

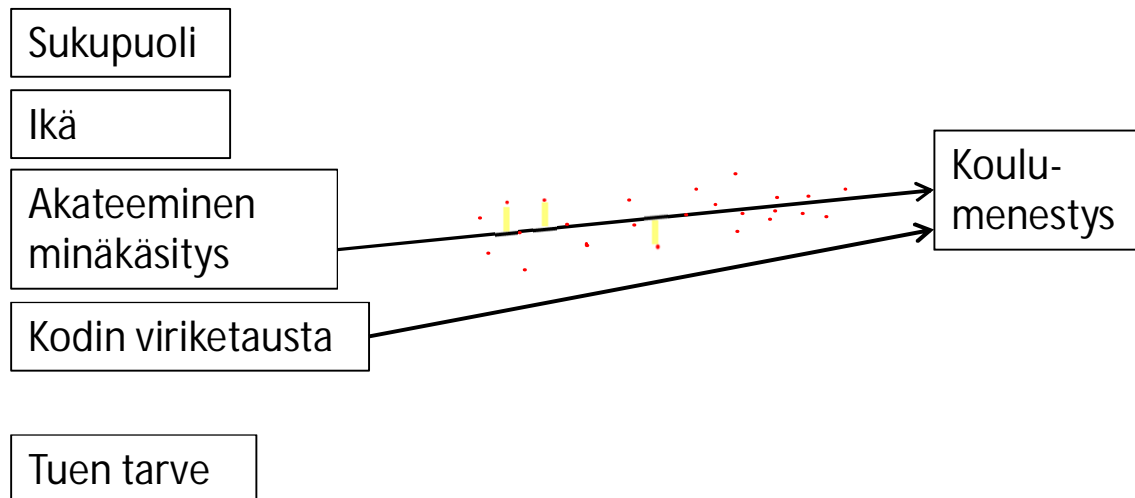
# Regressioanalyysi $R^2$

Regressioanalyysi kertoo kuinka malliin valitut selittävien muuttujien ( $X$ ) muutos ennustaa selitettävää muuttujaa ( $Y$ ). Muutoksen suuruus ilmoitetaan regressiokertoimella.

# Regressioanalyysi

Selittävät muuttujat

Selittävät muuttujat



## Regressioanalyysi $R^2$

- à multippelikoreelaatiokertoimen neliö
- à selittävä muuttuja (factor, dependent variable)
- à selittävät muuttujat (predicators, independent variable)

Lähtöoletuksia: selittäjät korreloivat kohtuullisesti selitettävien kanssa, mutteivat välttämättä toisensa kanssa  
(etenkään liikaa  $\rightarrow$  multikollineaarisuus)

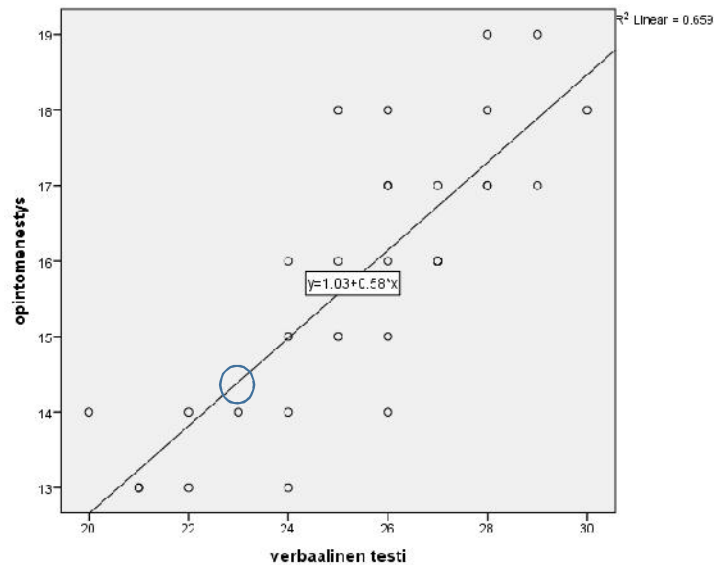
"Eksploratiivinen regressioanalyysi  $\rightarrow$  stepwise"  
"Vai teoriajohtoinen"  $\rightarrow$  otetaan malliin teorian kannalta oleellisimmat ja tarkastellaan mitkä pitävät paikkansa.

Kysytään:  
*Mitkä tekijät selittävät jatkuvaa muuttujaa?*

# Oletukset

- Mallin selittymättä jäävä osa  $\rightarrow$  jäännösvariانسsi, residuaalit ovat normaalisti jakautuneita ja hajonta homoskedastinen.
- Selitettävä muuttuja riittävästi normaalijakautunut
- Otoskoko: 50 + 8 kertaa selitettävien määrä  
 $\rightarrow$  3 selittäjää  $50 + 24 = 74$   
 $\rightarrow$  10 selittäjää = 130

- Kukin malliin tuleva muuttuja saa painokertoimen  $\beta$



Esim. verbaalisen testin  
pistemäärän 23 voi laskea  
opintomenestyksen pistemäärä  
 $1.03 + 0.58 \times 23 = 14.37$

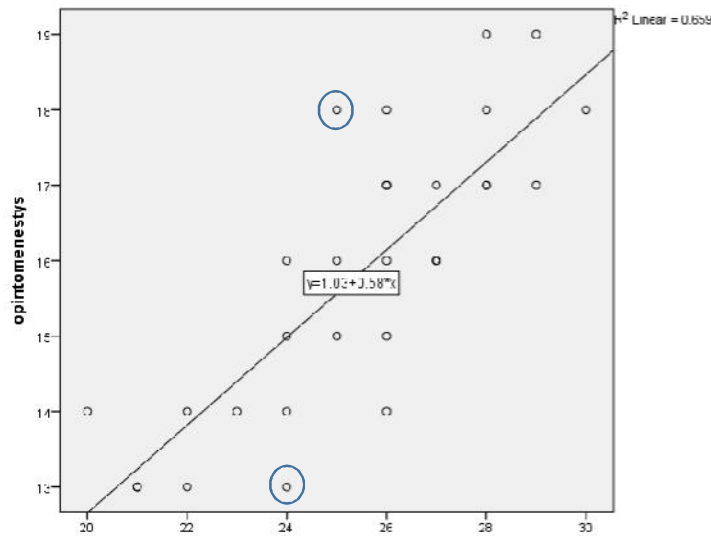
# Regressiomenetelmistä

- 1) Backward elimination (kaikki koneeseen poistetaan selittäjistä yksi kerrallaan, jäljelle jää vain tilastollisesti merkitsevät)
- 2) Forward selection (valitaan selittäjä joukosta mukaan voimakkain, sitten seuraava jolla on seuraavaksi voimakkain selitysvaikutus) → ei multikollineaarisuuden ongelmaa
- 3) Stepwise yhdistää lisäävän ja poistavan valinnan. Kutakin muuttujaa testataan prosessin aikana (poistetaan ja lisätään) tilastollisin perustein (esim.  $p < .05$ )
- 4) Enter (pakotettu) –menetelmä → standard (default –menetelmä), eli teoriasidonnainen "kaikki mukaan, mitä teoria väittää". Katsotaan mikä ei toimi ja mikä toimii.

# Mallin hyvyys

## a) $R^2$ (selitysaste)

## b) Poikkeavien havaintojen tarkastelu



-Normality plot

- Residuaalit ja ennustearvot -kuva
- Casewise diagnostics (residuaalin raja-arvo asetettavissa -3 ja 3)
- Durbin-Watson 1 – 3 välillä ok.

## c) Multikollinearisuus

- Kuntoisuusindeksi ( $X > 15$  ongelmia  $x > 30$  isoja)
- Tolerance sarake arvo suurempi kuin 1 -  $R^2$  -arvo.



Tutkimusesimerkki: Ruoho, K., & Hotulainen, R. (2009). Pädagogische Prophylaxe im Vorschulalter - eine Chance zur Verbesserung von Lern- und Lebensfolgen. Teoksessa (Breuer ym.), *Sprachenwerbsforschung im Spannungsfeld von Angewandter Linguistik und Pädagogik: [Internationalen Konferenz Pobierowo 2008 (31-45)*. Berlin: Trainmedia.

Pitkittäistutkimus: Mittauksia -89, -99, -05 (N = 90)

*Tabelle 5. Erklärende Faktoren der Selbstwertgefühls (Multiple Regression Analyse).*

Selitettävänä Itsetunto			
	Beta	t (85)	p-level
Sosiaalinen mk - 05	.34	3.92	p<0.001
Käytöksellinen mk - 05	.28	3.40	p<0.001
Ulkonäöllinen mk - 05	.23	2.69	p<0.01
Terveydellinen mk 05	.17	3.03	p<0.05
Lukemaan ja kirjoittamaan oppimisen valmiudet (VSM) - 89	.19	2.47	p<0.01
Adjusted R <sup>2</sup>	.51		
F-test	19.76		p<0.001
Std. Error of Estimate	.38		
Regress.	14.00	5 (df)	2.80
Residual	12.05	85 (df)	.14
Total	26.05		

Das Modell war statistisch signifikant: F (19,76)=19,76, p<.001. Außer <sup>9</sup>

# Raportointi yleensä taulukkojen avulla

LINEAARINEN REGRESSIO		
Selitettävänä muuttujana AIKOMUKSET		
n = 3430	Malli a	Malli b
<b>Kontrollimuuttujat</b>		
Ikä	-0,10***	-0,06***
Sukupuoli (mies)	0,17***	0,11***
Äidin yrittäjyys	0,14***	0,07***
Isän yrittäjyys	0,23***	0,10***
<b>Jatkuvat selittävät muuttujat</b>		
LÄHEISTEN MELIPITEET		0,39***
PYSTYVYYS YRITTÄJÄNÄ TOIMIMISEEN		0,17***
ASENTEET		0,27***
YRITTÄJÄMÄISYYS		0,06***
<b>Model fit statistics</b>		
Adjusted R2	0,124	0,574
F-statistics	122,0***	578,0***
F change		905,2***
* p<0,05 ** p<0,01 ***p<0,001		
Käytetty standardoituja kertoimia		

län kerroin on negatiivinen eli nuoremmilla on suuremmat aikomukset

Mitä suurempi kerroin, sitä parempi selittäjä. Kaikki mallissa mukana olevat kertoimet ovat til. merkitseviä

Mallin b selitysaste 57%

TILASTOLLINEN TUTKIMUS  
EDITA

(Tarja Heikkilää, 2014)