Aalto-yliopisto

MS-A0502 Todennäköisyyslaskennan ja tilastotieteen peruskurssi — Gripenberg, Kiiski
Harjoitus 5
24–28.11.2014, viikko 48

P1. Alla olevassa taulukossa on annettu joidenkin mittausasemien vedenkorkeuden kuukausikeskiarvot marraskuussa 1996 ja 2000:

	1	2	3	4	5	6	7	8	9	10
1996	78	77	159	110	192	157	95	93	143	42
2000	85	89	161	97	205	185	123	80	161	78

Onko näissä arvoissa merkitsevää eroa? Testaa merkitsevyystasolla 5%.
Vihje: Voit olettaa että mittausarvot jokaisella mittausasemalla ja joka kuukausi ovat normaalijakautuneita samalla varianssilla (tosin aika kyseenalainen oletus) mutta et voi olettaa, että odotusarvot eri mittausasemilla olisivat samat josta seuraa, ettei ole järkevää verrata vuosien 1996 ja 2000 eri asemien keskiarvojen keskiarvoja toisiinsa. Sen sijaan voi olla hyvä idea ensin laskea näiden vuosien mittaustulosten erotus joka asemalla. Mikä on nollahypoteesi?

Ratkaisu: Jos lasketaan mittaustulosten erotukset saadaan seuraava taulukko:

	1	2	3	4	5	6	7	8	9	10
”2000”-”1996”	7	12	2	-13	13	28	28	-13	18	36

Nollahypoteesiksi valitaan nyt, että vuosien 2000 ja 1996 mittausten erotukset ovat riippumattomia ja $\mathrm{N}(0,\sigma^2)$-jakautuneita koska ei ole annettu mitään tietoa, jonka perusteella voisi odottaa, että toinen marraskuu olisi ollut sateisempi kuin toinen.

Näiden erotusten keskiarvo om $\overline{x} = 11.8$ ja otosvarianssi 277.73.

Koska oletetaan että erotukset ovat normaalijakautuneita ja nollahypoteesin mukaan odotusarvo on 0 niin testimuuttuja on

$$W = \frac{\overline{X} - 0}{\sqrt{\frac{S^2}{n}}} \sim \mathrm{t}(10-1).$$

Testimuuttujan arvoksi tulee $w = \dfrac{11.8}{\sqrt{\frac{277.73}{10}}} = 2.2391$. Koska vaihtoehto nollahypoteesille on kaksipuolinen niin $p = \Pr(W \geq 2.2391 \text{ tai } W \leq -2.2391) = 2(1 - F_{\mathrm{t}(9)}(2.2391)) = 2 \cdot 0.026 = 0.052 > 0.05$ eikä nollahypoteesia hylätä merkitsevyystasolla 0.05.

P2. $\mathrm{N}(\mu,\sigma^2)$-jakautuneesta satunnaismuuttujasta on saatu otos $x_1, x_2, \ldots, x_{15}$. Nyt menetellään seuraavalla (väärällä) tavalla: Lasketaan havaintoarvoista aritmeettinen keskiarvo $\overline{x}$, jos $\overline{x} > 0$ niin nollahypoteesiksi valitaan $\mathrm{H}_0 : \mu \leq 0$ ja muissa tapauksissa valitaan nollahypoteesiksi $\mathrm{H}_0 : \mu \geq 0$, (ja virhe tehdään siinä, että nollahypoteesi riippuu havaintoarvoista). Sitten testataan nollahypoteesin paikkansapitävyyttä normaaliin tapaan merkitsevyystasolla 0.01 ja lasketaan otsosvarianssi s^2 ja testimuuttujan arvo $w = \frac{\overline{x}-0}{\sqrt{\frac{s^2}{n}}}$. Mikä on todennäköisyys, että nollahypoteesi hylätään jos todella pätee $\mu = 0$? Mikä on tämä todennäköisyys jos otoksen koko on n ja merkitsevyystaso on α?

Vihje: Määritä luku A siten, että nollahypoteesi hylätään jos $w > A$ (eli (A, ∞) sisältyy hylkäysalueeseen) tai $w < -A$ (eli $(-\infty, A)$ sisältyy hylkäysalueeseen) ja ota huomioon mikä nollahypoteesi on jos $w > 0$ jolloin myös $\overline{x} > 0$ ja vastaavasti mikä se on kun $w < 0$ jolloin $\overline{x} < 0$. Laske sitten todennäköisyys, että testimuuttuja on itseisarvoltaan suurempi kuin A jos $\mu = 0$.

Ratkaisu: Testimuuttujan arvo $w = \frac{\overline{x}-0}{\sqrt{\frac{s^2}{n}}}$ on positiivinen täsmälleen silloin kun $\overline{x} > 0$ eli kun nollahypoteesiksi valitaan $H_0 : \mu \leq 0$. Jos siis $w > 0$ niin nollahypoteesi hylätään jos $w > F^{-1}_{\mathrm{t}(14)}(0.99) = 2.6245$ ja jos $w < 0$ niin nollahypoteesiksi tulee $H_0 : \mu \geq 0$ jolloin nollahypoteesi hylätään jos $w \leq F^{-1}_{\mathrm{t}(14)}(0.01) = -2.6245$. Näin ollen nollahypoteesi hylätään jos testimuuttuja on itseisarvoltaan vähintään 2.6245 ja jos $\mu = 0$ niin tämä tapahtuu edellä olevien laskujen perusteella todennäköisyydellä 0.02 eikä todennäköisyydellä 0.01 niin kuin merkitsevyystason määritelmän mukaan pitäisi olla.

Tässä laskussa oli täysin epäoleellista että otoksen koko oli 10 ja merkitsevyystaso 0.01 koska jos koko olisi n ja merkitsevyystaso α niin luku 2.6245 korvattaisiin luvulla $t_\alpha(n-1) = F^{-1}_{\mathrm{t}(n)}(1-\alpha)$ ja tulos olisi, että nollahypoteesi hylätään todennäköisyydellä 2α eikä todennäköisyydellä α.

P3. Heität tavanomaista noppaa 150 kertaa ja saat n_k kertaa tulokseksi k alla olevan taulukon mukaisesti:

Silmäluku	1	2	3	4	5	6
n_k	35	15	16	35	31	18

Onko noppa virheetön, eli ovatko kaikki silmäluvut yhtä todennäköisiä? Testaa χ^2-testillä merkitsevyystasolla 0.01.

Ratkaisu: Nollahypoteesi on, että jokainen tulos esiintyy todennäköisyydellä $\frac{1}{6}$ (eikä tässä ole oikeastaan mitään valinnanvaraa tämän suhteen). Silloin $np_k = \dfrac{150}{6} = 25$ kaikilla k ja testimuuttujan $C = \sum_{k=1}^{m} \dfrac{(n_k - np_k)^2}{np_k}$ arvoksi tulee

$$c = \frac{(35-25)^2}{25} + \frac{(15-25)^2}{25} + \frac{(16-25)^2}{25} + \frac{(35-25)^2}{25} + \frac{(31-25)^2}{25} + \frac{(18-25)^2}{25} = 18.64.$$

Nyt p-arvoksi tulee $p = \Pr(C \geq 18.64) = 1 - F_{\chi^2(5)}(18.64) = 0.0022 < 0.01$ joten nollahypoteesi hylätään merkitsevyystasolla 0.01.

P4. Erään tuotteen valmistaja väittää, että heidän tuotteistaan korkeintaan 5% on viallisia. Asiakas testaa tuote-erän poimimalla otoksen, jonka koko on 400, ja testaa nollahypoteesinsä 1% merkitsevyystasolla. Mikä tämä nollahypoteesi on? Laske todennäkäisyys, että ostaja ei hylkää nollahypoteesiaan, eikä siten tätä tuote-erää, jos todellisuudessa todennäköisyys, että tuote on viallinen onkin 0.07? Käytä normaaliapproksimaatiota kuten ostajakin.

Vihje: Menettele esim. seuraavalla: Laske ensin luku A siten että ostaja hylkää nollahypoteesinsä jos erässä on vähintään A viallista tuotetta (eli oletat, että ostaja laskee binomijakautuneella satunnasimuuttujalla etkä Bernoulli-jakatuneiden satunnasimuuttujien keskiarvolla. Laske sitten todennäköisyys, että erässä on vähemmän kuin A viallista tuotetta jos todennäköisyys, että tuote on viallinen in 0.07. Oleta (tietenkin) että tapahtumat "tuote on viallinen" ovat riippumattomia.

Vastaus: 0.65

Ratkaisu: Jos p on todennäköisyys, että tuote on viallinen niin nollahypoteesi on $H_0 : p \leq 0.05$.

Nollahypoteesin (ääritapauksen $p = 0.05$) mukaan viallisten tuotteiden lukumäärä on Y on $\mathrm{Bin}(400, 0.05)$-jakautunut ja kun käytetään normaaliapproksimaatiota niin testimuuttujan $\frac{Y-\mathrm{E}(Y)}{\sqrt{\mathrm{Var}(Y)}}$ hylkäysalue on $[z_{0.01}, \infty)$ missä $z_{0.01} = F^{-1}_{\mathrm{N}(0,1)}(0.99)$ eli nollahypoteesi hylätään jos testimuuttuja saa arvon joka on suurempi $z_{0.01}$. Koska $\mathrm{E}(Y) = 0.05 \cdot 400$ ja $\mathrm{Var}(Y) = 0.05 \cdot 0.95 \cdot 400$ niin nollahypoteesi hylätään kun

$$Y \geq 0.05 \cdot 400 + \sqrt{0.05 \cdot 0.95 \cdot 400} \cdot 2.3263 = 30.1,$$

ja luvuksi A voimme valita 31.

Seuraavaksi meidän pitää laskea todennäköisyys, että jos $U \sim \mathrm{Bin}(400, 0.07)$ niin $U \leq 30$. Normaaliapproksimaatiolla saamme

$$\begin{aligned}\Pr(U \leq 30) &= \Pr\left(\frac{U-\mathrm{E}(U)}{\sqrt{\mathrm{Var}(U)}} \leq \frac{30-\mathrm{E}(U)}{\sqrt{\mathrm{Var}(U)}}\right) \\ &= \Pr\left(\frac{U-\mathrm{E}(U)}{\sqrt{\mathrm{Var}(U)}} \leq \frac{30-0.07\cdot 400}{\sqrt{0.07 \cdot 0.93 \cdot 400}}\right) = \Pr\left(\frac{U-\mathrm{E}(U)}{\sqrt{\mathrm{Var}(U)}} \leq 0.39193\right) \approx 0.65.\end{aligned}$$

P5. Oletetaan, että $X \sim \mathrm{N}(\mu, \sigma^2)$. Mikä on todennäköisyys, että nollahypoteesi $H_0 : \sigma^2 \geq 4$ hylätään merkitsevyystasolla 0.05 jos todellisuudessa $\sigma^2 = 2.5$ ja nollahypoteesia testataan otoksella, jonka koko on 40.

Vihje: Voit menetellä esim. seuraavalla tavalla: Määritä luku A siten, että nollahypoteesi hylätään jos havaittu otosvarianssi $s^2 < A$ (kun $n = 40$). Laske sitten todennäköisyys, että otosvarianssi $S^2 < A$ jos $X \sim \mathrm{N}(\mu, 2.5)$. Muista mitä normaalijakauman tapauksessa tiedetään otosvarianssin jakaumasta!

Vastaus: 0.62

Ratkaisu: Jos $X_j \sim \mathrm{N}(\mu, \sigma^2)$, $j = 1, \ldots, n$ niin tunnetusti $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

(a) Kun nollahypoteesia $\sigma^2 \geq 4$ testataan merkitsevyystasolla 0.01 niin todetaan ensin, että jos nollahypoteesin ääritapaus $\sigma^2 = 4$ on voimassa niin

$$\Pr\left(\frac{(40-1)S^2}{4} < F^{-1}_{\chi^2(39)}(0.05)\right) = 0.05,$$

joten kriittiseksi rajaksi A, siten että nollahypoteesi hylätään kun $S^2 < A$, tulee

$$A = \frac{4 \cdot F^{-1}_{\chi^2(39)}(0.05)}{39} = \frac{4 \cdot 25.695}{39} = 2.6354.$$

(b) Jos nyt varianssi todellisuudessa on $\sigma^2 = 2.5$ niin meidän pitää laskea todennäköisyys, että $S^2 < 2.6354$ eli

$$\begin{aligned}\Pr(S^2 < 2.6354) &= \Pr\left(\frac{(40-1)S^2}{2.5} < \frac{(40-1)\cdot 2.6354}{2.5}\right) \\ &= \Pr\left(\frac{(n-1)S^2}{\sigma^2} < 41.112\right) = F_{\chi^2(39)}(41.112) = 0.62.\end{aligned}$$
