

MTTTA1 Tilastomenetelmien perusteet
Luento 7.2.2019

4.1 Yksi selittävä muuttuja (täydennystä)

Regressiomalli

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

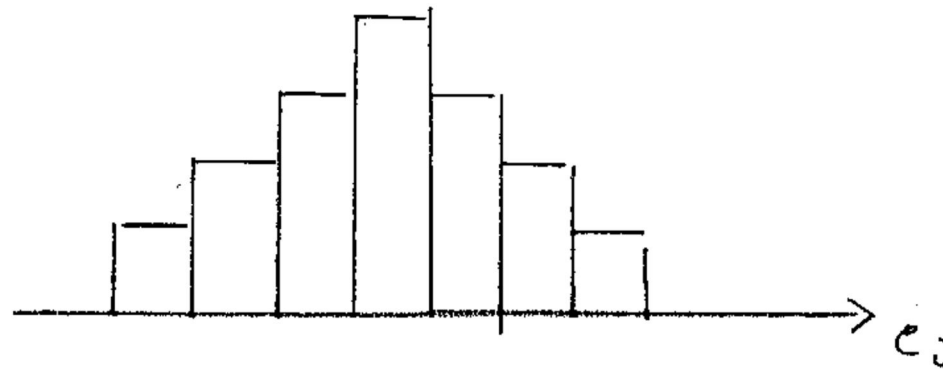
Regressiomallissa (1) oletetaan, että

- $\varepsilon_i \sim N(0, \sigma^2)$ ja
- ε_i :t ovat riippumattomia

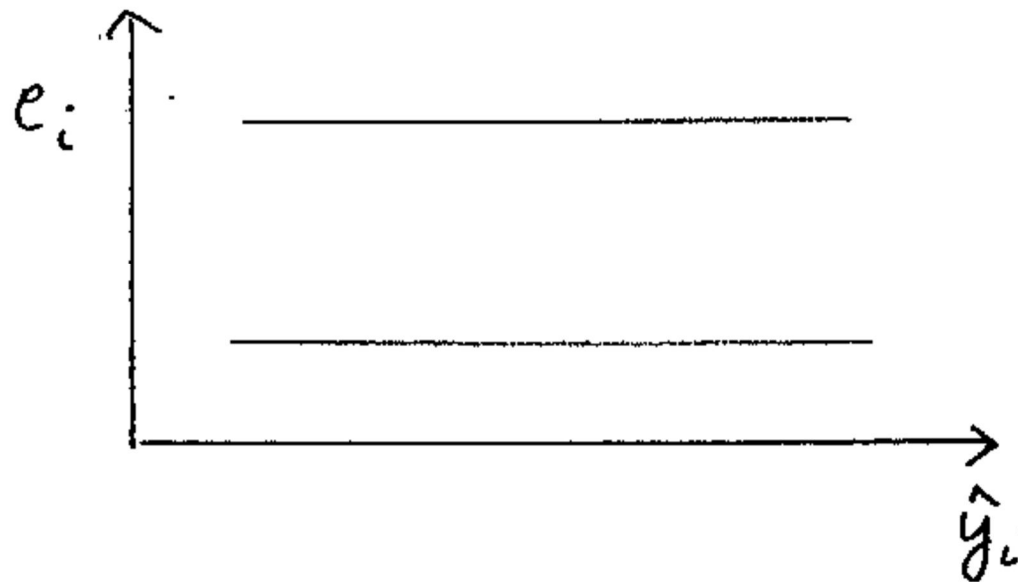
Näiden oletusten voimassaoloa tutkitaan residuaalien avulla. Koska satunnaisvirheistä ε_i ei ole havaintoja, niin estimoidaan niitä estimoidun mallin avulla lasketuilla residuaaleilla $e_i = y_i - \hat{y}_i = \hat{\varepsilon}_i$

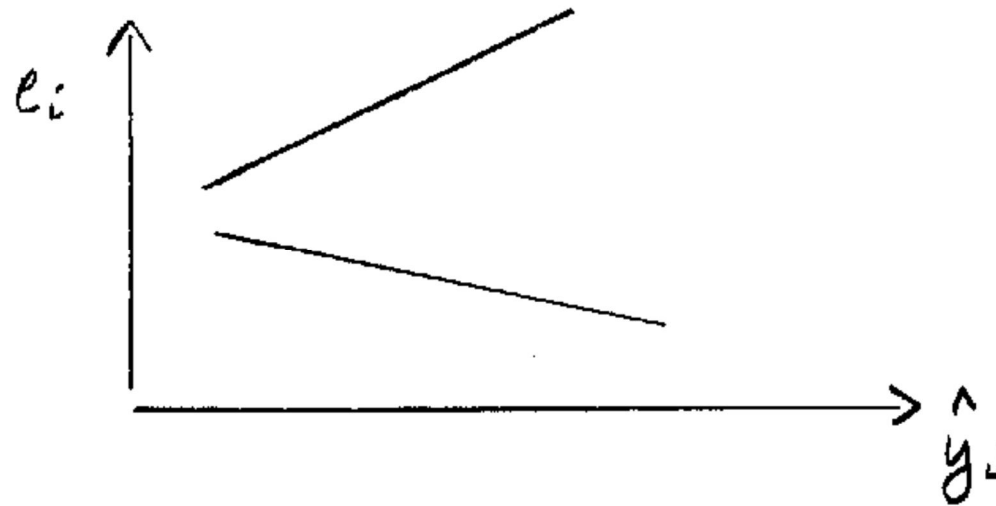
Tutkitaan normaalisuus-, vakiovarianssisuus- ja riippumattomuusoletuksia näiden residuaalien avulla. Voidaan käyttää graafisia esityksiä, esimerkiksi seuraavia:

- Normaalisuusoletuksen tutkiminen esim. histogrammin avulla



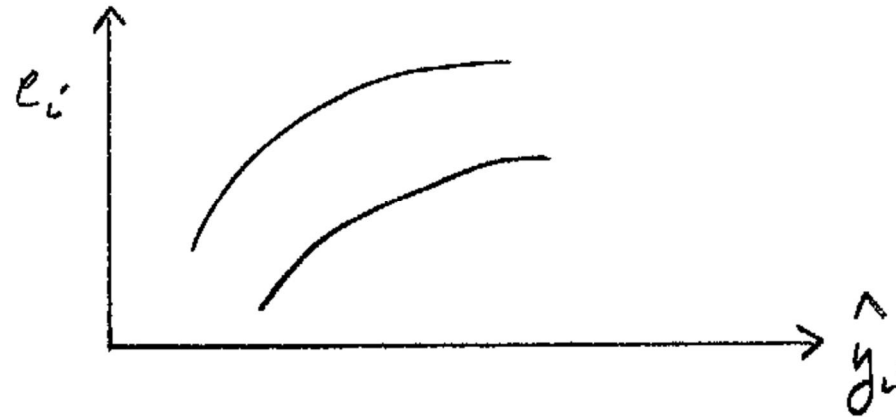
- Vakiovarianssisuuden ja riippumattomuuden tutkiminen pisteparvien avulla





Ei voida olettaa, että $\text{Var}(\varepsilon_i) = \sigma^2$, $i = 1, \dots, n$
(heteroskedastisuus).

- Mallin riittävyyden tutkiminen

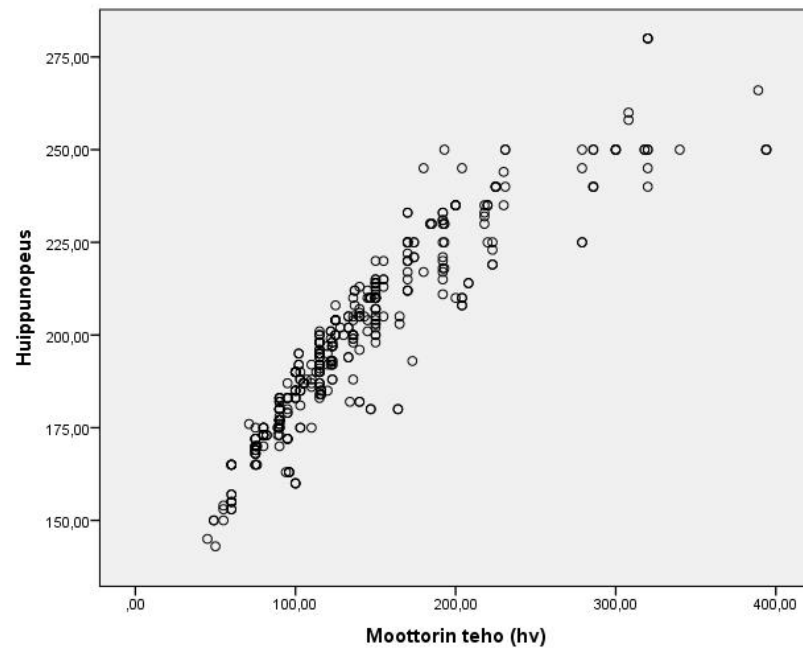


Esimerkki riittämättömästä mallista

Pisteparvissa voidaan käyttää x-akselilla myös selittäjää.

Esim. Autojen ominaisuuksia

$Y = \text{Huippunopeus}$, $x = \text{Teho}$



Coefficients^a

| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | |
|-------|-----------------------------|------------|---------------------------|------|---------|------|
| | B | Std. Error | Beta | | | |
| 1 | (Constant) | 148,577 | 1,193 | | 124,563 | ,000 |
| | Mootorin teho (hv) | ,361 | ,008 | ,914 | 46,370 | ,000 |

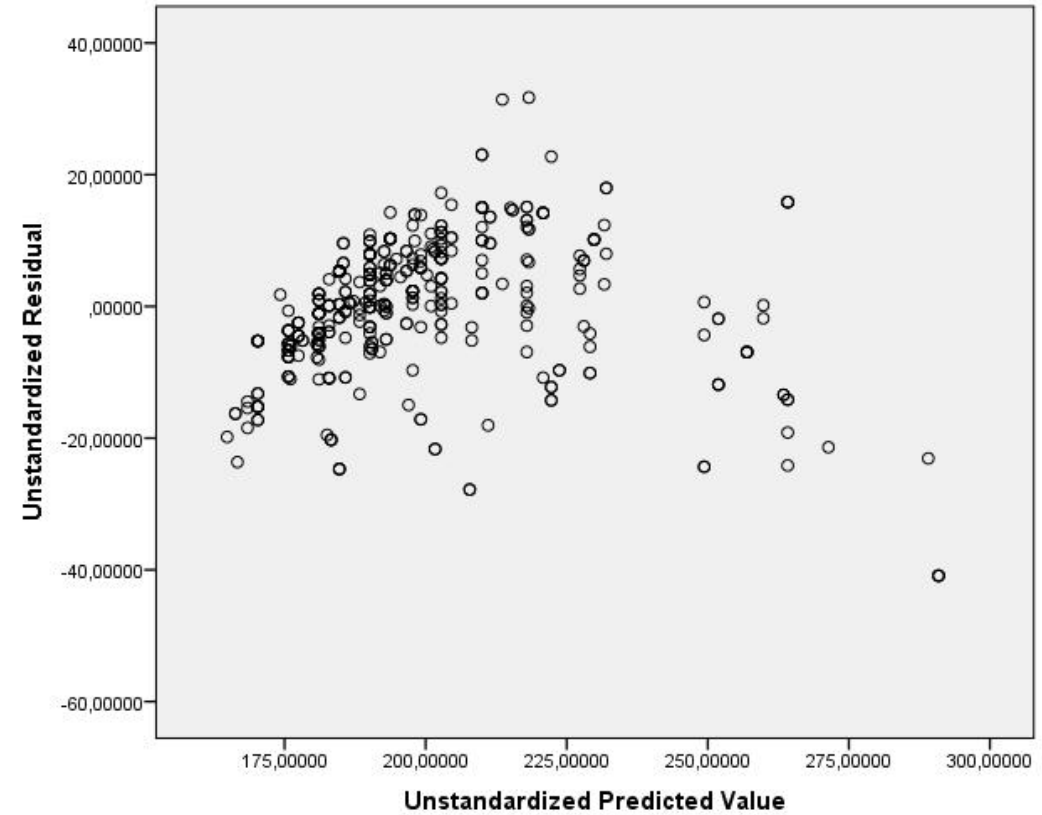
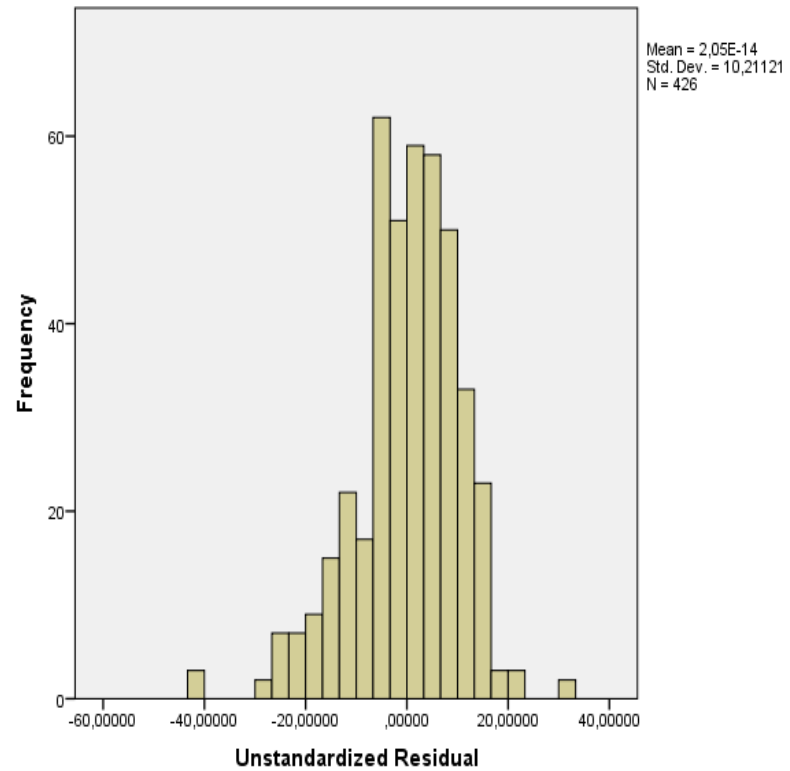
a. Dependent Variable: Huippunopeus

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | ,914 ^a | ,835 | ,835 | 10,22324 |

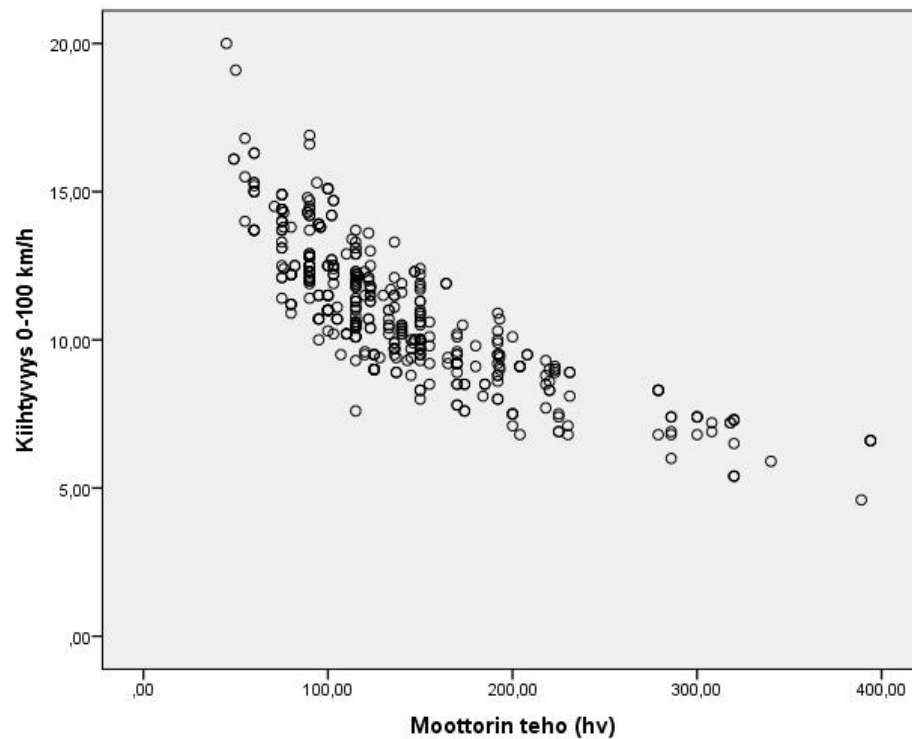
a. Predictors: (Constant), Mootorin teho (hv)

Jännöstarkastelut



Väärä mallin valinta

$Y = \text{Kiihtyvyyys}, x = \text{Teho}$

Coefficients^a

| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|-----------------------------|------------|---------------------------|--------|------|
| | B | Std. Error | Beta | | |
| 1 | (Constant) | 15,096 | ,162 | 93,342 | ,000 |
| | Moottorin teho (hv) | -,030 | ,001 | -,808 | ,000 |

a. Dependent Variable: Kiihtyvyyys 0-100 km/h

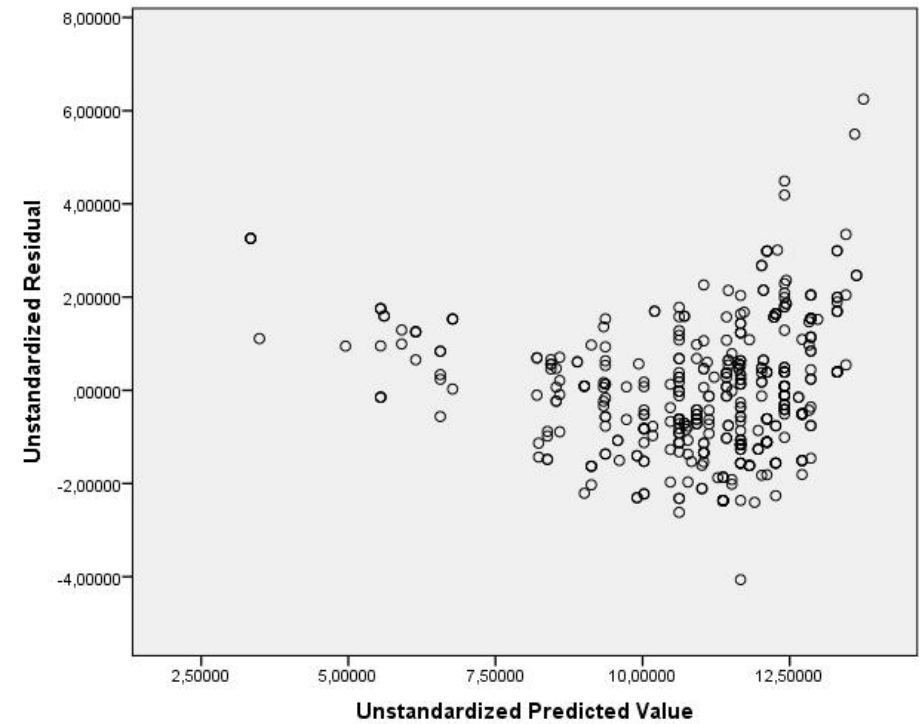
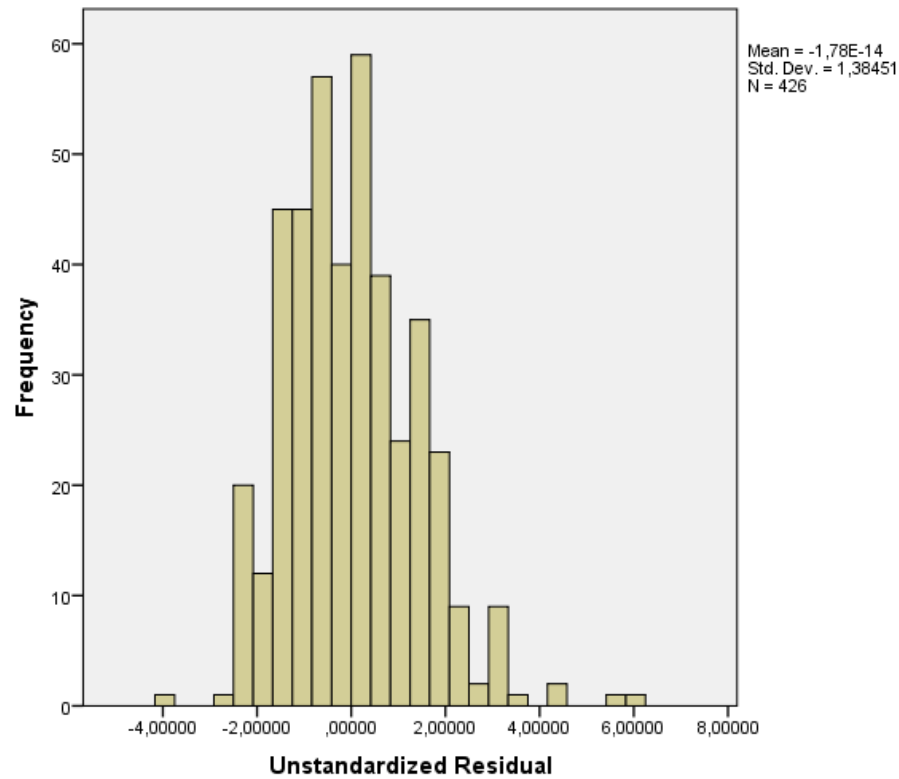
Model Summary^b

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | ,808 ^a | ,653 | ,652 | 1,38614 |

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

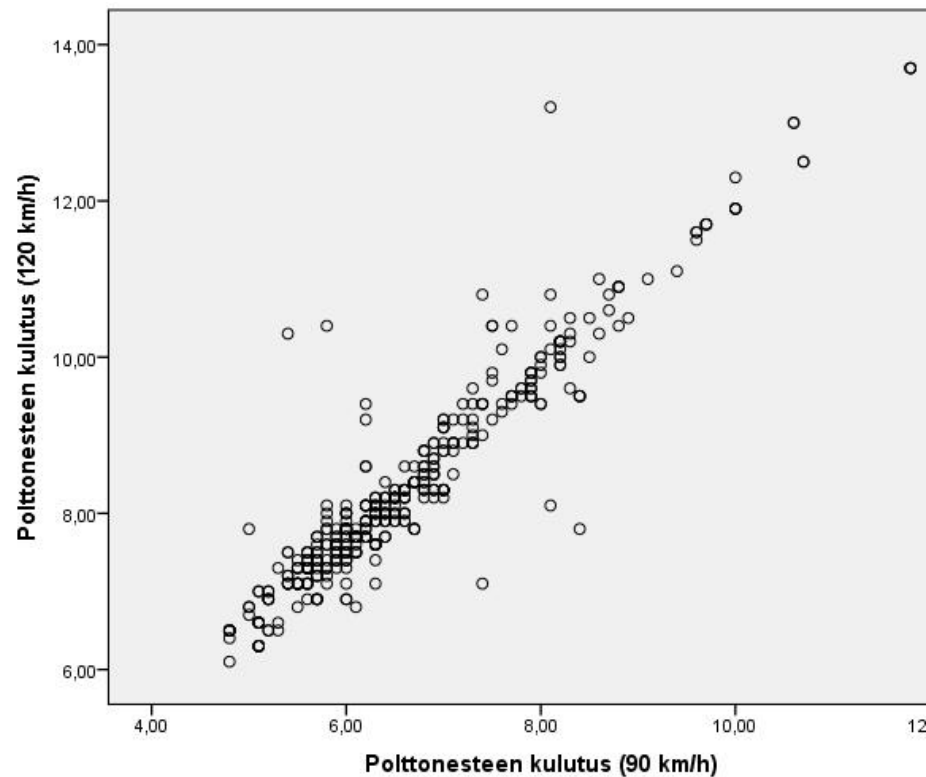
b. Dependent Variable: Kiihtyvyyys 0-100 km/h

Jännöstarkastelut



Väärä mallin valinta

$Y = \text{Kulutus } 120\text{km/h}$, $x = \text{Kulutus } 90 \text{ km/h}$

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)

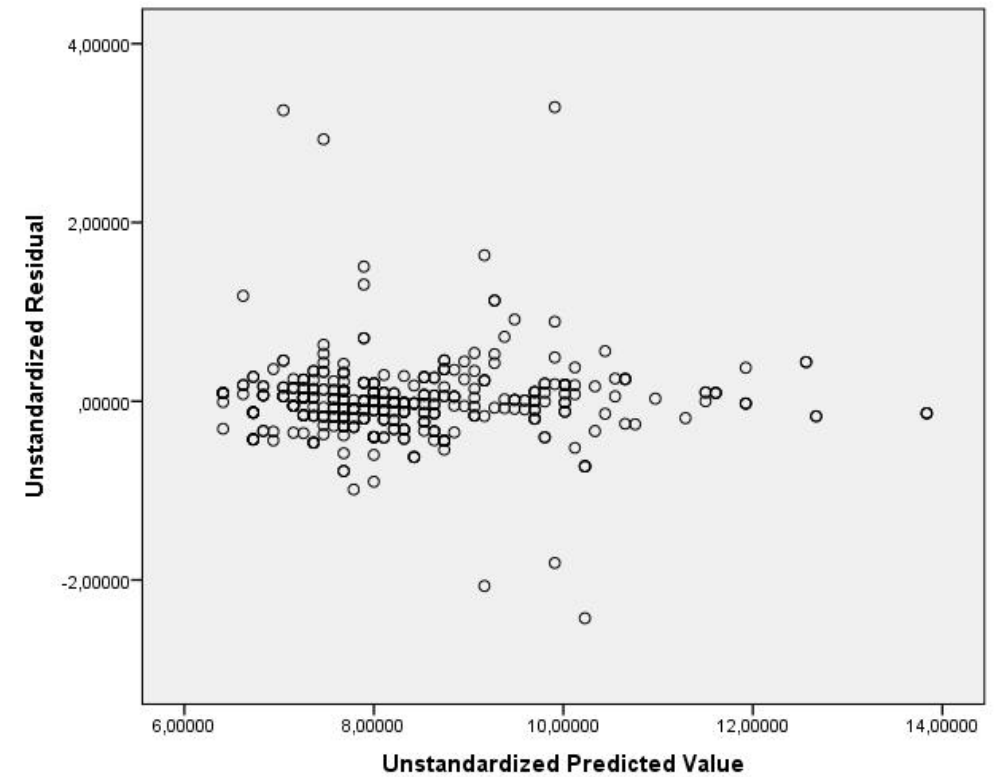
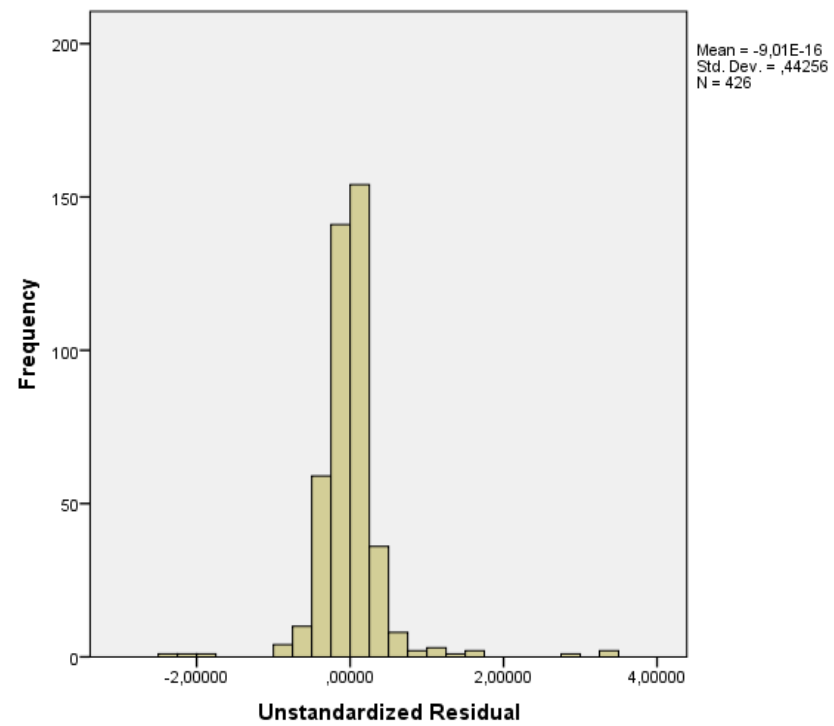
Model Summary^b

| 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)

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

Jännöstarkastelut

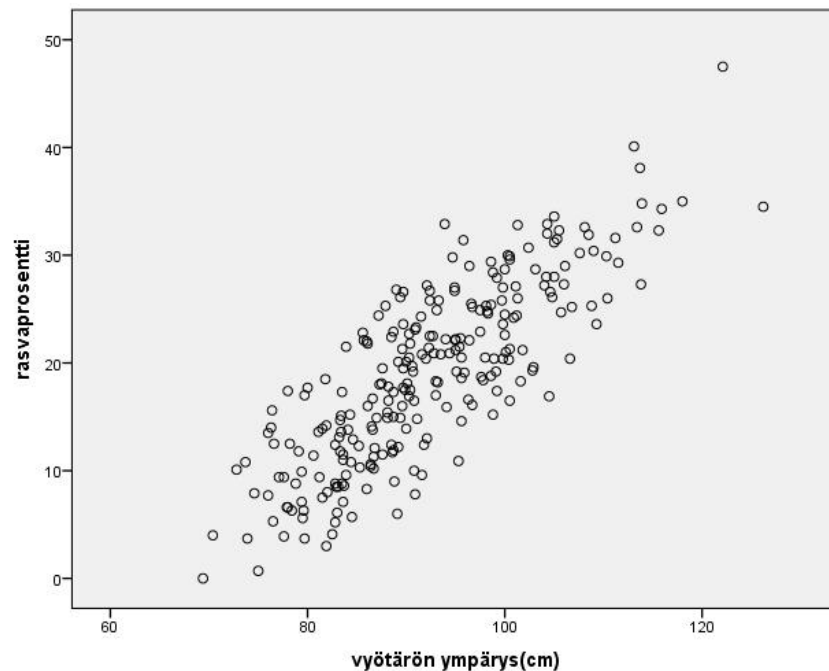


Esim. Aineisto Rasvaprosentti sivulla

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

y = rasvaprosentti

x = vyötärön ympärys



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

Model Summary

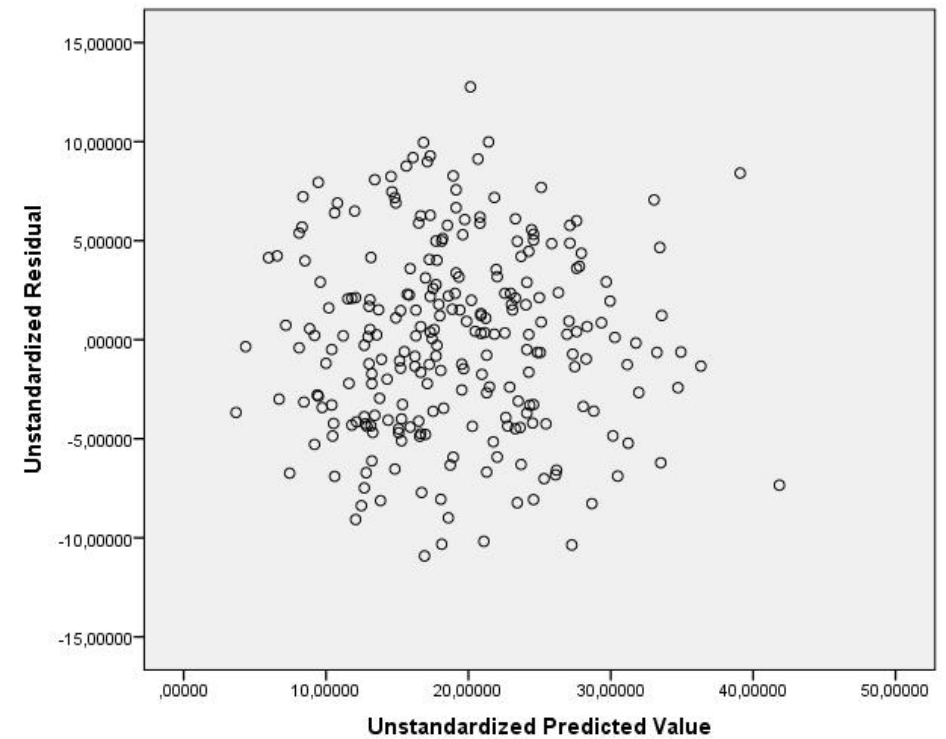
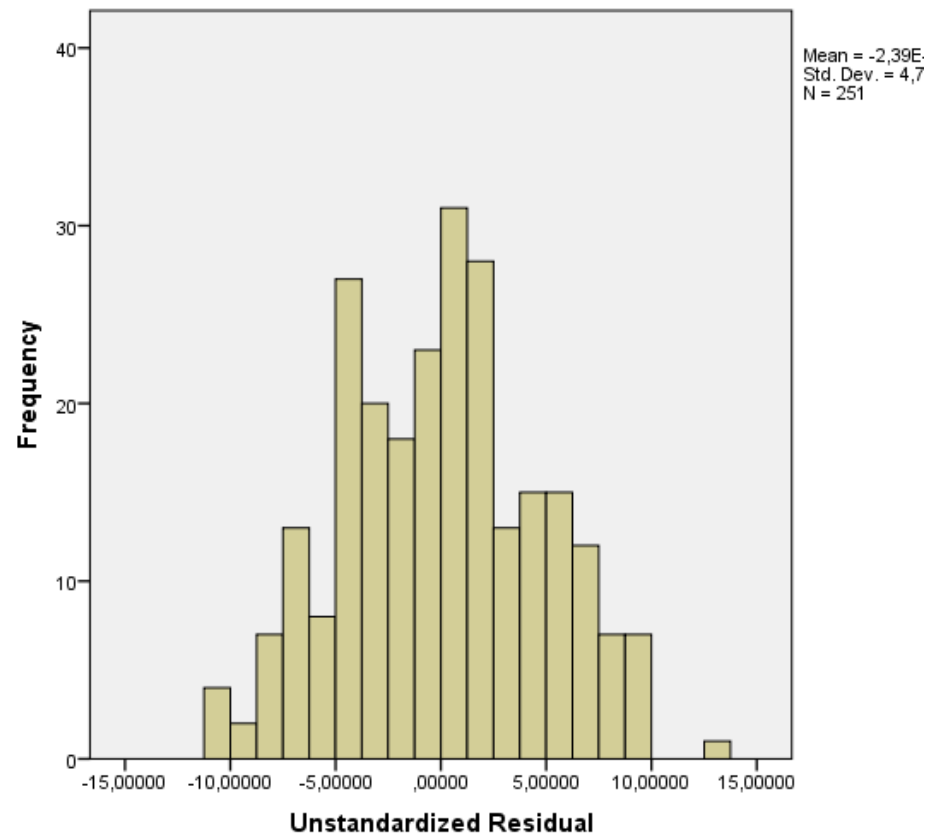
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | ,825 ^a | ,680 | ,679 | 4,717 |

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

Ks.

http://www.sis.uta.fi/tilasto/mttta1/kevat2015/RA_rasvaprosentti.pdf

Jäännöstarkastelut



4.2 Useampi selittävä muuttuja

Kaksi selittäjää (2-RA)

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2)$$

Malliin liittyvät oletukset

- $\varepsilon_i \sim N(0, \sigma^2)$ ja
- ε_i :t ovat riippumattomia

Estimointi

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 z_i,$$

Testaukset

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

$$t = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)} \sim t_{n-3}, \text{ kun } H_0 \text{ tosi}$$

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0$$

$$t = \frac{\hat{\beta}_2}{s(\hat{\beta}_2)} \sim t_{n-3}, \text{ kun } H_0 \text{ tosi}$$

$$H_0: \beta_0 = 0$$

$$H_1: \beta_0 \neq 0$$

$$t = \frac{\hat{\beta}_0}{s(\hat{\beta}_0)} \sim t_{n-3}, \text{ kun } H_0 \text{ tosi}$$

$$H_0: \beta_1 = \beta_2 = 0$$

H_1 : molemmat eivät nollia

$$F = \frac{MSR}{MSE} = \frac{\frac{SSR}{2}}{\frac{SSE}{n-3}} \sim F(2, n-3), \text{ kun } H_0 \text{ tosi}$$

Neliösummat

$$SST = SSR + SSE$$

$$MSR = SSR/2, \text{ MSE} = SSE/(n-3) = \hat{\sigma}^2$$

Selityskerroin

$$R^2 = SSR/SST$$

Esim. Aineisto Rasvaprosentti sivulla

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

$y = \text{rasva\%}$

$x_1 = \text{vyötärön ympäryys}$

$x_2 = \text{ikä}$

Regressioanalyysin tulokset

http://www.sis.uta.fi/tilasto/mttta1/kevat2015/RA_rasvaprosentti.pdf

Regressiomallissa

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

hypoteesin

$$H_0: \beta_1 = 0$$

testaaminen voidaan tehdä joko t-testillä tai F-testillä,
testisuureiden välinen yhteys

$$t^2 = \left(\frac{\hat{\beta}_1}{s(\hat{\beta}_1)} \right)^2 = F = \frac{MSR}{MSE}$$

Esim. Jalkapalloilijat
 $y = \text{paino}$
 $x = \text{pituus}$

| ANOVA ^a | | | | | | |
|--------------------|------------|----------------|-----|-------------|---------|-------------------|
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 4347,084 | 1 | 4347,084 | 320,705 | ,000 ^b |
| | Residual | 2060,325 | 152 | 13,555 | | |
| | Total | 6407,409 | 153 | | | |

a. Dependent Variable: Pelaajan paino
b. 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

$$t^2 = F$$