Statisztikai módszerek a várható érték, illetve a szórás becslésére korlátos mintából

Ha ismert a szórás, és $\bar{x}$ egy n elemű minta átlaga (n elég nagy, vagy a valószínűségi változó normális):

$$P\Big(\Big|\frac{\bar{x}-m}{\sigma/\sqrt{n}}\Big| < u\Big) \approx 2\Phi(u) - 1. \tag{1}$$

Felhasználása:

a) m értékének megbecslése $\bar{x}$, n és σ ismeretében;
b) Az $m = m_0$ feltevés hihetőségének eldöntése
c) n meghatározása ahhoz, hogy az átlag elég jól közelítse a várható értéket.

Ha nem ismert a szórás sem, csak a mintából kapott korrigált szórás:

$$P\Big(\Big|\frac{\bar{x}-m}{s^*/\sqrt{n}}\Big| < t\Big) = p, \tag{2}$$

ahol a p-hez tartozó t értéket a t(n-1)-eloszlás táblázatából tudjuk leolvasni (a t-eloszlás táblázatának $n-1$ indexű sora). A Közlekedésmérnöki példatárban a p-hez, a Ferenczi-féle példatárban az $1-p$-hez tartozó t értéket találjuk, az angol nyelvű lapon, pedig az $\frac{1-p}{2}$ értékhez tartozót.

Ha két különböző minta van, ismert elméleti szórással:

$$P\Big(\Big|\frac{\bar{x}_1 - \bar{x}_2 - (m_1 - m_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\Big| < u\Big) \approx 2\Phi(u) - 1. \tag{3}$$

Felhasználása:

annak eldöntése, hogy az elméleti várható értékek hihetően azonosak-e.

Ha két különböző minta van, de csak a tapasztalati szórásokat és a mintaátlagot ismerjük (és feltesszük, hogy az elméleti szórások azonosak):

$$P\left(\left|\frac{\bar{x}_1 - \bar{x}_2 - (m_1 - m_2)}{\sqrt{n_1 s_1^2 + n_2 s_2^2}}\right| \cdot \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} < t\right) = p, \tag{4}$$

ahol a p-hez rendelt t értéket a $t(n_1 + n_2 - 2)$-eloszlás táblázatából tudjuk leolvasni.

Ha csak a tapasztalati szórásokat ismerjük, és az elméleti szórások arányáról (egyenlőségéről) szeretnénk információt szerezni:

$$P\Big(\frac{1}{F} \le \frac{s_1^{*2}/\sigma_1^2}{s_2^{*2}/\sigma_2^2} \le F\Big) = p \qquad (5),$$

és itt a p-hez tartozó F értéket az F(n_1-1, n_2-1)-eloszlás táblázatából olvashatjuk le (mindkét példatárban csak $p = 0,95$-re van táblázat).

Az $x_1, \ldots, x_n$ mintára $\bar{x} = (x_1 + \ldots + x_n)/n$, $s^2 = (\sum x_i^2 - n\bar{x}^2)/n$, $s^{*2} = (\sum x_i^2 - n\bar{x}^2)/(n-1) = s^2 \cdot n/(n-1)$. A számológépünkön az s valószínűleg valójában s^*-ot adja. Ellenőrizzük!

Házi feladatok

1. (FXIV/5) Egy sorozatban gyártott terméknél 1000 db-nak ellenőrzik a valódi súlyát, és átlagnak 5100 dkg-ot kapnak. Tudjuk, hogy a szórás 10 dkg. Adjon meg intervallumot, amelybe a termék súlyának várható értéke 99%-os biztonsággal beleesik.

2. (FXIV/10) 15 izzónak megmérték az élettartamát, amely közelítőleg normális eloszlásúnak tekinthető. Az átlagra 1200 óra, az s^* empirikus szórásra 186 óra adódott. Adjon becslést $0,99$ biztonsággal az izzók élettartamának várható értékére!

3. (FXIV/11) Egy alkatrész átmérőjének várható értékére elfogadható-e az $m = 2$ hipotézis $0,9$ biztonsági szinten, ha 10 alkatrészt megvizsgálva a következő átmérőértékeket kapjuk: $1,99$; 2; $1,98$; $1,95$; $1,95$; $2,02$; $2,05$; $2,04$; $2,03$; $2,01$. Az átmérő eloszlására normális eloszlást feltételezhetünk.

4. (FXIV/12) Kétféle technológiával gyártanak egy terméket. 6–6 mintát megvizsgálnak a termékből és rendre a következő térfogatátlagokat, ill. szórásokat kapjuk: $\bar{x}_1 = 297$, $s_1^2 = 38,4$, $\bar{x}_2 = 310$, $s_2^2 = 46$. Van-e lényeges eltérés $0,99$ biztonsági szinten a két technológia között, ha normális eloszlásokat és egyenlő elméleti szórásokat feltételezünk?

5. (FXIV/13) Két technológiával gyártanak le egy bizonyos típusú terméket. 12 elemű minta alapján az első technológia esetén $s_1^{*2} = 31\,\text{mm}$, ill. a második technológiánál 15 elemű minta alapján $s_2^{*2} = 17\,\text{mm}$. Tehetünk-e ennek alapján lényeges különbséget $0,95$ biztonsági szinten a kétféle technológia között, ha a méretre normális eloszlást feltételezünk?

6. (FXIV/26) Egy valószínűségi változó sűrűségfüggvénye $f(x) = 2x$ $(0 \le x \le 1)$. Becsülje meg annak a valószínűségét, hogy 50 kísérleti eredmény átlaga
a) $\frac{1}{2}$ és $\frac{3}{4}$ közé esik!
b) az eloszlás várható értékének $0,1$ sugarú környezetébe esik!

7. (FXIV/29) Két üzemből érkezik ugyanazon típusú termékre 50–50 minta. Az egyik ötvennél a súlyukra $1,51\,\text{kg}$ átlagot mértek, $0,5\,\text{kg}$ szórással, a másik ötvennél $1,58\,\text{kg}$ átlagot, $0,35\,\text{kg}$ szórással. Van-e különbség $0,95$ biztonsági szinten a két üzemből érkező termékek súlya között?

Eredmények

1. Az (1) formulából: $p = 0,99 \Rightarrow u = 2,575$, $n = 1000$, $\bar{x} = 5100$, $\sigma = 10$, így $5099,19 = \bar{x} - u\sigma/\sqrt{n} \le m \le \bar{x} + u\sigma/\sqrt{n} = 5100,81$, azaz a várható érték $0,99$ valószínűséggel a $(5099,18; 5100,82)$ intervallumba esik.

2. A (2) formulából: $n = 15$, $\bar{x} = 1200$, $s^* = 186$, $p = 0,99 \Rightarrow t = 2,977$, és ebből $1057,03 \le m \le 1342,97$.

3. A (2) formulából: $\bar{x} = 2,002$, $s^* = 0,0349$, $p = 0,9 \Rightarrow t = 1,833$, és így a (2) formula szerint $|m - \bar{x}| < (1/\sqrt{10}) \cdot 0,0349 \cdot 1,833$, azaz $m = 2$ esetén az $1,980 < \bar{x} < 2,020$ egyenlőtlenségnek kell teljesülnie, s mivel $\bar{x} = 2,002$, ez teljesül is, vagyis az $m = 2$ feltevés elfogadható.

4. A (4) formulából: $n_1 = n_2 = 6$, $\bar{x}_1 = 297$, $\bar{x}_2 = 310$, $s_1^2 = 38,4$, $s_2^2 = 46$, $p = 0,99 \Rightarrow t == 3,169$. Ha $m_1 = m_2$, akkor a (4) formula szerint $\left|\frac{297-310}{\sqrt{6\cdot 38,4+6\cdot 46}}\right| \cdot \sqrt{30} < 3,169$, és ez valóban igaz, mert a bal oldal $\approx 3,164$. Tehát $0,99$ biztonsági szinten nincs különbség a két technológia között.

5. Az (5) formulából: $p = 0,95$, $n_1 = 12$, $n_2 = 15 \Rightarrow F = 2,57$, és $31/17$ $1/F$ és F között van, tehát 95%-os biztonsági szinten megegyeznek a szórások; nincs lényeges különbség a két technológia között.

6. Az adott sűrűségfüggvényből kiszámíthatjuk az elméleti várható értéket és szórást: $m = \frac{2}{3}$, $\sigma = \frac{1}{3\sqrt{2}}$.

a) Így felhasználva, hogy az azonos eloszlású független valószínűségi változók átlaga közelítően normális, $P(\frac{1}{2} < \bar{x} < \frac{3}{4}) \approx \Phi(\frac{1/12}{1/30}) - \Phi(\frac{-1/6}{1/30}) = \Phi(2,5) - \Phi(-5) \approx 0,9938 - 0 = 0,9938$.

b) $P(|\bar{x} - m| < 0,1) = 2\Phi(3) - 1 = 0,9974$.

7. A (4) formulából: $n_1 = n_2 = 50$, $\bar{x}_1 = 1,51$, $\bar{x}_2 = 1,58$, $s_1 = 0,5$, $s_2 = 0,35$, $p = 0,95 \Rightarrow t = 1,99$. $m_1 = m_2$ esetén

$$\left|\frac{-0,07}{4,316}\right| \cdot 49,497 \approx 0,803 < 1,99,$$

és ez valóban igaz, tehát $0,95$ biztonsági szinten nincs különbség.