



Tarja Heikkilä

Usean selittävän muuttujan regressioanalyysi


Yhden selittävän muuttujan regressioanalyysia on selvitetty kirjan luvussa 11, jonka esimerkissä 18 muodostettiin lapsen syntymäpainolle lineaarinen malli raskauden keston perusteella. Luvussa 12 on monimuuttujamenetelmien periaatteiden yhteydessä selvitetty lyhyesti myös usean selittävän muuttujan regressioanalyysia. Tässä tarkastellaan esimerkin avulla usean selittävän muuttujan lineaarisen regressioanalyysin suorittamista SPSS-ohjelmalla.

Esimerkki

Pyritään selvittämään lineaarisen regressioanalyysin avulla, mitkä tekijät selittävät parhaiten opiskelijoiden aikomuksia ryhtyä yrittäjäksi. Aineistossa (Entre-aineisto) on yli 3700 ammattikorkeakouluopiskelijan taustatiedot ja vastaukset yrittäjyyttä koskeviin kysymyksiin.

Otetaan regressioanalyysiin selittäviksi muuttujiksi ensimmäiselle ja jatkaville vuosikursseille yhteiset taustamuuttujat (malli a): sukupuoli, ikä, äidin ja isän yrittäjyys. Lisäksi otetaan jatkuviksi selittäviksi muuttujiksi läheisten mielipiteet yrittäjäksi ryhtymisestä, yrittäjyysasenteet, pystyvyysuskomus ja yrittäjämäisyys, jotka kaikki ovat useasta eri muuttujasta muodostettuja asiakokonaisuuksia (malli b).

Taustamuuttujista sukupuoli on koodattu: 1=mies, 0 = nainen. Ikä on ilmoitettu vuosina. Muuttujat äidin yrittäjyys ja isän yrittäjyys on koodattu: 1=on yrittäjä, 0=ei ole yrittäjä.



Jatkuvat selittäjät ovat useasta muuttujasta muodostettuja keskiarvomuuttujia:

AIKOMUKSET (asteikolla 1–7, yrittäjäksi ryhtyminen on 1 = ei ole lainkaan todennäköistä, 7 = on erittäin todennäköistä) sisältää seuraavat muuttujat:

- Kuinka todennäköistä on, että jatkat uraasi toisen palveluksessa (eli palkkatyössä) valmistumisesi jälkeen? (käännetty asteikko)
- Kuinka todennäköistä on, että tulet perustamaan oman yrityksen valmistumisesi jälkeen (tai opintojesi aikana)?
- Jos sinun pitäisi valita joko yrittäminen tai palkkatyö valmistumisesi jälkeen, kumman valitsisit? (1=palkkatyö, 7=yrittäminen)
- Kuinka todennäköistä on, että tulet olemaan suurimman osan työurastasi palkattuna johonkin yritykseen tai julkiseen organisaatioon (ilman kytköstä yrittäjyyteen)? (käännetty asteikko)
- Kuinka vakaa aikomus sinulla on tulla yrittäjäksi jossain vaiheessa työuraasi?
- Jos sinun pitäisi valita joko yrittäminen tai työttömyys valmistumisesi jälkeen, kumman valitsisit? (1=työttömyys, 7=yrittäminen)
- Kuinka todennäköistä on, että päädyt yrittäjäksi sukupolven- tai omistajavaihdoksen kautta valmistumisesi jälkeen (tai opintojesi aikana)?
- Kuinka todennäköistä on, että lähdet yrittäjäksi sen jälkeen kun olet hankkinut sopivan määrän työkokemusta?

LÄHEISTEN MIELIPITEET (asteikolla 1 – 7, 1=ei pitäisi, 7=pitäisi)

- Uskon että läheisimmät perheenjäseneni ajattelevat, että minun ei pitäisi/pitäisi tavoitella oman yrityksen perustamista ja yrittäjänä toimimista valmistumiseni jälkeen
- Uskon että parhaat kaverini ajattelevat, että minun ei pitäisi/pitäisi tavoitella oman yrityksen perustamista ja yrittäjänä toimimista valmistumiseni jälkeen
- Uskon että minulle erittäin tärkeät ihmiset ajattelevat, että minun ei pitäisi/pitäisi tavoitella oman yrityksen perustamista ja yrittäjänä toimimista valmistumiseni jälkeen.

Vastaavalla tavalla muodostettiin useasta eri väittämästä asteikolla 1–7 muuttujat *ASENTEET*, *PYSTYVYYS YRITTÄJÄNÄ TOIMIMISEEN* ja *YRITTÄJÄMÄISYYS*.

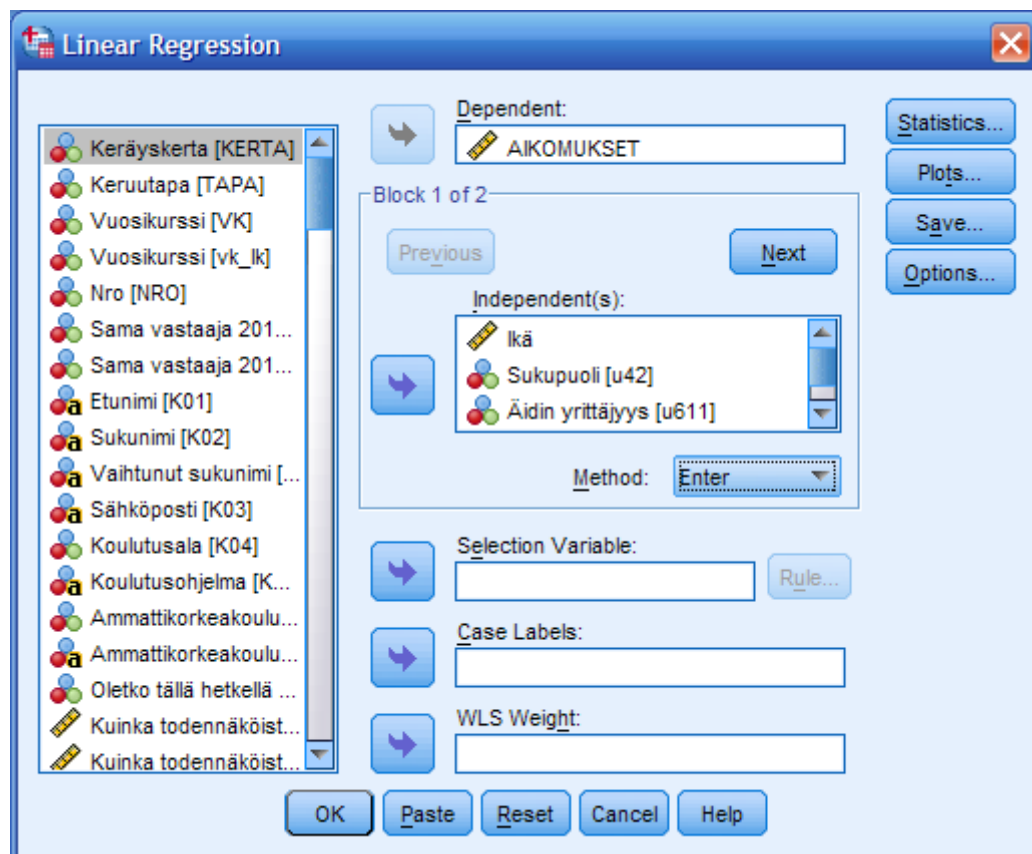
Ennen regressioanalyysiä asetettiin *Data / Select cases* -komennolla ehto, että tarkastellaan vain niiden opiskelijoiden vastauksia, jotka eivät ole perustamassa yritystä.

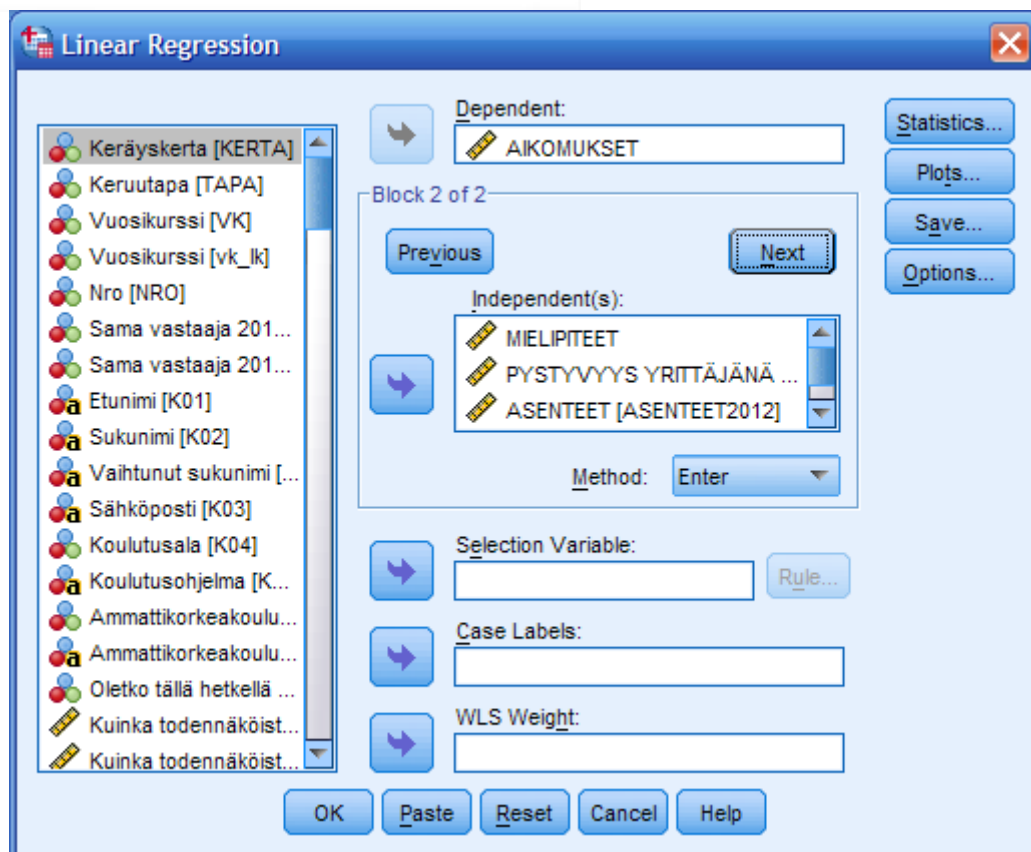
Valitaan mallin muodostamismenetelmäksi (*Method*) tässä *Enter*, joka jättää malliin kaikki valitut selittävät muuttujat. Menetelmäksi olisi voitu valita myös *Stepwise*, joka jättää lopulliseen malliin vain tilastollisesti merkitsevät selittävät muuttujat. Tässä tapauksessa olisi tulostunut kahdeksan peräkkäistä mallia, koska ohjelma olisi lopulliseen malliin ottanut kaikki selittävät, jotka ovat mukana seuraavassakin mallissa.

SPSS: *Analyze / Regression / Linear*

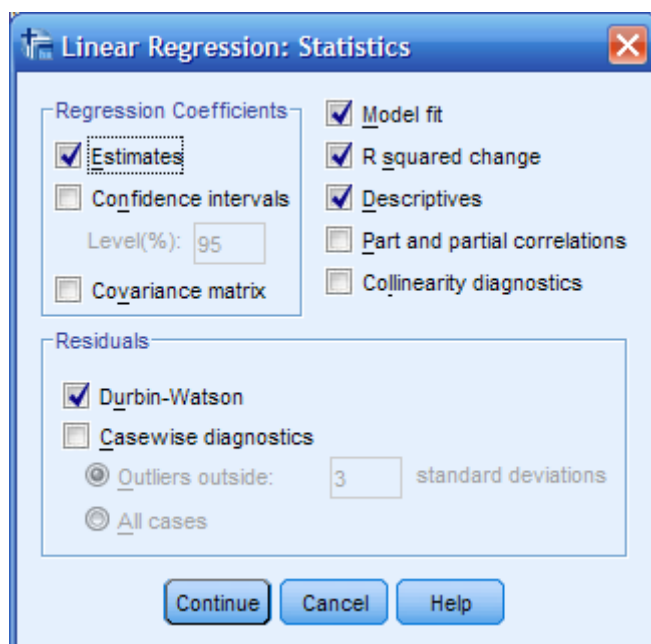
Siirretään *AIKOMUKSET*-muuttuja *Dependent*-laatikkoon.

Previous-kohtaan valitaan taustamuuttujat ja *Next*-painikkeella siirrytään valitsemaan jatkuvat selittäjät





Statistics -ikkunassa tehdään valinnat:



Saadaan runsaasti tulosteita, joista tähän on valittu tulkinnan kannalta tärkeimmät:

Descriptive Statistics

	Mean	Std. Deviation	N
AIKOMUKSET	3,2842	1,17257	3430
Ikä	22,83	4,715	3430
Sukupuoli	,42	,494	3430
Äidin yrittäjyys	,15	,356	3430
Isän yrittäjyys	,32	,467	3430
MIELIPITEET	3,75	1,140	3430
PYSTYVYYS YRITTÄJÄNÄ TOIMIMISEEN 2012	4,0266	,99859	3430
ASENTEET	4,8688	,84026	3430
YRITTÄJÄMÄISYYS	4,7743	,90257	3430

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,353 ^a	,125	,124	1,09764	,125	122,035	4	3425	,000	1,865
2	,758 ^b	,575	,574	,76551	,450	905,167	4	3421	,000	

a. Predictors: (Constant), Isän yrittäjyys, Sukupuoli, Äidin yrittäjyys

b. Predictors: (Constant), Isän yrittäjyys, Sukupuoli, Äidin yrittäjyys, YRITTÄJÄMÄISYYS, LÄHEISTEN MIELIPITEET, ASENTEET, PYSTYVYYS

c. Dependent Variable: AIKOMUKSET

Mallien selitysasteet

Katso arvot lopullisessa mallissa

Kun tämä ≈ 2 , ei esiinny multiko-linearisuutta. On OK.

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	588,117	4	147,029	122,035	,000 ^b
	Residual	4126,473	3425	1,205		
	Total	4714,591	3429			
2	Regression	2709,857	8	338,732	578,033	,000 ^c
	Residual	2004,733	3421	,586		
	Total	4714,591	3429			

a. Dependent Variable: AIKOMUKSET

b. Predictors: (Constant), Isän yrittäjyys, Ikä, Sukupuoli, Äidin yrittäjyys

c. Predictors: (Constant), Isän yrittäjyys, Ikä, Sukupuoli, Äidin yrittäjyys, YRITTÄJÄMÄISYYS, LÄHEISTEN MIELIPITEET, ASENTEET, PYSTYVYYS YRITTÄJÄNÄ TOIMIMISEEN

Mallit sopivat aineistoon.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3,425	,095		36,117	,000
	Sukupuoli	,396	,038	,166	10,411	,000
	Ikä	-,024	,004	-,098	-6,138	,000
	Äidin yrittäjyys	,451	,056	,137	8,088	,000
	Isän yrittäjyys	,573	,042	,228	13,504	,000
2	(Constant)	-1,111	,112		-9,948	,000
	Sukupuoli	,251	,027	,105	9,354	,000
	Ikä	-,015	,003	-,059	-5,172	,000
	Äidin yrittäjyys	,232	,039	,070	5,922	,000
	Isän yrittäjyys	,255	,030	,102	8,487	,000
	LÄHEISTEN MIELIPITEET	,406	,014	,395	29,230	,000
	ASENTEET	,379	,018	,271	20,825	,000
	PYSTYVYYS YRITTÄJÄNÄ TOIMIMISEEN	,196	,016	,167	12,309	,000
	YRITTÄJÄMÄISYYS	,074	,016	,057	4,642	,000

a. Dependent Variable: AIKOMUKSET

Kertoimet

p-arvot

Selitysasteeksi saatiin koko aineistossa 57 %. Kaikkien selittävien muuttujien kertoimet ovat tilastollisesti merkitseviä. Parhaaksi selittäjäksi osoittautui läheisten mielipiteet yrittäjäksi ryhtymisestä (suurin kerroin 0,395***). Yrittäjyysaikomuksia selittävät lisäksi yrittäjyysasenteet (kerroin 0,271***), pystyvyysuskomus yrittäjänä toimimiseen (0,167***), yrittäjämäisyys (0,057***), sukupuoli (0,105***), isän yrittäjyys (0,102***), äidin yrittäjyys (0,070***) ja ikä (-0,059***). Positiiviset kertoimet osoittavat esimerkiksi sen, että mitä enemmän läheiset olivat yrittäjäksi ryhtymisen kannalla, sitä suuremmat yrittäjyysaikomukset olivat. Sukupuolen suhteen miehillä oli suuremmat yrittäjyysaikomukset, koska kerroin on positiivinen ja miehen koodi (1) on suurempi kuin naisen (0). Iän negatiivinen kerroin puolestaan osoittaa, että nuoremmilla oli keskimäärin suuremmat yrittäjyysaikomukset kuin vanhemmilla vastaajilla.

Tulokset voisi koota seuraavanlaiseksi taulukoksi. Mallissa a on mukana vain taustamuuttujat. Siinä selitysaste on huono (12 %). Mallissa b on mukana sekä taustamuuttujat että jatkuvat selittävät muuttujat.

LINEAARINEN REGRESSIO Selitettävänä muuttujana AIKOMUKSET		
n = 3430	Malli a	Malli b
Kontrollimuuttujat		
Ikä	-0,10***	-0,06***
Sukupuoli (mies)	0,17***	0,11***
Äidin yrittäjyys	0,14***	0,07***
Isän yrittäjyys	0,23***	0,10***
Jatkuvat selittävät muuttujat		
LÄHEISTEN MIELIPITEET		0,39***
PYSTYVYYS YRITTÄJÄNÄ TOIMIMISEEN		0,17***
ASENTEET		0,27***
YRITTÄJÄMÄISYYS		0,06***
Model fit statistics		
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		

Jos olisi tarkasteltu erikseen ensimmäisen vuosikurssin ja jatkavien vuosikurssien yrittäjyysaikomuksia ja mukaan olisi otettu muitakin taustamuuttujia, olisi saatu 2. – 4. vuosikurssin opiskelijoilla (n=1541) selitysaste vähän paremmaksi (61 %). Tilastollisesti merkitseväksi selittäjäksi olisi tullut mukaan *työskentely opintojen ohella*.