Kaikissa tehtävissä käytetään testauksen rajana merkitsevyystasoa 5 %.

Aina keskiarvojen tarkastelu ei ole mielekästä tai mahdollista. Tässä tehtävässä testataan ovatko naisten ja miesten arvosanojen hajonnat samat kahden riippumattoman otoksen tilanteessa. Tehtävässä myös kerrottiin, että normaalijakaumaoletus on voimassa. Eli testattava hypoteesi on muotoa:

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Koska populaatioiden oletetaan olevan normaalisti jakautuneita, voidaan olettaa, että:

$$F_{hav} = \frac{S_1^2}{S_2^2} \sim F^{(n_1-1,n_2-1)}$$

Tehtävänannossa kerrottiin seuraavat asiat:

$n_1 = 31$ $S_1 = 0.15$
$n_2 = 31$ $S_2 = 0.29$
$\alpha = 0.05$

Sijoitetaan kaavaan:

$$F_{hav} = \frac{0.15^2}{0.29^2} \sim F^{(30,30)} \approx 0.27$$

p > 0.05 jos F_{hav} on välillä:

$$F_{0.025}^{(30,30)} > F_{hav} > F_{0.975}^{(30,30)}$$

Etsitään vapausasteella (30,30) olevat arvot (0,025) ja (0.975)

Arvoja ei löydy suoraan taulukosta, joten määrittelemme:

$$\boldsymbol{F_{0.025}^{(30,30)} \approx 2.07}$$

$$\boldsymbol{F_{0.975}^{(30,30)} = \frac{1}{F_{0.025}^{(30,30)}} \approx \frac{1}{2.07} \approx 0.48}$$

Koska $\boldsymbol{F_{hav} \approx 0.27}$, ei ole välillä

$$\boldsymbol{2.07 > F_{hav} > 0.48}$$

joten nollahypoteesi hylätään, koska $p<0.05$.

(Jos $\boldsymbol{F_{hav}}$ olisi laskettu $\boldsymbol{F_{hav} = \frac{0.29^2}{0.15^2} \sim F^{(30,30)} \approx 3.7}$, niin päätelmä olisi ollut sama.)

Vastaus: $p < 0.05$, eli aineisto ei tue väittämää, jonka mukaan populaatiossa miesten ja naisten asiaokaistyytyväisyyksien arvosanojen keskihajonnat ovat samoja.

Jatkoa tehtävään 1. Tiedetään, että:
a) Otosten jakaumat on poimittu normaalisti jakautuneesta perusjoukosta.
b) Oletus (kohdan 1 perusteella): populaatioiden varianssit ovat erisuuret.
c) Otosten koot ovat 31. (<50)

Yllämainittujen syiden takia käytetään t-jakaumaa testauksessa.

Ensin pitää laskea vapausaste:

$$C = \frac{\frac{S_1^2}{n_1}}{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_1}} = \frac{\frac{0.15^2}{31}}{\frac{0.15^2}{31} + \frac{0.29^2}{31}} \approx 0.211^2$$

$$\frac{1}{V} = \frac{C^2}{n_1 - 1} + \frac{(1-C)^2}{n_2 - 1} = \frac{0.211^2}{31-1} + \frac{(1-0.211)^2}{31-1} \approx 0.0222$$

$$V = 44.975 \approx 45$$

Sitten tehdään t-testi aivan viime viikon harjoitus mukaan:

$$t_{hav} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{3.9 - 4.5}{\sqrt{\frac{0.15^2}{31} + \frac{0.29^2}{31}}} \approx -10.23$$

$p = 2 * p\left(t^{(45)} > 10.23\right)$ **< 0.001 (2*0.0005)**

eli p < 0.05, joten nollahypoteesi ryhmien populaatiokeskiarvojen yhtäsuuruudesta hylätään.

Vielä pyydettiin määrittelemään luottamusväli:

$$\boldsymbol{V_{95} = \left[\overline{x_1} - \overline{x_2} - t_{\alpha/2}^{(v)} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} ; \overline{x_1} - \overline{x_2} + t_{\alpha/2}^{(v)} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right]}$$

$$= \left[\mathbf{3.9 - 4.5 - 2.01} \sqrt{\frac{\mathbf{0.15^2}}{\mathbf{31}} + \frac{\mathbf{0.29^2}}{\mathbf{31}}} ; \mathbf{3.9 - 4.5 + 2.01} \sqrt{\frac{\mathbf{0.15^2}}{\mathbf{31}} + \frac{\mathbf{0.29^2}}{\mathbf{31}}} \right]$$

$= [\mathbf{-0.71 ; -0.482}]$

Eli miesten ja naisten asiakastyytyväisyyden arvioiden ero on populaatiokeskiarvossa välillä Alaraja on -0.72 ja Yläraja -0.48 (alfan 5% virheen riskitaso). Eli naiset ovat keskimäärin tyytyväisempiä kuin miehet myös populaatiossa.

t-jakauman arvo 2.01 saadaan (2.000+2.021)/2 = 2.01
Voidaan myös käyttää vain arvoa 2.000 tai 2.021

Voidaan kuvitella, että ihmisten mielikuvat puolueista ja äänestystottumukset voivat vaihdella 6 kk sisällä. Ratkaistaan tehtävä yhden otoksen suhteellisen osuuden testillä. Ensin määritellään:

$$\boldsymbol{H_0: \pi = 0.123}$$
$$\boldsymbol{H_1: \pi \neq 0.123}$$
$$\boldsymbol{\pi_0 = 0.123}$$

$$\boldsymbol{p = \frac{135}{1000} = 0.135}$$

$$\boldsymbol{n = 1000}$$

Tarkastellaan ensin approksimaation hyvyyttä:

$$\boldsymbol{n * \pi_0 = 1000 * 0.123 = 123 \geq 5}$$
$$\boldsymbol{n * (1 - \pi_0) = 877 \geq 5}$$

Ehdot ovat ok, eli voidaan hyödyntää normaalijakaumaa

$$\boldsymbol{z_{hav} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0.135 - 0.123}{\sqrt{\frac{0.123(1-0.123)}{1000}}} = 1.16}$$

$$\boldsymbol{p = 2 * \left(1 - \Phi(1.16)\right) = 2 * (1 - 0.8770) = 0.246}$$

Koska p-arvo = 0.246 > 0.05 jää nollahypoteesi voimaan. Eli vaikka otoksen keskiarvo antoi eri kannatusluvut, menee se testauksen normaalin keskihajonnan vaihteluvälille, eikä ole perusteita olettaa, että kannatus on muuttunut 6 kk:n aikana.

Vielä pyydettiin määrittämään luottamusväli:

$$\boldsymbol{V_{95}} = \left[\boldsymbol{p} \quad - \mathbf{1.96} \sqrt{\frac{\boldsymbol{p}(\mathbf{1}-\boldsymbol{p})}{\boldsymbol{n}}} \; ; \boldsymbol{p} \quad + \mathbf{1.96} \sqrt{\frac{\boldsymbol{p}(\mathbf{1}-\boldsymbol{p})}{\boldsymbol{n}}}\right]$$

$$= \left[\mathbf{0.135} \quad - \mathbf{1.96} \sqrt{\frac{\mathbf{0.135}(\mathbf{1}-\mathbf{0.135})}{\mathbf{1000}}} \; ; \mathbf{0.135} \quad + \mathbf{1.96} \sqrt{\frac{\mathbf{0.135}(\mathbf{1}-\mathbf{0.135})}{\mathbf{1000}}}\right]$$

$$= [\mathbf{0.114}; \mathbf{0.156}]$$

Eli populaatiossa puolueen kannatuskunta on välillä 11.4 % – 15.6 % . (5 % virheen riski)

Tarkastellaan approksimaation hyvyyttä:

$$\boldsymbol{n} * \boldsymbol{p} = \mathbf{1000} * \mathbf{0.135} = \mathbf{135} \geq \mathbf{5}$$
$$\boldsymbol{n} * (\mathbf{1}-\boldsymbol{p}) = \mathbf{865} \geq \mathbf{5}$$

Ehdot ovat ok, eli voidaan hyödyntää normaalijakauman kertymäfunktiota.

Kyseessä on kahden riippumattoman otoksen suhteellisen osuuden testi.

$$H_0: \pi_1 = \pi_2$$
$$H_1: \pi_1 \neq \pi_2$$

$$n_1 = 35 \qquad p_1 = 0.41$$
$$n_2 = 45 \qquad p_2 = 0.51$$

$$n_{1*}p_1 = 35 * 0.41 = 14.35 \geq 5$$
$$n_1 * (1 - p_1) = 35 * (1 - 0.41) = 20.65 \geq 5$$
$$n_{2*}p_2 = 45 * 0.51 = 22.95 \geq 5$$
$$n_2 * (1 - p_2) = 45 * (1 - 0.51) = 22.05 \geq 5$$

Eli oletukset ok ja voidaan käyttää z-testiä.

$$p = \frac{n_1 * p_1 + n_2 * p_2}{n_1 + n_2} = \frac{35 * 0.41 + 45 * 0.51}{35 + 45} = 0.466$$

$$Z_{hav} = \frac{p_1 - p_2}{\sqrt{p(1-p) * (\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.41 - 0.51}{\sqrt{0.466(1 - 0.466) * (\frac{1}{35} + \frac{1}{45})}}$$

$$= -0.89$$

$$p = 2 * \left(1 - \Phi(0.89)\right) = 2 * (1 - 0.8133) = 0.3734$$

Koska p = 0.3734 > 0.05 niin nollahypoteesi jää voimaan. Eli Aineisto tuki väittämää, jonka mukaan ikäryhmien välillä ei ole eroa tuotteita tunnistavien osuuksissa perusjoukossa.

Määritellään vielä luottamusväli suhteellisien osuuksien erotukselle:

$$\boldsymbol{V_{95} = \left[p_1 - p_2 - z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\,; p_1 - p_2 + z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right]}$$

$$= \boldsymbol{\left[0.41 - 0.51 - 1.96\sqrt{\frac{0.41(1-0.41)}{35} + \frac{0.51(1-0.51)}{45}}\,; 0.41 - 0.51 + 1.96\sqrt{\frac{0.41(1-0.41)}{35} + \frac{0.51(1-0.51)}{45}}\right]}$$

$$= [\boldsymbol{-0.31\,;\ 0.119}]$$

Logon tunnistavien osuuksien ero ikäryhmien välillä on 32 % -yks. (vanhemmassa ikäryhmässä yleisempää) – 12 % -yks. (nuoremmassa yleisempää) (5 % virheen riski).

Luottamusväli pitää sisällään nollan. Eli voimme päätellä, että toinen ryhmä voi tunnistaa toista paremmin tuotteita tai päinvastoin.

$$\boldsymbol{H_0}: \boldsymbol{ei\ lineaarista\ riippuvuutta\ eli\ kerroin = 0}$$
$$\boldsymbol{H_1}: \boldsymbol{on\ lineaarista\ riippuvuutta\ eli\ kerroin \neq 0}$$

Tehtävänannosta tiedetään:

$\boldsymbol{n = 31}$ $\boldsymbol{r = 0.42}$ $\boldsymbol{V = 31 - 2 = 29}$

Koska populaatiot ovat normaalisti jakautuneita, voimme käyttää t-testiä.

$$\boldsymbol{t_{hav} = \frac{r * \sqrt{V}}{\sqrt{1 - r^2}} = \frac{0.42 * \sqrt{29}}{\sqrt{1 - 0.42^2}} = 2.492}$$

saadaan $\boldsymbol{p = 2 * p(t^{29} > 2.492)}$

eli 0.01<p<0.02

Ja koska p < 0.05 nollahypoteesi hylätään.

Eli palkan ja työkokemuksen välillä on lineaarista riippuvuutta populaatiossa. (alpha 0.05).

Otoskoko on suuri ja jälleen hyödynnetään normaaliapproksimaatiota:

$$\boldsymbol{p = \frac{9}{1000} = 0.009}$$

Luottamusväli:

$$\boldsymbol{V_{95} = \left[p - 1.96 * \sqrt{\frac{p*(1-p)}{n}}\,; p + 1.96 * \sqrt{\frac{p*(1-p)}{n}}\right] = [0.0031\,; 0.0149]}$$

Populaatiossa on viallisia tuotteita 0.31 % — 1.49 % (5 % virhetaso)

Testataan approksimaation hyvyys:

$$\boldsymbol{n * p = 1000 * 0.009 = 9 \geq 5}$$
$$\boldsymbol{n * (1-p) = 1000 * (1-0.009) = 991 \geq 5}$$

Kummatkin ehdot toteutuvat ja normaaliapproksimaation käyttö on tässä tapauksessa perusteltua.

Testataan ovatko osapopulaatiot suhteellisesti yhtä suuria:

$$H_0: Ikäryhmät\ populaatiossa\ ovat\ yhtäsuuria$$
$$H_1: Ikäryhmät\ populaatiossa\ eivät\ ole\ yhtäsuuria$$

Ikäryhmä	-18	19–35	36–50	50-
frekvenssit	52	75	39	42
odotettu.frek	52	52	52	52

Odotusarvo = (52+75+39+42)/4 = 208/4 = 52

Ehto: Jokaisen odotetun frekvenssin tulee olla suurempi kuin 5. (Täyttyy) Joten voimme käyttää Khiin neliötestiä.

$$\chi^2_{hav} = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} \sim \chi^2_{(k-1)}$$

Sijoitetaan arvot:

$$\chi^2_3 = \frac{(52-52)^2}{52} + \frac{(75-52)^2}{52} + \frac{(39-52)^2}{52} + \frac{(42-52)^2}{52}$$

$$= 0 + 10.17 + 3.25 + 1.92 = 15.35$$

$$0.001 < p = p\left(\chi^2_{(3)} > 15.35\right) < 0.005$$

Eli p < 0.05 joten nollahypoteesi hylätään.

Aineisto tuki 5 % merkitsevyystasolla olettamusta, jonka mukaan havaintoaineisto ei ole peräisin oletettua suhteellista jakaumaa noudattavasta populaatiosta (eli suhteellinen jakauma on jotain muuta kuin nollahypoteesissa väitettiin). Estimaatiksi sopii otoksesta laskettu prosenttijakauma:

Ikäryhmä	**-18**	**19–35**	**36–50**	**50-**
frekvenssit	**25 %**	**36 %**	**19 %**	**20 %**

19-35-vuotiaiden ikäryhmä kuuntelijakunnassa suurempi (36%>25%) kuin alun perin väitettiin. Vanhimpien ikäryhmien osuudet pienempiä.

Testataan ovatko osapopulaatiot suhteellisesti yhtä suuria:

$$H_0: \boldsymbol{Sukupuolen\ ja\ asenteen\ v}\text{ä}\boldsymbol{lill}\text{ä}\ \boldsymbol{ei\ ole\ eroa}$$
$$H_1: \boldsymbol{Sukupuolen\ ja\ asenteen\ v}\text{ä}\boldsymbol{lill}\text{ä}\ \boldsymbol{on\ eroa}$$

havaitut frekvenssit:

	Mies	Nainen	
Täysin eri mieltä	19	30	49
Jonkin verran eri mieltä	19	29	48
Jonkin verran samaa mieltä	23	22	45
Täysin samaa mieltä	30	26	56
	91	107	198

Odotetut frekvenssit:

	Mies	Nainen	
Täysin eri mieltä	22.52=91*49/198	26.48	49
Jonkin verran eri mieltä	22.06	25.94	48
Jonkin verran samaa mieltä	20.68	24.32=107*45/198	45
Täysin samaa mieltä	25.74	30.26	56
	91	107	198

Odotetut frekvenssit lasketaan kaavalla:

$$\boldsymbol{e_{ij}} = \frac{\boldsymbol{f_{\cdot i} f_{\cdot j}}}{\boldsymbol{n}}$$

Kaikki odotetut frekvenssit ovat yli 5 eli voidaan käyttää khiin nelilöjakaumaa.

$$\chi^2_{hav} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} \sim \chi^2_{(r-1)(s-1)}$$

Vapausaste V = (2-1)*(4-1) = 3

Sijoitetaan arvot:

$$\chi^2_3 = \frac{(19-22.52)^2}{22.52} + \frac{(30-26.48)^2}{26.48} + \cdots + \frac{(26-30.26)^2}{30.26}$$

$$= 3.59$$

$0.1 < p < 0.95$ **eli p > 0.05 joten nollahypoteesi jää voimaan.**

Eli asenteella ja sukupuolella ei ole riippuvuutta.

Aineisto tuki väittämää, jonka mukaan asenteella ja sukupuolella ei ole riippuvuutta perusjoukossa (p>0.05).