 35.0242.$$

Ha H_0 igaz, a próbastatisztika eloszlása khi-négyzet eloszlás $\nu = 19$ szabadsági fokkal.

A kritikus tartomány: $\mathcal{X} \geq \mathcal{X}^2_{19}(0.95) = 30.144$. A kapott érték beleesik,így elvetjük H_0-t. $\square$

1.10. Példa Kétfajta instant kávé oldódási idejét tesztelték, melyekből minden alkalommal azonos mennyiséget tettek 1 dl forrásban lévő vízbe. A kísérletek eredményeit az alábbi táblázat tartalmazza:

Kávé	Oldódási idő (másodperc)							
Mokka Makka	8.2	5.0	6.8	6.7	5.8	7.3	6.4	7.8
Koffe In	5.1	4.3	3.4	3.7	6.1	4.7		

a) Az oldódási időket normálisnak tételezve fel 95%-os szinten igazoljuk, hogy nincs különbség az oldódási idők szórása között!

b) Az a) pontbeli szinten vizsgáljuk meg azt az állítást, hogy a Mokka Makka kávé lassabban oldódik, mint a Koffe In!

Megoldás. x: a Mokka Makka oldódási ideje; y: a Koffe In oldódási ideje.
a)

$$H_0 : \sigma_x^2 = \sigma_y^2;$$
$$H_1 : \sigma_x^2 \neq \sigma_y^2.$$
$$\alpha = 0.05.$$

$n = 8,\ \sum x = 54,\ \sum x^2 = 372.1,$ így $\bar{x} = 6.75,\ s_x^{*2} = 1.0857.$

$m = 6,\ \sum y = 27.3,\ \sum y^2 = 129.05,$ így $\bar{y} = 4.55\ s_y^{*2} = 0.967.$

A próbastatisztika: $F^* = s_x^{*2}/s_y^{*2} = 1.0857/0.967 = 1.1228$ (a nagyobb érték kerül a számlálóba).

Ha H_0 igaz, a próbastatisztika eloszlása F-eloszlás $\nu_1 = 7$ és $\nu_2 = 5$ szabadsági fokokkal.

A kritikus tartomány: $F^* \geq F_{7,5}(0.975) = 6.853$. A kapott érték nem esik bele, elfogadjuk H_0-t.

b)

$$H_0 : m_x = m_y;$$
$$H_1 : m_x > m_y. \qquad \text{(egyoldali ellenhipotézis)}$$
$$\alpha = 0.05.$$

$$\hat{\sigma}^2 = \frac{(n-1)s_x^{*2} + (m-1)s_y^{*2}}{n+m-2} = \frac{7.5999 + 4.835}{12} = 1.0362.$$

A próbastatisztika:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\hat{\sigma}^2(\frac{1}{n} + \frac{1}{m})}} = \frac{6.75 - 4.55}{\sqrt{1.0362(\frac{1}{8} + \frac{1}{6})}} = 4.0018.$$

Ha H_0 igaz, a próbastatisztika eloszlása t-eloszlás $\nu = 12$ szabadsági fokkal.

A kritikus tartomány: $t \geq t_{12}(0.95) = 1.782$. A kapott érték beleesik, így elvetjük H_0-t. $\square$

1.11. Példa Az angliai New Dumber golflabdagyárában egy újfajta golflabda borítást fejlesztettek ki. A tesztek azt mutatták, hogy ez az új borítás jóval ellenállóbb, mint a hagyományos. Felmerült azonban a kérdés hogy az új borítás nem változtatja-e meg az átlagos ütéstávolságot. Ennek eldöntésére 42 labdát próbáltak ki, 26 hagyományosat és 16 labdát az újak közül. A labdákat géppel lőtték ki, elkerülve ezzel az emberi tényező okozta szóródást. A yardban mért ütéstávolságok összesítő adatait, mely távolságokat mindkét esetben normális eloszlásúnak tételezzük fel, az alábbi táblázat tartalmazza:

Borítás	Mintaelemszám	Mintaátlag	Korrigált empirikus szórásnégyzet
Hagyományos	26	271.4	35.58
Új	16	268.7	48.47

a) 90%-os szinten igazoljuk, hogy nincs különbség az ütéstávolságok szórása között!

b) Az a) pontbeli szinten vizsgáljuk meg, hogy az új borítás megváltoztatja-e az átlagos ütéstávolságot!

Megoldás. x: hagyományos borítás; y: az új borítás.
a)

$$H_0 : \sigma_x^2 = \sigma_y^2;$$
$$H_1 : \sigma_x^2 \neq \sigma_y^2.$$
$$\alpha = 0.1.$$

A próbastatisztika: $F^* = s_y^{*2}/s_x^{*2} = 48.47/35.58 = 1.3623$ (a nagyobb érték kerül a számlálóba).

Ha H_0 igaz, a próbastatisztika eloszlása F-eloszlás $\nu_1 = 15$ és $\nu_2 = 25$ szabadsági fokokkal.

A kritikus tartomány: $F^* \geq F_{15,25}(0.95) = 2.089$. A kapott érték nem esik bele, elfogadjuk H_0-t.
b)

$$H_0 : m_x = m_y;$$
$$H_1 : m_x \neq m_y.$$
$$\alpha = 0.1.$$

$$\hat{\sigma}^2 = \frac{(n-1)s_x^{*2} + (m-1)s_y^{*2}}{n+m-2} = \frac{889.50 + 727.05}{40} = 40.4138.$$

A próbastatisztika:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\hat{\sigma}^2(\frac{1}{n} + \frac{1}{m})}} = \frac{271.4 - 268.7}{\sqrt{40.4138(\frac{1}{26} + \frac{1}{16})}} = 1.3367.$$

Ha H_0 igaz, a próbastatisztika eloszlása t-eloszlás $\nu = 40$ szabadsági fokkal.

A kritikus tartomány: $|t| \geq t_{40}(0.95) = 1.684$. A kapott érték nem esik bele, elfogadjuk H_0-t. $\square$

1.12. Példa A Felsődörgicsei Sátorcövekgyár kilenc véletlenszerűen kiválasztott termékének hosszából számolt korrigált tapasztalati szórásnégyzet 63 mm^2. A konkurens Alsődörgicsei Cövek és Póznagyárban gyártott tizenhárom ugyancsak véletlenszerűen kiválasztott cövek esetén ez az érték 225 mm^2.

a) Döntsünk 90%-os szinten, van-e különbség a különböző gyárakból származó cövekek szórása között!

b) Milyen, az adatokra vonatkozó feltételekre van szükség, hogy az előző pontbeli hipotézisvizsgálat végrehajtható legyen!

1.3. Egy szempontú szórásanalízis

1.13. Példa Egy vizsgálat során azt próbálták kideríteni, hogy a diákok tanulási hatékonysága függ-e a tanulási szokásaiktól. Ennek érdekében a kísérletben részvevők kaptak egy szöveget, amit háromféle módszerrel memorizálhattak: csak olvasással, olvasva és aláhúzva a fontosabb részeket, olvasva és kijegyzetelve a lényeges dolgokat. Egy hét elteltével ugyanezek a diákok írtak egy felmérőt, amiben a kapott szöveg tartalmáról kérdezték őket. A felmérők eredményeinek (pontszámainak) összesítését az alábbi táblázat tartalmazza:

Tanulási módszer	A felmérő eredménye							
Olvas	15	14	18	13	11	14	13	
Olvas és aláhúz	16	20	18	17	14			
Olvas és jegyzetel	18	17	23	16	19	22	20	25

a) Töltse ki a szórásfelbontó táblázatot!

b) Hipotéziseit pontosan megfogalmazva döntsön 95%-os szinten, igaz-e hogy a tanulás módja befolyásolja annak hatékonyságát!

c) Mondjon legalább két, az adatokra vonatkozó feltételt, ami elengedhetetlen a hipotézisvizsgálat végrehajtásához!

Megoldás.

a)

Módszer	Összeg ($\sum x$)	Négyzet összeg ($\sum x^2$)
Olvas	98	1400
Olvas és aláhúz	85	1465
Olvas és jegyzetel	160	3268
Teljes	343	6133

$$n = 20.$$
$$SS = 6133 - \frac{343^2}{20} = 250.55;$$
$$SS_K = \left(\frac{98^2}{7} + \frac{85^2}{5} + \frac{160^2}{8}\right) - \frac{343^2}{20} = 134.55;$$
$$SS_B = SS - SS_K = 116.$$

Szóródás forrása	Négyzet-összeg	Szabadsági fok	Tapasztalati szórásnégyzet	F hányados
Külső	$SS_K = 134.55$	$\nu_1 = 2$	$MS_K = \frac{SS_K}{\nu_1} = 67.275$	$F = \frac{MS_K}{MS_B} = 9.8593$
Belső	$SS_B = 116$	$\nu_2 = 17$	$MS_B = \frac{SS_B}{\nu_2} = 6.8235$	
Teljes	$SS = 250.55$	$\nu = 19$		

b)

$$H_0 : \text{nincs különbség az átlagos pontszámok között;}$$
$$H_1 : \text{van különbség.}$$
$$\alpha = 0.05.$$

Ha H_0 igaz, az F próbastatisztika eloszlása F-eloszlás $\nu_1 = 2$ és $\nu_2 = 17$ szabadsági fokokkal.

A kritikus tartomány: $F \geq F_{2,17}(0.95) = 3.59$. A kapott érték beleesik,így elvetjük H_0-t.

c) A diákokat véletlenszerűen választjuk. A felmérők pontszámai mindhárom csoportban normális eloszlásúak. Az egyes csoportok szórásai megegyeznek. □

1.14. Példa Az Debreceni Egyetemen az egyik statisztika szemináriumvezető minden hétfőn, szerdán és pénteken autóval jár ki a Tócóskertből a város másik végén fekvő Kassai úti campusra. Otthonról mindig azonos időben indul el és ugyanazon az útvonalon autózik. Úgy érzi azonban, hogy a menetideje függ attól, hogy a hét melyik napján van órája. Ezért aztán márciusban, áprilisban és májusban véletlenszerűen kiválasztott 5-5 hétfőt, szerdát és pénteket és lejegyezte a menetidőket. Adatainak összegzését az alábbi táblázat tartalmazza:

Nap	Menetidő (x)					Összeg ($\sum x$)	Négyzet összeg ($\sum x^2$)
Hétfő	28	34	29	34	30	155	4837
Szerda	24	27	25	25	22	123	3039
Péntek	25	28	27	26	21	127	3255

a) Töltse ki a szórásfelbontó táblázatot!

b) Hipotéziseit pontosan megfogalmazva döntsön 99%-os szinten, igaz-e a szemináriumvezető sejtése!

Megoldás.

a)

$$n = 15, \ \sum\sum x = 405, \ \sum\sum x^2 = 11131.$$

$$SS = 11131 - \frac{405^2}{15} = 196;$$

$$SS_K = \left(\frac{155^2}{5} + \frac{123^2}{5} + \frac{127^2}{5}\right) - \frac{405^2}{15} = 121.6;$$

$$SS_B = SS - SS_K = 74.4.$$

Szóródás forrása	Négyzet-összeg	Szabadsági fok	Tapasztalati szórásnégyzet	F hányados
Külső Belső	$SS_K = 121.6$ $SS_B = 74.4$	$\nu_1 = 2$ $\nu_2 = 12$	$MS_K = \frac{SS_K}{\nu_1} = 60.8$ $MS_B = \frac{SS_B}{\nu_2} = 6.2$	$F = \frac{MS_K}{MS_B} = 9.8065$
Teljes	$SS = 196$	$\nu = 14$		

b)

H_0 : nincs különbség az átlagos menetidők között;

H_1 : van különbség.

$\alpha = 0.01.$

Ha H_0 igaz, az F próbastatisztika eloszlása F-eloszlás $\nu_1 = 2$ és $\nu_2 = 12$ szabadsági fokokkal.

A kritikus tartomány: $F \geq F_{2,12}(0.99) = 6.93$. A kapott érték beleesik, így elvetjük H_0-t. □

1.15. Példa Egy farmer négyféle korán termő paradicsomfajtát próbált ki. Földjét négy egyenlő részre osztotta, majd lejegyezte az egyes fajták terméshozamát. A következő évben ezt újra megismételte, majd a rákövetkezőben is. A harmadik év végén adatai átszámolta egy standardizált x mennyiségre, majd a kapott értékeket az alábbi táblázatban öszesítette:

	A fajta	B fajta	C fajta	D fajta
$\sum x$	-70	-33	4	59
$\sum x^2$	1990	964	487	1765

Döntsön 95%-os szinten, van-e eltérés a négy paradicsomfajta átlagos hozama között!

2. Nemparaméteres próbák

2.1. Binomiális próba

2.1. Példa Az egyik élemiszerbolt-hálózat üzleteibe érkező import baracknak eddig átlagosan 15%-a sérült meg szállítás közben. Miután beszállítót váltottak, az új szállítmányból megvizsgáltak 50 barackot. Ezek között 3 sérültet találtak. 95%-os szinten döntsön abban a kérdésben, megérte-e lecserélni a régi beszállítót!

Megoldás.

$$\begin{aligned} H_0 &: p = 0.15; \\ H_1 &: p < 0.15. \qquad \text{(egyoldali ellenhipotézis)} \\ \alpha &= 0.05. \end{aligned}$$

A próbastatisztika (a sérült barackok száma): $B = 3$.

Ha a H_0 igaz, B eloszlása $Bin(50, 0.15)$.

p-érték: $\mathsf{P}(B \leq 3) = 0.046 < 0.05$. Elvetjük H_0-t. □

2.2. Példa Egy felmérés során a háztartások mikrohullámú sütővel való ellátottságát vizsgálták. A véletlenszerűen kiválasztott 1000 háztartás 56%-ában találták meg a kérdéses háztartási gépet. 95%-os döntési szintet használva vizsgálja meg azt az állítást, miszerint a háztartásoknak kevesebb, mint 60%-a rendelkezik mikrohullámú sütővel!

Megoldás.

$$\begin{aligned} H_0 &: p = 0.6; \\ H_1 &: p < 0.6. \qquad \text{(egyoldali ellenhipotézis)} \\ \alpha &= 0.05. \end{aligned}$$

Jelölje B a mikrohullámú sütővel rendelkező háztartások számát. A nagy mintaelemszám miatt a próbastatisztika (folytonossági korrekcióval):

$$Z = \frac{B + 1/2 - np_0}{\sqrt{np_0(1-p_0)}} = \frac{560 + 1/2 - 600}{\sqrt{240}} = -2.5497.$$

Ha H_0 igaz, Z eloszlása közelítőleg standard normális.

A kritikus tartomány: $Z \leq Z(0.05) = -1.645$. A kapott Z érték beleesik, így elvetjük H_0-t.

Alternatív megoldás

p-érték: $\mathsf{P}(Z \leq -2.5497) = 0.0054 < 0.05$. Elvetjük H_0-t. □

2.3. Példa Péter talált egy elgörbült pénzérmét és kíváncsi volt, hogy ez befolyásolja-e a fej dobás valószínűségét. Tízszer feldobta az érmét, ami két alkalommal mutatott fejet.

a) Döntsön 95%-os szinten, hogy az elgörbült érmén azonos-e a fej, illetve az írás valószínűsége!

Péter keveselte a fej dobások számát és úgy gondolta, pontosabb eredményhez jut, ha a kísérletét többször ismétli meg. Ezért aztán még 390 alkalommal dobta fel az érmét, ami a 400 dobásból összesen 219 alkalommal mutatott fejet!

b) Az új adatok birtokában döntsön ismét 95%-os szinten az a) pontbeli kérdésről!

2.2. Előjel próba

2.4. Példa Egy mozitulajdonos állítása szerint az egy-egy rajzfilmre hetente eladott gyermekjegyek mediánja 300. Állításának alátámasztására kiválasztott 8, a moziban vetített rajzfilmet, és feljegyezte, hogy egy-egy filmre egy adott héten mennyi gyermekjegyet váltottak. A következő eredményeket kapta:

$$412,\ 232,\ 197,\ 454,\ 251,\ 114,\ 256,\ 318.$$

Hipotéziseit pontosan megfogalmazva, az előjel próba segítségével döntsön 90%-os szinten, igaz-e a mozitulajdonos állítása!

Megoldás.

$$H_0 : \mu = 300;$$
$$H_1 : \mu \neq 300.$$
$$\alpha = 0.1.$$

Az előjelek (érték - 300 előjele):

$$+ - - + - - - + .$$

A próbastatisztika (a + jelek száma): $B = 3$.
Ha H_0 igaz, B eloszlása $Bin(8, 0.5)$.

A kritikus tartomány: $B \leq c_a$ vagy $B \geq c_f$, ahol c_a a legnagyobb olyan egész, melyre

$$\sum_{k=0}^{c_a} \binom{8}{k} 0.5^8 \leq 0.05$$

és $c_f = 8 - c_a$. Esetünkben $c_a = 1,\ c_f = 7$. A kapott érték nem esik a kritikus tartományba, így elfogadjuk H_0-t.

Alternatív megoldás

p-érték: $\mathsf{P}(B \leq 3 \text{ vagy } B \geq 5) = 0.7266 > 0.1$. Elfogadjuk H_0-t. □

2.5. Példa Az alábbi adatok 12 *Turbo tudás* módszerrel felkészített hallgató vizsgapontszámait tartalmazzák (a maximális pontszám 50):

$$36 \quad 26 \quad 30 \quad 34 \quad 42 \quad 24 \quad 30 \quad 45 \quad 32 \quad 19 \quad 35 \quad 38.$$

Közismert, hogy a hagyományos módszerrel tanulók körében a pontok mediánja 30.

Az előjel próba segítségével döntsön 95%-os szinten, hogy az új módszerrel megszerzett pontok magasabb medián értékkel bírnak-e!

Megoldás.

$$H_0 : \mu = 30;$$
$$H_1 : \mu > 30. \qquad \text{(egyoldali ellenhipotézis)}$$
$$\alpha = 0.05.$$

Az előjelek (érték - 30 előjele):

$$+ \ - \ 0 \ + \ + \ - \ 0 \ + \ + \ - \ + \ + .$$

A próbastatisztika (a + jelek száma): $B = 7$.

Elhagyva a 0 különbségeket, ha H_0 igaz, B eloszlása $Bin(10, 0.5)$.

p-érték: $\mathsf{P}(B \geq 7) = 0.172 > 0.05$. Elfogadjuk H_0-t. □

2.6. Példa Egy olajtársaság állítása szerint az általuk az üzemanyagokhoz használt adalék lényegesen csökkenti a gépjárművek fogyasztását. Ennek ellenőrzésére 12 különböző középkategóriás benzinüzemű személygépkocsi fogyasztását tesztelték azonos körülmények között. A kapott eredmények az alábbiak (l/100 km):

$$6.8,\ 6.5,\ 6.9,\ 7.3,\ 6.4,\ 6.7,\ 7.0,\ 6.8,\ 6.9,\ 6.5,\ 7.2,\ 6.4.$$

Ismeretes, hogy a vizsgálatban résztvevő személygépkocsik fogyasztásának mediánja 7 l/100 km. Hipotéziseit pontosan megfogalmazva vizsgálja meg az olajtársaság állítását! Döntsön 95%-os szinten!

2.7. Példa Kétféle szójabab hozamát vizsgálva 12 parcellát megfeleztek, majd mindegyik parcella egyik felét az egyik, másik felét pedig a másik fajtával ültették be. A kilogrammban mért hozamokat az alábbi táblázat foglalja össze:

A fajta	142	124	133	151	121	127	135	141	149	150	151	132
B fajta	141	132	154	147	133	141	150	112	119	160	169	142

Hipotéziseit pontosan megfogalmazva döntsön 95%-os szinten, van-e különbség a két fajta hozama között!

Megoldás.

$$H_0 : \mu_{A-B} = 0; \quad \text{(a különbségek mediánja nulla)}$$
$$H_1 : \mu_{A-B} \neq 0.$$
$$\alpha = 0.05.$$

Különbség (d=A-B)	1	-8	-21	4	-12	-14	-15	29	30	-10	-18	-10
sign(d)	+	-	-	+	-	-	-	+	+	-	-	-

A próbastatisztika (a + jelek száma): $B = 4$.
Ha H_0 igaz, B eloszlása $Bin(12, 0.5)$.
A kritikus tartomány: $B \leq c_a$ vagy $B \geq c_f$, ahol c_a a legnagyobb egész, melyre

$$\sum_{k=0}^{c_a} \binom{12}{k} 0.5^{12} \leq 0.025$$

és $c_f = 8 - c_a$. Esetünkben $c_a = 2$, $c_f = 10$. A kapott érték nem esik a kritikus tartományba, így elfogadjuk H_0-t.

Alternatív megoldás

p-érték: $\mathsf{P}(B \leq 4 \text{ vagy } B \geq 8) = 0.3877 > 0.05$. Elfogadjuk H_0-t. □

2.8. Példa A *Roncsautó* című autós szaklap összehasonlította az azonos árfekvésű Skoda Sztrapacska és Lada Borscs 34 közös jellemzőjét. Az eredményt táblázatos formában is közölték, ahol a „+” jelentette, hogy az adott jellemző tekintetében a Skoda Sztrapacska bizonyult jobbnak, a „-”, hogy a Lada Borscs, a „0” pedig, hogy nincs különbség a két autó között. A táblázat összesítése: 21 „+ ”, 9 „-” és 4 „0”.

95%-os szintet alkalmazva vizsgálja meg a cseh autógyár állítását miszerint a Skoda Sztrapacska a jobb autó!

Megoldás.

H_0 : nincs különbség a „+” és „-” jelek száma között;

H_1 : több a „+” mint a „-”, azaz a Skoda a jobb. (egyoldali ellenhipotézis)

$\alpha = 0.05.$

A próbastatisztika (a + jelek száma): $B = 21$.
Elhagyva a 0 különbségeket, ha H_0 igaz, B eloszlása $Bin(30, 0.5)$.
p-érték: $\mathsf{P}(B \geq 21) = 0.021 < 0.05$. Elvetjük H_0-t. □

2.9. Példa Az előjelpróba segítségével döntsön 95%-os szinten, hogy van-e különbség az 1.8. Példában szereplő dolgozatok eredményei között!

2.3. Wilcoxon-féle előjeles rangösszeg próba

2.10. Példa Döntsön a Wilcoxon-féle előjeles rangösszeg próba segítségével, igazat állít-e a 2.4. Példában szereplő mozitulajdonos!

Megoldás.

$$H_0 : \mu = 300;$$
$$H_1 : \mu \neq 300.$$
$$\alpha = 0.1.$$

Eladott jegy (x)	412	232	197	454	251	114	256	318
$d = x - 300$	112	-68	-103	154	-49	-186	-44	18
rang $\big(\lvert d\rvert\big)$	6	4	5	7	3	8	2	1
sign(d)	+	-	-	+	-	-	-	+
Előjeles rang	6	-4	-5	7	-3	-8	-2	1

Próbastatisztika (a pozitív rangok összege): $T_+ = 14$. A mintaelemszám: $n = 8$.

Kritikus tartomány: $T_+ \leq T_a$ vagy $T_+ \geq 36 - T_a$. A T_a kritikus érték táblázatból határozható meg ($n = 8,\ \alpha = 0.1$, kétoldali ellenhipotézis), lásd Függelék. Esetünkben $T_a = 5$. A kapott érték nem esik a kritikus tartományba, így elfogadjuk H_0-t. □

2.11. Példa Döntsön a Wilcoxon-féle előjeles rangösszeg próba segítségével, igaz-e, hogy a 2.5. Példában szereplő *Turbo tudás* módszerrel felkészített hallgatók pontjainak mediánja nagyobb-e, mint a hagyományos módszerrel szerzett 30! A döntési szint legyen 95%-os!

Megoldás.

$$H_0 : \mu = 30;$$
$$H_1 : \mu > 30.$$
$$\alpha = 0.05.$$

Elhagyjuk a mediánnal egyező mintaelemeket.

Pontszám (x)	36	26	34	42	24	45	32	19	35	38
$d = x - 30$	6	-4	4	12	-6	15	2	-11	5	8
rang $(\lvert d\rvert)$	5.5	2.5	2.5	9	5.5	10	1	8	4	7
sign(d)	+	-	+	+	-	+	+	-	+	+
Előjeles rang	5.5	-2.5	2.5	9	-5.5	10	1	-8	4	7

Próbastatisztika (a pozitív rangok összege): $T_+ = 39$. A mintaelemszám: $n = 10$.

Kritikus tartomány: $T_+ \geq 55 - T_a$. A T_a kritikus érték táblázatból határozható meg ($n = 10,\ \alpha = 0.05$, egyoldali ellenhipotézis), lásd Függelék. Esetünkben $T_a = 10$. A kapott érték nem esik a kritikus tartományba, így elfogadjuk H_0-t. □

2.12. Példa Vizsgálja meg a 2.6. Példában szereplő olajtársaság állítását a Wilcoxon-féle előjeles rangösszeg próba segítségével! A döntési szint legyen 98%-os!

2.13. Példa A Wilcoxon-féle előjeles rangösszeg próba segítségével vizsgálja meg, van-e különbség a 2.7. Példában szereplő két szójabab hozama között! Hipotéziseit pontosan megfogalmazva döntsön 90%-os szinten!

Megoldás.

$$H_0 : \mu_{A-B} = 0; \quad \text{(a különbségek mediánja nulla)}$$
$$H_1 : \mu_{A-B} \neq 0.$$
$$\alpha = 0.1.$$

Különbség (d=A-B)	1	-8	-21	4	-12	-14	-15	29	30	-10	-18	-10
rang $(\lvert d \rvert)$	1	3	10	2	6	7	8	11	12	4.5	9	4.5
sign(d)	+	-	-	+	-	-	-	+	+	-	-	-
Előjeles rang	1	-3	-10	2	-6	-7	-8	11	12	-4.5	-9	-4.5

Próbastatisztika (a pozitív rangok összege): $T_+ = 26$. A mintaelemszám $n = 12$.

Kritikus tartomány: $T_+ \leq T_a$ vagy $T_+ \geq 78 - T_a$. A T_a kritikus érték táblázatból határozható meg ($n = 12$, $\alpha = 0.1$, kétoldali ellenhipotézis), lásd Függelék. Esetünkben $T_a = 17$. A kapott érték nem esik a kritikus tartományba, így elfogadjuk H_0-t. □

2.14. Példa Egy kísérlet során azt vizsgálták, hogy a rendszeres sportolás milyen hatással van a gyerekek pulzusszámára. 16 gyereket vontak be a kísérletbe, akik közül 8 versenyszerűen sportol, a másik nyolc pedig nem rendszeresen sportoló egészséges gyermek. Ez utóbbiakat úgy választották ki, hogy minden sportoló gyereknek legyen egy nem sportoló párja, akinek nagyjából azonos a kora, testmagassága, tömege és testfelszíne. Az alábbi táblázat a mért pulzusszámokat tartalmazza:

Pár	1	2	3	4	5	6	7	8
Nem sportoló	90	85	75	120	95	105	100	95
Sportoló	95	75	75	85	80	80	85	75

A Wilcoxon-féle előjeles rangösszeg próba segítségével vizsgálja meg, igaz-e, hogy a sportoló gyerekek pulzusa lassabban ver, mint a nem sportoló társaiké! Döntsön 97.5%-os szinten!

Megoldás. x: sportoló gyerek pulzusszáma; y: nem sportoló gyerek pulzusszáma.

$H_0 : \mu_{y-x} = 0;$ (pulzusszámok különbségének mediánja nulla)

$H_1 : \mu_{y-x} > 0.$ (egyoldali ellenhipotézis)

$\alpha = 0.025.$

Elhagyjuk a megegyező értékeket.

Különbség (d=y-x)	-5	10	35	15	25	15	20
rang $(\lvert d\rvert)$	1	2	7	3.5	6	3.5	5
sign(d)	-	+	+	+	+	+	+
Előjeles rang	-1	2	7	3.5	6	3.5	5

Próbastatisztika (a pozitív rangok összege): $T_+ = 27$. A mintaelemszám: $n = 7$.

Kritikus tartomány: $T_+ \geq 28 - T_a$. A T_a kritikus érték táblázatból határozható meg ($n = 7,\ \alpha = 0.025$, egyoldali ellenhipotézis), lásd Függelék. Esetünkben $T_a = 2$. A kapott érték a kritikus tartományba esik, így elvetjük H_0-t. □

2.15. Példa Egy motorosklub vezetője azt állítja, hogy tagjai által hetente a nyeregben töltött kilométerek mediánja 250 km.

20 véletlenszerűen kiválasztott klubtag lejegyezte, hány kilométert motorozott egy adott héten. A kapott adatokat az alábbi táblázat tartalmazza:

248	257	282	230	262	270	240	249	290	302
259	200	291	287	400	150	261	286	308	192

a) Hipotéziseit pontosan megfogalmazva az előjel próba segítségével döntsön 95%-os szinten, igaza van-e a klub vezetőjének!

A fenti mintában szereplő első tíz klubtagot megkérték, jegyezzék le azt is, menynyit motoroztak a következő héten. Az eredményeket az alábbi táblázat tartalmazza:

Motoros	1	2	3	4	5	6	7	8	9	10
1. hét	248	257	282	230	262	270	240	249	290	302
2. hét	260	292	198	235	280	284	301	270	288	324

b) Hipotéziseit pontosan megfogalmazzva döntsön 95%-os szinten

i) az előjel próba,

ii) a Wilcoxon-féle előjeles rangösszeg próba segítségével,

hogy a klubtagok a második héten többet motoroztak-e, mint az elsőn!

c) értékelje a b) pontban kapott eredményeket!

2.4. Mann–Whitney próba

2.16. Példa A Csajágóröcsögei Vegyipari Kombinát gépkezelői közül néhányat továbbképzésre küldtek annak érdekében, hogy munkájuk során kevesebb hibát vétsenek. A tanfolyam eredményességét vizsgálandó 6, a tanfolyamot már elvégzett, és 13 még előtte álló gépkezelőnek ugyanazt a feladatot adták és feljegyezték a végrehajtás során vétett hibáik számát.

Tanfolyam után	11	9	4	7	6	2							
Tanfolyam előtt	3	17	12	13	21	29	5	1	15	19	16	14	10

a) Hipotéziseit pontosan megfogalmazva egy alkalmas nemparaméteres próba segítségével döntsön 95%-os szinten, volt-e haszna a tanfolyamnak!

b) Mondjon legalább egy feltételt, amire figyelni kellett a kísérletben résztvevő gépkezelők kiválasztásánál!

Megoldás. x: tanfolyam előtti pontszám; y: tanfolyam utáni pontszám.
a) A kérdés eldöntésére a Mann–Whitney próbát alkalmazzuk.

$$H_0 : \mu_x = \mu_y; \quad \text{(hibák számának mediánjai megegyeznek)}$$
$$H_1 : \mu_x < \mu_y. \quad \text{(egyoldali ellenhipotézis)}$$
$$\alpha = 0.05.$$

Az egyesített rendezett minta (aláhúzással jelölve a tanfolyamot elvégzők adatait):

$$1,\ \underline{2},\ 3,\ \underline{4},\ 5,\ \underline{6},\ \underline{7},\ \underline{9},\ 10,\ \underline{11},\ 12,\ 13,\ 14,\ 15,\ 16,\ 17,\ 19,\ 21,\ 29.$$

Az aláhúzott elemek rangjai: 2, 4, 6, 7, 8, 10.

A rangösszeg: $U' = 37$.

A próbastatisztika ($m = 6,\ n = 13$): $U = U' - m(m+1)/2 = 37 - 21 = 16$.

A kritikus tartomány: $U \leq U_{6,13}(0.95) = 19$. A kapott érték beleesik, így elvetjük H_0-t.

b) A kísérletben résztvevő gépkezelőket véletlenszerűen választják ki! □

2.17. Példa A 2.4. Példában szereplő mozitulajdonos arra is kíváncsi, van-e különbség a családi filmek és az akciófilmek nézőszáma között. A kérdés eldöntése érdekében feljegyezte $6-6$ véletlenszerűen kiválasztott családi, illetve akciófilm nézettségi adatait, melyek az alábbiak:

Családi film	232	197	454	451	114	256
Akciófilm	285	278	149	283	119	196

Alkalmas kétmintás nemparaméteres próbát használva segítsen a mozitulajdonosnak! Döntsön 90%-os szinten!

Megoldás. x: családi film nézőszáma; y: akciófilm nézőszáma.

A kérdés eldöntésére a Mann–Whitney próbát alkalmazzuk.

$$H_0 : \mu_x = \mu_y; \quad \text{(nézőszámok mediánjai megegyeznek)}$$
$$H_1 : \mu_x \neq \mu_y.$$
$$\alpha = 0.1.$$

Az egyesített rendezett minta (aláhúzással jelölve a családi filmek adatait):

$$\underline{114},\ 119,\ 149,\ 196,\ \underline{197},\ \underline{232},\ \underline{256},\ 278,\ 283,\ 285,\ \underline{451},\ \underline{454}.$$

Az aláhúzott elemek rangjai: 1, 5, 6, 7, 11, 12.

A rangösszeg: $U' = 42$.

A próbastatisztika ($m = 6,\ n = 6$): $U = U' - m(m+1)/2 = 42 - 21 = 21$.

A kritikus tartomány: $U \leq U_{6,6}(0.95) = 7$ vagy $U \geq nm - U_{6,6}(0.95) = 29$. A kapott érték nem esik bele, így elfogadjuk H_0-t. □

2.18. Példa Mit tud mondani az 1.10. Példában szereplő két kávéfajta oldódási idejéről, ha az oldódási időkről nem tételezzük fel, hogy normális eloszlásúak? Döntsön 95%-os szinten!

2.5. Khi-négyzet próbák

2.19. Példa Egy újonnan kifejlesztett müzli ötféle magot (A, B, C, D és E) tartalmaz, melyek százalékos megoszlása a terméken lévő tájékoztató szerint 35%, 25%, 20%, 10%, illetve 10%. Egy véletlenül kiválasztott zacskóban az alábbi mennyiségi megoszlást találtuk:

Összetevő	A	B	C	D	E
Szem (darab)	184	145	100	68	63

Döntsön 90%-os szinten, hogy a minta összetétele megfelel-e a csomagoláson feltüntetettnek!

Megoldás.

H_0 : az összetétel megfelel a csomagoláson feltüntetettnek;

H_1 : az összetétel nem felel meg a csomagoláson feltüntetettnek.

$\alpha = 0.1.$

$N = 560,\ r = 5,\ p_1 = 0.35,\ p_2 = 0.25,\ p_3 = 0.20,\ p_4 = 0.10,\ p_5 = 0.10.$

Megfigyelt érték (k_ℓ):	184	145	100	68	63
Várt érték (Np_ℓ):	196	140	112	56	56

A próbastatisztika:

$$\mathcal{X}^2 = \sum_{\ell=1}^{r} \frac{(k_\ell - Np_\ell)^2}{Np_\ell} = 5.6454.$$

Ha H_0 igaz, a próbastatisztika aszimptotikus eloszlása khi-négyzet eloszlás $\nu = 4$ szabadsági fokkal.

A kritikus tartomány: $\mathcal{X}^2 \geq \mathcal{X}_4^2(0.9) = 7.779$. A kapott érték nem esik bele, így elfogadjuk H_0-t. □

2.20. Példa Egy számítógép segítségével 12, a $[-6, 6]$ intervallumon vett egyenletes eloszlásból származó véletlen számot generáltunk, majd ezt még 99 alkalommal megismételtük. A száz darab mintaátlag eloszlását az alábbi táblázatban összesítettük:

	Megfigyelt gyakoriság
$(-\infty, -0.6745)$	26
$[-0.6745, 0)$	21
$[0, 0.6745)$	27
$[0.6745, \infty)$	26

a) Vizsgálja meg 95%-os szinten azt a hipotézist, hogy a mintaátlagok a négy felsorolt intervallum mindegyikébe azonos valószínűséggel esnek!

b) $-0.6745,\ 0$ és 0.6745 a standard normális eloszlás alsó kvartilise, mediánja, illetve felső kvartilise. Fejtse ki, milyen kapcsolatban áll az előző pontban kapott eredmény a központi határeloszlás tétellel!

Megoldás.

a)

$$H_0 : p_\ell = 0.25,\ \ell = 1, 2, 3, 4;$$
$$H_1 : \exists\, \ell \in \{1, 2, 3, 4\},\ p_\ell \neq 0.25.$$
$$\alpha = 0.05.$$

$N = 100,\ r = 4.$

Megfigyelt gyakoriság (k_ℓ):	26	21	27	26
Várt gyakoriság (Np_ℓ):	25	25	25	25

A próbastatisztika:

$$\mathcal{X}^2 = \sum_{\ell=1}^{r} \frac{(k_\ell - Np_\ell)^2}{Np_\ell} = 0.8800.$$

Ha H_0 igaz, a próbastatisztika aszimptotikus eloszlása khi-négyzet eloszlás $\nu = 3$ szabadsági fokkal.

A kritikus tartomány: $\mathcal{X}^2 \geq \mathcal{X}_3^2(0.95) = 7.815$. A kapott érték nem esik bele, így elfogadjuk H_0-t.

b) Egy 12 elemű, a $[-6, 6]$ intervallumon vett egyenletes eloszlásból származó minta esetén a mintaátlag várható értéke 0, szórása pedig 1. A khi-négyzet próba alapján a mintaátlagok eloszlása a központi határeloszlás tételből adódó standard normális eloszlás. $\square$

2.21. Példa Egy elsőéves programozó informatikus hallgatónak házi feladatként egy olyan programot kellett írnia, mely egyenletes eloszlás szerint generál véletlen számokat az $1, 2, \ldots, 15$ halmazból. Jelölje X az első hárommal osztható szám megjelenéséig generált véletlen számok számát (beleértve az utolsó hárommal osztható számot is). Ha a véletlenszám generátor jól működik, akkor X geometriai eloszlású, azaz

$$\mathsf{P}(X = \ell) = p(1-p)^{\ell-1}, \qquad \ell = 1, 2, \ldots,$$

ahol p annak a valószínűsége, hogy a generált szám osztható hárommal, vagyis $p = 1/3$. Az alábbi táblázat az X változó 160 megfigyelt értékét tartalmazza:

A generált egészek száma (x)	1	2	3	4	5	6	7	8	>8
Gyakoriság (k)	63	34	28	13	9	7	2	4	0

a) Számítsa ki a mintaátlagot!

b) Döntsön 95%-os szinten arról, hogy a minta a $p = 1/3$ paraméterű geometriai eloszlásból származik-e!

c) Döntsön 90%-os szinten arról, hogy a minta egyáltalán geometriai eloszlásból származik-e!

Megoldás.

a)

$$\bar{x} = \frac{\sum kx}{\sum x} = \frac{\sum kx}{N} = \frac{400}{160} = 2.5.$$

b)

H_0 : a minta 1/3 paraméterű geometriai eloszlásból származik;

H_1 : a minta nem a fenti eloszlásból származik.

$\alpha = 0.05.$

A modell:

$$p_\ell = \mathsf{P}(X = \ell) = 0.3(0.7)^{\ell-1}, \qquad \ell = 1, 2, \ldots.$$

Megfigyelt gyak. (k_ℓ):	63	34	28	13	9	7	2	4	0
Várt gyak. (Np_ℓ):	48	33.6	23.52	16.46	11.52	8.07	5.65	3.95	9.23

Az utolsó előtti kategóriában a várt gyakoriság túl kicsi (< 5), így ezt a kategóriát összevonjuk az előtte lévővel ($r = 8$, az összevont kategóriában a megfigyelt gyakoriság 6, a várt gyakoriság: 9.6).

A próbastatisztika:

$$\mathcal{X}^2 = \sum_{\ell=1}^{r} \frac{(k_\ell - Np_\ell)^2}{Np_\ell} = 18.1919.$$

Ha H_0 igaz, a próbastatisztika aszimptotikus eloszlása khi-négyzet eloszlás $\nu = 7$ szabadsági fokkal.

A kritikus tartomány: $\mathcal{X}^2 \geq \mathcal{X}_7^2(0.9) = 14.067$. A kapott érték beleesik, így elvetjük H_0-t.

c)

H_0 : a minta geometriai eloszlásból származik;

H_1 : a minta nem geometriai eloszlásból származik.

$\alpha = 0.1.$

Az ismeretlen p paraméter maximum likelihood becslése: $1/\bar{x} = 1/2.5 = 0.4$.

A becsült modell:

$$p_\ell = \mathsf{P}(X = \ell) = 0.4(0.6)^{\ell-1}, \qquad \ell = 1, 2, \ldots.$$

Megfigyelt gyak. (k_ℓ):	63	34	28	13	9	7	2	4	0
Várt gyak. (Np_ℓ):	64	38.4	23.04	13.82	8.29	4.98	2.99	1.79	2.69

Az utolsó négy kategóriában a várt gyakoriságok túl kicsik (< 5), így ezeket a kategóriákat összevonjuk ($r = 6$, az utolsó kategóriában a megfigyelt gyakoriság 13, a várt gyakoriság: 12.45).

A próbastatisztika:

$$\mathcal{X}^2 = \sum_{\ell=1}^{r} \frac{(k_\ell - Np_\ell)^2}{Np_\ell} = 1.7213.$$

Ha H_0 igaz, a próbastatisztika aszimptotikus eloszlása khi-négyzet eloszlás $\nu = 4$ szabadsági fokkal (egy paramétert becsültünk).

A kritikus tartomány: $\mathcal{X}^2 \geq \mathcal{X}_4^2(0.9) = 7.779$. A kapott érték nem esik bele, így elfogadjuk H_0-t. $\square$

2.22. Példa Egy biológus megvizsgálta azt az elméletet miszerint egy bizonyos rovarfaj napban kifejezett élettartama a $[0, 20]$ intervallumon vett egyenletes eloszlással modellezhető. A kutatásai során kapott adatokat az alábbi táblázat összesíti:

Élettartam (a legközelebbi egész napra kerekítve)	0–2	3–5	6–10	11–20
Rovarok száma	38	53	75	112

a) Vizsgálja meg, a kapott élettartamak eloszlása tényleg megfelel-e az elmélet által meghatározottnak! Döntsön 95%-os szinten!

A kutató azt is megfigyelte, hogy az általa vizsgált rovarok egyike sem élt tovább 16 napnál. Úgy döntött tehát, hogy új elméletet állít fel, miszerint az élettartamot a $[0, 16]$ intervallumon vett egyenletes eloszlás modellezi.

b) Döntsön 95%-os szinten, vajon az adatok alátámasztják-e ezt az elméletet!

2.23. Példa Egy botanikus hallgató úgy gondolta, hogy egy bizonyos növényfajta a füves réteken véletlenszerűen szétszórt helyeken bukkan fel. Kutatásai során megszámolta a növény egy véletlenszerűen kiválasztott egy négyzetméteres négyzetben (kvadráns) előforduló egyedeinek a számát, majd-e kísérletet többször is megismételte. Az így kapott megfigyeléseit az alábbi táblázatban összegezte:

A növények száma	0	1	2	3	4	5	6	legalább 7
Gyakoriság	9	24	43	34	21	15	2	0

a) Az adatokból számítsa ki a vizsgált növény egyedeinek egy négyzetméterre eső átlagos számát!

A szakkönyvek szerint a fenti jellegű megfigyelési eredmények Poisson eloszlással modellezhetők.

b) Döntsön 95%-os szinten, vajon a Poisson modell megfelelően illeszkedik-e a hallgató által kapott adatokra!

2.24. Példa Egy kutatócsoport azt vizsgálta, van-e összefüggés egy bizonyos betegség lefolyásának súlyossága és a betegek életkora között. A vizsgálat során 200 beteg adatait gyűjtötték össze, majd azokat csoportosították a betegség súlyossági foka és a paciens életkora szerint. Eredményül az alábbi táblázatot kapták:

		Életkor		
		40 alatti	40–60	60 fölötti
Lefolyás	enyhe	41	34	9
	közepes	25	25	12
	súlyos	6	33	15

Hipotéziseit pontosan megfogalmazva döntsön 99%-os szinten, van-e összefüggés a betegek életkora és a betegség lefolyásának súlyossága között!

Megoldás.

$$H_0 : \text{nincs összefüggés;}$$
$$H_1 : \text{van összefüggés.}$$
$$\alpha = 0.01.$$

$r = s = 3,\ N = 200.$

Megfigyelt gyakoriságok ($k_{ij},\ i = 1, \ldots, r,\ j = 1, \ldots, s$):

$$\begin{array}{ccc|c} 41 & 34 & 9 & 84 \\ 25 & 25 & 12 & 62 \\ 6 & 33 & 15 & 54 \\ \hline 72 & 92 & 36 & 200 \end{array}$$

Várt gyakoriságok ($k^*_{ij} = k_{i\cdot}k_{\cdot j}/N,\ i = 1, \ldots, r,\ j = 1, \ldots, s$):

$$\begin{array}{ccc|c} 30.24 & 38.64 & 15.12 & 84 \\ 22.32 & 28.52 & 11.16 & 62 \\ 19.44 & 24.84 & 9.72 & 54 \\ \hline 72 & 92 & 36 & 200 \end{array}$$

A próbastatisztika:

$$\mathcal{X}^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(k_{ij} - k^*_{ij})^2}{k^*_{ij}} = 22.5230.$$

Ha H_0 igaz, a próbastatisztika aszimptotikus eloszlása khi-négyzet eloszlás $\nu = (r-1)(s-1) = 4$ szabadsági fokkal.

A kritikus tartomány: $\mathcal{X}^2 \geq \mathcal{X}^2_4(0.99) = 13.277$. A kapott érték beleesik, így elvetjük H_0-t. $\square$

2.25. Példa A Szváziföldi Gyáriparosok Szövetségének elnöke egy interjúban a vállalatvezetők véleményéről beszélt abban a kérdésben, hogy Szváziföld csatlakozzon-e az Európai Unióhoz. A nyilatkozó azt állította, az integráció támogatottsága függ attól, hogy az illető vezető mekkora vállalat élén áll. Az elnök állítását ellenőrizendő egy közvéleménykutató cég kikérte közel háromszáz véletlenszerűen kiválasztott vállalat első emberének véleményét a kérdésről. Az eredményeket az alábbi táblázat tartalmazza.

	A vállalat mérete		
	Nagy	Közepes	Kicsi
Támogatja	13	24	76
Ellenzi	7	26	143

a) Döntsön 99%-os szinten, hogy az adatok alátámasztják-e a Szövetség elnökének állítását!

A későbbi adatelemzések során kiderült, hogy az egyik kérdezőbiztos hibázott, mivel egy vállalatot kifelejtett az összesítésből. Így azon közepes méretű vállalatok száma, melyek vezetője támogatja a csatlakozást 25-re módosult.

b) Az újabb adatot felhasználva döntsön ismét 99%-os szinten!

Hivatkozások

[1] Bolsev, L. N., Szmirnov, N. V., *Matematikai statisztikai táblázatok.* Nauka, Moszkva, 1965 (orosz nyelven).

[2] Fazekas I. (szerk.), *Bevezetés a matematikai statisztikába.* Kossuth Egyetemi Kiadó, Debrecen, 2000.

[3] Hunyadi L., Mundruczó Gy., Vita L., *Statisztika.* Aula Kiadó, Budapest, 1996.

[4] Lehmann, E. L. *Nonparametrics: statistical methods based on ranks.* Holden-Day, San Francisco, 1975.

[5] Lukács O., *Matematikai statisztika példatár.* Műszaki Könyvkiadó, Budapest, 1987.

[6] Mogyoródi J., Michaletzky Gy. (szerk.), *Matematikai statisztika.* Nemzeti Tankönyvkiadó, Budapest, 1995.

[7] Pratt, J. W., Gibbons, H. D., *Concepts of Nonparametric Theory.* Springer, New York, 1981.

[8] Vincze I., Varbanova M. *Nemparaméteres matematikai statisztika.* Akadémiai kiadó, Budapest, 1993.

Függelék

A Wilcoxon-féle előjeles rangösszeg próba kritikus értékei

n	α					$\frac{n(n+1)}{2}$
	0.01	0.025	0.05	0.10	0.20	
7	0	2	3	5	8	28
8	1	3	5	8	11	36
9	3	5	8	10	14	45
10	5	8	10	14	18	55
11	7	10	13	17	22	66
12	9	13	17	21	27	78

A Wilcoxon-féle előjeles rangösseg próba T_a kritikus értékei egyoldali próba esetén. Kétoldali próbánál az $\alpha/2$-höz tartozó értéket kell tekinteni.