

Tuntitehtävä 1)

Markkinatutkimuksella haluttiin selvittää kuinka suuri osuus valmistuvista kauppatieteen maistereista haluaisi työskennellä eräässä yrityksessä. Kahdessa satunnaisotoksessa haastateltiin 50 kpl naisia ja miehiä 60 kpl. Naisista 80 % haluaisi työskennellä yrityksessä ja miehistä 70 %. Testaa onko sukupuolten välillä eroa. Tarkastele myös approksimoinnin hyvyttä.

Kyseessä on kahden riippumattoman otoksen suhteellisen osuuden testi.

$$H_0: \pi_1 = \pi_2$$

$$H_1: \pi_1 \neq \pi_2$$

$$n_1 = 50 \quad p_1 = 0.80$$

$$n_2 = 60 \quad p_2 = 0.70$$

$$n_1 * p_1 = 50 * 0.80 = 40 \geq 5$$

$$n_1 * (1 - p_1) = 50 * (1 - 0.80) = 10 \geq 5$$

$$n_2 * p_2 = 60 * 0.70 = 42 \geq 5$$

$$n_2 * (1 - p_2) = 60 * (1 - 0.70) = 18 \geq 5$$

Eli otoskoko on riittävä approksimaatioon ja voidaan käyttää z-testiä.

$$p = \frac{n_1 * p_1 + n_2 * p_2}{n_1 + n_2} = \frac{50 * 0.80 + 60 * 0.70}{50 + 60} = 0.745$$

$$Z_{hav} = \frac{p_1 - p_2}{\sqrt{p(1-p) * (\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.80 - 0.70}{\sqrt{0.745(1 - 0.745) * (\frac{1}{50} + \frac{1}{60})}} \\ = 1.198$$

$$p = 2 * (1 - \Phi(1.20)) = 2 * (1 - 0.8849) = 0.2302$$

Koska $p > 0.05$, joten nollahypoteesi jää voimaan. Eli sukupuolten välillä ei voida havainta eroa.

Tuntiesimerkki 2)

Lomakohteessa Balilla on matkajia neljästä eri pohjoismaisesta kaupungista. Testataan ovatko osapopulaatiot suhteellisesti yhtä suuria:

H_0 : Kaupunkiryhmät Balilla ovat yhtäsuuria
 H_1 : Kaupunkiryhmät Balilla eivät ole yhtäsuuria

Ikäryhmä	Turku	Pori	Helsinki	Tukholma
frekvenssit	100	200	150	120
odotettu.frek	142.5	142.5	142.5	142.5

Odotettu matkailijamäärä jos ryhmien välillä ei ole eroa = $(100+200+150+120)/4 = 142.5$

Ehto: Odotetuista frekvensseistä yli 20 % tulee olla suurempi kuin 5 ja yksikään ei ole alle 1. (Täyttyy) Joten voimme käyttää Khiin neliötestiä:

Vapauaste: $k - 1$, eli ryhmien määrä -1. Nyt : $4 - 1 = 3$.

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

Sijoitetaan arvot:

$$\begin{aligned}\chi^2_3 &= \frac{(100 - 142.5)^2}{142.5} + \frac{(200 - 142.5)^2}{142.5} + \frac{(150 - 142.5)^2}{142.5} \\ &\quad + \frac{(120 - 142.5)^2}{142.5} \\ &= 12.68 + 23.20 + 0.39 + 3.55 = 39.82\end{aligned}$$

katsotaan todennäköisyys taulukosta: $p < 0.001$

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:

Turku	Pori	Helsinki	Tukholma
17.5 %	35.1 %	26.3 %	21.1 %

Tehtävä 5)

Testataan ovatko osapopulaatiot suhteellisesti yhtä suuria. Alla on lueteltu erään otoksen miesten ja naisten lempi elokuvagenrejä.

	Mies	Nainen
Komedia	50	30
Kauhu	40	40
Jännitys	40	30

H_0 : Sukupuolen ja asenteen välillä ei ole eroa

H_1 : Sukupuolen ja asenteen välillä on eroa

havaitut frekvenssit:

	Mies	Nainen	Summa
Komedia	50	30	80
Kauhu	40	40	80
Jännitys	40	30	70
Summa	130	100	230

Odotetut frekvenssit:

	Mies	Nainen
Komedia	$80 \cdot 130 / 230 = 45.2$	$80 \cdot 100 / 230 = 34.8$
Kauhu	$80 \cdot 130 / 230 = 45.2$	$80 \cdot 100 / 230 = 34.8$
Jännitys	$70 \cdot 130 / 230 = 39.6$	$70 \cdot 100 / 230 = 30.4$

Odotetut frekvenssit lasketaan kaavalla:

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

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)*(2-1) = 1

Vapausasteessa
oli virhe, korjaukset
merkitty punaisella.

vapausaste = [taulukon rivimäärä-1 * sarakemäärä-1]

Sijoitetaan arvot:

$$\chi^2 = \frac{(50 - 45.2)^2}{45.2} + \frac{(30 - 34.8)^2}{34.8} + \dots + \frac{(30 - 30.4)^2}{30.4}$$
$$= 2.56$$

**0.95 > p > 0.1 eli p > 0.05 joten nollahypoteesi jää voimaan.
Eli lempi elokuvagenren ja sukupuolen välillä ei ole riippuvuutta.**