

MTTTP1, luento 20.9.2018

## KERTAUSTA JA TÄYDENNYSTÄ

Tunnusluvut

### 1) Sijainnin tunnuslukuja

Keskilukuja

- moodi (Mo)
- mediaani (Md)
- keskiarvo, kaava (1)

Muita sijainnin tunnuslukuja

- ala- ja yläkvartiili, muut fraktiilit

## 2) Vaihtelua mittaavia tunnuslukuja

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/luentorunko.pdf#page=32>

- Varianssi, kaava (2)

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/kaavat.pdf>

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Mittaa muuttujan arvojen keskittymistä keskiarvon ympärille, sallittu kvantitatiivisen muuttujan yhteydessä.

- Keskihajonta, kaava (3)

$$s_x = \sqrt{s_x^2}$$

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

[http://www.sis.uta.fi/tilasto/mhttp1/syksy2018/esimerkit\\_kaavoihin.pdf](http://www.sis.uta.fi/tilasto/mhttp1/syksy2018/esimerkit_kaavoihin.pdf)

Esim. 5.1.28. Otosvarianssin laskeminen tenttipisteistä  
95, 86, 78, 90, 62, 73, 89

$$s^2 = ((95-81,9)^2 + (86-81,9)^2 + \dots + (89-81,9)^2) / (7-1) = 132,5$$
$$s = 11,5.$$

Esim. 5.1.35. Normaalijakauma

[http://www.sis.uta.fi/tilasto/tiltp7/moniste\\_6.pdf](http://www.sis.uta.fi/tilasto/tiltp7/moniste_6.pdf)

Esim. Laskuri <http://vassarstats.net/>, jossa keskiarvon ja varianssin lasku <http://vassarstats.net/vsmisc.html>

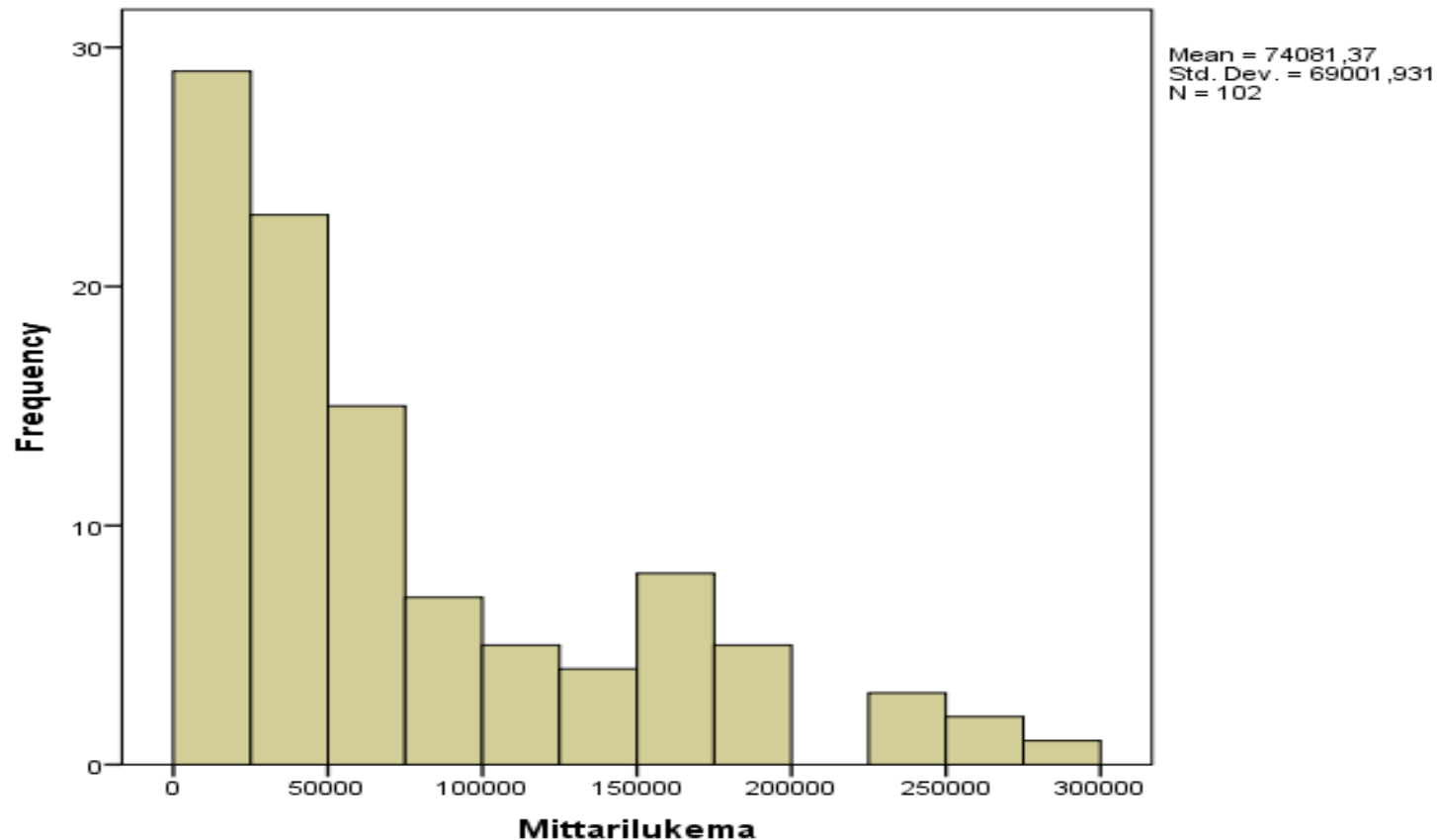
### 3) Muita tunnuslukuja

Voidaan mitata esim. jakauman vinoutta ja huipukkuutta

Esim.

Sivulta \_

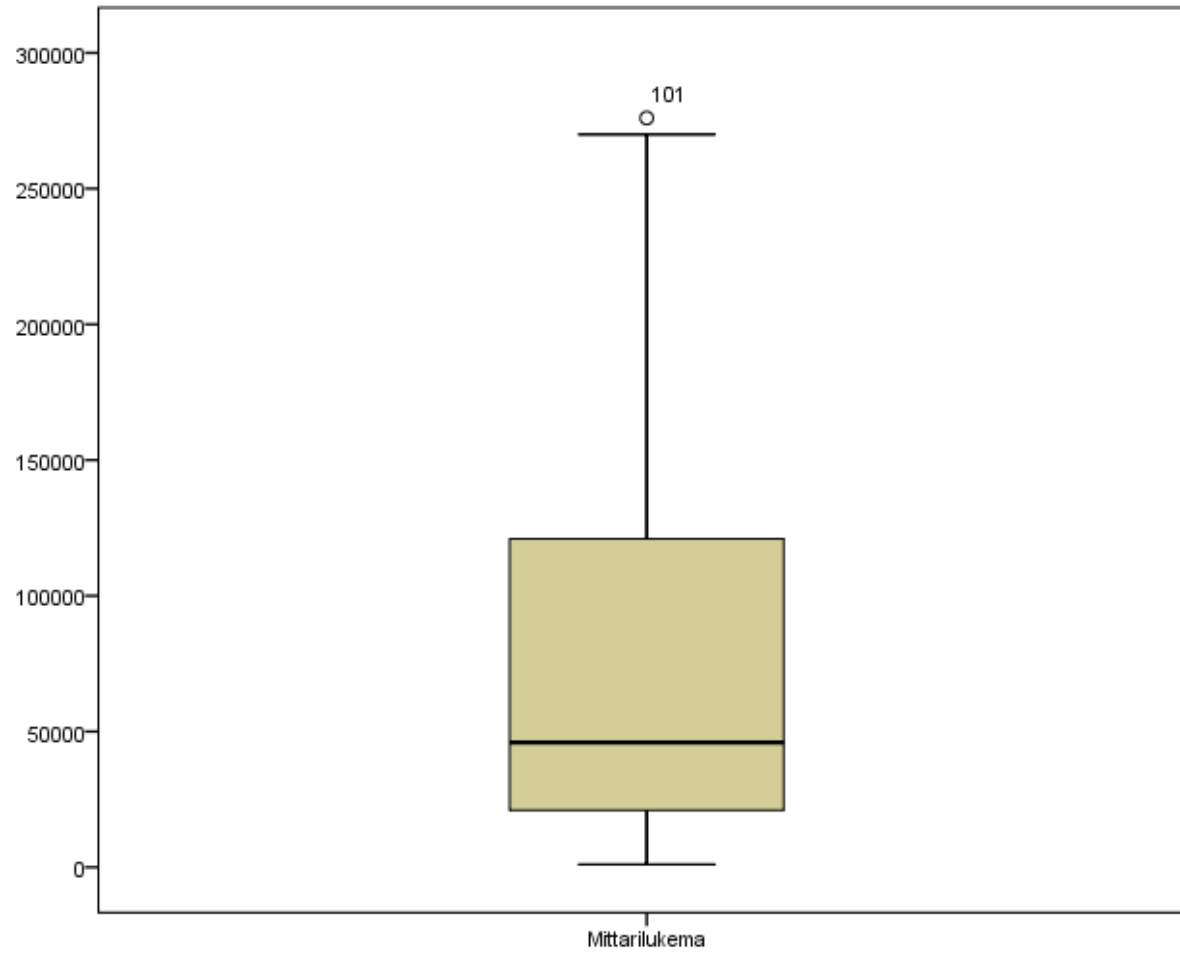
<https://coursepages.uta.fi/mhttp1/esimerkkiaineistoja/>  
aineisto Toyota.sav, jossa Toyota Avensis -farmariautoja  
vuosilta 2007 - 2009, oikotie.fi -sivustolta 2.2.2010.



## Statistics

### Mittarilukema

N	Valid	102
	Missing	0
Mean (keskiarvo)		74081,37
Median (mediaani)		46000,00
Std. Deviation (keskihajonta)		69001,931
Percentiles	25	21000,00
	50	46000,00
	75	121250,00



**Moottorin tilavuus**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,6	5	4,9	4,9	4,9
	1,8	25	24,5	24,5	29,4
	2,0	55	53,9	53,9	83,3
	2,2	17	16,7	16,7	100,0
	Total	102	100,0	100,0	

Taulukosta laskettuna:

$$Md = 2,0,$$

$$Mo = 2,0,$$

Keskiarvo

$$(1,6 \cdot 5 + 1,8 \cdot 25 + 2,0 \cdot 55 + 2,2 \cdot 17) / 102 = 1,965$$



## Statistics

### Moottorin tilavuus

N	Valid	102
	Missing	0
Mean		1,965
Median		2,000
Std. Deviation		,1526
Percentiles	25	1,800
	50	2,000
	75	2,000

## Väri

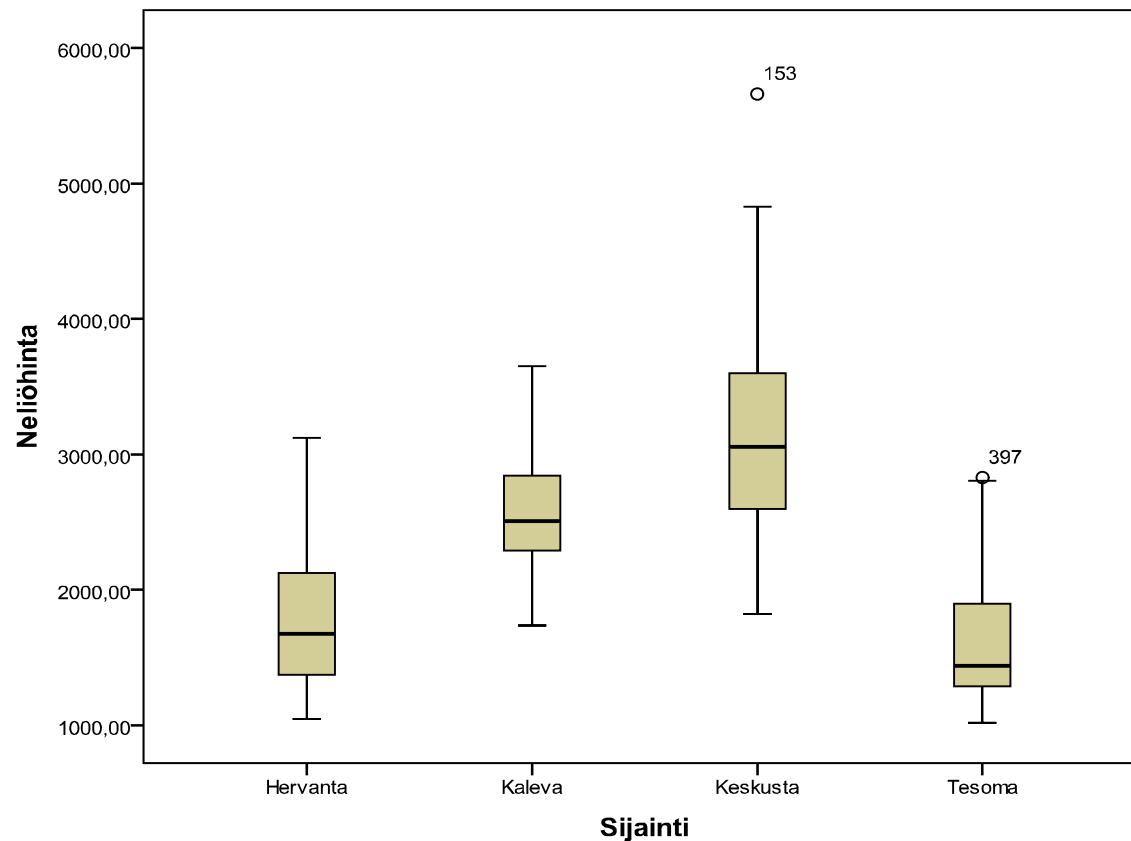
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Harmaa	10	9,8	12,7	12,7
	Hopea	43	42,2	54,4	67,1
	Musta	15	14,7	19,0	86,1
	Punainen	5	4,9	6,3	92,4
	Sininen	2	2,0	2,5	94,9
	Ruskea	4	3,9	5,1	100,0
	Total	79	77,5	100,0	
Missing	System	23	22,5		
Total		102	100,0		

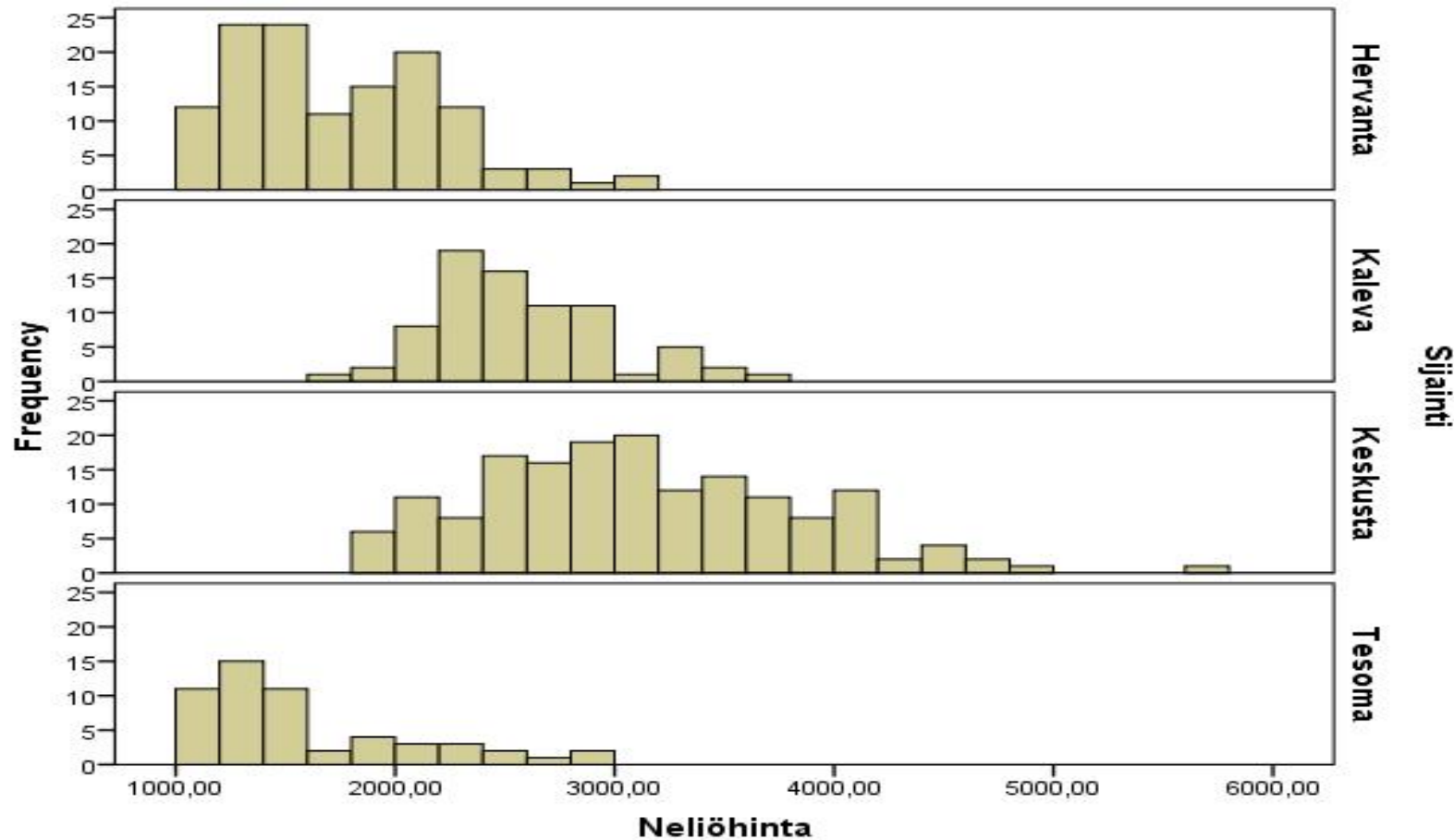
Hopeisia on eniten (54,4 %).

Esim. 5.1.24, 5.1.25, 5.1.27

Myytyjen kerrostaloasuntojen neliöhintoja Tampereella,  
sivun

[https://coursepages.uta.fi/mttp1/esimerkkiaineistoja/](https://coursepages.uta.fi/mttp1/esimerkkiaineistoja/aineisto_Tre_myydyt_asunnot_2012.sav)  
aineisto Tre\_myydyt\_asunnot\_2012.sav





	<u>Hervanta</u>	<u>Kaleva</u>	<u>Keskusta</u>	<u>Tesoma</u>
Keskiarvo	1753	2569	3118	1593
Mediaani	1677	2510	3058	1438
Keskihajonta	457	394	712	484

Esim. 5.1.30. Lisäaineen vaikutus teräksen kovuusindeksiin

Tuote-erä	1	2	3	4	5	6	7	8	9	10
Lisäaine A	22	26	29	22	31	34	31	20	33	34
Lisäaine B	27	25	31	27	29	41	32	27	32	34
Erotus	-5	1	-2	-5	2	-7	-1	-7	1	0

Laskurilla <http://vassarstats.net/vsmisc.html> erotuksen keskiarvon ja varianssin lasku

Lineaarinen muunnos muuttujalle  $x$

$$y_i = ax_i + b, \quad i = 1, 2, \dots, n$$

- vaikutus keskiarvoon

$$\bar{y} = a\bar{x} + b$$

mittayksikkö vaikuttaa keskiarvon

- vaikutus keskihajontaan

$$s_y = |a|s_x$$

mittayksikkö vaikuttaa keskihajontaan

Muuttujan  $x$  standardointi

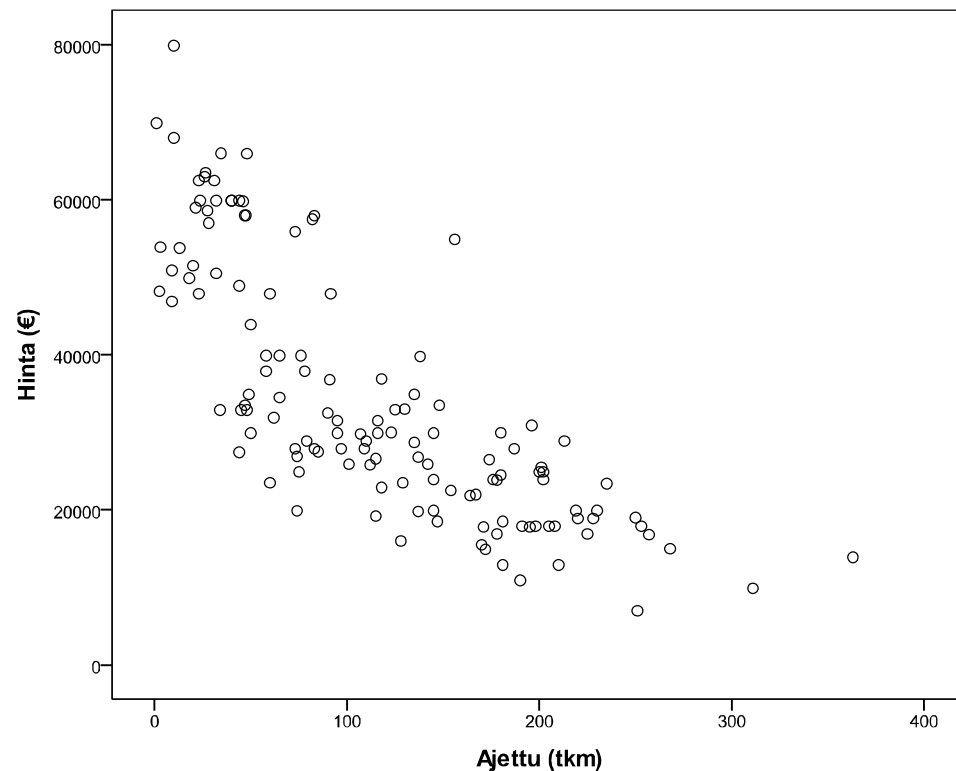
$$z_i = \frac{x_i - \bar{x}}{s}$$

## 5.2 Kaksiulotteinen jakauma

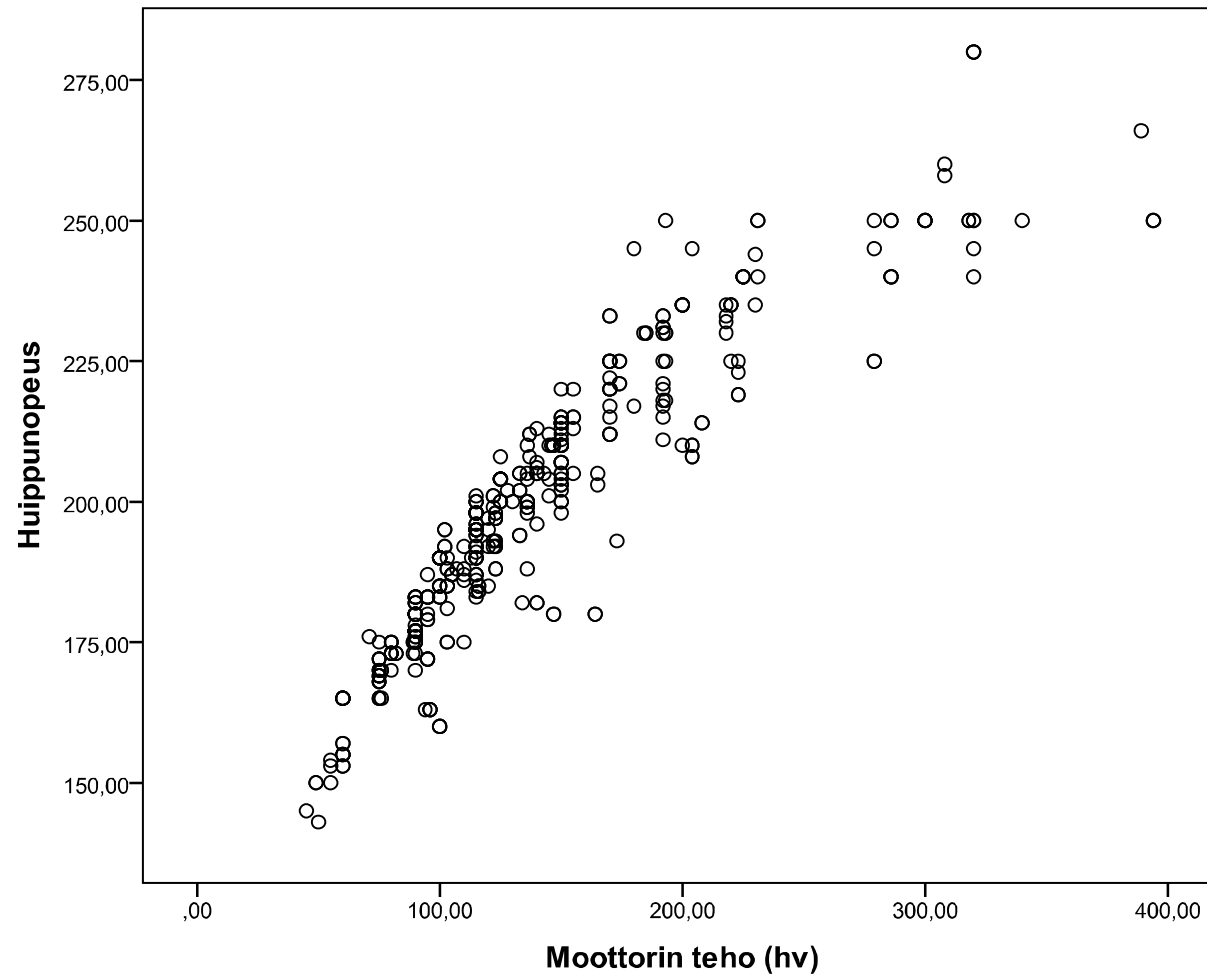
### 5.2.1 Pisteparvi

Esim. Auton hinta ja ajetut kilometrit, aineistona Audi A6 –henkilöautoja sivulta

<https://coursepages.uta.fi/mttp1/esimerkkiaineistoja/>



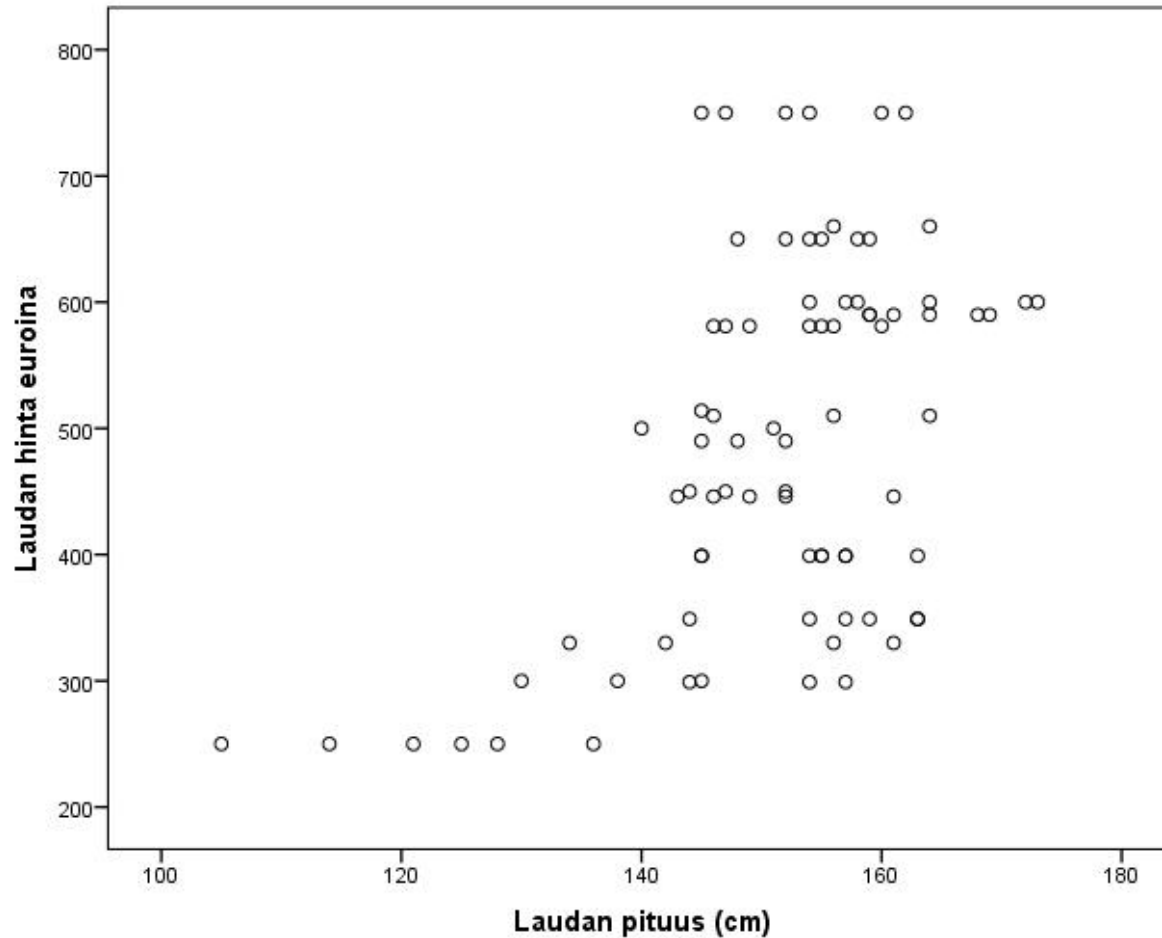
Esim. Auton huippunopeus ja teho, aineistona auto94.sav  
(mikroluokissa)



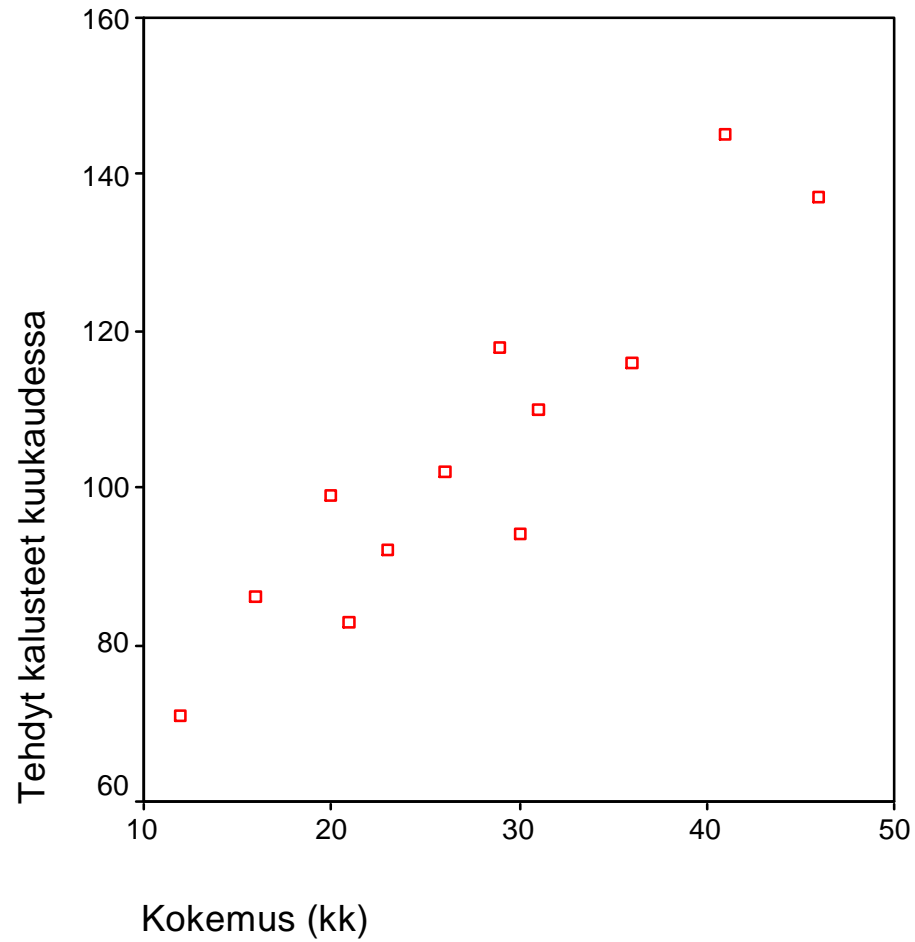


Esim. Lumilaudan hinta ja pituus, aineistona Lumilaudat.sav sivulta

<https://coursepages.uta.fi/mhttp1/esimerkkiaineistoja/>



## Esim. 5.2.2. Tehdyt kalusteet ja työntekijän kokemus,



Kaksiulotteisessa jakaumassa tarkastellaan kahta muuttujaa samanaikaisesti. Tutkitaan muuttujien välisiä riippuvuussuhteita.

Pisteparvi on graafinen esitys, jos selitettävä muuttuja kvantitatiivinen.

## 5.2.2 Ristiintaulukko

Esim. Miesten, naisten ja lasten lumilaudat valmistusmaittain, aineistona Lumilaudat.sav sivulla <https://coursepages.uta.fi/mttp1/esimerkkiaineistoja/>

**Kenelle lauta on tarkoitettu \* Merkki luokiteltuna maan mukaan Crosstabulation**

Count		Merkki luokiteltuna maan mukaan		
		Kanada	USA	Total
Kenelle lauta on tarkoitettu	Miehet	22	29	51
	Naiset	11	7	18
	Lapset	7	4	11
Total		40	40	80

## Esim. 5.2.5. Automallien koot valmistusmaittain

		Valmistusmaa			
		USA	Eur.	Japani	
	Iso	36	4	2	42
Koko	Kesk.	53	17	54	124
	Pieni	26	19	92	137
		115	40	148	303

Koko-muuttujan ehdolliset prosenttijakaumat eli koko-muuttujan prosentuaaliset jakaumat erikseen valmistusmaittain:

		Valmistusmaa		
		USA	Eur.	Japani
	Iso	31,30	10,00	1,35
Koko	Kesk.	46,09	42,50	36,49
	Pieni	22,61	47,50	62,16
		100,00	100,00	100,00

## Esim. 5.2.6. Markkinointisuunnitelma tavaratalon koon mukaan

		Suunnitelma		
		on	ei	yht.
Henkilöstö- määrä	alle 100	13	10	23
	100 – 500	18	12	30
	yli 500	32	6	38

Markkinointisuunnitelman olemassaolon ehdolliset prosenttijakaumat (koon mukaan)

		Suunnitelma	
		on	ei
Henkilöstö- määrä	alle 100	56,6	43,5
	100 – 500	60,0	40,0
	yli 500	84,2	15,8

Esim. Miesten, naisten ja lasten lumilaudat valmistusmaittain, aineistona Lumilaudat.sav sivulla <https://coursepages.uta.fi/mhttp1/esimerkkiaineistoja/>

**Kenelle lauta on tarkoitettu \* Merkki luokiteltuna maan mukaan Crosstabulation**

% within Merkki luokiteltuna maan mukaan

		Merkki luokiteltuna maan mukaan		
		Kanada	USA	Total
Kenelle lauta on tarkoitettu	Miehet	55,0%	72,5%	63,7%
	Naiset	27,5%	17,5%	22,5%
	Lapset	17,5%	10,0%	13,8%
Total		100,0%	100,0%	100,0%



Esim. Asunnon kunto sijainnin mukaan, aineistona

Tre\_myydyt\_asunnot\_2010 sivulla

<https://coursepages.uta.fi/mhttp1/esimerkkiaineistoja/>

### Huoneiston kunto \* Sijainti Crosstabulation

Count		Sijainti				Total
		Keskusta	Kaleva	Hervanta	Härmälä	
Huoneiston kunto	huono	3	1	1	0	5
	hyvä	102	31	91	60	284
	tydyttävä	39	21	37	7	104
Total		144	53	129	67	393

Onko tässä korjattavaa?

## Ristiintaulukko yleisesti

		$x$				
		$E_1$	$E_2$	...	$E_J$	
	$F_1$	$f_{11}$	$f_{12}$	...	$f_{1J}$	$f_{1\cdot}$
	$F_2$	$f_{21}$	$f_{22}$	...	$f_{2J}$	$f_{2\cdot}$
$y$	·	·	·		·	·
	·	·	·		·	·
	·	·	·		·	·
	$F_I$	$f_{I1}$	$f_{I2}$	...	$f_{IJ}$	$f_{I\cdot}$
		$f_{\cdot 1}$	$f_{\cdot 2}$	...	$f_{\cdot J}$	$f_{\cdot\cdot} = n$

Tutkitaan  $y$ :n riippuvuutta  $x$ :stä vertaamalla  $y$ :n ehdollisia prosenttijakaumia.

## 6 AIKASARJOISTA

Määritelmä

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/luentorunko.pdf#page=51>

Graafinen esitys

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/luentorunko.pdf#page=51>

Esimerkkejä luentomonisteen esimerkeissä 6.1.1.- 6.1.6.

## Harjoitustyön riippuvuustarkastelut

Ks.

<http://www.sis.uta.fi/tilasto/mhttp1/syksy2018/htyop118.pdf#page=4>

### Riippuvuustarkastelu 1

Tarkastelu, jossa  $y$  (selitettävä) on kvantitatiivinen ja  $x$  (selittäjä) kvalitatiivinen. Menetelminä laatikko-jana-kuvio, ryhmäkeskiarvot, muut tarvittavat tunnusluvut sekä päättely riippumattomien otosten  $t$ -testi avulla (käytetään näitä jokaista). Jos ryhmiä enemmän kuin kaksi, niin  $t$ -testissä voi tarkastella niistä kahta tai yhdistellä ryhmiä sopivasti.

Ks. luentomonisteen esimerkit 5.1.25, 5.1.26, 7.7.9, 7.7.11.

## Riippuvuustarkastelu 2

Tarkastelu, jossa sekä  $y$  että  $x$  kvalitatiivisia (kvantitatiiviset voi myös luokitella), selitettävä muuttuja eri kuin riippuvuustarkastelussa 1. Menetelmänä ristiintaulukko ja  $\chi^2$ -riippumattomuustesti.

Ks. luentomonisteen esimerkit 7.7.6, 7.7.7, 7.7.8.