

MTTTA1 Tilastomenetelmien perusteet
Luento 29.1.2019

Kertausta ja täydennystä
 χ^2 -riippumattomuustesti

Ristiintaulukon perusteella riippumattomuuden
testaaminen

H_0 : X ja Y ovat riippumattomia

H_1 : X ja Y ovat riippuvia

Ristiintaulukkoa

		x				
		1	2	...	J	
y	1	f_{11}	f_{12}	...	f_{1J}	$f_{1\cdot}$
	2	f_{21}	f_{22}	...	f_{2J}	$f_{2\cdot}$
	\vdots	\vdots	\vdots		\vdots	\vdots
	I	f_{I1}	f_{I2}	...	f_{IJ}	$f_{I\cdot}$
		$f_{\cdot 1}$	$f_{\cdot 2}$...	$f_{\cdot J}$	n

Jos H_0 on tosi, niin

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \sim \chi^2(I-1)(J-1)$$

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

Nyt H_0 hylätään riskitasolla α , jos

$$\chi_{hav.}^2 > \chi_{\alpha, (I-1)(J-1)}^2$$

Jos $I = 2$ ja $J = 2$ (nelikenttä), niin testisuure voidaan laskea myös kaavalla

$$\chi^2 = \frac{n(f_{11}f_{22} - f_{12}f_{21})^2}{f_{\cdot 1}f_{\cdot 2}f_{1\cdot}f_{2\cdot}}$$

Esim. 3.2.3 Naisten ja miesten
tenttimenestyminen

<http://www.sis.uta.fi/tilasto/mttta1/kevat2019/luentorunko.pdf#page=31>

	Miehet	Naiset	Yht.
Hylätty	34	15	49
Hyväksytty	59	23	82
Yht.	93	38	131

H_0 : ei riippuvuutta

$$\chi_{hav.}^2 = \frac{(34 \cdot 23 - 59 \cdot 15)^2 \cdot 131}{93 \cdot 38 \cdot 49 \cdot 82} = 0,09787 < 3,84 = \chi_{0,05;1}^2$$

H_0 hyväksytään, ei riippuvuutta.

Esim. Tutkimuksessa vertailtiin erään kasvaimen yleisyyttä kahdella rottalajilla A ja B. Valittiin satunnaisesti molemmista ryhmistä 100 samanikäistä rottaa. Rotat pidettiin samankaltaisissa olosuhteissa vuoden ajan. Vuoden seurannan jälkeen kasvain löytyi 25:ltä lajin A rotalta ja 15:ltä lajin B rotalta. Onko kasvaimen yleisyys samanlaista molemmilla lajeilla?

H_0 : ei riippuvuutta

	Laji A	Laji B	
On kasvain	25	15	40
Ei kasvainta	75	85	160
	100	100	200

$$\chi_{hav.}^2 = \frac{(25 \cdot 85 - 75 \cdot 15)^2 \cdot 200}{100 \cdot 100 \cdot 40 \cdot 160} = 3,125 < 3,84 = \chi_{0,05;1}^2$$

H_0 hyväksytään, yleisyys samanlaista.

$$P(\chi_1^2 > 3,125) = 0,0771,$$

ks.

http://onlinestatbook.com/2/calculators/chi_square_prob.html

Laskureita

http://www.physics.csbsju.edu/stats/contingency_NROW_NCOLUMN_form.html

<http://vassarstats.net/newcs.html>

Missä mennään?

Menetelmien valinnasta

<http://www.fsd.uta.fi/menetelmaopetus/menetelma/menetelmatyypit.html>

Luku 4

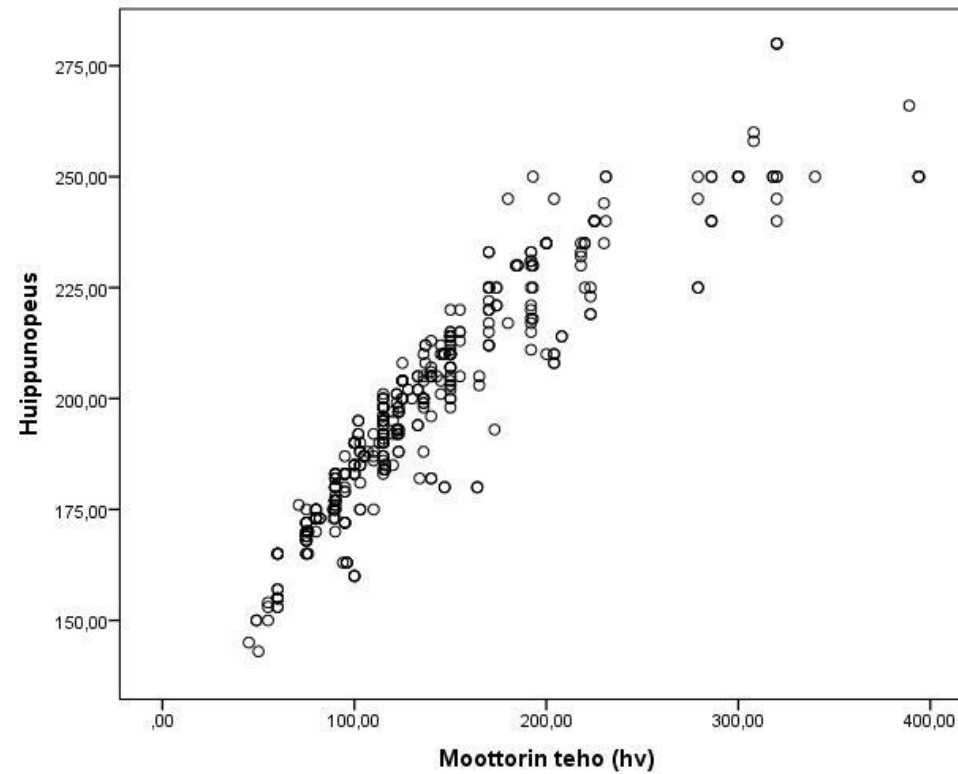
Regressioanalyysi

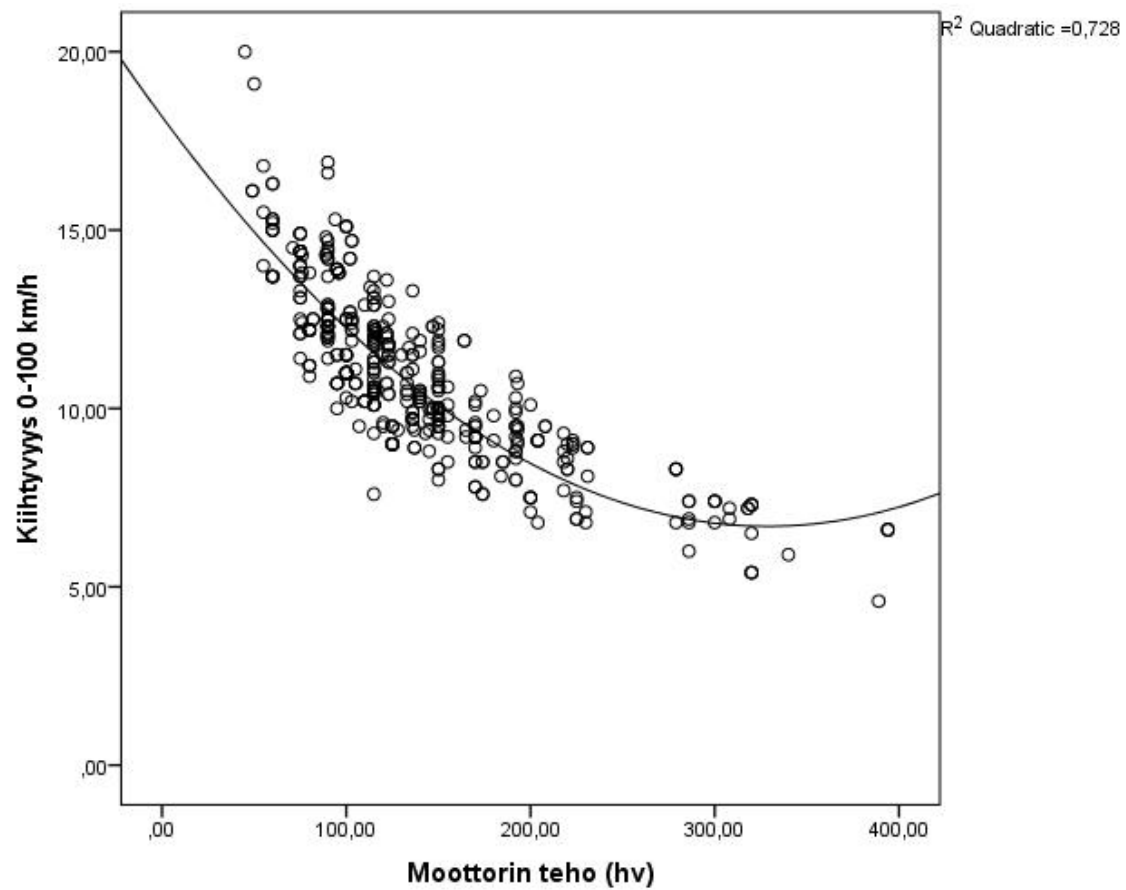
Voidaanko y :n vaihtelua selittää samanaikaisesti useammalla muuttujalla?

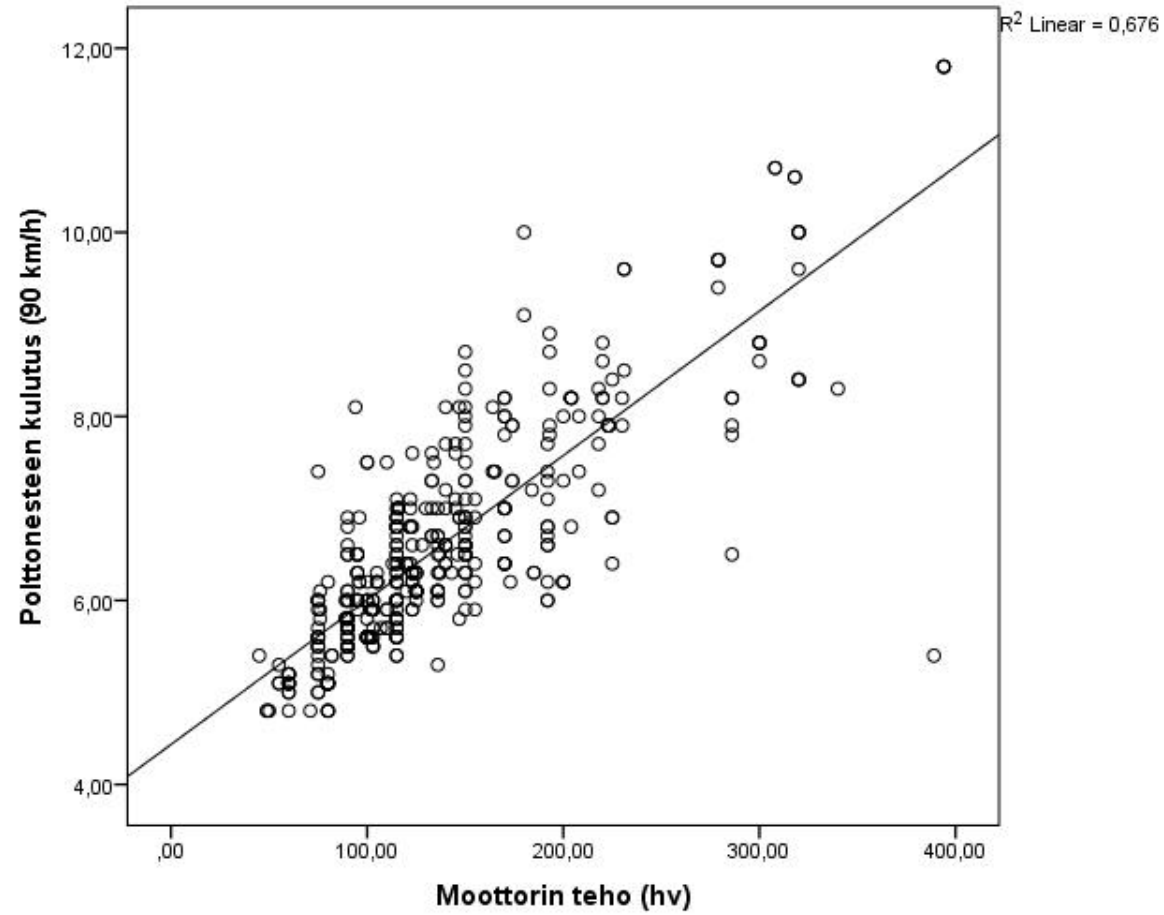
Voidaanko tätä riippuvuutta mallintaa?

Tarkastellaan tilanteita, joissa sekä selitettävä että selittäjät ovat kvantitatiivisia.

Esim. Erilaisia pisteparvia, tilastoyksikkönä auto







Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,822 ^a	,676	,675	,69270

a. Predictors: (Constant), Moottorin teho (hv)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	4,435	,081		54,873	,000
	Moottorin teho (hv)	,016	,001	,822	29,744	,000

a. Dependent Variable: Polttonesteen kulutus (90 km/h)

Malli

$$\text{Kulutus} = \beta_0 + \beta_1 \text{Teho} + \varepsilon$$

Estimoidaan mallin parametrit β_0 ja β_1 .

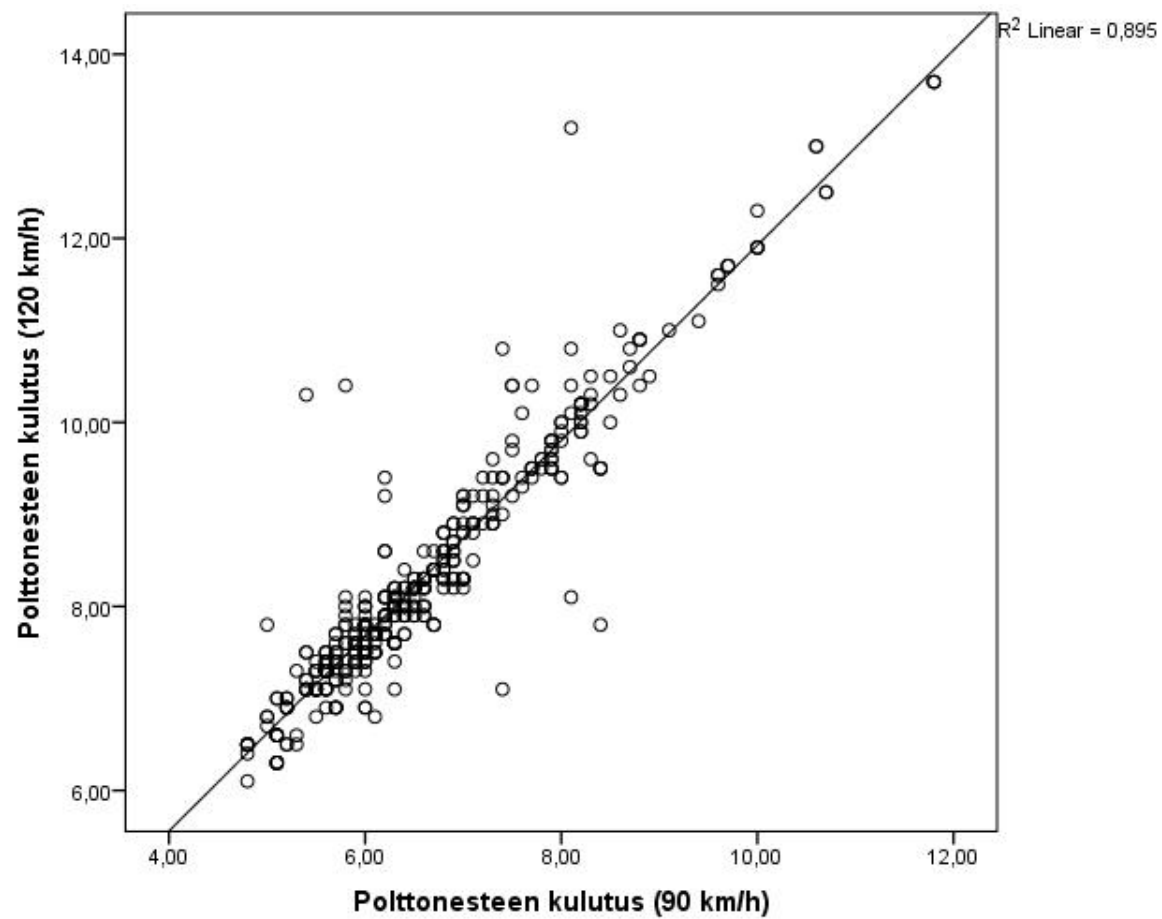
Saadaan

$$\hat{\beta}_0 = 4,435, \hat{\beta}_1 = 0,016$$

Pisteparveen sovitetun suoran yhtälö

$$y = 4,435 + 0,016x,$$

$y =$ Kulutus, $x =$ Teho



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,946 ^a	,895	,894	,44309

a. Predictors: (Constant), Polttonesteen kulutus (90 km/h)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1,316	,119		11,059	,000
	Polttonesteen kulutus (90 km/h)	1,061	,018	,946	59,997	,000

a. Dependent Variable: Polttonesteen kulutus (120 km/h)

Merkitään

$Y =$ Polttonesteen kulutus (120 km/h)

$x =$ Polttonesteen kulutus (90 km/h)

Malli

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Estimoidaan mallin parametrit β_0 ja β_1 .

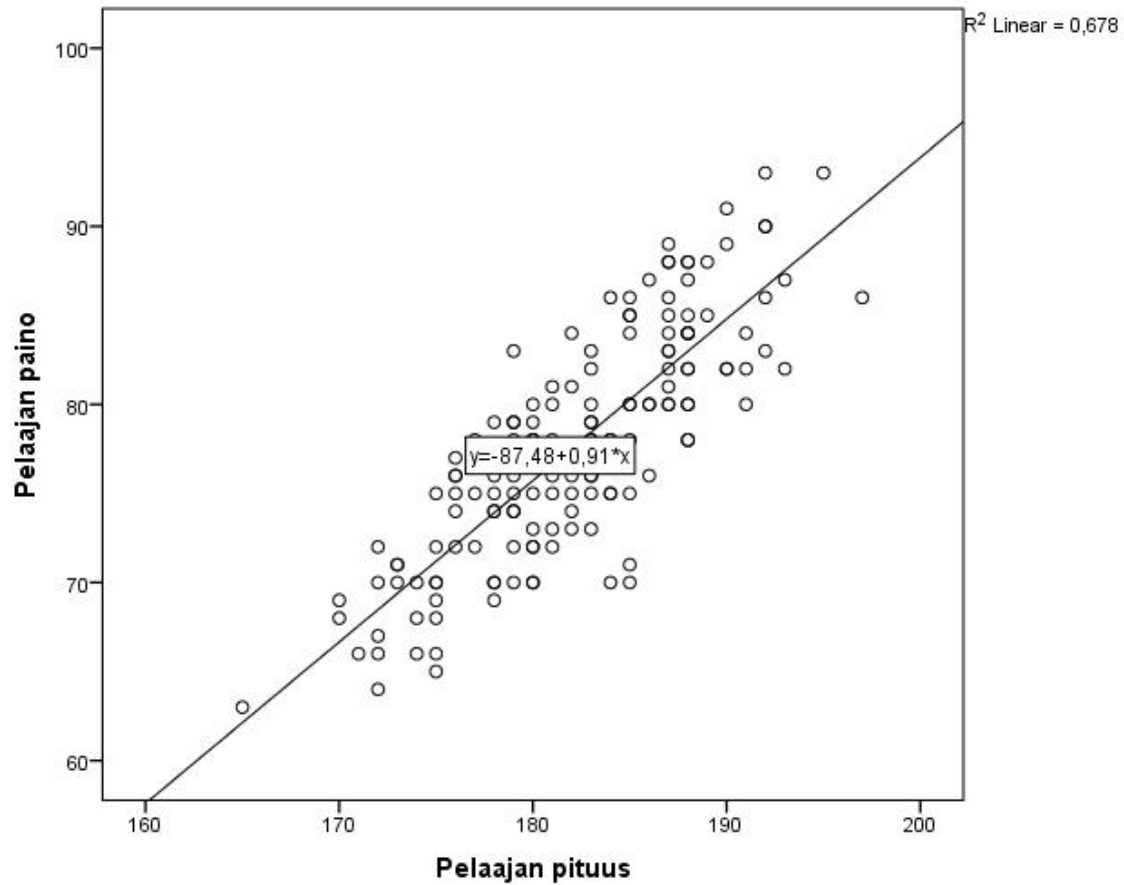
Saadaan

$$\hat{\beta}_0 = 1,316, \hat{\beta}_1 = 1,061$$

Pisteparveen sovitetun suoran yhtälö

$$y = 1,316 + 1,061x$$

Esim. Aineisto Jalkapalloilijat_2006 sivulta
<https://coursepages.uta.fi/mttp1/esimerkkiaineistoja/>



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,824 ^a	,678	,676	3,682

a. Predictors: (Constant), Pelaajan pituus

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-87,476	9,227		-9,480	,000
	Pelaajan pituus	,907	,051	,824	17,908	,000

a. Dependent Variable: Pelaajan paino

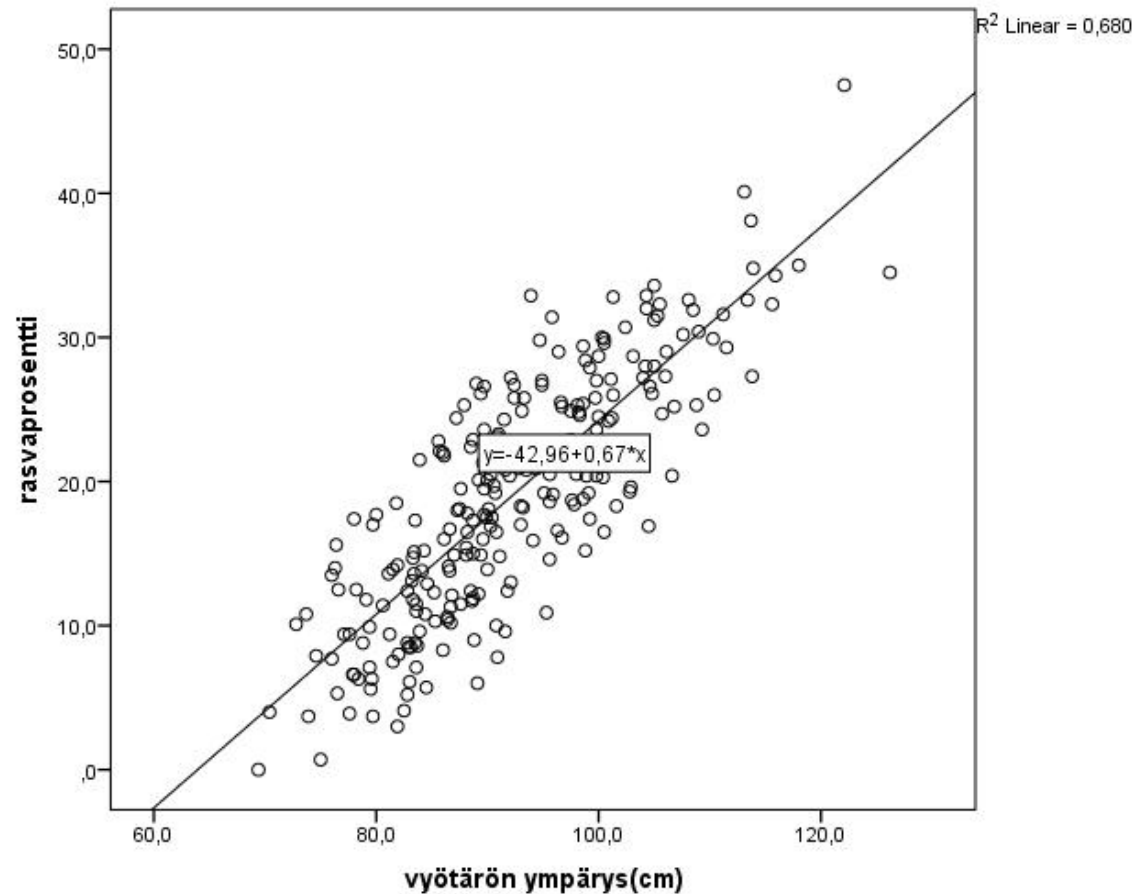
Malli ja estimoinnin tulos:

$$Paino = \beta_0 + \beta_1 \cdot Pituus + \varepsilon$$

$$\widehat{Paino} = -87,476 + 0,907 \cdot Pituus$$

Esim. Aineisto Rasvaprosentti sivulta

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



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,825 ^a	,680	,679	4,7174

a. Predictors: (Constant), vyötärön ympärys(cm)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-42,958	2,713		-15,833	,000
	vyötärön ympärys(cm)	,672	,029	,825	23,006	,000

a. Dependent Variable: rasvaprosentti

Malli ja estimoinnin tulos:

$$\text{Rasva\%} = \beta_0 + \beta_1 \cdot \text{Vyötärön ymp.} + \varepsilon$$

$$\widehat{\text{Rasva\%}} = -42,958 + 0,672 \cdot \text{Vyötärön ymp.}$$