[MTTTP5] Tilastollisen päättelyn perusteet, syksy 2018

HARJOITUS 7 viikko 50

Ratkaisuja

1. $H_0 : \pi = 75$, $H_1 : \pi > 75$. Kun H_0 on tosi, niin $Z = \dfrac{p - 75}{\sqrt{75(100-75)/n}} \overset{likimain}{\sim} N(0,1)$. Nyt

$Z_{hav} = \dfrac{80 - 75}{\sqrt{75(100-75)/300}} = 2$. Yksisuuntaiseen testiin liittyvä pienin riskitaso, jolla nollahypoteesisi voidaan hylätä on P(Z > 2) = 1 -P(Z ≤ 2) = 1 - Φ(2) = 1 – 0,9772 = 0,0228. Jos kiinnitetään riskitaso, joka on suurempi kuin 2,28 %, niin tehdään päätelmä, että on tapahtunut muutos. Mutta jos valitaan esim. 1 %:n riskitaso, niin tehdään päätelmä, että ei muutosta. Testin voi tehdä myös kaksisuuntaisena, jolloin p-arvo on 0,0456. Tällöin esim. 5 %:n riskitasolla päätellään muutosta tapahtuneen, mutta ei esim. 3 %:n riskitasolla

2. $H_0 : \pi = 25$, $H_1 : \pi \neq 25$. Kun H_0 on tosi, niin $Z = \dfrac{p - 25}{\sqrt{25(100-25)/n}} \overset{likimain}{\sim} N(0,1)$. Nyt

$Z_{hav} = \dfrac{15 - 25}{\sqrt{25(100-25)/200}} = -3{,}266$. P(Z ≤ -3,266) = 1 - Φ(3,266) = 1 – 0,9995 = 0,0005, joten pienin ristitaso, jolla H_0voidaan hylätä on 2x0,0005 = 0,001. Kasvitieteilijän on syytä muuttaa käsitystään.

3.
H_0: $\mu_A = \mu_B$
$H_1 : \mu_A \neq \mu_B$

Jos H_0 tosi, niin $Z = \dfrac{\overline{X}_A - \overline{X}_B}{\sqrt{\dfrac{0{,}005^2}{100} + \dfrac{0{,}005^2}{100}}} \sim N(0,1)$. Nyt $z_{hav} = \dfrac{0{,}002}{\sqrt{\dfrac{0{,}005^2}{100} + \dfrac{0{,}005^2}{100}}} = 2{,}828$, $z_{0,01/2} = 2{,}57$, joten H_0 hylätään 1 %:n riskitasolla. Päätellään, että eivät enää tuota keskimäärin samanpituisia komponentteja. Pienin ristitaso, jolla H_0 voidaan hylätä, on 2(1-Φ(2,828)) = 0,0046.

4.
Hypoteesit $H_0 : \mu_{hyvä} = \mu_{tyydyttävä}$, $H_1 : \mu_{hyvä} \neq \mu_{tyydyttävä}$

Käytetty riippumattomien otosten t -testiä, jolloin $t = \dfrac{\overline{Y}_H - \overline{Y}_T}{S\sqrt{\dfrac{1}{n_H} + \dfrac{1}{n_T}}}$, missä

$$S^2 = \frac{(n_H - 1)S_H^2 + (n_T - 1)S_T^2}{n_H + n_T - 2} = \frac{SS_H + SS_T}{n_H + n_T - 2}$$

$$s^2 = \frac{(91-1)\cdot 428{,}81497^2 + (37-1)\cdot 393{,}10922^2}{91+37-2} = 175497{,}3 \text{ , joten}$$

$$t_{hav} = \frac{1815{,}6627 - 1636{,}1252}{418{,}9\sqrt{\frac{1}{91}+\frac{1}{37}}} = 2{,}198 \text{ —>a)-kohta}$$

Keskiarvojen erotuksen sekä keskiarvojen erotuksen estimoidun keskivirheen saa myös tuloksista suoraan:

Mean Difference 179,53750

Std. Error Difference = 81,68055= $418{,}9\sqrt{\frac{1}{91}+\frac{1}{37}}$

Vapausasteet 91+37-2 = 126 —>b)-kohta.

Koska $t_{0,025,\ 91+37-2} \approx$ 1,98 ja $t_{0,01,\ 91+37-2} \approx$ 2,358, niin $0{,}01 < p/2 < 0{,}025$ ja $0{,}02 < p < 0{,}05$ —>c)-kohta.

Jos valitaan 5 %:n ristitaso, niin H_0 hylätään, päätellään eroja olevan. Jos valitaan 1 %:n riskitaso, päätellään, että ei eroja.

Riippumattomien otosten t-testissä oletetaan, että on riippumattomat satunnaisotokset kahdesta normaalijakaumasta, joiden varianssit tuntemattomia, mutta yhtä suuria. Varianssien yhtäsuuruutta testataan Levenen testillä. Nollahypoteesi on, että varianssit samat. Hyväksytään nollahypoteesi, koska p-arvo on 0,262>0,05.

5. Jos oletetaan, että päivä- ja työtyötä tekevät ovat eri henkilöitä, niin on kyseessä riippumattomat otokset.

Havaintomatriisissa 20 riviä, kaksi muuttujaa

Poissaolot	Työvuoro
15	1
10	1
.	
12	1
8	2
9	2
.	
3	2

$H_0 : \mu_{yö} = \mu_{päivä}$, $H_1 : \mu_{yö} > \mu_{päivä}$. Kun H_0 on tosi, niin $t = \dfrac{\overline{Y}_{yö} - \overline{Y}_{päivä}}{S\sqrt{\frac{1}{n_{yö}}+\frac{1}{n_{päivä}}}} \sim t(n_{yö}+n_{päivä}-2)$,

missä $S^2 = \dfrac{(n_{yö}-1)S^2_{yö} + (n_{päivä}-1)S^2_{päivä}}{n_{yö}+n_{päivä}-2} = \dfrac{SS_{yö}+SS_{päivä}}{n_{yö}+n_{päivä}-2}$

Nyt $n_{yö} = n_{päivä}$ =10, $\bar{y}_{yö}$= 90 / 10 = 9, $\bar{y}_{päivä}$=50 / 10 = 5, $SS_{yö}$= 90, $SS_{päivä}$ = 128, joten s^2= (90+128) / (10+10-2) = 12,111, josta s = 3,48.

$t_{hav} = \dfrac{9-5}{3{,}48\sqrt{\frac{1}{10}+\frac{1}{10}}} = 2{,}57$. Nyt $t_{0{,}01,\,18}$= 2,552, $t_{0{,}005,\,18}$ = 2,878. Koska t_{hav}> 2,552, niin H_0 hylätään 1 %:n riskitasolla. Koska 2,552 < t_{hav} < 2,878, niin pienin riskitaso, jolla H_0 voidaan hylätä (yksisuuntaisessa testissä) on 0,005 < p < 0,01, itse asiassa p ≈ 0,01.

6. a) 9
   b) 5
   c) $\sqrt{90}/\sqrt{9}$ = 3,162
   d $\sqrt{128}/\sqrt{9}$ = 3,771
   e) 9-5= 4
   f) 1,556 (=$3{,}48\sqrt{\frac{1}{10}+\frac{1}{10}}$)
   g) 10+10-2=18
   h) 2,57

7. Riippumattomien otosten t-testi kuten tehtävissä 4 ja 5.

SPSS-ohje: Tallenna ensin havaintomatriisi, johon muuttujat viljelymenetelmä ja sato, siis 16 tilastoyksikköä. Riippumattomien otosten t-testi Analyze-> Compare Means-> Independent Samples t-test … (Test Variable(s) = Sato, Grouping Variable = Viljelymenetelmä), ks. http://cs.uef.fi/statistics/newspss/index.php/fi/7.

Saat tuloksen: ks. http://www.sis.uta.fi/tilasto/mtttp3/kevat2014/harj_4_ratk_t_6_spss.pdf
Nyt $H_0 : \mu_A = \mu_B$, $H_1 : \mu_A \neq \mu_B$. Testisuureen arvoksi saat -2,888 ja p-arvoksi 2-suuntaisessa testissä 0,012. Päätellään eroja olevan, jos tarkastellaan 5%:n riskitasolla (0,05>p) mutta ei enää 1%:n riskitasolla (0,01<p). Jos suoritat testin 1-suuntaisena, niin p-arvo on 0,012/2.

8. Muodosta aluksi neliövuokra-muuttuja.

Riippumattomien otosten t-testi Analyze - > Compare Means - > Independent Samples T-test. Ks. myös http://cs.uef.fi/statistics/newspss/index.php/fi/7, valitse ryhmittelymuuttujaksi Alue (Keskusta=19, Hervanta=8).