Matematika A4 - Valószínűségszámítás 13. gyakorlat
Hipotézisvizsgálat, gyakorlás

Hipotézisvizsgálat

Legyen $X \sim N\left(\mu, \sigma^2\right)$, ahol μ ismeretlen, σ ismert és μ_0 egy adott érték. Az ismeretlen várható értékről szeretnénk eldönteni különböző feltevéseket (**nullhipotézis**) valamilyen α **szignifikancia-szint** mellett. A legkisebb α értéket, ami mellett már el tudjuk utasítani a nullhipotézist, **p-érték**nek nevezzük.

Kétoldalas A várható érték megegyezik-e az adott μ_0-val?
Nullhipotézis: $H_0 : \mu = \mu_0$.
Alternatív hipotézis: $H_1 : \mu \neq \mu_0$.
Elfogadjuk H_0-t, ha μ_0 benne van a **kétoldalú Z-próbá**nál látott konfidenciaintervallumban. Ez ekvivalens azzal, hogy $Z = \frac{\sqrt{n}\cdot(\overline{\mathcal{X}_n}-\mu_0)}{\sigma}$ esetén $|Z| \leq z_{\alpha/2}$. A p-érték az az α érték, amire $|Z| = z_{\alpha/2}$.
Ha ez nem teljesül, akkor elvetjük H_0-t.
Ismeretlen σ esetén a T-próbánál látott konfidenciaintervallumot kell használni. Vagyis, ha $T = \frac{\sqrt{n}\cdot(\overline{\mathcal{X}_n}-\mu_0)}{\hat{\sigma}}$, akkor elfogadjuk a nullhipotézist, ha $|T| \leq t_{\alpha/2,n-1}$. A p-érték az az α érték, amire $|T| = t_{\alpha/2,n-1}$.

Egyoldalas A várható érték kisebb (nagyobb), mint az adott μ_0?
Nullhipotézis: $H_0 : \mu \leq \mu_0$ (vagy $\mu \geq \mu_0$).
Alternatív hipotézis: $H_1 : \mu > \mu_0$ (vagy $\mu < \mu_0$).
Ha $H_0 : \mu \leq \mu_0$, akkor elfogadjuk H_0-t, ha $\overline{\mathcal{X}_n} < \mu_0 + \frac{z_\alpha \sigma}{\sqrt{n}}$. Ha $U = \frac{\sqrt{n}\cdot(\overline{\mathcal{X}_n}-\mu_0)}{\sigma}$, akkor ez ekvivalens azzal, hogy $U \leq z_\alpha$. A p-érték az az α, amire egyenlőség van.
$H_0 : \mu \geq \mu_0$ esetén elfogadáshoz az kell, hogy $\overline{\mathcal{X}_n} > \mu_0 - \frac{z_\alpha \sigma}{\sqrt{n}}$. Vagy ezzel ekvivalensen H_0 elfogadásához az kell, hogy $U \leq z_\alpha$, ahol most $U = \frac{\sqrt{n}\cdot(\mu_0-\overline{\mathcal{X}_n})}{\sigma}$.

1. feladat Vásárlói panaszok érkeznek, hogy egy élelmiszerboltban az 1 kg-os feliratú kenyér súlya kevesebb. Szeretnénk korrekt módon kivizsgálni az ügyet. Kiszállunk az üzletbe, megmérünk n véletlenszerűen kiválasztott kenyeret, $X_1, \ldots, X_n$ a minta. Legyen $n = 25$ és a realizációban azt találjuk, hogy az átlaguk 0.98 kg. Tudjuk, hogy a kenyerek tömegének szórása $\sigma = 0.05$.

(a) Döntsük el 0.05 és 0.01 szignifikanciával, hogy csalnak-e a boltban, vagyis, hogy a kenyér tömegének várható értéke eléri-e az 1 kg-ot! Mennyi a p-érték?

(b) Mit mondhatunk, ha azt szeretnénk eldönteni, hogy pontosan 1 kg a várható érték?

(c) És, ha szintén azt szeretnénk eldönteni, hogy pontosan 1 kg várható érték, de a szórást most nem ismerjük, csak a mintából tudjuk becsülni, aminek torzítatlan szórása 0.06?

Megoldásvázlat

(a) $\mu_0 = 1$ és $H_0 : \mu \geq \mu_0$ ($H_1 : \mu < \mu_0$).
$\alpha = 0.05$ esetén $\mu_0 - \frac{z_\alpha \sigma}{\sqrt{n}} = 1 - \frac{1.64\cdot 0.05}{5} = 0.984 > 0.98 = \overline{\mathcal{X}_{25}}$, így elvetjük a nullhipotézist.

$\alpha = 0.01$ esetén $\mu_0 - \frac{z_\alpha \sigma}{\sqrt{n}} = 1 - \frac{2.33\cdot 0.05}{5} = 0.977 < 0.98 = \overline{\mathcal{X}_{25}}$, így elfogadjuk a nullhipotézist.

$U = \frac{\sqrt{n}\cdot(\mu_0-\overline{\mathcal{X}_n})}{\sigma} = \frac{5\cdot(1-0.98)}{0.05} = 2$. Mivel $z_{0.01} > U > z_{0.05}$, amiből szintén az következik, hogy 99%-os szignifikanciával elfogadjuk H_0t, de 95%-ossal elutasíthatjuk.

$\Phi(U) = \Phi(2) = 0.9772 = 1 - \alpha$, ha $\alpha = 0.0228$, tehát a p-érték 0.0228. Így ennél kisebb α esetén elfogadjuk H_0-t, ennél nagyobb α esetén elutasítjuk.

(b) Most $H_0 : \mu = \mu_0$ ($H_1 : \mu \neq \mu_0$).
A megfelelő konfidenciaintervallumok:
$\alpha = 0.05$: $\left(0.98 - \frac{1.96\cdot 0.05}{5}, 0.98 + \frac{1.96\cdot 0.05}{5}\right) = (0.9604, 0.9996)$
$\alpha = 0.01$: $\left(0.98 - \frac{2.57\cdot 0.05}{5}, 0.98 + \frac{2.57\cdot 0.05}{5}\right) = (0.9543, 1.0057)$

Jól látszik, hogy az első esetben $\mu_0 = 1$ nincs benne az intervallumban, míg második esetben benne van, így $\alpha = 0.05$ esetén elutasítjuk, $\alpha = 0.01$ esetén elfogadjuk H_0-t.

$Z = \frac{\sqrt{n}\cdot(\overline{X}_n - \mu_0)}{\sigma} = \frac{5\cdot(0.98-1)}{0.05} = -2$. Ekkor $z_{0.005} > |Z| > z_{0.025}$, amiből ugyanazt a következtetést vonhatjuk le.

$|Z| = z_{\alpha/2}$, ha $\alpha = 0.0456$, így most a p-érték 0.0456. Így ennél kisebb α esetén elfogadjuk H_0-t, ennél nagyobb α esetén elutasítjuk.

(c) Továbbra is $H_0 : \mu = \mu_0$ ($H_1 : \mu \neq \mu_0$).

A megfelelő konfidenciaintervallumok:

$\alpha = 0.05$: $\left(0.98 - \frac{2.064\cdot 0.06}{5}, 0.98 + \frac{2.064\cdot 0.06}{5}\right) = (0.955, 1.005)$

$\alpha = 0.01$: $\left(0.98 - \frac{2.797\cdot 0.06}{5}, 0.98 + \frac{2.797\cdot 0.06}{5}\right) = (0.946, 1.014)$

Jól látszik, hogy a mindkét esetben $\mu_0 = 1$ benne van az intervallumban, így elfogadjuk H_0-t.

$T = \frac{\sqrt{n}\cdot(\overline{X}_n - \mu_0)}{\hat{\sigma}} = \frac{5\cdot(0.98-1)}{0.06} = -1.67$. Ekkor $t_{0.005,24} > t_{0.025,24} > |T|$, amiből ugyanazt a következtetést vonhatjuk le.

$|T| = t_{\alpha/2,24}$, ha $\alpha = 0.108$, így most a p-érték 0.108. Így ennél kisebb α esetén elfogadjuk H_0-t, ennél nagyobb α esetén elutasítjuk.

2022. január 12. vizsga

1) Használt alkatrészeket különböző dobozokban tárolnak. Az első dobozban 50 alkatrész van, 10%-a selejtes (a többi még használható), a második dobozban 100 alkatrész van, 20%-a selejtes, a harmadikban pedig 1000 és ennek 30%-a selejtes.

a) Az első dobozból húzok ötször visszatevés nélkül, mi a valószínűsége, hogy lesz benne selejtes? (3p)

b) A második dobozból húzok tízszer visszatevéssel, mi a valószínűsége, hogy a kihúzottak között lesz legalább 2 selejtes? (3p)

c) A harmadikból húzok 300-szor visszatevéssel, adj közelítést annak a valószínűségére, hogy pontosan 60 selejtes lesz közöttük! (3p)

d) Összeöntjük a három doboz tartalmát. Mi a valószínűsége, hogy ha egy alkatrészt választunk, az selejtes lesz? Ha szintén összeöntés után választottunk egy selejtest, az az első dobozból való volt? (3p)

2) Béla és Géza különböző buszokkal (tehát egymástól függetlenül) érkeznek megbeszélt találkozójukra 13 és 14 óra között, egyenletes eloszlás szerint.

a) Max 20 percet hajlandóak egymásra várni, mi a valószínűsége, hogy találkoznak? (5p)

b) Mi lesz két Uni(13, 14) valváltozó összegének sűrűségfüggvénye? (6p)

c) Ha Béla odaért 13:20-ra akkor mi a valószínűsége, hogy Géza 13:50 előtt érkezik meg? (3p)

d) Mi lesz a később érkező időpontjának sűrűségfüggvénye? (3p)

3) Spike és Jet fejvadászok, a sikeres megbízások (bounty-k) száma, amiket havonta együtt teljesítenek Poisson(5) eloszlásúak.

a) Mi a valószínűsége, hogy két bounty között több, mint egy hét telik el? (3p)

b) Ha az első héten begyűjtöttek 3-at, mi a valószínűsége, hogy a 2. hét végéig legalább 2-t fognak? (3p)

c) a 4. bounty-t a 2. hét végéig gyűjtik be? (3p)

4) Legyen X, Y együttes sűrűségfüggvénye $f(x, y) = e^{x+y}$ a $0 < x < a$, $0 < y < a$ négyzeten (mindenhol máshol nulla).

a) Mennyi a? (4p)

b) Független-e X és Y? (Indokolj!) (3p)

c) Mennyi $\mathbb{P}(Y > X)$? (2p)

d) Mennyi $\mathbb{E}(X)$ és $\mathbb{E}(X|Y)$? (3p)

e) Mennyi $\mathrm{CORR}(X, Y)$? (1p)

5) Egy cinkelt érmét dobálunk, a fej valószínűsége p. A következő adatsorunk van az első fej megjelenésére: (2; 3; 1; 2; 3; 3; 2; 1; 4; 4; 4; 2; 3; 2; 1; 4).

a) Adj Maximum Likelihood becslést a p-re! (4p)

b) Valaki megmondja, hogy $p = 0,4$. Mi a valószínűsége, hogy a 3. fej a 8. dobásnál lesz? (3p)

c) Legyen $p = 0,4$ továbbra is. Ha az első 5 dobásnál nem volt fej, mi a valószínűsége, hogy a 7.-nél lesz először? (2p)