

MTTTP1, luento 19.3.2019

KERTAUSTA

- Varianssi, kaava (2)

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

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

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

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

$$\bar{x} = (-5 + 1 - 2 - 5 + 2 - 7 - 1 - 7 + 1 + 0) / 10 = -2,3$$

Lisäaineiden vaikutuksessa teräksen kovuuteen ei eroja, jos erotuksen keskiarvo riittävän lähellä nolaa. Päättely testauksen avulla.

$$s^2 = ((-5 + 2,3)^2 + (1 + 2,3)^2 + \dots + (0 + 2,3)^2) / (10 - 1) = 11,79$$

$$s = 3,4.$$

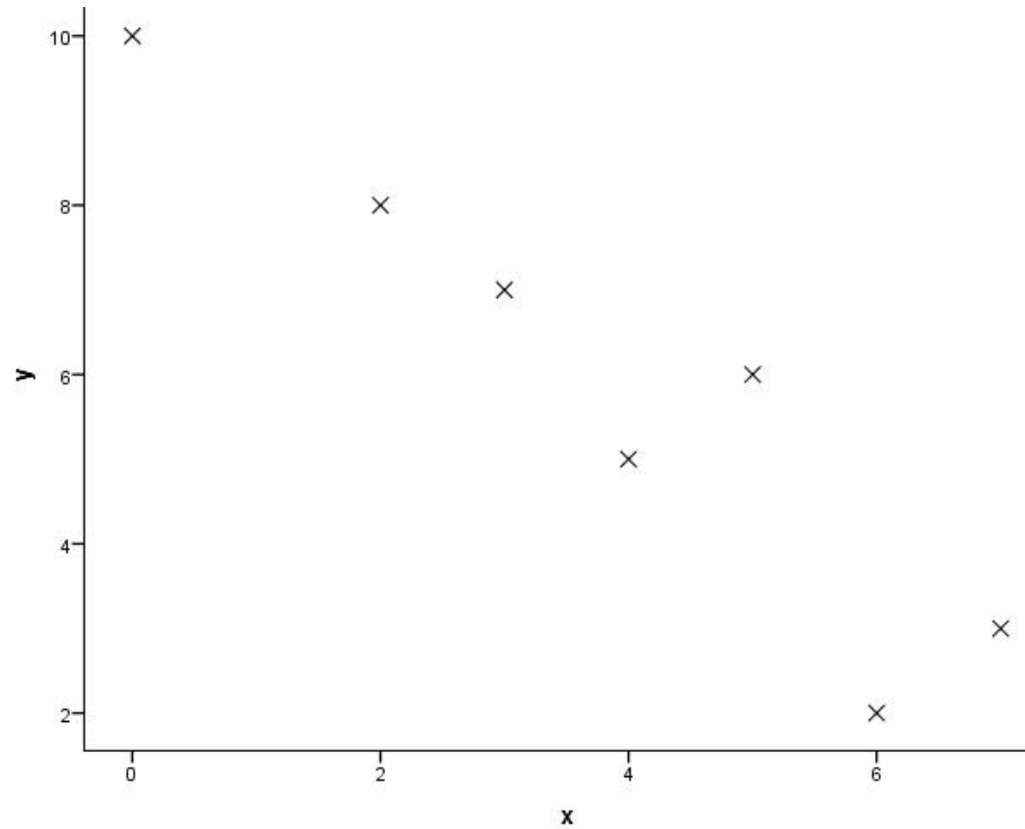
- Muuttujan x standardointi

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

Esim. 5.1.32. Opiskelija osallistui tentteihin A ja B saaden pisteet 25 ja 24. Tentissä A tuloksen keskiarvo oli 20 ja keskihajonta 4. Vastaavat luvut tentissä B olivat 20 ja 2. Kummassa tentissä opiskelija menestyi suhteellisesti paremmin? Standardoidut arvot ovat $5/4$ ja 2 , joten menestyminen tentissä B oli parempi.

- Kaksiulotteinen jakauma

Pisteparvi, graafinen esitys



5.2.2 Ristiintaulukko (jatkuu)

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

Tutkitaan koon riippuvuutta valmistusmaasta vertaamalla koon prosenttijakaumia valmistusmaittain.

On eroja, $p < 0,001$

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	43,4
	100 – 500	60,0	40,0
	yli 500	84,2	15,8

On eroja, $p = 0,031$

Esim. Miesten, naisten ja lasten lumilaudat valmistusmaittain, aineistona Lumilaudat.sav sivulla <https://coursepages.uta.fi/mtt1/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

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%

Ei eroja, $p = 0,263$

Esim. Esim. Asunnon kunto sijainnin mukaan,
 aineistona Tre_myydyt_asunnot_2010 sivulla
<https://coursepages.uta.fi/mhttp1/esimerkkiaineistoja/>

		Sijainti				
		Keskusta	Kaleva	Hervanta	Härmälä	Total
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

		Sijainti				
		Keskusta	Kaleva	Hervanta	Härmälä	Total
Huoneiston kunto	Huono tai tyydyttävä	29,2%	41,5%	29,5%	10,4%	27,7%
	Hyvä	70,8%	58,5%	70,5%	89,6%	72,3%
Total		100,0%	100,0%	100,0%	100,0%	100,0%

On eroja, $p = 0,002$

Esim. Toyota Avensis –farmariautoja,
nelikenttä (2x2-taulukko)

		Vaihteisto		Total
		Automaatti	Manuaali	
Polttoaine	Bensiini	12	29	41
	Diesel	7	51	58
Total		19	80	99

On eroja, $p = 0,032$

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.

5.2.3 Kaksiulotteisen jakauman tunnuslukuja

- Mitataan kahden muuttujan välistä riippuvuuden voimakkuutta
- Ristiintaulukosta kontingenssikerroin
- Kvantitatiivisista muuttujista lineaarisen riippuvuuden voimakkuuden mittari korrelaatiokerroin (r)
- Järjestysasteikollisilla muuttujilla järjestyskorrelaatiokertoimet

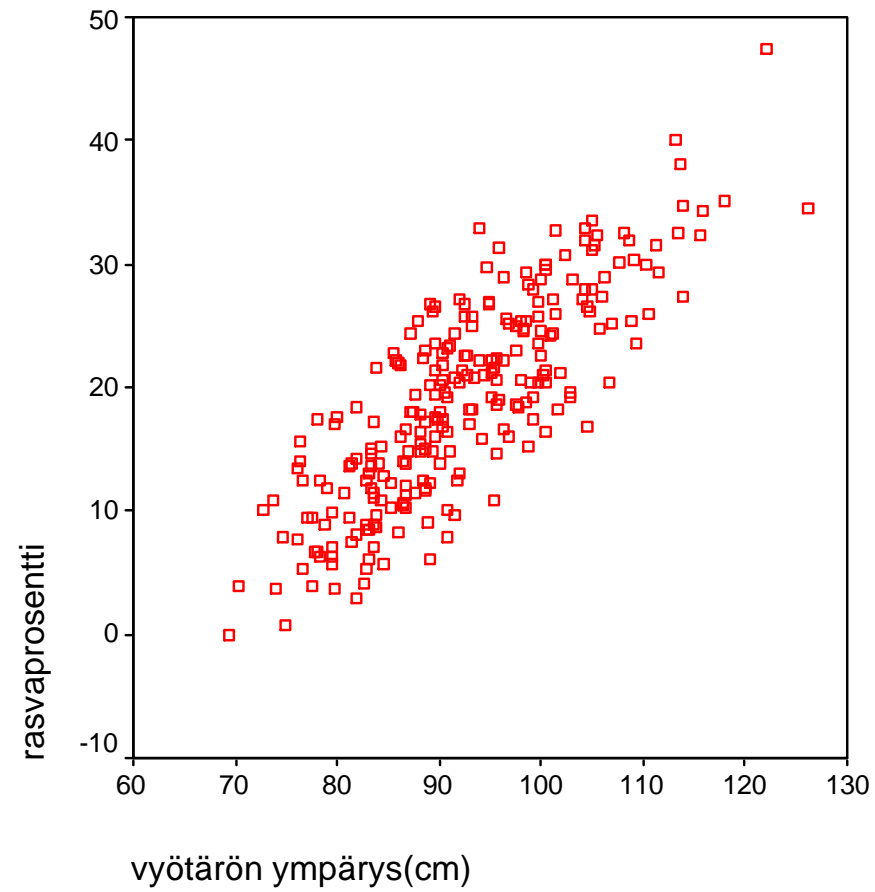
Korrelaatiokerroin r

Mittaa kahden kvantitatiivisen muuttujan välistä lineaarista riippuvuutta, sen voimakkuutta. Mittaa sitä, miten tiiviisti pisteparven pisteet ovat sijoittuneet pisteparveen sovitettavan suoran ympärille.

Ominaisuuksia

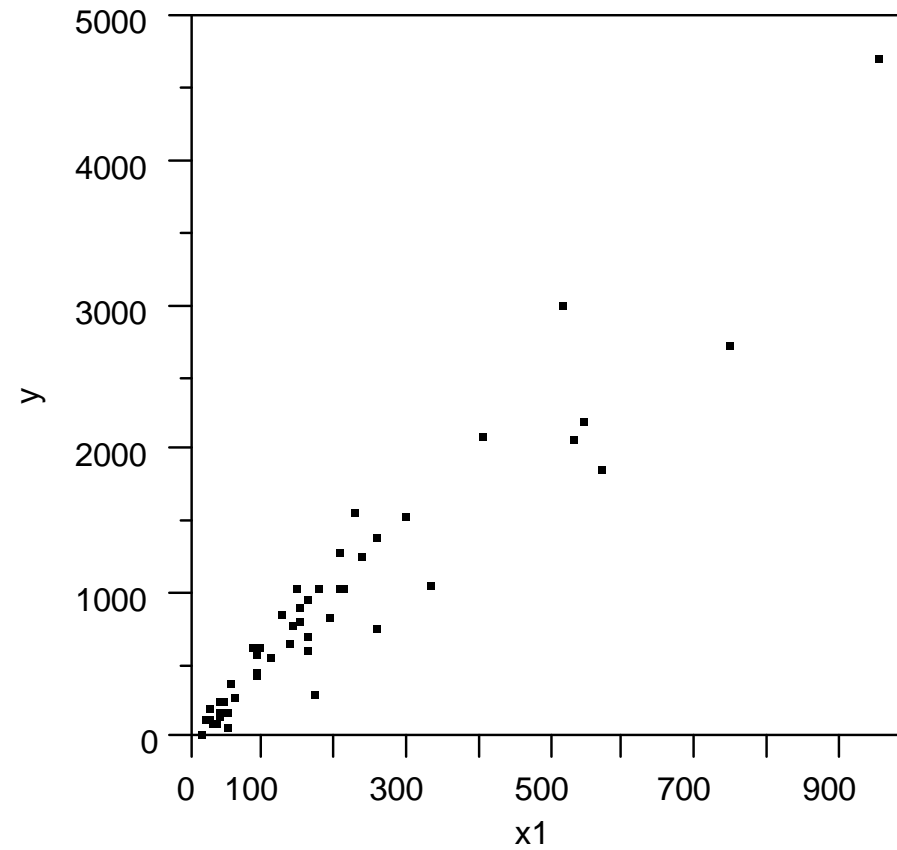
- $-1 \leq r \leq 1$
- $r = 1$, jos kaikki pisteet samalla nousevalla suoralla
- $r = -1$, jos kaikki pisteet samalla laskevalla suoralla
- $r \approx 0$, jos ei lineaarista riippuvuutta

Esim. 5.2.8.



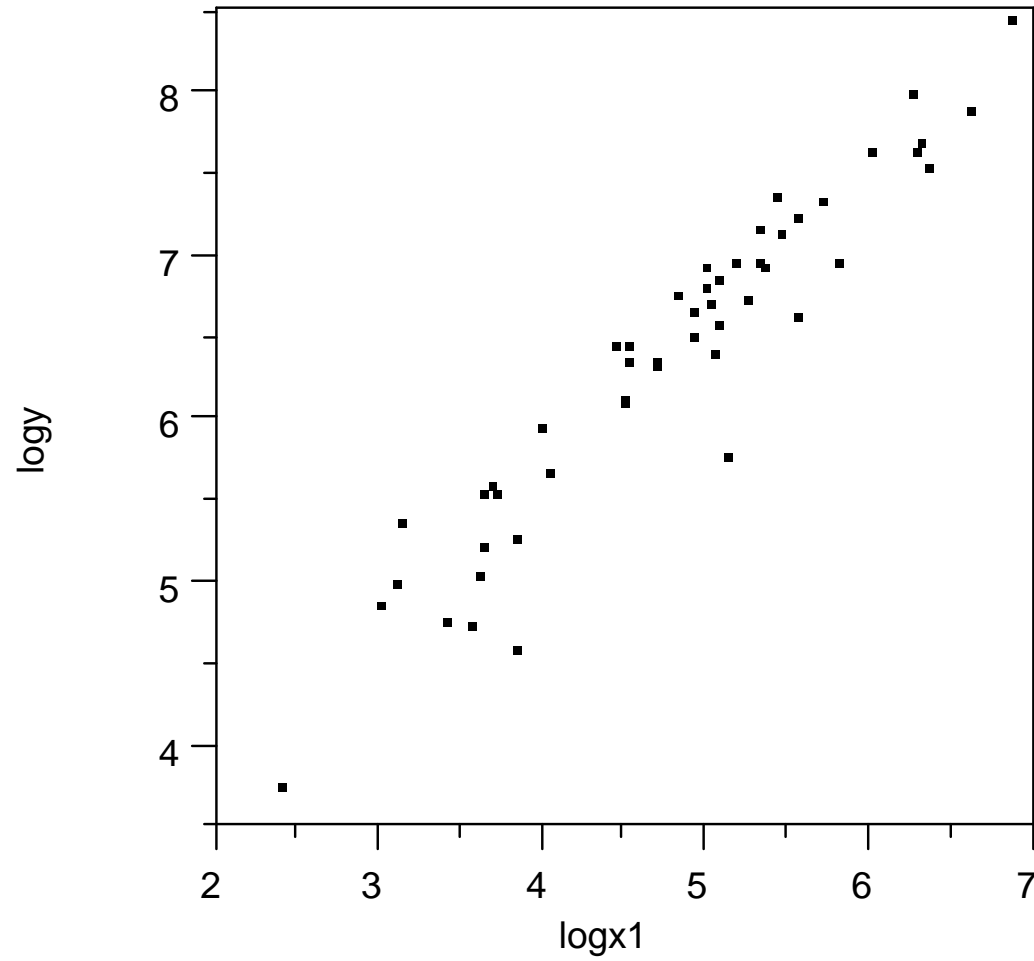
$$r = 0,825$$

Esim. 5.2.10.



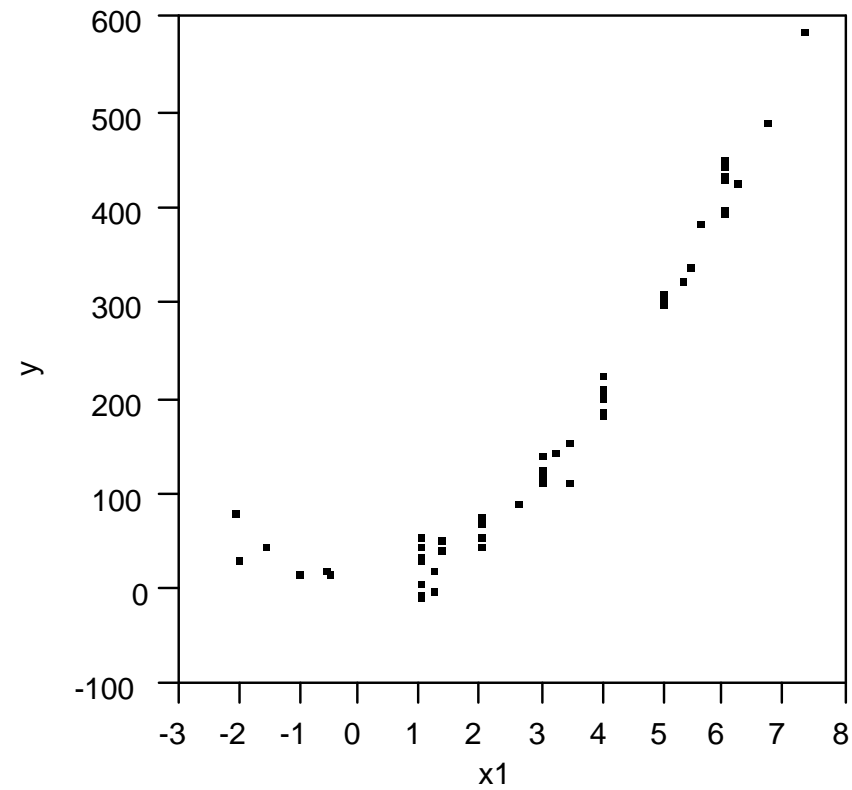
$$r = 0,9559$$

Esim. 5.2.11.

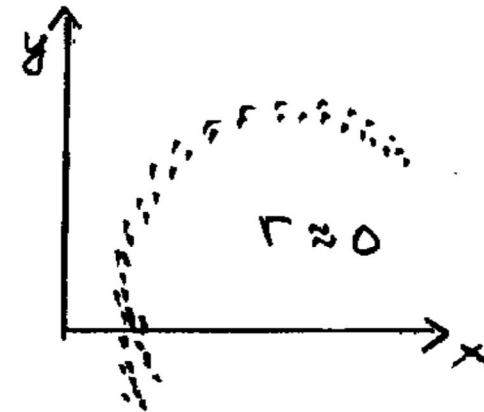
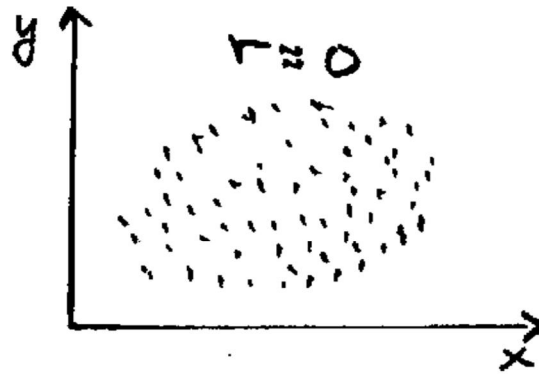
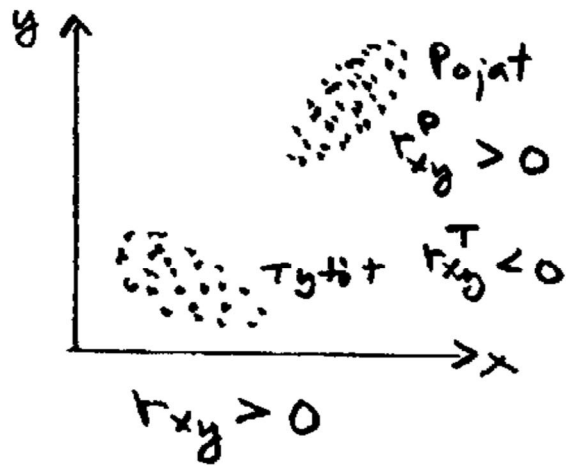
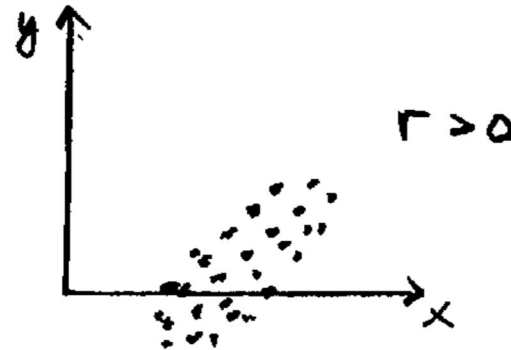


$$r = 0,9537$$

Esim. 5.2.12. Riippuvuutta, joka ei lineaarista.



Esim. Pisteparvia ja arviot korrelaatiokertoimista



$$|r_{xy}| \leq 1$$

Esim. 5.2.13. Pisteparvia ja korrelaatiokertoimia

http://www.sis.uta.fi/tilasto/tiltp7/moniste_8.pdf

Esim. 5.2.17. Korrelaatiomatriisi, CTESTI-aineisto

Correlations

		ika	pituus	paino	cooper
ika	Pearson Correlation	1	,807**	,768**	,399**
	Sig. (2-tailed)		,000	,000	,000
	N	152	152	152	152
pituus	Pearson Correlation	,807**	1	,892**	,236**
	Sig. (2-tailed)	,000		,000	,003
	N	152	153	153	153
paino	Pearson Correlation	,768**	,892**	1	,102
	Sig. (2-tailed)	,000	,000		,210
	N	152	153	153	153
cooper	Pearson Correlation	,399**	,236**	,102	1
	Sig. (2-tailed)	,000	,003	,210	
	N	152	153	153	153

** . Correlation is significant at the 0.01 level (2-tailed).

Korrelaatiokertoimen laskukaava kaavakokoelman kaava
(4)

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

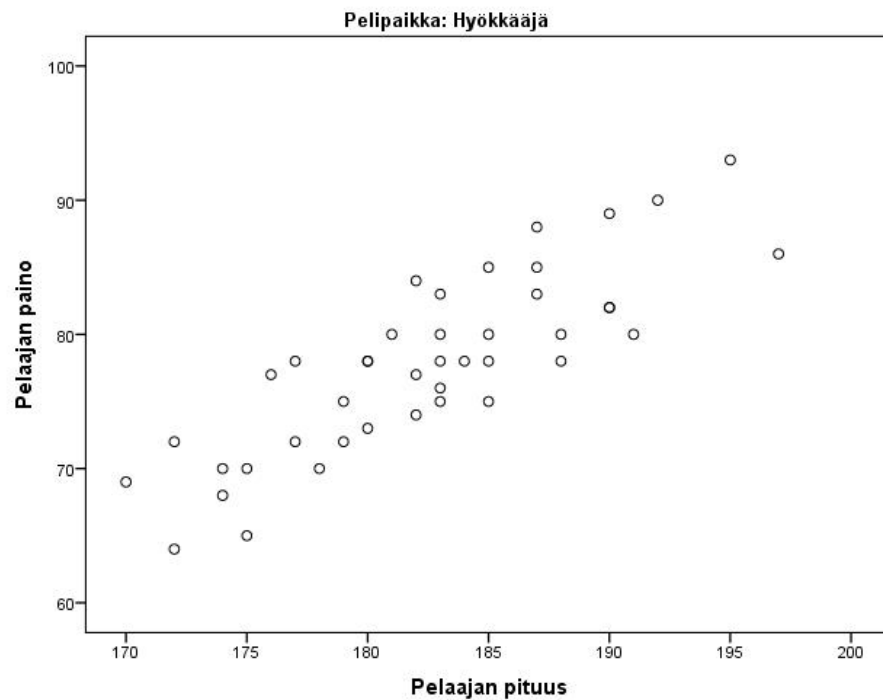
ks. myös

http://www.sis.uta.fi/tilasto/mttp1/syksy2018/esimerkit_kaavoihin.pdf

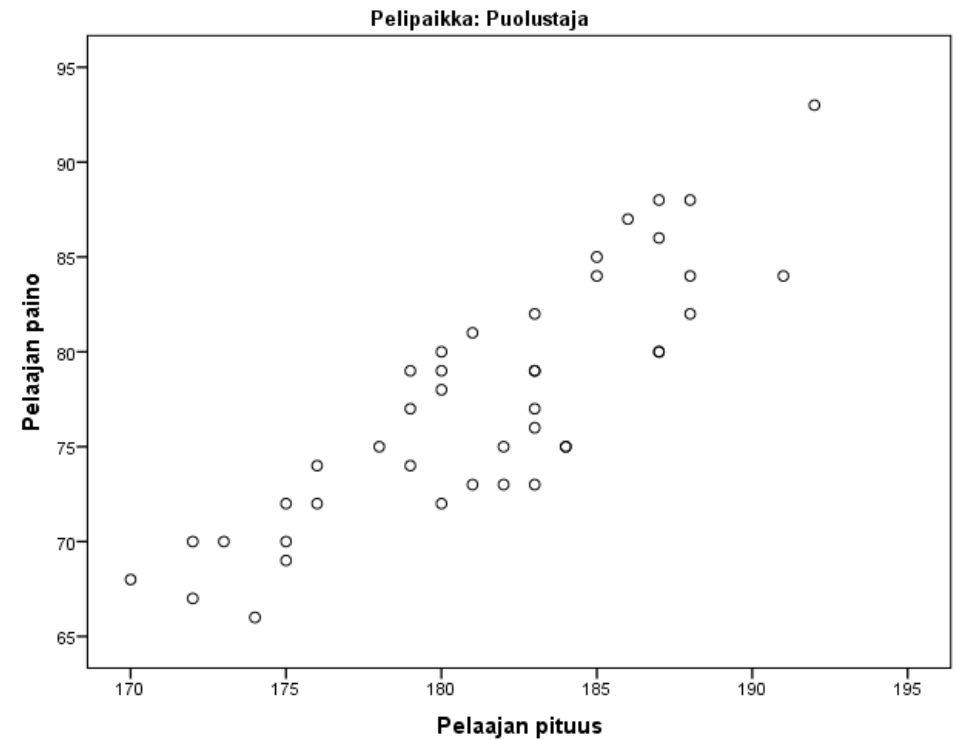
Esim. 5.2.14. Mittayksikön vaihto ei vaikuta korrelaatiokertoimeen, ks. lineaarisen muunnoksen vaikutus korrelaatiokertoimeen

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/luentorunko.pdf#page=47>

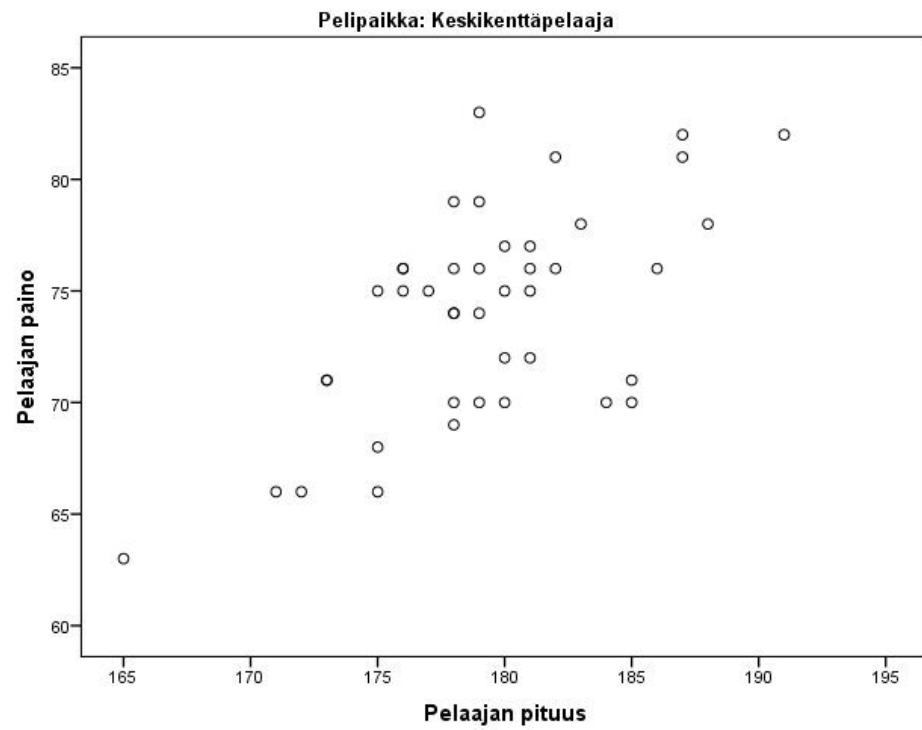
Esim. 5.2.16. Korrelaatiokertoimet pelipaikoittain, ehdolliset korrelaatiot



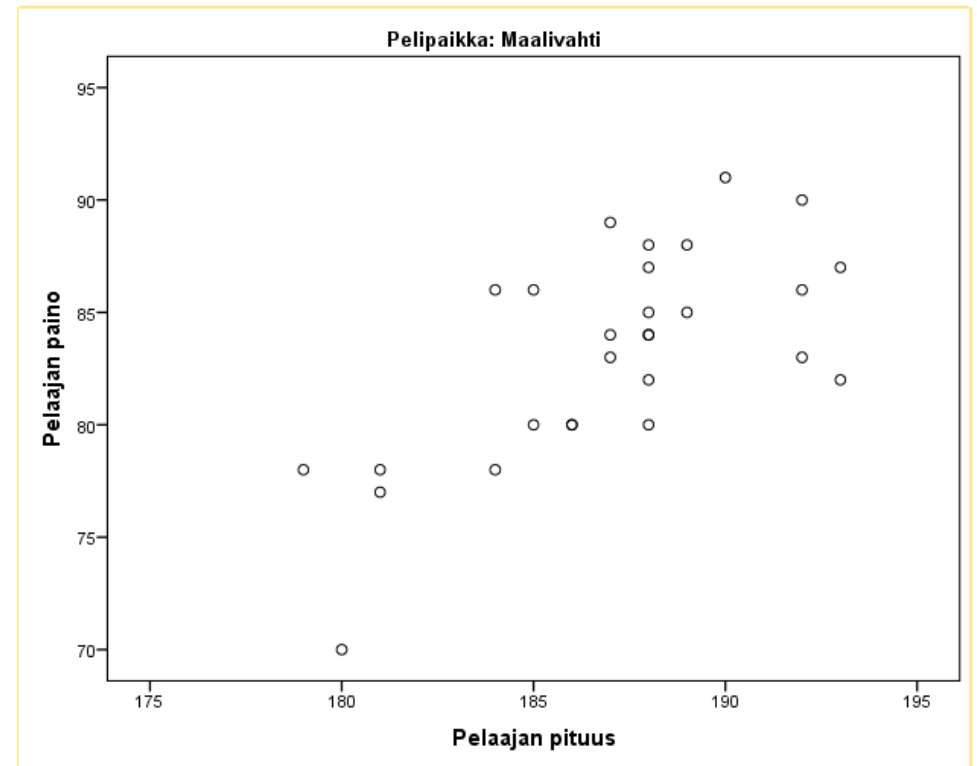
$r = 0,84$, $n = 42$



$r = 0,86$, $n = 42$



$r = 0,62, n = 42$



$r = 0,68, n = 28$

Esim. 5.2.17. Osittaiskorrelaatiokertoimet ikä vakioituna,
CTESTI-aineisto

Correlations

Control Variables			cooper	paino	pituus
ika	cooper	Correlation	1,000	-,349	-,160
		Significance (2-tailed)	.	,000	,050
		df	0	149	149
	paino	Correlation	-,349	1,000	,719
		Significance (2-tailed)	,000	.	,000
		df	149	0	149
	pituus	Correlation	-,160	,719	1,000
		Significance (2-tailed)	,050	,000	.
		df	149	149	0

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

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/htyop118.pdf#page=4>

Riippuvuustarkastelu 1

y (selitettävä) on kvantitatiivinen ja x (selittäjä) kvalitatiivinen

- laatikko-jana-kuvio
- ryhmäkeskiarvot, muut tarvittavat tunnusluvut
- päättely riippumattomien otosten t-testi avulla

Riippuvuustarkastelu 2

y ja x kvalitatiivisia (kvantitatiiviset voi luokitella), selitettävä muuttuja eri kuin riippuvuustarkastelussa 1

- ristiintaulukko
- χ^2 -riippumattomuustesti.