Harjoitus 9

1. Odotusarvojen erotuksen testi (kahden keskiarvon vertaaminen)

X = Elinaika seerumihoidolla; $X \sim N(\mu_x, \sigma^2)$

Y = Elinaika ilman seerumihoitoa; $Y \sim N(\mu_y, \sigma^2)$

Hypoteesit:

$H_0 : \mu_x = \mu_y$ (seerumihoito ei ole tehokasta)

$H_1 : \mu_x > \mu_y$ (hoito on tehokasta)

Testimuuttuja:

$$T = \frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{1}{n} + \frac{1}{m}} \cdot \sqrt{\frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}}} \sim t, n+m-2$$

H_0:n hylkäysalue:

$\hat{T} > r_0 = 1{,}895$ (raja taulukosta 2, yksisuuntainen testi, p= 0,05 ja $f = n + m - 2 = 7$)

Testimuuttujan arvo:

Havainnot: $\bar{x} = 2,86,\ s_x = 1,9705,\ n = 5,$

$\bar{y} = 2,\ s_y = 1,1916,\ m = 4$

$$\hat{T} = \frac{\bar{x} - \bar{y} - \overbrace{(\mu_x - \mu_y)}^{=0}}{\sqrt{\frac{1}{n} + \frac{1}{m}} \cdot \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}} \approx 0,762$$

Johtopäätös: H_0 on voimassa, ts. seerumihoito ei ole tehokasta.

(Graafisen laskimen käyttäjillä p-arvo = 0,235)

2. Samaa näytettä on tutkittu ennen ja jälkeen säilytyksen, joten mittaustulokset ovat pareittain riippuvia $\Rightarrow$ Parittainen vertailu

X_1 = Sorbaattihapon jäämä ennen säilytystä

X_2 = Sorbaattihapon jäämä säilytyksen jälkeen

Muodostetaan erotukset $D_i = X_{1i} - X_{2i} \quad i = 1, \cdots, 8$

X_1	224	270	400	444	590	660	1400	680
X_2	116	96	239	329	437	597	689	576
D	108	174	161	115	153	63	711	104

Tehdään tälle aineistolle yhden otoksen T-testi: $D \sim N(\mu, \sigma^2)$

Hypoteesit:

$H_0 : \mu = \mu_0 = 0$ (säilytys ei vaikuta)

$H_1 : \mu \neq \mu_0 = 0$ (säilytys vaikuttaa)

Testimuuttuja:

$$T = \frac{\overline{D} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t, n-1$$

Hylkäämisalue: $|\hat{T}| > r_0 = 2,365$ (r_0 taulukosta 2, kun p= 0,05 ja $f = 7$)

Testimuuttujan arvo: havainnoista saadaan $\overline{d} = 198,625$, $s = 210,165$ ja $n = 8$.

$\hat{T} = \frac{\overline{d} - \mu_0}{\frac{s}{\sqrt{n}}} \approx 2,67 > r_0$

Johtopäätös:

H_0 hylätään, ts. säilytys vaikuttaa.

(HUOM! Fysikaalisen perustietämyksen takia olisi perusteltua tehdä testi yksisuuntaisenakin: Vastahypoteesi $H_1 : \mu > \mu_0 = 0$, H_0 :n hylkäämisalue $\hat{T} > r_0 = 1,895$. H_0 hylätään tässäkin tapauksessa)

(p-arvo yksisuuntaisessa testissä 0,016, kaksisuuntaisessa testissä 0,032)

3. Yhteensopivuustesti

X=Kaupungin väkiluvun ensimmäinen numero

Benfordin jakauma: $P(X=x) = \log_{10}(1+\frac{1}{x})$

$p_1 = P(X=1) = 0,30103, \quad p_2 = P(X=2) = 0,17609, \quad p_3 = 0,12494, \quad p_4 = 0,0969,$

$p_5 = 0,07918, \quad p_6 = 0,06695, \quad p_7 = 0,05799, \quad p_8 = 0,05115, \quad p_9 = 0,04575.$

X	1	2	3	4	5	6	7	8	9	yht.
f_i	107	55	39	22	13	18	13	23	15	305=n
np_i	91,814	53,707	38,107	29,558	24,150	20,419	17,688	15,602	13,956	

$n = 305 > 50$ ja jokainen $np_i > 5$, joten testin ehdot voimassa.

Hypoteesit:

H_0: Aineisto noudattaa Benfordin lakia

H_1: Aineisto ei noudata Benfordin lakia

Testimuuttuja:

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} \sim \chi_{k-1}^2,$$

missä $k = 9$.

Riskitaso: α = 0,05.

Hylkäysalue: $\chi_0^2 > r_0$ = 15,507 (raja taulukosta 3, kun $f = 8$).

Testimuuttujan arvo:

$$\chi_0^2 = \frac{(107 - 91,814)^2}{91,814} + \ldots + \frac{(15 - 13,956)^2}{13,956} \approx 14,76 < r_0$$

Johtopäätös: H_0 jää voimaan, ts. aineisto noudattaa Benfordin lakia.

(p-arvo = 0,064)

4. Yhteensopivuustesti

	A	B	AB	O	Yht.
havaittu f_i	89	18	12	81	200
oletettu np_i	$0,41 \cdot 200 = 82$	20	8	90	200

$n = 200 > 50$ ja jokainen $np_i > 5$, joten testin ehdot voimassa.

Hypoteesit:

H_0: Havainnot vastaavat ilmoitettuja osuuksia.

H_1: Havainnot eivät vastaa ilmoitettuja osuuksia.

Testimuuttuja:

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} \sim \chi_{k-1}^2.$$

Riskitaso: $\alpha = 0{,}05$.

Hylkäysalue: $\chi_0^2 > r_0 = 7{,}815$ (raja taulukosta 3, kun $f = k - 1 = 3$).

Testimuuttujan arvo:

$$\chi_0^2 = \frac{(89-82)^2}{82} + \ldots + \frac{(81-90)^2}{90} \approx 3,70 < 7,815.$$

Johtopäätös: H_0 jää voimaan.

5. Riippumattomuustesti

Menetelmä	A	B	C	D	Yht.
Johtava	31	42	22	25	120
Viallinen	19	8	28	25	80
Yhteensä	50	50	50	50	200

$\hat{p}_1 = \frac{120}{200}$, $\hat{p}_2 = \frac{80}{200}$, $\hat{q}_1 = \hat{q}_2 = \hat{q}_3 = \hat{q}_4 = \frac{50}{200}$.

Pienin $n\hat{p}_i\hat{q}_j = 20 > 5$ ja $n = 200 > 50$, joten testin ehdot voimassa.

Hypoteesit:

H_0: Valmistusmenetelmillä ei ole vaikutusta levyn suprajohtavuuteen (ts. johtavuus on riippumatonta valmistusmenetelmästä)

H_1: Valmistusmenetelmillä on vaikutusta levyn suprajohtavuuteen.

Testimuuttuja:

$$\chi_0^2 = \sum_{i=1}^{k}\sum_{j=1}^{l} \frac{(f_{ij} - n\hat{p}_i\hat{q}_j)^2}{n\hat{p}_i\hat{q}_j} \sim \chi^2_{(k-1)(l-1)}$$

Riskitaso: $\alpha = 0{,}05$.

Hylkäysalue: $\chi_0^2 > r_0 = 7{,}815$ (raja taulukosta 3, kun $f = (k-1)(l-1) = (2-1)(4-1) = 3$)

Testimuuttujan arvo:

$$\chi_0^2 = \frac{(31 - 200 \cdot \frac{120}{200} \cdot \frac{50}{200})^2}{200 \cdot \frac{120}{200} \cdot \frac{50}{200}} + \ldots + \frac{(25 - 200 \cdot \frac{80}{200} \cdot \frac{50}{200})^2}{200 \cdot \frac{80}{200} \cdot \frac{50}{200}} = 19,5 > 7,815.$$

Johtopäätös: H_0 hylätään ja H_1 on voimassa ts. valmistusmenetelmillä on vaikutusta levyn suprajohtavuuteen.