

[MTTTP1] TILASTOTIETEEN JOHDANTOKURSSI, kevät 2019

<https://coursepages.uta.fi/mtttp1/kevat-2019/>

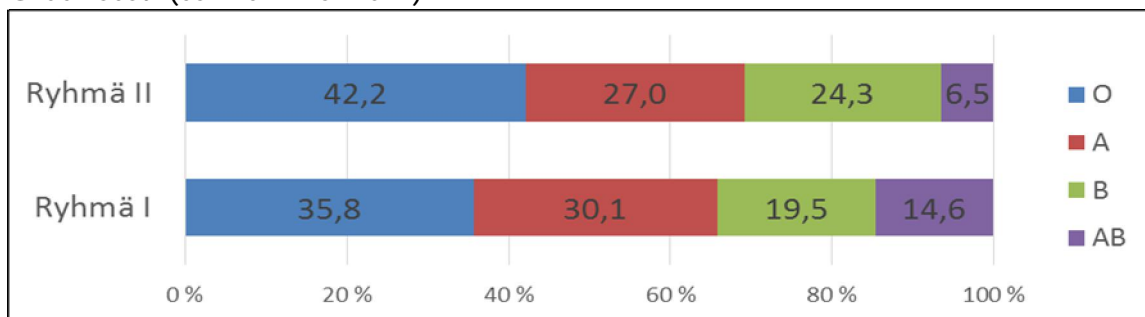
HARJOITUS 2 RATKAISUJA

Joitain ratkaisuja

- Esimerkiksi
 - AVIOSUH, koska luokitteluasteikollinen muuttuja,
 - LAPSLKM, koska kvantitatiivinen, diskreetti saaden vain muutaman arvon, KUNTO, koska järjestysasteikollinen,
 - LAPSPAI, LAPSPIT, koska suhdeasteikollisia.
- Lasketaan tutkittavan (=veriryhmä) muuttujan ehdolliset (väestöryhmittäin) prosenttijakaumat ja saadaan:

	Veriryhmä			
	O	A	B	AB
Ryhmä I	35,8%	30,1%	19,5%	14,6%
Ryhmä II	42,2%	27,0%	24,3%	6,5%

Graafisesti (esimerkiksi näin):



Vertaillaan näitä prosenttijakaumia. Jakaumat näyttävät poikkeavan toisistaan, joten voidaan sanoa, että veriryhmän jakaumissa on eroja väestöryhmien välillä.

Esimerkiksi näyttäisi siltä, että yleisimpään veriryhmään väestöryhmässä II kuuluu suhteellisesti enemmän kuin ryhmässä I, mutta toisaalta harvinaisempaa ryhmään kuuluvia on suhteessa vähemmän.

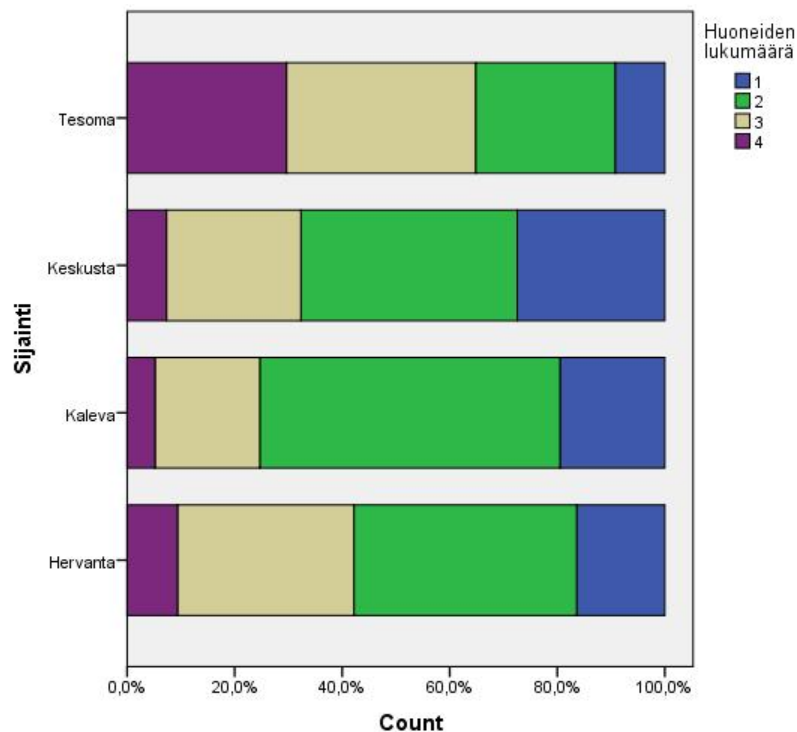
- $$\bar{x} = (0 \cdot 1 + 1 \cdot 6 + 2 \cdot 10 + 3 \cdot 16 + 4 \cdot 19 + 5 \cdot 5 + 6 \cdot 8 + 7 \cdot 3 + 8 \cdot 1 + 9 \cdot 1) / 70 = 261 / 70 \approx 3,7.$$
 Alle keskiarvon jäätiin $(1 + 6 + 10 + 16) = 33$ vuonna, joka on n. 47 %. Tätä voidaan pitää todennäköisyytenä sille, että seuraavana vuonna jäädään alle keskiarvon. Mediaani on neljä.
- On käytetty frekvenssihistogrammeja. Oikea valinta, koska neliöhinta-muuttuja on kvantitatiivinen. Huomataan, että keskustassa ja ei-keskustassa olevien asuntojen neliöhintojen jakaumat erilaisia. Ne ovat eri kohdassa ja muodoltaankin erilaiset. Keskustassa jakauma on "ylempänä", siis kalliimmat neliöhinnat. Ei-keskustassa jakauma selvästi vino oikealle

6. Koska tutkitaan sijainnin vaikutusta huoneiden lukumäärään, on ensin laskettava ehdolliset prosentuaaliset jakaumat huoneiden lukumäärästä. Saadaan jakaumat (SPSS-tuloste, jakaumat löytyvät myös 14.9. luentokalvoista):

% within Sijainti

		Sijainti				Total
		Hervanta	Kaleva	Keskusta	Tesoma	
Huoneiden lukumäärä	1	16,4%	19,5%	27,4%	9,3%	20,3%
	2	41,4%	55,8%	40,2%	25,9%	41,6%
	3	32,8%	19,5%	25,0%	35,2%	27,7%
	4	9,4%	5,2%	7,3%	29,6%	10,4%
Total		100,0%	100,0%	100,0%	100,0%	100,0%

Yksi vaihtoehto graafiseksi esitykseksi on:



Huomataan, että % -osuudet poikkeavat toisistaan, joten sijainnilla näyttäisi olevan vaikutusta asunnon kokoon. Voidaan esimerkiksi todeta, että 65 % Tesoman asunnoista on kaksioita suurempia, Kalevassa tämä osuus vain 25 %.

7. Kaleva:
 Keskiarvo = $(1 \cdot 15 + 2 \cdot 43 + 3 \cdot 15 + 4 \cdot 4) / 77 \approx 2,1$
 Md = 2

Tesoma:
 Keskiarvo = $(1 \cdot 5 + 2 \cdot 14 + 3 \cdot 19 + 4 \cdot 16) / 54 \approx 2,9$
 Md = 3

Kuten tehtävässä 6 voidaan todeta Kalevan ja Tesoman huoneiden lukumäärissä olevan eroja. Vertailemalla keskiarvoja huomataan, että Kalevassa on keskimäärin pienempiä asuntoja.

8. a) Keskiarvo 8.

Keskistetyt arvot :

$$8-8=0$$

$$9-8=1$$

$$6-8=-2$$

$$7-8=-1$$

$$10-8=2$$

Näistä keskiarvo $(0+1-2-1+2)/5 = 0/5 = 0$. Keskistetyn muuttujan keskiarvo on aina nolla, ks. b)-kohta.

Keskistys SPSS:llä: Transform-> Compute Variable -> Numeric Expression
 pisteet-8

b) Muuttujan x keskistys $z_i = x_i - \bar{x}$, $i = 1, 2, \dots, n$. Keskistetyn muuttujan keskiarvo on nolla.

Todistus: Muuttujan x keskistys $z_i = x_i - \bar{x}$, $i = 1, 2, \dots, n$.

$$\begin{aligned} \bar{z} &= \frac{1}{n} \sum_{i=1}^n z_i = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) = \frac{1}{n} \left(\sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^n x_i - n \cdot \bar{x} \right) = \frac{1}{n} \sum_{i=1}^n x_i - \frac{1}{n} \cdot n \cdot \bar{x} = \bar{x} - \frac{1}{n} \cdot n \cdot \bar{x} = 0 \end{aligned}$$