Gyakorló feladatok_alapok

a Kísérletek tervezése és értékelése c. tárgyból

Eloszlások

Normális eloszlás

1. példa

Egy $\mu = 12.35$ várható értékű és $\sigma^2 = 10^{-4}$ varianciájú normális eloszlásból 9 elemű mintát veszünk. Számítsuk ki, milyen szimmetrikus intervallumban lesz 95 %-os valószínűséggel a minta átlagértéke!

Megoldás

$P\left(\overline{x}_a < \overline{x} < \overline{x}_f\right) = 0.95$ [1]

$$P\left(\frac{\overline{x}_a - \mu}{\sigma/\sqrt{n}} < \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} < \frac{\overline{x}_f - \mu}{\sigma/\sqrt{n}}\right) = P\left(z_a < z < z_f\right) = 0.95$$

$$z_a = -z_f \qquad z_f = z_{0.975} = 1.96$$

$$\overline{x}_a = \mu - z_a\,\sigma/\sqrt{n} = 12.35 - 1.96 \cdot 10^{-2}/3 = 12.343; \qquad \overline{x}_f = 12.35 + 1.96 \cdot 10^{-2}/3 = 12.357$$

$$P\left(12.343 < \overline{x} < 12.357\right) = 0.95$$

2. példa

A 300 mm névleges hosszúságú dobozokat gyártó gépsoron készült dobozok hosszmérete 299.5 mm körül ingadozik, varianciájának négyzetgyöke 0.8 mm. Elfogadva, hogy a dobozok méretének eloszlása normális, a dobozok hány %-ának nem megfelelő a mérete (selejt), ha a tűréshatár 300.0 ± 2.0 mm?

Megoldás

$$1 - P\left(298 < x < 302\right) = 1 - 0.9696 = 0.0304$$

A *z*-eloszlás táblázata a $P\left(3.125 < z\right)$ valószínűséget nem tartalmazza (*z*=3.09 a legnagyobb érték a táblázatban), ezért a számolás során azzal a közelítéssel élünk, hogy $P\left(302 < x\right) = P\left(3.125 < z\right) \approx 0$.

3. példa

Egy régóta gyártott elektromos kondenzátor élettartama normális eloszlásúnak tekinthető, σ = 225 h. Véletlenszerűen kivéve 30 elemű mintát, az átlagos élettartamra 1407.5 h adódott. Adjuk meg a várható élettartam 99%-os konfidencia intervallumát.

Megoldás

$$P\left(1301.5 < \mu < 1513.5\right) = 0.99$$

[1] A példák megoldásánál az intervallumok megadásakor az egyenlőség jelet az egyszerűség kedvéért elhagyjuk. Ezt azért tehetjük meg, mert folytonos eloszlású véletlen változókról van szó.

4. példa

Feltételezhetjük, hogy a névlegesen 1kg-os csomagolású kristálycukor tényleges súlyeloszlása közelítőleg normális, várható értéke 1.000 kg, és σ= 0.01 kg. A cukorral töltött zacskók hány százalékának lesz a súlya 0.985 kg alatt?

Megoldás

$$P(x \le 0.985) = P(z \le -1.5) = 0.06681$$

5. példa

A 6 mm névleges átmérőjű golyókat gyártó gépsoron készült golyók átmérőjének várható értéke 5.998 mm, varianciájának négyzetgyöke 0,006 mm. Elfogadva, hogy a golyók átmérőjének eloszlása normális, a golyók hány %-ának nem megfelelő a mérete (selejt), ha a tűréshatár 6.000±0.009 mm?

Megoldás

$$F(1.833) = 0.9664$$
$$F(-1.167) = 1 - F(1.167) = 1 - 0.8784 = 0.1216$$
A selejtarány:
$$(1 - 0.9664) + 0.1216 = 0.0336 + 0.1216 = 0.1552$$

t-eloszlás

6. példa

Egy normális eloszlásból vett 9 elemű minta elemeinek átlagértéke $\bar{x} = 18.5$, szórásnégyzete $s^2 = 2.35 \cdot 10^{-4}$. Milyen szimmetrikus intervallumban van 95 %-os valószínűséggel a sokaság μ várható értéke? (Adjunk 95 %-os konfidencia-intervallumot a várható értékre!)

Megoldás

$$P\left(-t_{\alpha/2} < t < t_{\alpha/2}\right) = 1 - \alpha; \qquad \alpha = 0.05; \qquad \nu = n - 1 = 8$$
$$P\left(-t_{\alpha/2} < \frac{\bar{x} - \mu}{s/\sqrt{n}} < t_{\alpha/2}\right) = 1 - \alpha$$
$$P\left(\bar{x} - t_{\alpha/2}\, s/\sqrt{n} < \mu < \bar{x} + t_{\alpha/2}\, s/\sqrt{n}\right) = 1 - \alpha$$
$$t_{0.025}(8) = 2.306 \qquad \frac{\sqrt{2.35 \cdot 10^{-4}} \cdot 2.306}{\sqrt{9}} = 0.012$$
$$P(18.5 - 0.012 < \mu < 18.5 + 0.012) = 0.95$$
$$P(18.488 < \mu < 18.512) = 0.95$$

7. példa

Az MM-Liner karton négyzetmétertömegére az alábbi mérési adatokat kapták (5, egymástól független mintavétel) 230.8 g/m^2; 230.6 g/m^2; 229.9 g/m^2; 229.7 g/m^2; 230.4 g/m^2. Elfogadva, hogy az adatok normális eloszlásúak, adja meg a négyzetmétertömeg várható értékének 95%-os konfidencia intervallumát!

Megoldás

$$P\left(229.70 < \mu < 230.86\right) = 0.95$$

8. példa

Ismételt méréseket végeztünk egy szállítmány hatóanyagtartalmának meghatározására: 24.88; 24.92; 24.67; 25.21; 25.28. Elfogadva, hogy az adatok normális eloszlásúak, adja meg, hogy a tényleges hatóanyagtartalom mely érték felett található 95%-os valószínűséggel!

Megoldás

$$P(24.753 \le \mu) = 0.95$$

Khi-négyzet eloszlás[2]

9. példa

Számítsuk ki, hogy egy $\mu = 12.35$ várható értékű és $\sigma^2 = 10^{-4}$ varianciájú normális eloszlásból 9 elemű minta korrigált tapasztalati szórásnégyzete milyen alsó és felső határ közé esik 95 %-os szimmetrikus valószínűséggel!

Megoldás

$$P\left(s_a^2 < s^2 < s_f^2\right) = 0.95$$

$$P\left(\frac{s_a^2 \nu}{\sigma^2} < \frac{s^2 \nu}{\sigma^2} < \frac{s_f^2 \nu}{\sigma^2}\right) = 0.95; \qquad \nu = n - 1$$

$$P\left(\chi_a^2 < \chi^2 < \chi_f^2\right) = 0.95$$

$$F\left(\chi^2 \le \chi_f^2\right) = 0.975 \to \chi_f^2 = \chi_{0.025}^2(8) = 17.535;$$

$$F\left(\chi^2 \le \chi_a^2\right) = 0.025 \to \chi_a^2 = \chi_{0.975}^2(8) = 2.180$$

$$s_a^2 = \frac{\chi_a^2 \cdot \sigma^2}{\nu} = \frac{2.18 \cdot 10^{-4}}{8} = 2.725 \cdot 10^{-5}; \qquad s_f^2 = \frac{\chi_f^2 \cdot \sigma^2}{\nu} = \frac{17.535 \cdot 10^{-4}}{8} = 2.192 \cdot 10^{-4}$$

$$P\left(2.725 \cdot 10^{-5} < s^2 < 2.192 \cdot 10^{-4}\right) = 0.95$$

[2] A fejezet néhány feladatában a khi-négyzet eloszlás mellett, a *z*- és *t*-eloszlást is kell használni.

10. példa

Négy egymástól független hajlítási merevség mérés eredménye: 7.2; 7.8; 8,0; 7.7. Elfogadva, hogy az adatok normális eloszlásúak, adja meg a várható érték (μ) 99%-os és σ^2 95%-os konfidencia intervallumát.

Megoldás

$P(6.68 < \mu < 8.67) = 0.99$ $\qquad$ $P(0.0372 < \sigma^2 < 1.609) = 0.95$

11. példa

Egy karton hajlítási merevségére szálirányban tíz, egymástól független mérést végeztek. A 10 mérés átlaga és a szórás: $\bar{x} = 23.5;\ s = 0.4$. Elfogadva, hogy az adatok normális eloszlásúak, adja meg a várható érték (μ) 90%-os és σ^2 95%-os konfidencia intervallumát.

Megoldás

$P(23.27 < \mu < 23.73) = 0.90$ $\qquad$ $P(0.0757 < \sigma^2 < 0.5333) = 0.95$

12. példa

Egy kartonpapír négyzetmétertömegének várható értéke 250 g/m^2, varianciája $\sigma^2 = 1.0\left(\text{g/m}^2\right)^2$. Feltételezve, hogy a négyzetmétertömeg normális eloszlású, milyen intervallumban lesz 95%-os valószínűséggel egy 5 elemű, véletlen minta

a) számtani középértéke,

b) korrigált tapasztalati szórásnégyzete?

Megoldás

a) $\bar{x}_a = 249.1$ $\qquad$ $\bar{x}_f = 250.9$

b) $s_f^2 = 2.786$ $\qquad$ $s_a^2 = 0.121$

13. példa

Négy egymástól független pH mérés eredménye: 7.90, 7.94, 7.91 és 7.93. Elfogadva, hogy a mérési hiba normális eloszlású

a) adjon 95%-os felső határt a pH mérés várható értékére!

b) adjon 90%-os alsó határt a pH mérés varianciájára!

Megoldás

a) $P(\mu < 7.942) = 0.95$

b) $P(1.607 * 10^{-4} < \sigma^2) = 0.90$

14. példa

Egy szerves oldat nedvességtartalmát tíz, egymástól független méréssel határoztuk meg. A 10 mérés átlaga és a szórás: $\bar{x} = 0.552$ g/kg, $s = 0.037$ g/kg. A mérési hiba normális eloszlásúnak tekinthető.

a) Milyen valószínűséggel lesz a nedvességtartalom a 0.59 g/kg-os érték felett?

b) Milyen érték alatt van a nedvességtartalom varianciája 95%-os valószínűséggel?

Megoldás

$$P(\mu > 0.59) = 0.005$$
$$P(\sigma^2 < 0.0037) = 0.95$$

F-eloszlás

15. példa

Egyazon normális eloszlású sokaságból két mintát veszünk. Az első 6 elemű, a második 10 elemű. Milyen szimmetrikus intervallumban lesz a két szórásnégyzet hányadosa 90 %-os valószínűséggel?

Megoldás

$$P\left(F_a < F < F_f\right) = 0.90; \qquad \nu_1 = 5;\ \nu_2 = 9$$

$$P\left(F_a < \frac{s_1^2}{s_2^2} < F_f\right) = 0.90$$

$$F_f = F_{0.05}(5,9) = 3.48 \quad F_a = F_{0.95}(5,9) = \frac{1}{F_{0.05}(9,5)} = \frac{1}{4.77} = 0.210$$

$$P\left(0.210 < \frac{s_1^2}{s_2^2} < 3.48\right)$$

Hipotézisvizsgálat

A hipotézisvizsgálathoz kapcsolódó példák két csoportra oszthatók. Az egyik példatípusnál a feladat szövege egyértelműen közli, hogy mi a vizsgálandó nullhipotézis. A másik, öszszetettebb feladattípusnál a szakmai problémát írjuk le, ott az olvasó feladat eldönteni, hogy mi kerüljön a nullhipotézisbe.

z-próba

16. példa

Ismert varianciájú ($\sigma^2 = 1600$) normális eloszlású sokaságból vett 64 elemű minta elemeinek középértéke 136.5. Megvizsgálandó 0.01-os szignifikanciaszinten az a nullhipotézis, hogy a sokaság várható értéke 130.

Megoldás:

$H_0 : \mu = \mu_0 = 130 \qquad H_1 : \mu \neq \mu_0 = 130$

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{136.5 - 130}{\sqrt{1600/64}} = 1.3; \qquad z_{krit} = z_{0.995} = 2.58;$$

Elfogadási tartomány: $-z_{\text{krit}} < z_0 < z_{\text{krit}}$.

Elfogadjuk a nullhipotézist, mivel $-z_{\text{krit}} < z_0 < z_{\text{krit}}$ teljesül. Az adatok nem mondanak ellen annak a feltételezésnek, hogy a sokaság várható értéke 130.

t-próba

17. példa

A 320 μm névleges vastagságú kartonpapírból 5 mintát vettek. A mért vastagságok számtani középértéke 317.8 μm, szórása 2.1 μm. Elfogadva, hogy a kartonpapír vastagságának eloszlása normális, megvizsgálandó α = 0.05 szignifikanciaszinten az a nullhipotézis, hogy a papír vastagság várható értéke legalább 320.0 μm!

Megoldás:

$H_0 : \mu \geq \mu_0 = 320 \qquad H_1 : \mu < \mu_0 = 320$

$$t_0 = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{317.8 - 320}{2.1/\sqrt{5}} = -2.34; \qquad -t_{krit} = -t_{0.05}(4) = -2.132;$$

Elfogadási tartomány: $-t_{\text{krit}} < t_0$.

Elutasítjuk a nullhipotézist, mivel t_0 az elfogadási tartományon kívül esik.
Tehát az adatok 5%-os szignifikanciaszinten ellentmondanak annak, hogy a papír vastagság várható értéke legalább 320.0 μm. Az adatok 5%-os szignifikanciaszinten azt bizonyítják, hogy a papírvastagság várható értéke 320.0 μm alatt van.

18. példa

Egy készítmény hatóanyagtartalmát négy ismételt méréssel határozták meg. A tapasztalati szórásnégyzet 0.5476, a négy mérési eredmény számtani középértéke pedig 0.452%. Alátámasztják-e ezek az eredmények azt a gyanút, hogy a készítmény hatóanyagtartalma kevesebb mint 0.5%? α=0.05 szignifikanciaszinten vizsgálja a kérdést!

Megoldás

$H_0: \quad \mu \geq \mu_0 = 0.5 \qquad H_1: \quad \mu < \mu_0 = 0.5 \qquad$ (baloldali ellenhipotézis)

$t_0 = -0.130 \qquad -t_{krit} = -t_{0.05}(3) = -2.353$

Elfogadási tartomány: $-t_{krit} < t_0$, elutasítási tartomány: $t_0 < -t_{krit}$, tehát a nullhipotézist elfogadjuk. Az adatok nem bizonyítják, hogy a hatóanyagtartalom 0.5% alatt van.

19. példa

Egy bizonyos oldat szennyeződését vizsgálva, az oldatból vett 7 db minta elemzési eredményei a következők (%-ban): 7.18, 7.17, 7.12, 7.13, 7.14, 7.15 és 7.16. A megengedett szennyeződés max. 7.13%. Elfogadva, hogy az adatok normális eloszlásúak, $\alpha = 0.05$ szignifikanciaszinten vizsgálja azt a nullhipotézist, hogy az oldat szennyeződése a határérték alatt van.

Megoldás

$H_0: \quad \mu \le \mu_0 = 7.13 \quad H_1: \quad \mu > \mu_0 = 7.13$ (jobboldali ellenhipotézis)

$t_0 = 2.450 \qquad t_{krit} = t_{0.05}(6) = 1.943$

Elfogadási tartomány: $t_0 < 1.943$, elutasítási tartomány: $1.943 < t_0$.

Tehát a nullhipotézist elutasítjuk. Az adatok 5%-os szignifikanciaszinten bizonyítják, hogy a szennyezés túllépte a határértéket.

A fordítva feltett nullhipotézist ($H_0: \quad \mu \ge \mu_0 = 7.13$) elfogadnánk. Ez még önmagában nem bizonyítaná azt, hogy a szennyezés túllépte a határértéket. (Ez alapján még nem lenne jogos megbírságolni azt a céget, aki kibocsátotta a szennyvizet.) Csak annyit tudnánk mondani, hogy nem bizonyítható, hogy a szennyezés koncentrációja a határérték alatt volt.

Khi-négyzet próba

20. példa

Normális eloszlású sokaságból vett 17 elemű minta szórásnégyzete 0.24. Megvizsgálandó $\alpha = 0.05$ szignifikanciaszinten az a nullhipotézis, hogy a variancia értéke legfeljebb 0.18.

Megoldás

$H_0: \sigma^2 \le \sigma_0^2 = 0.18 \qquad H_1: \sigma^2 > \sigma_0^2 = 0.18$

$$\chi_0^2 = \frac{s^2 \nu}{\sigma_0^2} = \frac{0.24 \cdot 16}{0.18} = 21.33$$

$\chi_0^2 = \frac{0.24 \cdot 16}{\sigma^2} \cdot \frac{\sigma^2}{0.18} = \chi^2 \frac{\sigma^2}{0.18}$; ha H_1 igaz, $\frac{\sigma^2}{0.18} > 1 \rightarrow \chi_0^2$ jobbra eltolódik, tehát egyoldali fölső határa van a nullhipotézis elfogadási tartományának:

$\chi_{krit}^2 = \chi_{0.05}^2(16) = 26.296$.

Elfogadási tartomány: $\chi_0^2 < \chi_{krit}^2$.

Elfogadjuk a nullhipotézist, mivel $\chi_0^2 < \chi_{kit}^2$ teljesül. Az adatok nem mondanak ellent annak a feltételezésnek, hogy a sokaság varianciája legfeljebb 0.18.

21. példa

Egy szerves anyag víztartalmát 10 ismételt méréssel meghatározva a mérések átlaga: $\bar{x} = 0.452$ g/kg, a szórás: $s = 0.37$ (g/kg)^2. Elfogadva, hogy az adatok közelítőleg normális eloszlásúak,

a) vizsgálja meg 5%-os szignifikanciaszinten azt a nullhipotézist, hogy a víztartalom nem haladja meg a 0.5 g/kg értéket;

b) vizsgálja meg 5%-os szignifikanciaszinten azt a nullhipotézist, hogy a víztartalom mérési módszerének varianciája (σ^2) nem nagyobb, mint 0.16 (%)^2.

Megoldás

a) $H_0: \quad \mu \le \mu_0 = 0.5$ $\qquad H_1: \quad \mu > \mu_0$ $\qquad$ (jobboldali ellenhipotézis)

$t_0 = -0.410 \qquad t_{krit} = t_{0.05}(9) = 1.833$

Elfogadási tartomány: $t_0 < 1.833$, elutasítási tartomány: $1.833 < t_0$, tehát a nullhipotézist elfogadjuk. Tehát az adatok nem bizonyítják, hogy a víztartalom meghaladja a 0.5%-ot.

b) $H_0: \quad \sigma^2 \le \sigma_0^2 = 0.16$ $\qquad H_1: \quad \sigma^2 > \sigma_0^2$ $\qquad$ (jobboldali ellenhipotézis)

$\chi_0^2 = 7.70 \qquad \chi_{krit}^2 = \chi_{0.05}^2(9) = 16.919$

Mivel $\chi_0^2 < \chi_{krit}^2$, ezért elfogadjuk a nullhipotézist. Az adatok nem mondanak ellent annak a feltételezésnek, hogy a variancia 0.16 (%)^2 alatt van.

F-próba

22. példa

Két független, normális eloszlásból vett mintánk van, az egyik 11 elemű, korrigált tapasztalati szórásnégyzete 0.76; a másik 14 elemű, szórásnégyzete 0.38. Megvizsgálandó α = 0.05 szignifikanciaszinten az a nullhipotézis, hogy az első minta mögött álló sokaság varianciája legfeljebb akkora, mint a másik sokaság varianciája.

Megoldás

$H_0: \sigma_1^2 \le \sigma_2^2 \qquad H_1: \sigma_1^2 > \sigma_2^2$

$F_0 = \dfrac{0.76}{0.38} = 2.00 \qquad F_{krit} = F_{0.05}(10, 13) = 2.67$;

Elfogadási tartomány: $F_0 < F_{krit}$.

Elfogadjuk a nullhipotézist, mivel $F_0 < F_{\mathrm{krit}}$ teljesül. Tehát az adatok nem mondanak ellent annak a feltételezésnek, hogy az első sokaság varianciája legfeljebb akkora, mint a másodiké. (Vagy: az adatok nem bizonyítják, hogy az első sokaság varianciája nagyobb, mint a másodiké.)

23. példa

Két független, normális eloszlásból vett mintánk van, az egyik 9 elemű, szórásnégyzete 14.4; a másik 6 elemű, szórásnégyzete 20.5. Megvizsgálandó 0.1-es szignifikanciaszinten, hogy a varianciák egyenlők-e vagy nem.

Megoldás

$H_0: \sigma_1^2 = \sigma_2^2 \qquad H_1: \sigma_1^2 \neq \sigma_2^2$

$F_0 = \dfrac{20.5}{14.4} = 1.424 \qquad F_{0.05}(5, 8) = 3.69$;

Elfogadási tartomány: $F_{0.95}(5, 8) < F_0 < F_{0.05}(5, 8)$.

Elfogadjuk a nullhipotézist, mivel $F_0 < F_{0.05}(5,8)$.[3] Az adatok nem mondanak ellent annak a feltételezésnek, hogy a két sokaság varianciája megegyezik.

24. példa

Egy automata gépen készülő alkatrészek jellemző méretére a következő adatokat mérték (zárójelben az előfordulások száma):

3.0 (2); 3.5 (6); 3.8 (9); 4.4 (7); 4.5 (1)

Az előírás szerint a gyártás bizonytalanságára jellemző variancia nem haladhatja meg a $\sigma^2 = 0.1$ értéket. Felmerült a gyanú, hogy a variancia meghaladja az előírt értéket. Bizonyítják-e az adatok ezt a gyanút 0.05-os szignifikanciaszinten?

Megoldás

$s^2 = 0.1975; \quad \nu = 25 - 1 = 24$;

$H_0: \sigma^2 \le \sigma_0^2 = 0.1 \qquad H_1: \sigma^2 > \sigma_0^2 = 0.1$

$\chi_0^2 = \dfrac{s^2 \nu}{\sigma_0^2} = \dfrac{0.1975 \cdot 24}{0.1} = 47.4; \quad \chi_0^2 = \dfrac{0.1975 \cdot 24}{\sigma^2} \cdot \dfrac{\sigma^2}{0.1} = \chi^2 \dfrac{\sigma^2}{0.1}$;

ha H_1 igaz, $\dfrac{\sigma^2}{0.1} > 1 \rightarrow \chi_0^2$ jobbra eltolódik, tehát egyoldali fölső határa van a nullhipotézis elfogadási tartományának: $\chi_{krit}^2 = \chi_{0.05}^2(24) = 36.415$.

Elfogadási tartomány: $\chi_0^2 < \chi_{krit}^2$.

Elutasítjuk a nullhipotézist, mivel $\chi_0^2 > \chi_{\mathrm{krit}}^2$, az adatok 0,05 szignifikanciaszinten vizsgálva ellentmondanak a nullhipotézisnek.
Az adatok 0.05-os szignifikanciaszinten alátámasztják azt a gyanút, hogy a variancia meghaladja a határértéket. Ez azt jelenti, hogy kisebb mint 5% annak az esélye, hogy ilyen (0.1975) vagy ennél még nagyobb szórásnégyzetű mérési eredményeket kapunk, ha $\sigma^2 = 0.1$.[4] Minél kisebb ez a valószínűség 5%-nál, annál inkább kételkedhetünk az kezdeti feltételezésünk (nullhipotézis) teljesülésében.

---

[3] Az alsó határnak való megfelelés biztos teljesül, mivel $F_{0.95}(5, 8)$ biztos kisebb mint 1 (az *F*-eloszlás természetéből adódóan) és a próbastatisztikát úgy írtuk fel, hogy az nagyobb legyen, mint 1. (Részletes magyarázat található az *F*-eloszlás használatáról a honlapról letölthető könyvkivonatban.)

[4] Szoftverrel számolva ennek a valószínűségnek (elsőfajú hiba elkövetésének valószínűsége) a pontos értéke is megadható (*p* érték).

Kétmintás t-próba

25. példa

Két független minta (x és y) adatai a következők (zárójelben az előfordulások száma):

x:	3.4 (2);	3.5 (3);	3.7 (4);	3.9 (1)
y:	3.2 (2);	3.4 (2);	3.6 (8);	

Vizsgáljuk meg 0.025-es szignifikanciaszinten, hogy a két minta mögött álló sokaság várható értéke egyenlő-e vagy nem.

Megoldás

$\bar{x} = 3.6$; $\bar{y} = 3.5$; $s_x^2 = 0.0267$; $s_y^2 = 0.02545$

Kétmintás t-próbát kell alkalmazni, de először F-próbát kell végezni a varianciák egyenlőségére!

F-próba

$H_0 : \sigma_x^2 = \sigma_y^2$ $H_1 : \sigma_x^2 \neq \sigma_y^2$

$F_0 = \frac{0.0267}{0.02545} = 1.049$; $F_{0.05}(9,11) = 2.90$

Elfogadási tartomány: $F_{0.95}(9,11) < F_0 < F_{0.05}(9,11)$. Mivel a próbastatisztika felírásából következően $F_0 > 1$, elég csak a felső határt ellenőrizni. (Ld. a 23. példa.)

$F_0 < F_{0.05}(9,11)$, tehát elfogadjuk a feltételezést, hogy a két minta mögött álló sokaság varianciája megegyezik.

Az egyesített szórásnégyzet:

$$s^2 = \frac{0.0267 \cdot 9 + 0.02545 \cdot 11}{9+11} = 0.026$$

Kétmintás t-próba

$H_0 : \mu_x = \mu_y$ $H_1 : \mu_x \neq \mu_y$

$$t_0 = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(n_x^{-1} + n_y^{-1}\right)}} = \frac{3.6 - 3.5}{\sqrt{0.026\left(\frac{1}{10} + \frac{1}{12}\right)}} = 1.448$$

$t_{krit} = t_{0.025/2}(20) = 2.423$.

Elfogadási tartomány: $-t_{\text{krit}} < t_0 < t_{\text{krit}}$.

Elfogadjuk a nullhipotézist, mivel $-t_{\text{krit}} < t_0 < t_{\text{krit}}$ teljesül. Az adatok 0.025-es szignifikanciaszinten vizsgálva nem mondanak ellent annak, hogy a két sokaság várhatóértéke megegyezik.

26. példa

Az MCM 300 g/m^2 négyzetmétertömegét 5 mintavételből határozták meg. Az 5 mérés számtani középértéke 307.3 g/m^2, szórása 0.8 g/m^2. A Grafopack 300 g/m^2 kartonpapír szállítmányból 6 mintát véve az átlag 303.4 g/m^2, a szórás 0.6 g/m^2.
Elfogadva, hogy az adatok normális eloszlású sokaságból származnak, 5%-os szignifikanciaszinten vizsgálja meg, hogy a két szállítmány négyzetméter tömegének várható értéke különbözik-e egymástól?

Megoldás

$\bar{x}_M = 307.3$; $\bar{x}_G = 303.4$; $s_M^2 = 0.6400$; $s_G^2 = 0.3600$

Kétmintás *t*-próbát kell alkalmazni, de először *F*-próbát kell végezni a varianciák egyenlőségére!

F-próba

$H_0 : \sigma_M^2 = \sigma_G^2 \qquad H_1 : \sigma_M^2 \neq \sigma_G^2$;

$F_0 = \frac{0.64}{0.36} = 1.78 < F_{0.05}(4,5) = 5.19$; →Elfogadjuk a feltételezést, miszerint a két minta mögött álló sokaság varianciája megegyezik.

Az egyesített szórásnégyzet:

$$s^2 = \frac{0.64 \cdot 4 + 0.36 \cdot 5}{5+4} = 0.484$$

Kétmintás t-próba

$H_0 : \mu_M = \mu_G \qquad H_1 : \mu_M \neq \mu_G$

$$t_0 = \frac{\bar{x}_M - \bar{x}_G}{\sqrt{s^2\left(n_M^{-1} + n_G^{-1}\right)}} = \frac{307.3 - 303.4}{\sqrt{0.484\left(\frac{1}{5} + \frac{1}{6}\right)}} = 9.252$$

$t_{krit} = t_{0.05/2}(9) = 2.262$.

Elfogadási tartomány: $-t_{\text{krit}} < t_0 < t_{\text{krit}}$.

Elutasítjuk a nullhipotézist, mivel a próbastatisztika értéke (9.252) kívül esik a $-t_{\text{krit}} ... t_{\text{krit}}$ elfogadási intervallumon. Az adatok alapján elutasítjuk azt a feltételezést, hogy a két szállítmány négyzetméter tömegének várható értéke megegyezik.

27. példa

Egy kémiai reakciót 10 alkalommal az A receptúra szerint, 10 alkalommal pedig a tovább-fejlesztett B receptúra szerint hajtottak végre. Az alábbi kitermelés adatokat kapták:

A recept	B recept
54.6	74.9
45.8	78.3
57.4	80.4
40.1	58.7
56.3	68.1
51.5	64.7
50.7	66.5
64.5	73.5
52.6	81.0
48.6	73.7

Az elvárások szerint a javított (B) recept szerinti reakcióvezetés minimálisan 15 egységgel nagyobb kitermelést eredményez! Alátámasztják ezt a mérési adatok, ha a vizsgálatot α=0.05 szignifikanciaszinten végezzük?

Megoldás

Kétmintás *t*-próbát kell alkalmazni, de először *F*-próbát kell végezni a varianciák egyenlőségére!

F-próba

$H_0: \sigma_A^2 = \sigma_B^2; \quad H_1: \sigma_A^2 \neq \sigma_B^2;$

$F_0 = \dfrac{53.38}{44.9} 1.189 < F_{krit} = F_{0.05}(9,9) = 3.18$ →Elfogadjuk a feltételezést, miszerint a két minta mögött álló sokaság varianciája megegyezik.

Az egyesített szórásnégyzet:

$$s^2 = \frac{53.38 \cdot 9 + 44.9 \cdot 9}{9+9} = 49.14$$

Kétmintás t-próba

$H_0: \quad \mu_B - \mu_A \leq (\mu_B - \mu_A)^0 = 15$

$H_1: \quad \mu_B - \mu_A > (\mu_B - \mu_A)^0 = 15$ (baloldali ellenhipotézis)

$$t_0 = \frac{\bar{x}_B - \bar{x}_A - (\mu_B - \mu_A)^0}{\sqrt{s^2\left(n_A^{-1} + n_B^{-1}\right)}} = \frac{71.98 - 52.21 - 15}{\sqrt{49.14(1/10 + 1/10)}} = 1.522, \quad t_{krit} = t_{0.05}(18) = 1.734$$

Elfogadási tartomány: $t_0 < 1.734$, elutasítási tartomány: $1.734 < t_0$, a nullhipotézist tehát elfogadjuk, mivel t_0 az elfogadási tartományba esik. Az adatok nem bizonyítják (α=0.05 szignifikanciaszinten), hogy a javított (B) recept szerinti reakcióvezetés 15 vagy több egységgel nagyobb kitermelést eredményez.

Páros t-próba

28. példa

8 különböző mintát elemeztek kétféle módszerrel. Az egyes mintákra kapott eredmények az alábbi táblázatban találhatók.

Minta sorszáma	1	2	3	4	5	6	7	8
x (*első módszer*)	15	20	16	22	24	14	18	20
y (*második módszer*)	15	22	14	25	29	16	20	24

Megvizsgálandó 0.05-os szignifikanciaszinten, hogy a két módszerrel kapott eredmények különböznek-e!

Megoldás

$d_i = y_i - x_i \qquad \bar{d} = 2 \quad s_d = 2.2039 \qquad \nu = 7$

$H_0: \quad E[d] = 0 \qquad H_1: \quad E[d] \neq 0$ (kétoldali ellenhipotézis)

$t_0 = \dfrac{\bar{d} - 0}{s_d / \sqrt{n}} = \dfrac{2}{2.2039/\sqrt{8}} = 2.567\,; \qquad t_{krit} = t_{0.025}(7) = 2.365\,;$

Elfogadási tartomány: $-t_{\mathrm{krit}} < t_0 < t_{\mathrm{krit}}$.

Elutasítjuk a nullhipotézist, mivel $t_0 > t_{\mathrm{krit}}$. Az eredmények ellentmondanak annak a feltételezésnek, hogy a két módszerrel kapott eredmények várhatóértékei megegyeznek. Vagy: az eredmények azt bizonyítják (0.05-os szignifikanciaszinten), hogy a két módszerrel végzett elemzés várhatóértéke különböző.

29. példa

A kartonpapír vastagságának mérését végző laboratóriumban két személy dolgozik (A és B). A beérkező kartonpapír szállítmányból vett minta vastagságát mindkét személy megméri. A 7 véletlenszerűen kiválasztott minta mért vastagságát (μm) az alábbi táblázat tartalmazza:

minta	**A**	**B**
1	395.4	396.5
2	364.8	366.2
3	272.0	274.3
4	255.7	256.0
5	385.6	387.6
6	312.5	314.5
7	328.7	333.3

Elfogadva, hogy az adatok normális eloszlású sokaságból származnak, 5%-os szignifikanciaszinten vizsgálja meg, hogy a két személy által mért adatok szignifikánsan különböznek-e egymástól?

Megoldás

$d_i = B_i - A_i \qquad \bar{d} = 1.96 \qquad s_d = 1.348 \qquad \nu = 6$

$t_0 = 3.847; \quad t_{krit} = t_{0.05/2}(6) = 2.447;$

Elfogadási tartomány: $-t_{\text{krit}} < t_0 < t_{\text{krit}}$.

Elutasítjuk a nullhipotézist, mivel a próbastatisztika értéke (3.842) kívül esik a $-t_{\text{krit}} ... t_{\text{krit}}$ elfogadási intervallumon. A két személy méréseinek várhatóértéke szignifikánsan különbözik.

30. példa

Két laboratórium (A és B) munkáját úgy hasonlították össze, hogy 4 csomag dohány nikotintartalmát mindkét laboratóriummal megmérették. Minden esetben a csomagból kivett mintát megfelezve, felét az A laborba, másik felét pedig a B laborba küldték vizsgálatra. Az eredményeket a következő táblázat tartalmazza (kódolt egységben).

Labor	Dohány			
	a	b	c	d
A	26	24	28	27
B	28	31	23	29

Megvizsgálandó α=0.05 szignifikanciaszinten, hogy a két labor torzít-e egymáshoz képest!

Megoldás

$H_0: \ E[d] = 0 \qquad H_1: \ E[d] \neq 0 \qquad$ (kétoldali ellenhipotézis)

$t_0 = 0.608 \qquad t_{krit} = t_{0.025}(3) = 3.182$

Elfogadási tartomány: $-t_{\text{krit}} < t_0 < t_{\text{krit}}$.

Mivel $-t_{\text{krit}} < t_0 < t_{\text{krit}}$ teljesül, ezért a nullhipotézist elfogadjuk. Az adatok nem bizonyítják, hogy a két labor torzít egymáshoz képest.

Diszkrét eloszlások és közelítésük normális eloszlással [5]

31. példa

Minimálisan hány elemű mintákat kell vennünk, ha azt akarjuk, hogy 99% valószínűséggel találjunk legalább 1 hibás darabot, vagyis $P(D>0) \geq 0.99$, amennyiben a sokaságban a selejtarány 3% (p=0.03)?

[5] Nem lesz a zárthelyiben.

Megoldás

A binomiális eloszlás összefüggéseivel számolunk:

$$P(D>0)=1-P(D=0)=1-\binom{n}{0}\cdot 0.03^0\cdot 0.97^n=1-0.97^n=0.99$$

$$n=\frac{\ln 0.01}{\ln 0.97}=151.2$$

Tehát a szükséges mintaelemszám 152.

32. példa

Tételezzük fel, hogy a gyártott papírdobozokon átlagosan 2 nyomdai hiba van. Hány doboz tartozzék egy mintába, hogy a mintában 95%-os valószínűséggel legalább 1 hibát találjunk?

Megoldás

A Poisson-eloszlás összefüggéseivel számolunk. Jelölje *r* az egy mintába tartozó dobozok számát.

$$P(k\geq 1)=1-P(k=0)=0.95$$

$$\lambda=2r\,,\qquad P(k=0)=\frac{\lambda^0 e^{-\lambda}}{0!}=e^{-2r}$$

$$1-e^{-2r}=0.95\,,\qquad \ln(1-0.95)=-2r\,,\qquad r=-\frac{\ln(0.05)}{2}=-\frac{-2{,}9957}{2}=1.498\,.$$

$$r=2 \text{ esetén } P(k\geq 1)=1-e^{-2\cdot 2}=1-0.0183=0.9817$$

33. példa

Egy gyártási folyamatból n_1=50 elemű mintát vettek, a mintában 13 selejtes darabot találtak. Egy későbbi időpontban szintén történt mintavétel, az n_2=40 elemű mintában a selejtes darabok száma 4 volt. Eldöntendő, hogy megváltozott-e a folyamat selejtaránya, vagy csak a véletlen ingadozás okozza a különbséget.

Megoldás

$H_0 : p_1 = p_2;\qquad H_1 : p_1 \neq p_2$

A két mintában a tapasztalati selejtarány és az egyesített becslés:

$$\hat{p}_1=\frac{13}{50}=0.26\,;\qquad \hat{p}_2=\frac{4}{40}=0.10\,;\qquad \hat{p}=\frac{13+4}{50+40}=0.189\,.$$

$$z_0=\frac{0.26-0.10}{\sqrt{0.189(1-0.189)\left(\frac{1}{50}+\frac{1}{40}\right)}}=1.927$$

Az elfogadási tartomány 0.05 kétoldali szignifikanciaszinten (-1.96, 1.96), a próbastatisztika kiszámított értéke ezen belül van, bár közel az elutasítási határhoz, tehát elfogadjuk a nullhipotézist. Az eltérő selejtarány a véletlen ingadozásnak tulajdonítható.

34. példa

Egy gyártási folyamatból 10 alkalommal vettek 100 – 100 elemű mintát. A 10 mintából számított 10 selejtarány számtani középértéke (az átlagos selejtarány) 0.10, a selejtarány korrigált tapasztalati szórásnégyzete 0.12. A gyártás későbbi szakaszában 8 alkalommal vettek mintát, a minták elemszáma minden esetben 100 volt. A 8 mintából számított átlagos selejtarány 0.08, a selejtarány korrigált tapasztalati szórásnégyzete 0.085. Döntsük el 0.05-os szinten, hogy a selejtarány a két szakaszban azonos-e!

Megoldás

$$H_0 : E(\bar{p}_I) = E(\bar{p}_{II}); \; H_1 : E(\bar{p}_I) \neq E(\bar{p}_{II})$$

A két szakasz több mintából áll, a több mintából kiszámolható átlagos selejtarány a centrális határeloszlási tétel értelmében akkor is jó közelítéssel normális eloszlású, ha az egy minta selejtaránya még nem lenne elég közel a normális eloszláshoz (p kicsi vagy nagy, n nem elég nagy). Ráadásul itt a varianciát sem kell az ismeretlen p paraméter mintabeli becsléséből számolnunk, hanem több ismétlés lévén, tapasztalati szórásnégyzetet használhatunk. Ekkor viszont nem u-, hanem t-próbát végzünk.

$$t_0 = \frac{\bar{p}_I - \bar{p}_{II}}{\sqrt{\frac{s_1^2}{n_I} + \frac{s_2^2}{n_{II}}}} = \frac{0.1 - 0.08}{\sqrt{\frac{0.12}{10} + \frac{0.085}{8}}} = 0.133$$

A $10 + 8 - 2 = 16$ szabadsági fokszámhoz és α=0.05 szignifikanciaszinthez tartozó kritikus érték 2.12. Mivel a próbastatisztika talált értéke ez alatt van, elfogadjuk a nullhipotézist, mely szerint a két szakaszban a selejtarány csak a véletlen ingadozás miatt különbözik.

Vegyes feladatok az eloszlások és próbák témakörből

A fejezetben szereplő példák megoldása a fejezet végén található. A *-gal jelölt feladatok kicsit összetettebbek a többinél. Ehhez hasonló számolásra zárthelyin nem kell számítani, de megértésük segíti az anyag megértését.

35. példa

1. Egy profi futócipő súlya 340g körül ingadozik 200 g^2 varianciával.
(a) Mi a valószínűsége annak, hogy egy cipő nehezebb, mint 370g?
(b) Mekkora lehet a megengedett variancia a gyártás során, ha azt kívánjuk elérni, hogy a cipők 99.9%-a könnyebb legyen 370g-nál?
(c) Ha a variancia 200 g^2 marad, mekkora legyen a várhatóérték, ha azt kívánjuk elérni, hogy a cipők 99.9%-a könnyebb legyen 370g-nál?

36. példa

Egy printerrel nyomtatott festékpont átmérője normális eloszlást követ 0.05 mm várhatóértékkel és 10^{-4} mm^2 varianciával.
(a) Mi a valószínűsége annak, hogy egy pont átmérője nagyobb lesz, mint 0.065 mm?
(b) Mi a valószínűsége annak, hogy egy pont átmérője 0.04 és 0.065 mm közé esik?
(c) Milyen intervallumban lesz a pontok 99%-ának átmérője?

37. példa

Egy kémiai reakcióban a kitermelés normális eloszlást követ. Előzetes adatokból ismert, hogy az ingadozás szigmája 3. Az elmúlt öt nap során a gyártás az alábbi kitermelésekkel ment: 91.6, 88.75, 90.8, 89.95 és 91.3. Adjon 95%-os konfidencia intervallumot a kitermelés várhatóértékére.

38. példa

Autó motorjába használt tömítőgyűrűket gyártanak. A tömítőgyűrűk átmérője normális eloszlást követ, a szigma 0.001 mm. 15 tömítőgyűrűt megvizsgálva azt találták, hogy az átlagos átmérő 74.036 mm.

(a) Adjon 99%-os kétoldali konfidencia intervallumot a tömítőgyűrű átmérőjére!

(b) Adjon 99%-os alsó határt a tömítőgyűrű átmérőjére! Hasonlítsa össze a számított értéket az (a) feladatrészben kapott alsó határral.

39. példa

A TV képcsöveket úgy ellenőrzik, hogy mérik a rajtuk átfolyó áramerősséget. 10 képcső lemérésével kapott átlagos áramerősség 317.2 μA, a szórás 15.7 μA.

(a) Adjon 99%-os konfidencia intervallumot az áramerősségre. Kell-e bármilyen feltételezésnek teljesülnie a számoláshoz az áramerősség értékek eloszlásával kapcsolatban?

(b) Igazolják-e az adtok 5%-os szignifikanciaszinten, hogy az áramerősség várhatóértéke meghaladja a 300 μA-t?

40. példa

Őszibarack konzervekben használt szirup cukortartalma normális eloszlást követ. 10 konzerv cukortartalmát megmérve az találták, hogy az átlagos cukortartalom 32.4 g, a szórás 1.8 g.

(a) Adjon 90%-os felső határt a cukortartalom várhatóértékére.

(b) Adjon 90%-os felső határt a cukortartalom varianciájára.

41. példa

Egy bizonyos elemtípus élettartama közelítőleg normális eloszlást követ, az ingadozás szigmája 1.25 óra. 10 elem élettartamát megmérve az átlagos élettartam 40.5 órának adódott.

(a) Ezek a mérési eredmények bizonyítják-e azt az állítást, hogy az elemtípus várható élettartama meghaladja a 40 órát? A kérdés 5%-os szignifikanciaszinten vizsgálandó.

(b)* Mekkora a statisztikai próbához tartozó p-érték?

(c)* Mekkora a másodfajú hiba valószínűsége az (a) feladatrészben, ha az élettartam várhatóértéke 42 óra?

(d)* Magyarázza meg, hogyan lehetett volna válaszolni az (a) feladatrész kérdésére a megfelelő konfidencia intervallum kiszámolása alapján!

42. példa

A *Medicine and Science in Sports and Exercise* cikke egy olyan kísérletről számol be, amelyben elektrostimuláció hatását vizsgálták jégkorong játékosok teljesítményére. Megmérték, hogy a kísérletben résztvevő 17 játékos mennyi idő alatt korcsolyázik le 10 métert. Az eredmények szórása 0.09 s volt. Előzetes tanulmányokból ismert, hogy a 10 m megtételéhez szükséges idő szigmája elektrostimuláció nélkül 0.75 s. A kísérlet eredményei alapján állíthatjuk-e azt 5%-os szignifikanciaszinten, hogy az elektrostimuláció hatására csökken a játékosok teljesítményének ingadozása?

43. példa

Egy szakaszos reaktorban kétféle katalizátorral végzik el ugyanazt a reakciót. 12 sarzsot gyártanak az A katalizátorral, az átlagos kitermelés 86%, a szórás 3%. 15 sarzsot gyártanak a B katalizátorral, az átlagos kitermelés 89% a szórás 2%. Bizonyítják-e az adatok azt a feltételezést, hogy a B katalizátorral nagyobb a kitermelés? 5%-os szignifikanciaszinten végezzük el az elemzést.

44. példa

Klinikai kísérletek során egypetéjű ikerpárok intelligenciahányadosát vizsgálták. (*Brain size, head size, and intelligence quotient in monozygotic twins, Neurology, 1998, Vol. 50, pp. 1246–1252.*) Az alábbi táblázat tartalmazza a kísérleti eredményeket:

Ikerpár sorszáma	**Születési sorrend: 1**	**Születési sorrend: 2**
1	6.08	5.73
2	6.22	5.8
3	7.99	8.42
4	7.44	6.84
5	6.48	6.43
6	7.99	8.76
7	6.32	6.32
8	7.6	7.62
9	6.03	6.59
10	7.52	7.67

Az adatok alapján 1%-os szignifikanciaszinten mondhatjuk-e azt, hogy az intelligencia függ a születési sorrendtől? Milyen feltételnek kell teljesülnie a választott statisztikai módszer alkalmazhatóságához?

Megoldások

35. példa
(a) 0.0169
(b) 94.28 g^2
(c) 326.4 g

36. példa
(a) 0.0668
(b) 0.775
(c) 0.0242…0.0758

37. példa
87.85, 93.11

38. példa
(a) 74.0353 … 74.0367
(b) 74.035

39. példa
(a) 301.06, 333.34
(b) H_0: $\mu \leq \mu_0 = 300$; $t_0 = 3.46 > 1.833$, elutasítjuk H_0-t.

40. példa
(a) 33.187
(b) 7.0

41. példa
(a) H_0: $\mu \leq \mu_0 = 40$; $z_0 = 1.26 < 1.65$, elfogadjuk H_0-t, nem látjuk bizonyítva, hogy az élettartam nagyobb mint 40 óra
(b)* 0.1038
(c) 0.000325
(d)* $39.85 \leq \mu$

42. példa
H_0: $\sigma \geq 0.75$, $\chi_0^2 = 0.23 < 7.962$ Elutasítjuk H_0-t, az adatok 5%-os szignifikanciaszinten bizonyítják, hogy csökken az ingadozás mértéke.

43. példa
$F_0 = 2.25 < 2.57 = F(11,14)$, a varianciák egyesíthetők; H_0: $\mu_B - \mu_A \leq 0$, $t_0 = 3.11 > 1.708$, elutasítjuk a nullhipotézist, az adatok bizonyítják, hogy B katalizátorral nagyobb a kitermelés.

44. példa
$t_0 = -0.366$; $-3.25 < -0.366 < 3.25$, elfogadjuk a nullhipotézist, az adatok nem bizonyítják, hogy az intelligenciahányados függ a születés sorrendjétől.