

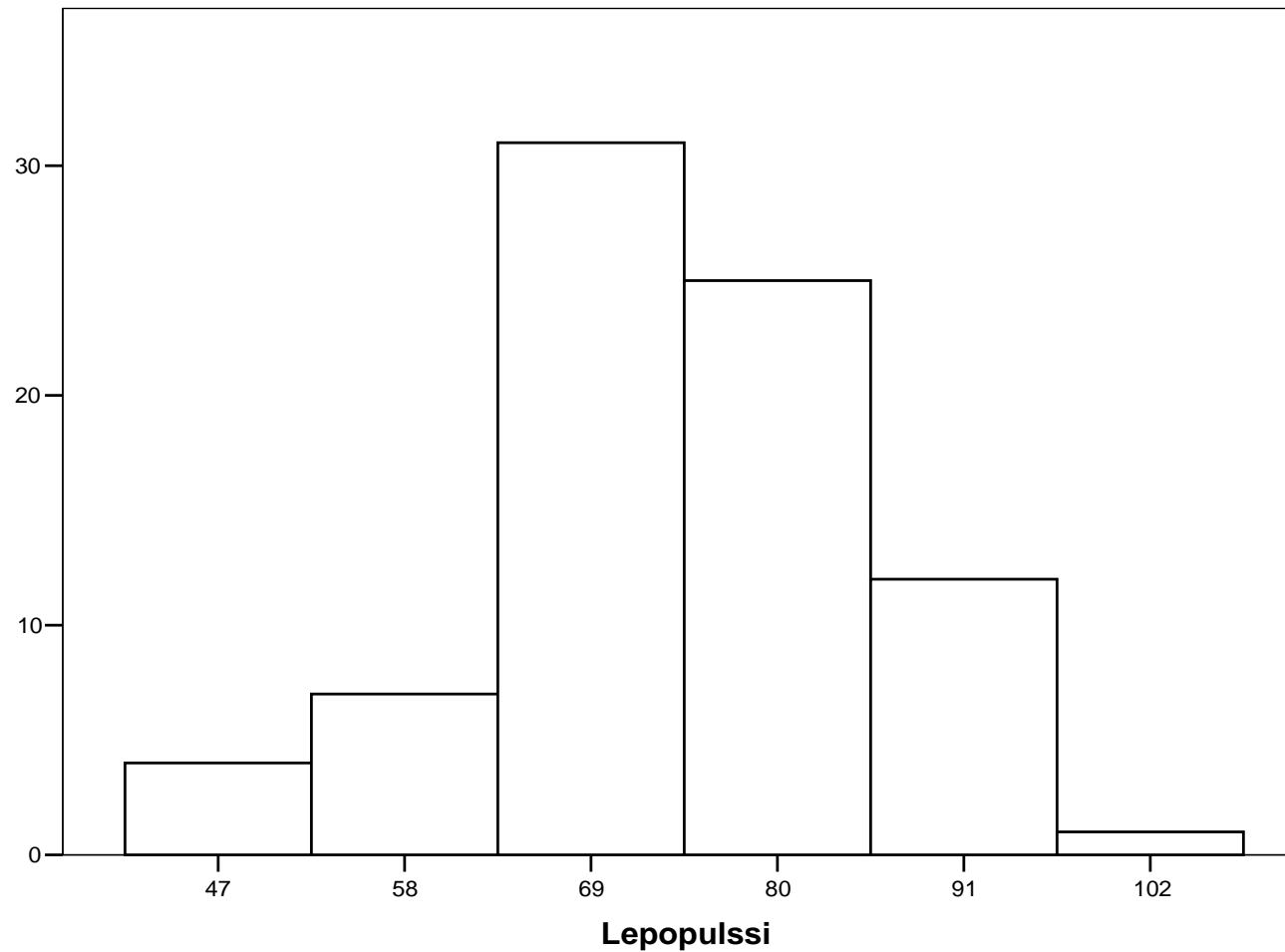
MTTTP1, luento 12.3.2019

## KERTAUSTA

Esim. Pulssi-muuttujan frekvenssijakauma, aineisto  
luentomoniste liite 4

<u>pyöristetyt</u> <u>luokkarajat</u>	<u>todelliset</u> <u>luokkarajat</u>	<u>luokka-</u> <u>keskus</u>	<u>frekvenssi</u>
42–52	41,5–52,5	47	4
53–63	52,5–63,5	58	7
64–74	63,5–74,5	69	31
75–85	74,5–85,5	80	25
86–96	85,5–96,5	91	12
97–107	96,5–107,5	102	1

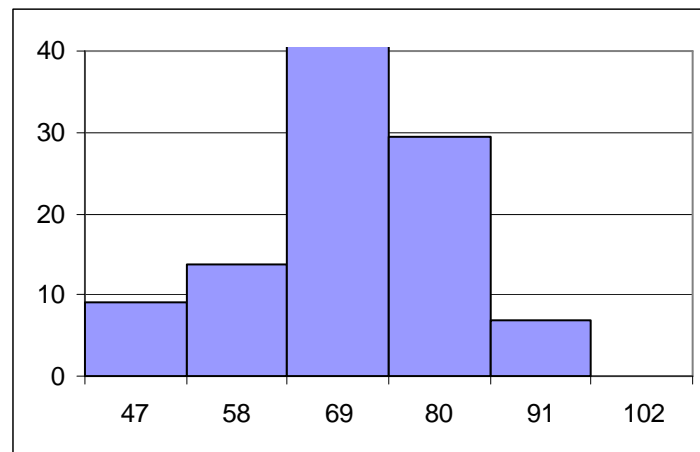
# Graafinen esitys frekvenssihistogrammi



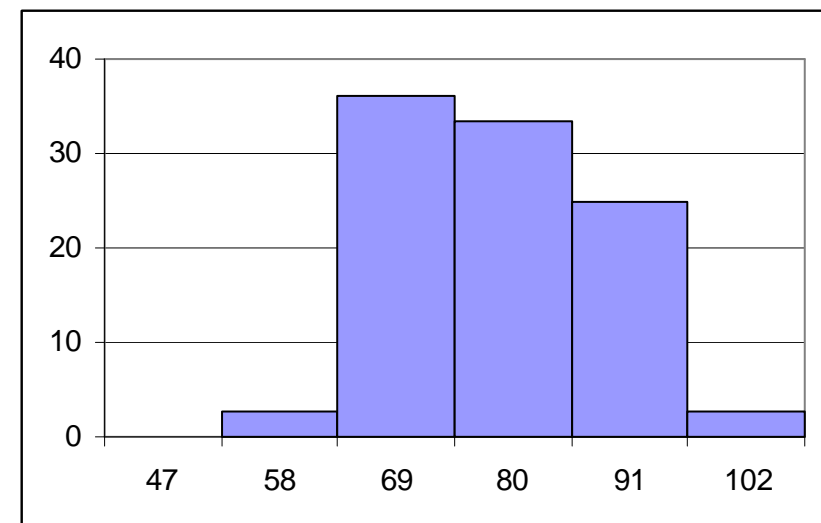
Huom. Piirretään todellisista luokkarajoista

Esim. 5.1.13. Pulssi-muuttujan frekvenssihistogrammit miehillä ja naisilla esimerkin 5.1.7 taulukosta, piirretään käyttäen prosentuaalisia frekvenssejä, jotta jakaumien vertailu olisi paremmin mahdollista.

Lepopulssin jakauma  
miehillä



Lepopulssin jakauma  
naisilla



Esim. Tampereelle 2009 myytyjä pieniä (alle 35 m<sup>2</sup>)  
asuntoja, aineisto Tre\_myydyt\_asunnot\_2009.sav  
sivulla  
<https://coursepages.uta.fi/mttp1/esimerkkiaineistoja/>

Muuttujat: Neliöt, Hinta, Rakennusvuosi, Sijainti,  
Kunto, Neliöhinta = Hinta/Neliöt

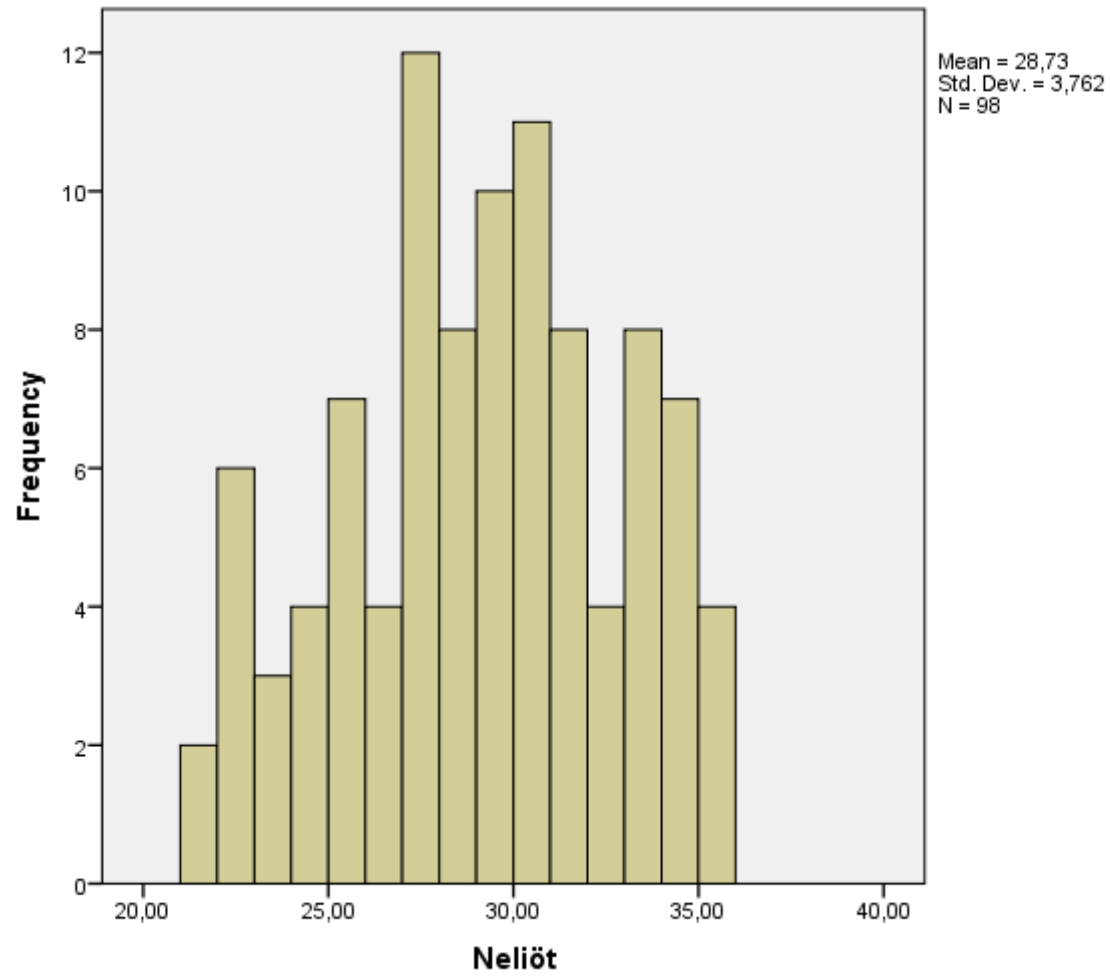
## Jakaumia (SPSS-ohjelman tuottamia taulukoita ja kuvia)

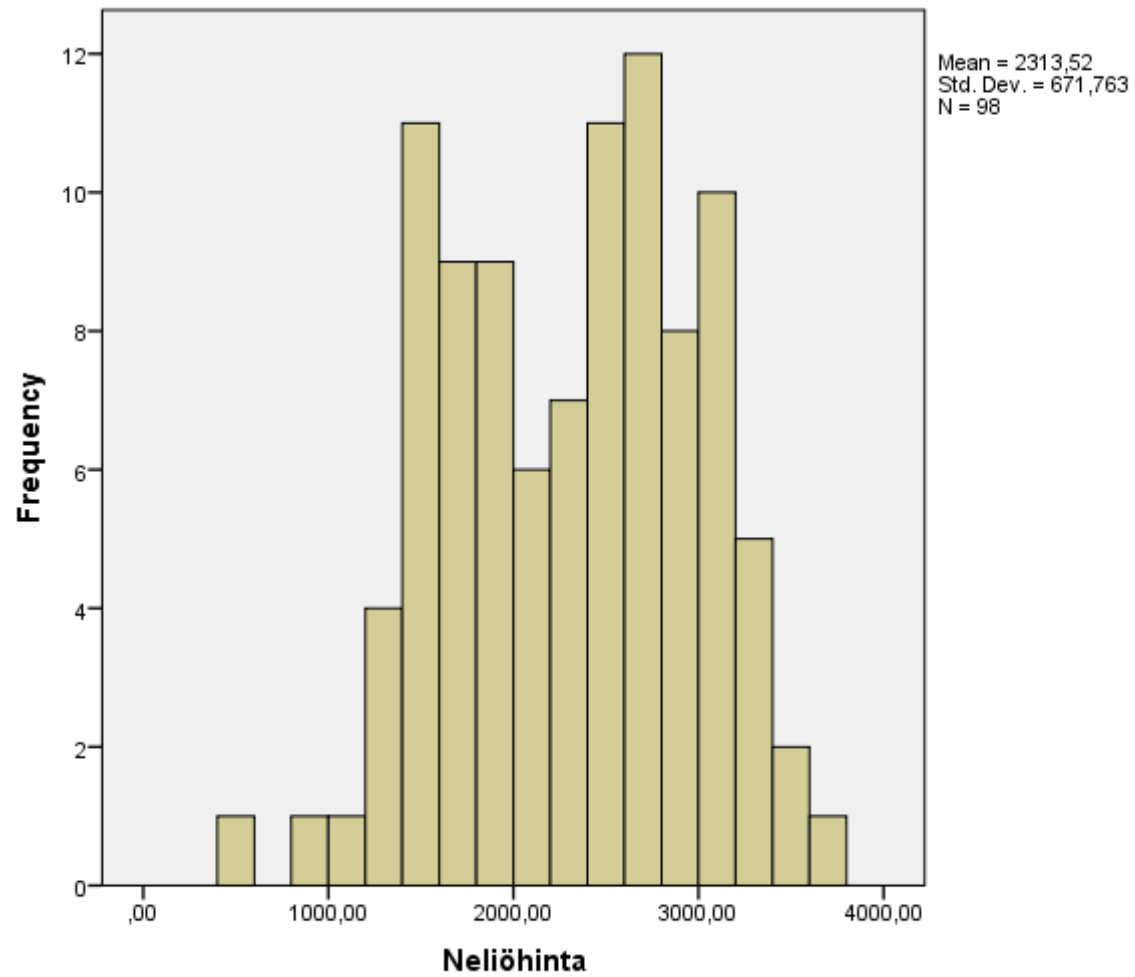
### Sijainti

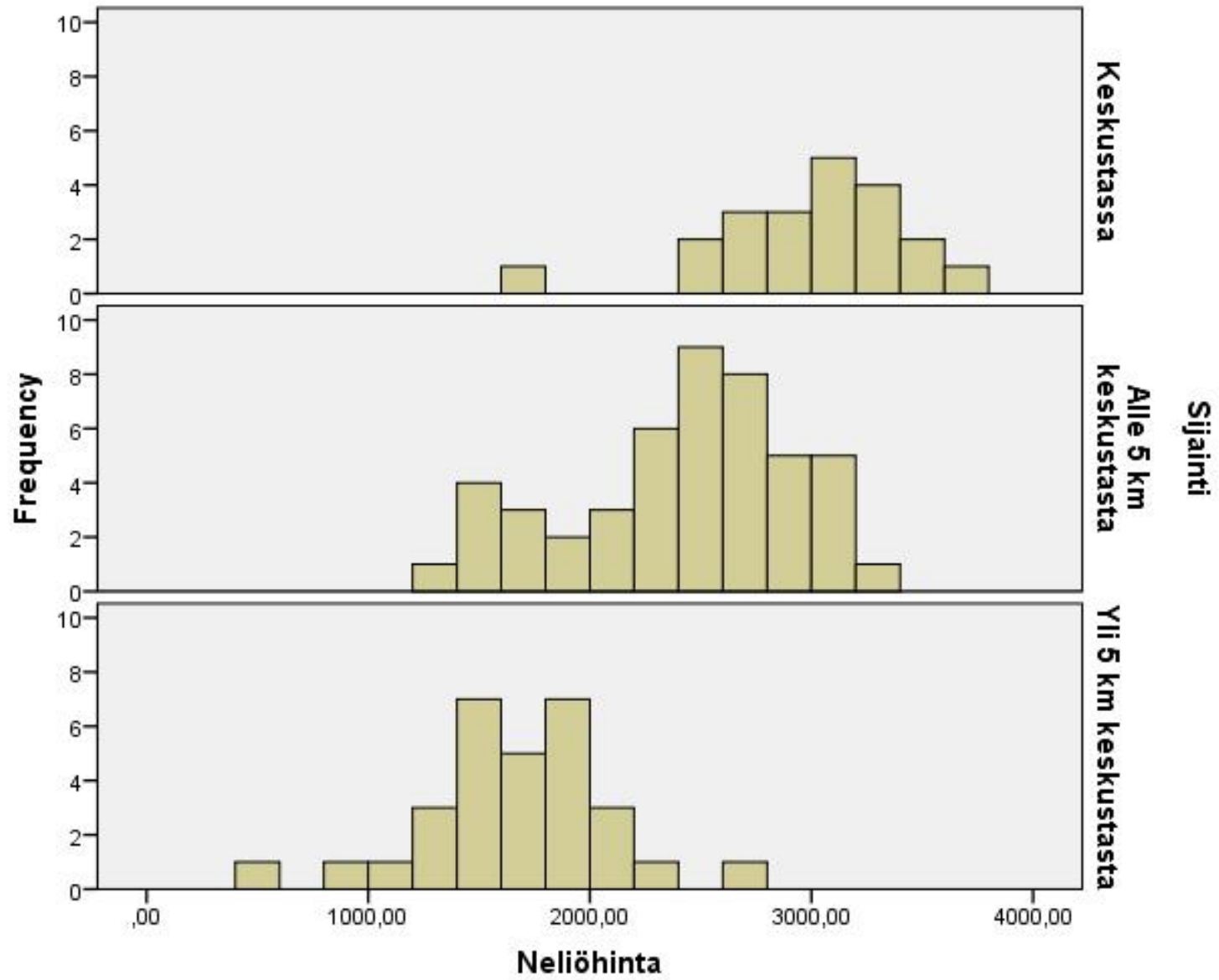
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Keskustassa	21	21,4	21,4	21,4
Alle 5 km keskustasta	47	48,0	48,0	69,4
Yli 5 km keskustasta	30	30,6	30,6	100,0
Total	98	100,0	100,0	

### Kunto

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Hyvä	32	32,7	32,7	32,7
Tyydyttävä	44	44,9	44,9	77,6
Huono	22	22,4	22,4	100,0
Total	98	100,0	100,0	









## 5.1.3 Yksiulotteisen jakauman tunnuslukuja

### Tunnusluvut

- kuvataan jakaumaa muuttujan arvoista lasketulla (tai arvojen avulla määritellyllä) luvulla
- kuvataan jakauman sijaintia, vaihtelua, vinoutta, huipukkuutta, jne.
- mitta-asteikko määrittää tunnusluvun valinnan

### 1) Sijainnin tunnuslukuja

#### Keskilukuja

- moodi (Mo)
- mediaani (Md), järjestysasteikollisuus
- keskiarvo, kvantitatiivisuus

## Muita sijainnin tunnuslukuja

- ala- ja yläkvartiili, muut fraktiilit, järjestysasteikollisuus
- laatikko-jana –kuvio muodostetaan kvartiilien avulla

## 2) Vaihtelua mittaavia tunnuslukuja

- varianssi, keskihajonta, kvantitatiivisuus
- variaatiokerroin, suhdeasteikollisuus

## 3) Muita tunnuslukuja

- erilaisia vinous- ja huipukkuuskertoimia

## 1) Sijainnin tunnuslukuja

## Keskilukuja

- Moodi ( $M_o$ ) on se muuttujan arvo, joka esiintyy useimmin tai se luokka, jossa on eniten havaintoja

Esim. Lapsen sisarusten lukumäärä, esim. 5.1.29

Sisarusten lukumäärä Frekv.

0	56	$M_o = 0$
1	39	
2	13	
3	10	
4	5	
5	2	
6	1	
<hr/>		
Yht.	126	

- Mediaani ( $Md$ ) on sellainen muuttujan arvo, jota pienempiä ja suurempia arvoja on yhtä paljon. Muuttujan oltava vähintään järjestysasteikollinen.

Esim. 5.1.14. Tenttipisteet: 95, 86, 78, 90, 62, 73, 89  
 $Md = 86$

Esim. 5.1.29. Sisarusten lukumäärä  
 $Md = 1$

- Keskiarvo, kaava (1),  
<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/kaavat.pdf> ,  
vaaditaan kvantitatiivisuus

Muuttujan  $x$  arvot tilastoyksiköittäin  $x_1, x_2, \dots, x_n$ , tällöin

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Esim. Etäisyydet, joista lepakot löysivät hyönteisiä, ks.  
Selityksiä ja esimerkkejä kaavoihin

[http://www.sis.uta.fi/tilasto/mttp1/syksy2018/esimerkit\\_kaavoihin.pdf](http://www.sis.uta.fi/tilasto/mttp1/syksy2018/esimerkit_kaavoihin.pdf)

Esim. 5.1.15. Keskiarvo tenttipisteistä

$$(95 + 86 + 78 + 90 + 62 + 73 + 89) / 7 = 81,9$$

Esim. 5.1.16. Lepopulssin keskiluvut, mediaani 74 ja keskiarvo 73,75.

SPSS-tulos:

Statistics		
Pulssi		
N	Valid	80
	Missing	0
Mean		73,7500
Median		74,0000
Std. Deviation		11,12814

Esim. 5.1.29. Sisarusten lukumäärän keskiarvo,  
keskiarvo frekvenssijakaumasta

<u>Sisarusten lukumäärä</u>	<u>Frekv.</u>
0	56
1	39
2	13
3	10
4	5
5	2
6	1
<hr/>	
Yht.	126

$\bar{x} = (0 \cdot 56 + 1 \cdot 39 + 2 \cdot 13 + 3 \cdot 10 + 4 \cdot 5 + 5 \cdot 2 + 6 \cdot 1) / 126 = 1,04$ . Aineistossa lapsella on keskimäärin 1,04 sisarusta.

Muuttujan  $x$  keskistäminen

$$x_i - \bar{x}, i = 1, 2, \dots, n$$

Keskiarvo ryhmäkeskiarvojen avulla

$$\bar{x} = (n_1\bar{x}_1 + n_2\bar{x}_2 + \dots + n_k\bar{x}_k)/(n_1 + n_2 + \dots + n_k)$$

Esim. 5.1.20. Lepopulssin keskiarvo miehillä ja naisilla

Pulssi				
Sukupuoli	Mean	Std. Deviation	N	Median
Mies	70,6364	11,27684	44	70,0000
Nainen	77,5556	9,80800	36	77,5000
Total	73,7500	11,12814	80	74,0000

$$73,7500 = (44 \cdot 70,6364 + 36 \cdot 77,5556) / 80$$



Esim. 5.1.21. Voidaanko sadon määrää selittää käytetyllä viljelymenetelmällä?

satomäärä = selitettävä, riippuva muuttuja (y)  
 viljelymenetelmä = selittävä, riippumaton muuttuja (x)

Esim. 5.1.24. Voidaanko neliöhintaa selittää sijainnilla?

y = neliöhinta

x = sijainti

Neliöhinta			
Sijainti	Mean	N	Std. Deviation
Hervanta	1752,6063	127	456,78817
Kaleva	2569,1688	77	394,44750
Keskusta	3118,4512	164	712,47930
Tesoma	1593,3333	54	484,13026
Total	2412,0213	422	855,94477

Esim. 5.1.17. Keskiluvut symmetristen ja vinojen jakaumien tapauksessa,  
[http://www.sis.uta.fi/tilasto/tiltp7/moniste\\_4.pdf](http://www.sis.uta.fi/tilasto/tiltp7/moniste_4.pdf)

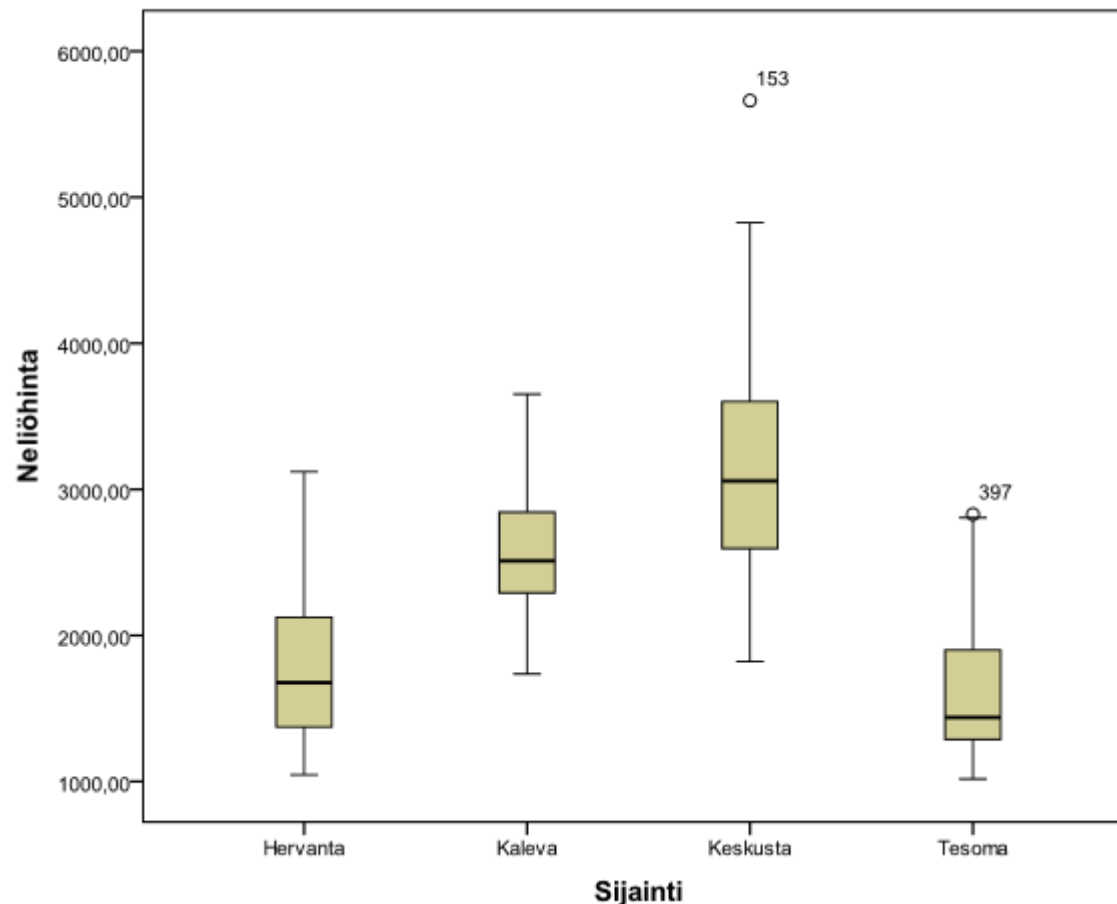
Muita sijainnin tunnuslukuja

- Ala- ja yläkvartiili, muut fraktiilit

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/luentorunko.pdf#page=29>

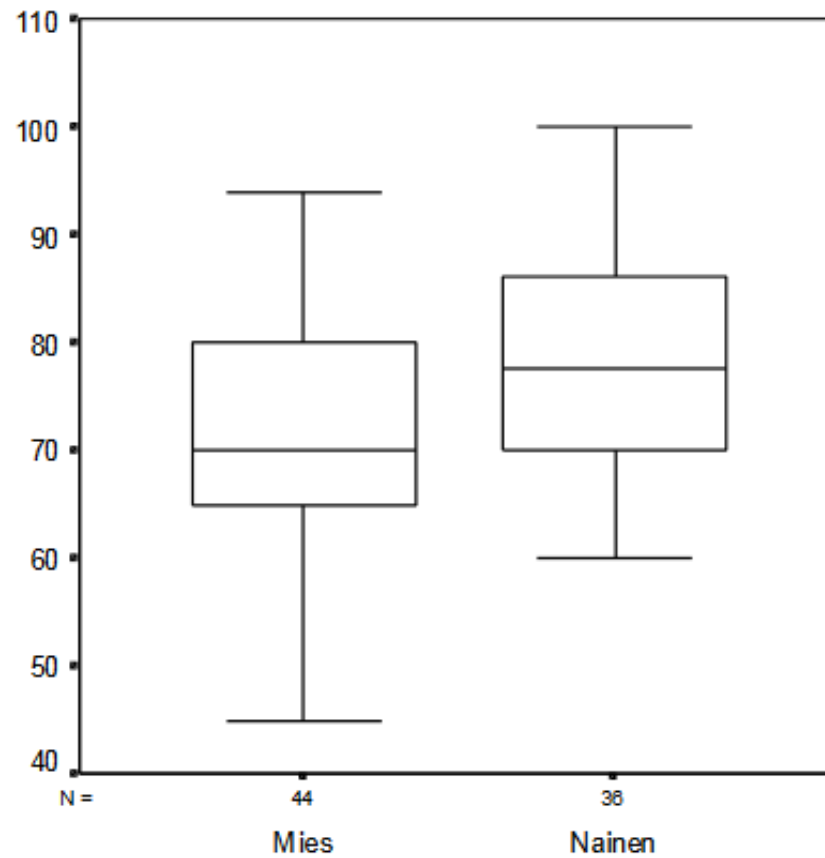
Laatikko-jana –kuvio muodostetaan ala- ja yläkvartiilin sekä mediaanin avulla.

## Esim. 5.1.25. Neliöhinta sijainnin mukaan



Laatikoissa keskimäinen viiva on mediaanin kohdalla, ylimmäinen yläkvartiilin ja alimmainen alakvartiilin kohdalla. Janojen ylä- ja alarajat ovat suurimpien ja pienimpien arvojen kohdalla (ellei ole kovin poikkeavia arvoja).

## Esim. 5.1.26. Lepopulssi miehillä ja naisilla



Esim. Cooperin testin tulokset luokka-asteittain, CTESTI-aineisto, ks. muuttujien esittely luento 7.3.

