

Erkki K. Laitinen, professori
Laskentatoimen ja rahoituksen yksikkö

Luento löytyy netistä:

<http://www.uwasa.fi/laskentatoimi/henkilökunta/laitinenerkkik/>

Luennon aiheet:

- 1. Faktorianalyysi**
- 2. Logistinen regressioanalyysi**

1. FAKTORIANALYYSI

Luennon sisältö:

- 1.1. Faktorianalyysin idea**
- 1.2. Faktorianalyysin vaiheet**
- 1.3. Faktorianalyysin muodot**
- 1.4. Peruskäsitteet**
- 1.5. Esimerkit**

survival181203.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples

	Name	Type	Label	Values	Missing	Columns	Align	Measure
1	lytunnus	Numeric		None	None	11	Right	Nominal
2	ymimi	String		None	None	72	Left	Nominal
3	tol	Numeric		None	None	8	Right	Nominal
4	henkluok	Numeric	Henkilömäärä kpl	None	None	8	Right	Nominal
5	postinro	Numeric		None	None	8	Right	Nominal
6	kunta	Numeric	Kuntakoodi	None	None	8	Right	Nominal
7	maakunta	Numeric		None	None	8	Right	Nominal
8	laani	Numeric		None	None	8	Right	Nominal
9	haikoodi	String		None	None	11	Left	Nominal
10	rekprm	Date		None	None	16	Right	Nominal
11	hvuosi	Numeric		None	None	8	Right	Scale
12	hkuukaus	Numeric		None	None	8	Right	Scale
13	maksuhai	Numeric	Maksuhairio = 1, Ei hairiota = 0	None	None	8	Right	Scale
14	konkurs	Numeric	Konkurssi = 1, häiriötön = 0	None	None	8	Right	Scale
15	saneer	Numeric	Saneeraus = 1, häiriötön = 0	None	None	8	Right	Scale
16	muumak	Numeric	Muu häiriö kuin konkurssi, san	None	None	8	Right	Scale
17	ikamaks	Numeric	Yrityksen ikä maksuhäiriövuon	None	None	8	Right	Scale
18	selvit	Numeric	Selvitystila = 1, 0 = häiriötön	None	None	8	Right	Scale
19	groups	Numeric	1=häir+ei puutu+ei sel, 2=terve	None	None	8	Right	Scale
20	bigroup	Numeric	1 =häir+ei puutu+ei sel, 2=terv	None	None	8	Right	Scale
21	rakent	Numeric		None	None	8	Right	Scale
22	valmist	Numeric		None	None	8	Right	Scale
23	kauppa	Numeric		None	None	8	Right	Scale
24	hotrav	Numeric		None	None	8	Right	Scale
25	liikevar	Numeric		None	None	8	Right	Scale
26	palvelu	Numeric		None	None	8	Right	Scale
27	toimiala	Numeric	tol: 1=val, 2=rak, 3=kaup, 4=m	None	None	8	Right	Scale
28	year	Numeric		None	None	8	Right	Scale
29	vuosi	Numeric		None	None	8	Right	Scale
30	liikev	Numeric		None	None	8	Right	Scale
31	Inliikev	Numeric		None	None	8	Right	Scale
32	sinntu	Numeric		None	None	8	Right	Scale

Factor

SPSS Processor is ready

start | empirinenlu... | kyselysuomi... | auditqualific... | 2 Netscp | 2 Microsof... | survival181... | Output2 - S... | FI | 12:01

survival181203.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

Factor Analysis

Variables:

Selection Variable:

Value...

Descriptives... Extraction... Rotation... Scores... Options...

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1 lytunnus	Numeric	11	0		None	None	11	Right	Nominal
2 yritys	Numeric	72	0		None	None	72	Left	Nominal
3 tol	Numeric	8	0		None	None	8	Right	Nominal
4 he	Numeric	8	0		None	None	8	Right	Nominal
5 po	Numeric	8	0		None	None	8	Right	Nominal
6 ku	Numeric	8	0		None	None	8	Right	Nominal
7 ma	Numeric	8	0		None	None	8	Right	Nominal
8 laa	Numeric	8	0		None	None	8	Right	Nominal
9 ha	Numeric	11	0		None	None	11	Left	Nominal
10 rek	Numeric	16	0		None	None	16	Right	Nominal
11 hvu	Numeric	8	0		None	None	8	Right	Scale
12 hku	Numeric	8	0		None	None	8	Right	Scale
13 ma	Numeric	8	0		None	None	8	Right	Scale
14 ko	Numeric	8	0		None	None	8	Right	Scale
15 sa	Numeric	8	0		None	None	8	Right	Scale
16 mu	Numeric	8	0		None	None	8	Right	Scale
17 ika	Numeric	8	0		None	None	8	Right	Scale
18 selvit	Numeric	8	2	Selvitystila = 1, 0 = häiriötön	None	None	8	Right	Scale
19 groups	Numeric	8	2	1=häir+ei puutu+ei sel, 2=terve	None	None	8	Right	Scale
20 bigroup	Numeric	8	2	1=häir+ei puutu+ei sel, 2=terv	None	None	8	Right	Scale
21 rakent	Numeric	8	2		None	None	8	Right	Scale
22 valmist	Numeric	8	2		None	None	8	Right	Scale
23 kauppa	Numeric	8	2		None	None	8	Right	Scale
24 hotrav	Numeric	8	2		None	None	8	Right	Scale
25 liikevar	Numeric	8	2		None	None	8	Right	Scale
26 palvelu	Numeric	8	2		None	None	8	Right	Scale
27 toimiala	Numeric	8	2	tol: 1=val, 2=rak, 3=kaup, 4=m	None	None	8	Right	Scale
28 year	Numeric	4	0		None	None	8	Right	Scale
29 vuosi	Numeric	8	2		None	None	8	Right	Scale
30 liikev	Numeric	8	2		None	None	8	Right	Scale
31 Inliikev	Numeric	8	2		None	None	8	Right	Scale
32 suntu	Numeric	8	2		None	None	8	Right	Scale

SPSS Processor is ready

start | empiirinenlu... | kyselysuomi... | auditqualific... | 2 Netscp | 2 Microsof... | survival181... | Output2 - S... | FI | 12:01

1.1. Faktorianalyysin idea

- havaintoaineisto: havaintoja enemmän kuin muuttujia, $n > m$
- useita keskenään korreloivia kvantitatiivisia muuttujia
- tavoitteena löytää havaintoyksikön ominaisuuksia kuvaavasta muuttujajoukosta:
 - piileviä yhdenmukaisuuksia: faktoreita
 - = keskenään korreloivia muuttujia ja niiden painotettuja yhdistelmiä
 - tulkita faktorit ja käyttää niitä jatkoanalyyseissä

Havaintomatriisi:

Havainto	Muuttuja 1	Muuttuja 2	...	Muuttuja m
1				
2				
...				
n				

- esimerkiksi 1:

- mitataan havaittua ympäristön epävarmuutta (PEU, perceived environmental uncertainty)
- kuinka hyvin pystytään ennustamaan kehitystä 7 eri osa-alueella (tuotteet, hinnat, työvoima, teknologia, ...)
- vastaukset asteikolla 1 = ei yhtään ... 5 = täysin
- = construct = "konstruktio" = mittaristo
- yhdistetään vastausten informaatio yhdeksi faktoriksi
- käytetään tätä faktoria selittämään laskentajärjestelmien käyttöönottoa

- esimerkiksi 2:

- yrityksen taloudellista tilaa mitataan 20 tunnusluvulla
- näistä tunnusluvuista löytyy viisi erillistä ulottuvuutta
- yhdistetään tunnuslukujen informaatio viiteen faktoriin
- valitaan kutakin faktoria vastaava paras tunnusluku
- tunnuslukujen määrä karsiutuu viiteen

- perusidea siis:

- faktorianalyysi tutkii muuttujien välisiä suhteita**
- korrelaatioiden perusteella muuttujat pelkistetään muutamaaan harvaan faktoriin**
- ne muuttujat, joiden välillä on voimakas korrelaatio ja jotka ovat käsitteellisesti lähellä toisiaan, mittaavat samaa taustalla olevaa ulottuvuutta, faktoria**

1.2. Faktorianalyysin vaiheet

- hypoteesien (teorian) kehittäminen ilmiön ulottuvuuksista**
 - määrä**
 - luonne**
 - riippuvuus**
- tutkittavan havaintoyksikön ja niiden lukumäärän (n) valinta**
- mitattavien muuttujien (ominaisuuksien) valinta: m**
- muuttujien mittaaminen**
- korrelaatioiden tutkiminen**
- faktorointi**
- faktoreiden (dimensioiden) lukumäärästä päättäminen**
- faktoreiden rotatointi**
- faktoreiden tulkinta**
- faktoripistemäärien muodostaminen**
- jatkoanalyysien tekeminen**

1.3. Faktorianalyysin muodot

- 1) eksploratiivinen faktorianalyysi (EFA)

- muuttujien korrelaatorakenteen selittäminen (exploratory factor analysis)
- aineiston dimensioiden etsintä
- muuttujien määrän vähentäminen
 - ei hypoteeseja
 - "aineistolähtöinen"
 - eksploratiivinen

- perusidea siis:

- pyritään kuvaamaan aineistoa muuttujia harvalukuisemmilla faktoreilla
- pyritään etsimään muuttujajoukosta faktoreita, jotka pystyvät selittämään havaittujen muuttujien vaihtelua ilman, että tutkijalla on etukäteen vahvoja odotuksia löydettävien faktoreiden määrästä tai niiden tulkinnasta

- 2) konfirmatorinen faktorianalyysi (CFA)

- faktorien lukumäärää ja rakennetta koskevien hypoteesien testaus (confirmatory factor analysis)
- aineiston dimensioiden testaus
 - hypoteesien, teorian testaus
 - "rakenneteorialähtöinen"
 - konfirmatorinen

- perusidea siis:

- tutkijalla on jo etukäteen teorian pohjalta muodostettu käsitys aineiston faktorirakenteesta

- analyysin tehtävänä on joko varmistaa tai kumota tämä käsitys empiirisen aineiston pohjalta

- 3) rakenneyhtälömallianalyysi (SEM)

- faktoreiden välisten kausaalisten suhteiden testaus regressioanalyysin avulla (structural equation models)

- dimensioiden välisten kausaalisuhteiden testaus

- kausaalimallin testaus

- "kausaaliteorialähtöinen"

- SEM: rakenneyhtälöanalyysi

- faktorit, kausaalisuus, regressioanalyysi

- AMOS, LISREL, PLS

- perusidea siis:

- konfirmatorisen faktorianalyysin laajennus, jossa yhdistyvät faktori- ja regressioanalyysi

- regressioanalyysin avulla tutkitaan faktoreiden välistä kausaalisuutta

1.4. Peruskäsitteet

- faktori (factor)

- piilomuuttuja: sitä ei voida suoraan havainnoida, vaan sen olemassaolo päätellään ainoastaan havaittujen muuttujien avulla

- hypoteettinen konstruktio tai teoreettinen käsite, jonka olemassaolo päätellään konkreettisista havainnoista

- käytännössä faktorin muodostaa joukko muuttujia, jotka korreloivat vahvasti keskenään, mutta vähän muiden muuttujien kanssa.

- muuttujan saama lataus faktorilla (loading)

- muuttujan ja faktorin välinen korrelaatio ($-1 \leq \text{lataus} \leq 1$)

- jos korrelaatio on itseisarvoltaan suuri, sanotaan muuttujan latautuvan faktorille
- latauksen neliö vastaa siten selitysosuutta eli kertoo, kuinka suuren osan tietyn muuttujan vaihtelusta laskettu faktori pystyy kattamaan
- faktorilatausmatriisi: kunkin faktorianalyysissä mukana olleen muuttujan korrelaatio laskettujen faktoreiden kanssa

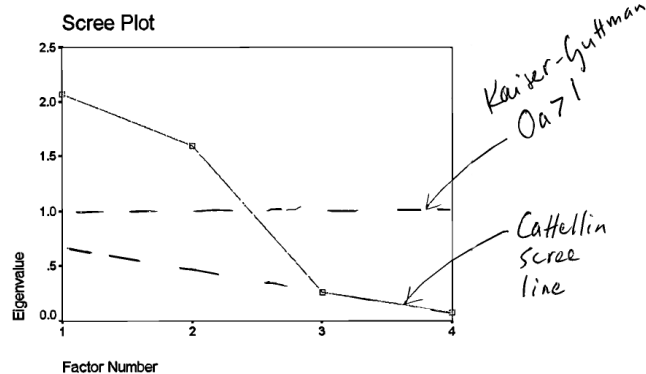
	Faktorit:			
Muuttujat:	F_1	...	F_k	Summa
X_1	lataus_{11}^2		lataus_{1k}^2	komm_1
X_2	lataus_{21}^2		lataus_{2k}^2	komm_2
..				
X_m	lataus_{m1}^2		lataus_{mk}^2	komm_m
Summa	omarvo_1	...	omarvo_k	

- muuttujan kommunaliteetti (communality)

- latausten neliöiden (selityssasteiden) summa rivin (faktoreiden) yli
- osoittaa kuinka paljon faktorit yhteensä selittävät muuttujan vaihtelusta

- faktorin ominaisarvo (eigenvalue)

- latausten neliöiden (selityssasteiden) summa sarakkeen (muuttujien) yli
- kun jaetaan muuttujien lukumäärällä m , saadaan faktorin selityssaste
- se osoittaa kuinka suuren osan faktori selittää kaikkien muuttujien vaihtelusta
- kriteeri faktorin mukaanotolle
- faktoreiden selityssasteiden summa: koko mallin selityssaste
- faktoreiden lukumäärä perustuu ominaisarvoon:
 - Cattelin scree-testi: käännepiste
 - Kaiser-Guttman: < 1



- havaintoyksikön faktoripistemäärä (factor score)

- havainnon arvo uudella tiivistetyllä muuttujalla, faktorilla
- painotettu keskiarvo alkuperäisten muuttujien standardoiduista arvoista
- painoina faktorilataukset
- faktoripistemäärien keskiarvo = 0

- faktoriratkaisun rotatointi (rotation)

- tulkinnan helpottaminen
- kierretään faktorien muodostamaa koordinaatistoa muuttujajoukossa siten, että saadaan jonkin kriteerin mukaan paras ratkaisu
- eräs kriteeri parhaalle ratkaisulle on "yksinkertainen rakenne" (simple structure), jossa kukin muuttuja latautuu selvästi vain yhdellä faktorilla
- esimerkiksi varimax: maksimoidaan latausten neliöiden varianssit kullakin faktorilla ("varimax")
 - pyritään saamaan samalle faktorille mahdollisimman suuria ja mahdollisimman pieniä latauksia, jolloin niiden

(latausten faktorimatriisin sarakkeilla) vaihtelu, varianssi, on niin suuri kuin mahdollista

- selitysosuuksien summa pysyy ennallaan
- selitysosuuden jakautuminen faktoreille muuttuu
- rotaatio voi olla suorakulmainen (orthogonal) tai vino (oblique)
- suorakulmaisessa rotaatiossa faktorit toisistaan riippumattomia
- vinokulmaisessa ei riippumattomia

Rotatointimenetelmät (ROTATION-optio)	
Optio	Menetelmä
VARIMAX	<i>Varimax</i> -rotatointi
QUARTIMAX	<i>Quartimax</i> -rotatointi
EQUAMAX	<i>Equamax</i> -rotatointi
ORTHOMAX	yleinen <i>ortomax</i> -rotatointi GAMMA-option määrittämällä painolla.
PARSIMAX	Suorakulmainen <i>Parsimax</i> -rotatointi
HK	<i>Harris-Kaiser II. orthoblique</i> -rotatointi
PROMAX	<i>Promax</i> -rotatointi
NONE	ei rotatoida

- estimointimenetelmä

- menetelmä, jonka avulla faktoriratkaisu estimoidaan

- esimerkiksi:

- suurimman uskottavuuden-menetelmä (maximum likelihood factor analysis)
- pääakselimenetelmä (principal factor analysis)
- pääkomponenttianalyysi (principal component analysis)

1.5. ESIMERKKI: KASVIHUONEYRITYSTEN TALOUDELLISET DIMENSIOT

1.5.1. HAVAINTOAINEISTO

Yritykset: 178 kasvihuonealan yritystä

Muuttujat:

OMAVARAS	OMAVARAISUUSASTE
VPTAKA	RAHOITUSTULOS / VPO
MYSAKN	MYyntISAAMISTEN KIERTOAIKA
OVELKN	OSTOVELKOJEN KIERTOAIKA
RAHTULPR	RAHOITUSTULOS-%
QUICKRAT	QUICK RATIO
QURRAT	CURRENT RATIO
OMAPOPR	OMAN PÄÄOMAN TUOTTO-%
TULOSPR	TULOS-% LIIKEVAIHDOSTA
JALARPR	JALOSTUSARVO-%
KKATEPR	KÄYTTÖKATE-%
TUOTTOPR	PÄÄOMAN TUOTTO-%
LVAIHTO	LIIKEVAIHTO
TASELSU	TASEEN LOPPUSUMMA
LVKASVU	LIIKEVAIHDON KASVU-%
TASKASVU	TASEEN KASVU-%

1.5.2. TUTKIMUKSEN ONGELMA

a) Kysymys: mitä dimensioita aineistosta löytyy?
miten voisit karsia muuttujien lukumäärää?

- exploratiivinen faktorianalyysi

b) Mahdollista: miettiä rakenneteoriaa etukäteen

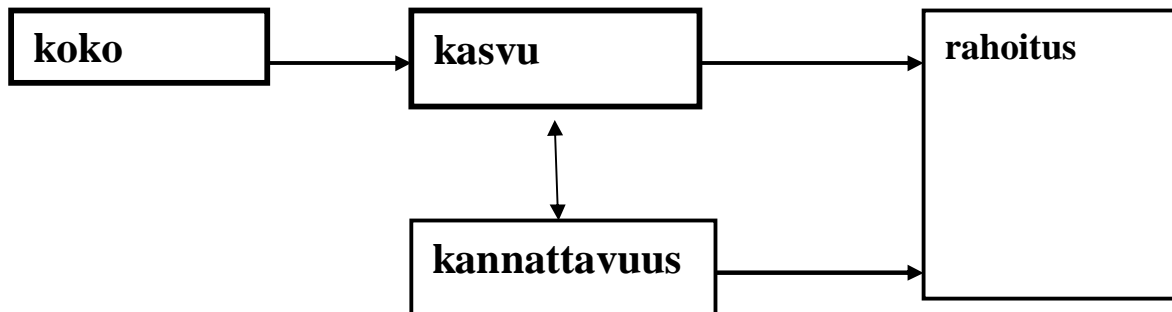
- tilinpäätösanalyysin teoria
- dimensiot: koko, kasvu, kannattavuus, rahoitus

Havaintoaineiston muuttujat	Mahdolliset faktorit
	1. koko
	2. kasvu
	3. kannattavuus
	4. rahoitus

- konfirmatorinen faktorianalyysi

c) Mahdollista: miettiä kausaalista teoriaa etukäteen

- tilinpäätösanalyysin ja rahoituksen teoria
- hypoteesit:
 - samat faktorit
 - koko vaikuttaa kasvuun
 - kasvu ja kannattavuus vaikuttavat rahoitukseen
 - kasvu ja kannattavuus vaikuttavat toisiinsa



- rakenneyhtälömallitarkastelu (SEM)

1.5.3. ANALYYSIN VAIHEET

VAIHE a): KESKIVARVOT JA HAJONNAT

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
OMAVARAS	178	31.78615	48.70233	5658	-303.94737	95.20548
VPTAKA	178	56.65970	106.42031	10085	-121.42857	790.47619
MYSAKN	178	25.57108	23.56540	4552	0	148.82831
OVELKN	176	85.20791	95.74225	14997	0	618.21875
RAHTULPR	178	10.79082	44.04406	1921	-540.90909	46.65511
QUICKRAT	178	1.76675	1.94753	314.48142	0.05338	11.52632
QURRAT	178	2.44682	2.47333	435.53338	0.18727	15.36667
OMAPOPR	178	13.09422	628.93216	2331	-6733	4720
TULOSPR	178	6.65255	45.81985	1184	-559.09091	61.19196
JALARPR	178	33.74673	35.38668	6007	-368.18182	88.28423
KKATEPR	178	15.83956	44.04115	2819	-527.27273	64.78336
TUOTTOPR	178	14.20573	33.62078	2529	-263.04348	98.13412
LVAIHTO	178	6757	7628	1202825	22.00000	41386
TASELSU	178	7331	11794	1304962	46.00000	75547
LVKASVU	134	18.00430	50.64057	2413	-39.05160	363.63636
TASKASVU	134	14.43853	51.29733	1935	-49.33333	516.59824

- outlierit

VAIHE b): KORRELAATIOT

Pearson Correlation Coefficients
 Prob > |r| under H0: Rho=0
 Number of Observations

	OMAVARAS	VPTAKA	MYSAKN	OVELKN
OMAVARAS	1.00000	0.44955	0.03004	-0.23038
OMAVARAISUUSASTE		<.0001	0.6906	0.0021
	178	178	178	176
VPTAKA	0.44955	1.00000	-0.09906	-0.22721
RAHOITUSTULOS / VPO	<.0001		0.1883	0.0024
	178	178	178	176
MYSAKN	0.03004	-0.09906	1.00000	0.10481
MYNTISAAMISTEN KIERTOAIKA	0.6906	0.1883		0.1662
	178	178	178	176
OVELKN	-0.23038	-0.22721	0.10481	1.00000
OSTOVELKOJEN KIERTOAIKA	0.0021	0.0024	0.1662	
	176	176	176	176
RAHTULPR	0.26201	0.27373	0.03554	-0.00099
RAHOITUSTULOS-%	0.0004	0.0002	0.6376	0.9896
	178	178	178	176
QUICKRAT	0.49671	0.60357	-0.06427	-0.37307
QUICK RATIO	<.0001	<.0001	0.3940	<.0001
	178	178	178	176
QURRAT	0.42518	0.46816	-0.05672	-0.36298
CURRENT RATIO	<.0001	<.0001	0.4520	<.0001
	178	178	178	176
OMAPOPR	0.01021	0.01786	-0.01309	-0.10857
OMAN PÄÄOMAN TUOTTO-%	0.8924	0.8129	0.8623	0.1515
	178	178	178	176
TULOSPR	0.28672	0.31434	0.03485	-0.03606
TULOS-% LIIKEVAIHDESTA	0.0001	<.0001	0.6442	0.6347
	178	178	178	176
JALARPR	0.26124	0.23555	0.11120	0.01125
JALOSTUSARVO-%	0.0004	0.0015	0.1395	0.8822
	178	178	178	176
KKATEPR	0.25808	0.28905	0.04630	0.03364
KÄYTTÖKATE-%	0.0005	<.0001	0.5394	0.6576
	178	178	178	176
TUOTTOPR	0.38987	0.44584	0.00331	-0.01311
PÄÄOMAN TUOTTO-%	<.0001	<.0001	0.9651	0.8629
	178	178	178	176
OMAVARAS	0.26201	0.49671	0.42518	0.01021
OMAVARAISUUSASTE	0.0004	<.0001	<.0001	0.8924
	178	178	178	178

VPTAKA	0.27373	0.60357	0.46816	0.01786
RAHOITUSTULOS / VPO	0.0002	<.0001	<.0001	0.8129
	178	178	178	178
MYSAKN	0.03554	-0.06427	-0.05672	-0.01309
MYNTISAAMISTEN KIERTOAIKA	0.6376	0.3940	0.4520	0.8623
	178	178	178	178
OVELKN	-0.00099	-0.37307	-0.36298	-0.10857
OSTOVELKOJEN KIERTOAIKA	0.9896	<.0001	<.0001	0.1515
	176	176	176	176
RAHTULPR	1.00000	0.09949	0.04010	-0.02138
RAHOITUSTULOS-%		0.1864	0.5952	0.7769
	178	178	178	178
QUICKRAT	0.09949	1.00000	0.87178	0.01498
QUICK RATIO	0.1864		<.0001	0.8427
	178	178	178	178
QURRAT	0.04010	0.87178	1.00000	0.04866
CURRENT RATIO	0.5952	<.0001		0.5190
	178	178	178	178
OMAPOPR	-0.02138	0.01498	0.04866	1.00000
OMAN PÄÄOMAN TUOTTO-%	0.7769	0.8427	0.5190	

- runsaasti: vaikea saada kokonaiskuvaa

VAIHE c): FAKTORIT

The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

OMINAISARVOT

Eigenvalues of the Correlation Matrix: Total = 16 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	4.96671205	2.61272777	0.3104	0.3104
2	2.35398428	0.11418907	0.1471	0.4575
3	2.23979521	0.96521782	0.1400	0.5975
4	1.27457740	0.19050532	0.0797	0.6772
5	1.08407208	0.12759342	0.0678	0.7449
6	0.95647866	0.17309962	0.0598	0.8047
7	0.78337904	0.14781791	0.0490	0.8537
8	0.63556112	0.07423185	0.0397	0.8934
9	0.56132927	0.15663105	0.0351	0.9285
10	0.40469822	0.13249718	0.0253	0.9538
11	0.27220104	0.03722613	0.0170	0.9708
12	0.23497492	0.15153509	0.0147	0.9855
13	0.08343982	0.01415125	0.0052	0.9907
14	0.06928857	0.00980479	0.0043	0.9950
15	0.05948379	0.03945926	0.0037	0.9987
16	0.02002453		0.0013	1.0000

4 factors will be retained by the NFACTOR criterion.

- faktorien määrä rajattiin neljään

FAKTORIT PERUSMUODOSSA

Factor Pattern

		Factor1	Factor2
OMAVARAS	OMAVARAISUUSASTE	0.55770	0.34672
VPTAKA	RAHOITUSTULOS / VPO	0.72096	0.18749
MYSAKN	MYYNNTISAAMISTEN KIERTOAIKA	-0.03108	0.15005
OVELKN	OSTOVELKOJEN KIERTOAIKA	-0.42951	-0.27512
RAHTULPR	RAHOITUSTULOS-%	0.85558	-0.34611
QUICKRAT	QUICK RATIO	0.63066	0.63677
QURRAT	CURRENT RATIO	0.51847	0.69719
OMAPOPR	OMAN PÄÄOMAN TUOTTO-%	0.06125	-0.00526
TULOSPR	TULOS-% LIIKEVAIHDOSTA	0.90579	-0.25120
JALARPR	JALOSTUSARVO-%	0.60902	-0.33189
KKATEPR	KÄYTTÖKATE-%	0.83757	-0.41816
TUOTTOPR	PÄÄOMAN TUOTTO-%	0.75900	-0.26834
LVAIHTO	LIIKEVAIHTO	0.13880	0.44777
TASELSU	TASEEN LOPPUSUMMA	0.13788	0.50493
LVKASVU	LIIKEVAIHDON KASVU-%	-0.05158	-0.28288
TASKASVU	TASEEN KASVU-%	0.19555	-0.36530
		Factor3	Factor4
OMAVARAS	OMAVARAISUUSASTE	-0.17903	0.25322
VPTAKA	RAHOITUSTULOS / VPO	-0.42454	-0.05627
MYSAKN	MYYNNTISAAMISTEN KIERTOAIKA	0.60590	0.00829
OVELKN	OSTOVELKOJEN KIERTOAIKA	0.07665	0.06567
RAHTULPR	RAHOITUSTULOS-%	0.20839	-0.16697
QUICKRAT	QUICK RATIO	-0.27622	0.08084
QURRAT	CURRENT RATIO	-0.35773	0.06380
OMAPOPR	OMAN PÄÄOMAN TUOTTO-%	-0.04065	-0.14246
TULOSPR	TULOS-% LIIKEVAIHDOSTA	0.12447	-0.09494
JALARPR	JALOSTUSARVO-%	0.37158	-0.12641
KKATEPR	KÄYTTÖKATE-%	0.17316	-0.15709
TUOTTOPR	PÄÄOMAN TUOTTO-%	-0.06442	0.18882
LVAIHTO	LIIKEVAIHTO	0.77197	0.16167
TASELSU	TASEEN LOPPUSUMMA	0.78425	0.17761
LVKASVU	LIIKEVAIHDON KASVU-%	-0.05388	0.71642
TASKASVU	TASEEN KASVU-%	-0.05985	0.69848

OMINAISARVOT

Variance Explained by Each Factor

Factor1	Factor2	Factor3	Factor4
4.9667120	2.3539843	2.2397952	1.2745774

KOMMUNALITEETIT

Final Commuality Estimates: Total = 10.835069

OMAVARAS	VPTAKA	MYSAKN	OVELKN	RAHTULPR	QUICKRAT
0.52741197	0.73833471	0.39066671	0.27035681	0.92312471	0.88603714
QURRAT	OMAPOPR	TULOSPR	JALARPR	KKATEPR	TUOTTOPR
0.88692424	0.02572725	0.90805785	0.63511282	0.93104876	0.68788826
LVAIHTO	TASELSU	LVKASVU	TASKASVU		
0.84184612	0.92055483	0.59883885	0.66313792		

ROTATOINTI: VARIMAX

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2	3	4
1	0.80175	0.59640	0.03547	0.01568
2	-0.52128	0.68182	0.43737	-0.26852
3	0.24989	-0.38799	0.88676	-0.02594
4	-0.15171	0.16999	0.14530	0.96280

Rotated Factor Pattern

		Factor1	Factor2
OMAVARAS	OMAVARAISUUSASTE	0.18325	0.68152
VPTAKA	RAHOITUSTULOS / VPO	0.38274	0.71297
MYSAKN	MYNTISAAMISTEN KIERTOAIKA	0.04701	-0.14990
OVELKN	OSTOVELKOJEN KIERTOAIKA	-0.19175	-0.46231
RAHTULPR	RAHOITUSTULOS-%	0.94379	0.16504
QUICKRAT	QUICK RATIO	0.09240	0.93120
QURRAT	CURRENT RATIO	-0.04682	0.93421
OMAPOPR	OMAN PÄÄOMAN TUOTTO-%	0.06330	0.02449
TULOSPR	TULOS-% LIIKEVAIHDESTA	0.90267	0.30450
JALARPR	JALOSTUSARVO-%	0.77333	-0.02873
KKATEPR	KÄYTTÖKATE-%	0.95661	0.12053
TUOTTOPR	PÄÄOMAN TUOTTO-%	0.70366	0.32680
LVAIHTO	LIIKEVAIHTO	0.04626	0.11605
TASELSU	TASEEN LOPPUSUMMA	0.01637	0.15241
LVKASVU	LIIKEVAIHDON KASVU-%	-0.01604	-0.08095
TASKASVU	TASEEN KASVU-%	0.22629	0.00952

Rotated Factor Pattern

		Factor3	Factor4
OMAVARAS	OMAVARAISUUSASTE	0.04946	0.16408
VPTAKA	RAHOITUSTULOS / VPO	-0.27707	-0.08221
MYSAKN	MYNTISAAMISTEN KIERTOAIKA	0.60302	-0.04851
OVELKN	OSTOVELKOJEN KIERTOAIKA	-0.05805	0.12839
RAHTULPR	RAHOITUSTULOS-%	0.03950	-0.05981
QUICKRAT	QUICK RATIO	0.06768	-0.07610
QURRAT	CURRENT RATIO	0.01536	-0.10838
OMAPOPR	OMAN PÄÄOMAN TUOTTO-%	-0.05688	-0.13374
TULOSPR	TULOS-% LIIKEVAIHDESTA	0.01884	-0.01298
JALARPR	JALOSTUSARVO-%	0.18757	-0.03267
KKATEPR	KÄYTTÖKATE-%	-0.02245	-0.03032
TUOTTOPR	PÄÄOMAN TUOTTO-%	-0.12013	0.26742
LVAIHTO	LIIKEVAIHTO	0.90881	0.01757
TASELSU	TASEEN LOPPUSUMMA	0.94697	0.01724
LVKASVU	LIIKEVAIHDON KASVU-%	-0.06924	0.76631
TASKASVU	TASEEN KASVU-%	-0.10442	0.77520

Tulkinta:

1. Faktori: KANNATTAVUUS

MITTARI: käyttökatte-%

2. Faktori: TAKAISINMAKSUKYKY

MITTARI: current ratio

3. Faktori: KOKO

MITTARI: taseen loppusumma

4. Faktori: KASVU

MITTARI: taseen kasvu-%

OMINAISARVOT

Variance Explained by Each Factor

Factor1	Factor2	Factor3	Factor4	
4.0014911	3.2349205	2.2446919	1.3539654	: 16 = m
25.0 %	20.2 %	14.0 %	8.4 %	

KOMMUNALITEETIT

Final Communalities Estimates: Total = 10.835069

OMAVARAS	VPTAKA	MYSAKN	OVELKN	RAHTULPR	QUICKRAT
0.52741197	0.73833471	0.39066671	0.27035681	0.92312471	0.88603714
QURRAT	OMAPOPR	TULOSPR	JALARPR	KKATEPR	TUOTTOPR
0.88692424	0.02572725	0.90805785	0.63511282	0.93104876	0.68788826
LVAIHTO	TASELSU	LVKASVU	TASKASVU		
0.84184612	0.92055483	0.59883885	0.66313792		

FAKTORIPISTEMÄÄRIEN LASKEMINEN

Scoring Coefficients Estimated by Regression

Squared Multiple Correlations of the Variables with Each Factor

Factor1	Factor2	Factor3	Factor4
1.0000000	1.0000000	1.0000000	1.0000000

Standardized Scoring Coefficients

		Factor1	Factor2
OMAVARAS	OMAVARAISUUSASTE	-0.03687	0.23218
VPTAKA	RAHOITUSTULOS / VPO	0.03419	0.20691
MYSAKN	MYYNTISAAMISTEN KIERTOAIKA	0.02837	-0.06412
OVELKN	OSTOVELKOJEN KIERTOAIKA	-0.00767	-0.13578
RAHTULPR	RAHOITUSTULOS-%	0.25788	-0.05588
QUICKRAT	QUICK RATIO	-0.07965	0.31880
QURRAT	CURRENT RATIO	-0.11820	0.33467
OMAPOPR	OMAN PÄÄOMAN TUOTTO-%	0.02347	-0.00613
TULOSPR	TULOS-% LIIKEVAIHDESTA	0.22703	0.00178
JALARPR	JALOSTUSARVO-%	0.22831	-0.10423
KKATEPR	KÄYTTÖKATE-%	0.26582	-0.07149
TUOTTOPR	PÄÄOMAN TUOTTO-%	0.15228	0.04976
LVAIHTO	LIIKEVAIHTO	-0.00987	0.03420
TASELSU	TASEEN LOPPUSUMMA	-0.02320	0.05064
LVKASVU	LIIKEVAIHDON KASVU-%	-0.03697	0.01676
TASKASVU	TASEEN KASVU-%	0.02265	0.02120

Standardized Scoring Coefficients

		Factor3	Factor4
OMAVARAS	OMAVARAISUUSASTE	0.02639	0.15556
VPTAKA	RAHOITUSTULOS / VPO	-0.13451	-0.05670
MYSAKN	MYYNTISAAMISTEN KIERTOAIKA	0.26849	-0.01797
OVELKN	OSTOVELKOJEN KIERTOAIKA	-0.01635	0.07875
RAHTULPR	RAHOITUSTULOS-%	0.00527	-0.08636
QUICKRAT	QUICK RATIO	0.02267	-0.00638
QURRAT	CURRENT RATIO	-0.00112	-0.02555
OMAPOPR	OMAN PÄÄOMAN TUOTTO-%	-0.03288	-0.10635
TULOSPR	TULOS-% LIIKEVAIHDESTA	-0.00175	-0.04164
JALARPR	JALOSTUSARVO-%	0.07539	-0.06001
KKATEPR	KÄYTTÖKATE-%	-0.02106	-0.07032
TUOTTOPR	PÄÄOMAN TUOTTO-%	-0.04842	0.17638
LVAIHTO	LIIKEVAIHTO	0.40825	0.06254
TASELSU	TASEEN LOPPUSUMMA	0.42554	0.06792
LVKASVU	LIIKEVAIHDON KASVU-%	0.00741	0.57390
TASKASVU	TASEEN KASVU-%	-0.01055	0.57060

JATKOANALYYSI

RYHMÄT SUURIMMAN FAKTORIPISTEMÄÄRÄN MUKAAN

The FREQ Procedure

FAKTORI	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	86	48.31	86	48.31
2	32	17.98	118	66.29
3	35	19.66	153	85.96
4	25	14.04	178	100.00

- klusterointi

KORRELAATIOT JA LISÄMUUTTUJA: IKÄ

The CORR Procedure

5 Variables: Factor1 Factor2 Factor3 Factor4 IKA

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Factor1	133	0	1.00000	0	-5.13677	3.22851
Factor2	133	0	1.00000	0	-2.36459	4.70156
Factor3	133	0	1.00000	0	-1.48571	4.87272
Factor4	133	0	1.00000	0	-3.17283	7.38774
IKA	178	16.77528	9.60373	2986	4.00000	34.00000

Pearson Correlation Coefficients

Prob > |r| under H0: Rho=0

Number of Observations

	Factor1	Factor2	Factor3	Factor4	IKA
Factor1	1.00000	0.00000	0.00000	0.00000	0.11874
		1.0000	1.0000	1.0000	0.1734
	133	133	133	133	133
Factor2	0.00000	1.00000	0.00000	0.00000	0.10677
	1.0000		1.0000	1.0000	0.2213
	133	133	133	133	133
Factor3	0.00000	0.00000	1.00000	0.00000	0.44451
	1.0000	1.0000		1.0000	<.0001
	133	133	133	133	133
Factor4	0.00000	0.00000	0.00000	1.00000	-0.11510
	1.0000	1.0000	1.0000		0.1871
	133	133	133	133	133
IKA	0.11874	0.10677	0.44451	-0.11510	1.00000
	0.1734	0.2213	<.0001	0.1871	
	133	133	133	133	178

FAKTORIANALYYSIN JA REGRESSIOANALYYSIN YHDISTÄMINEN

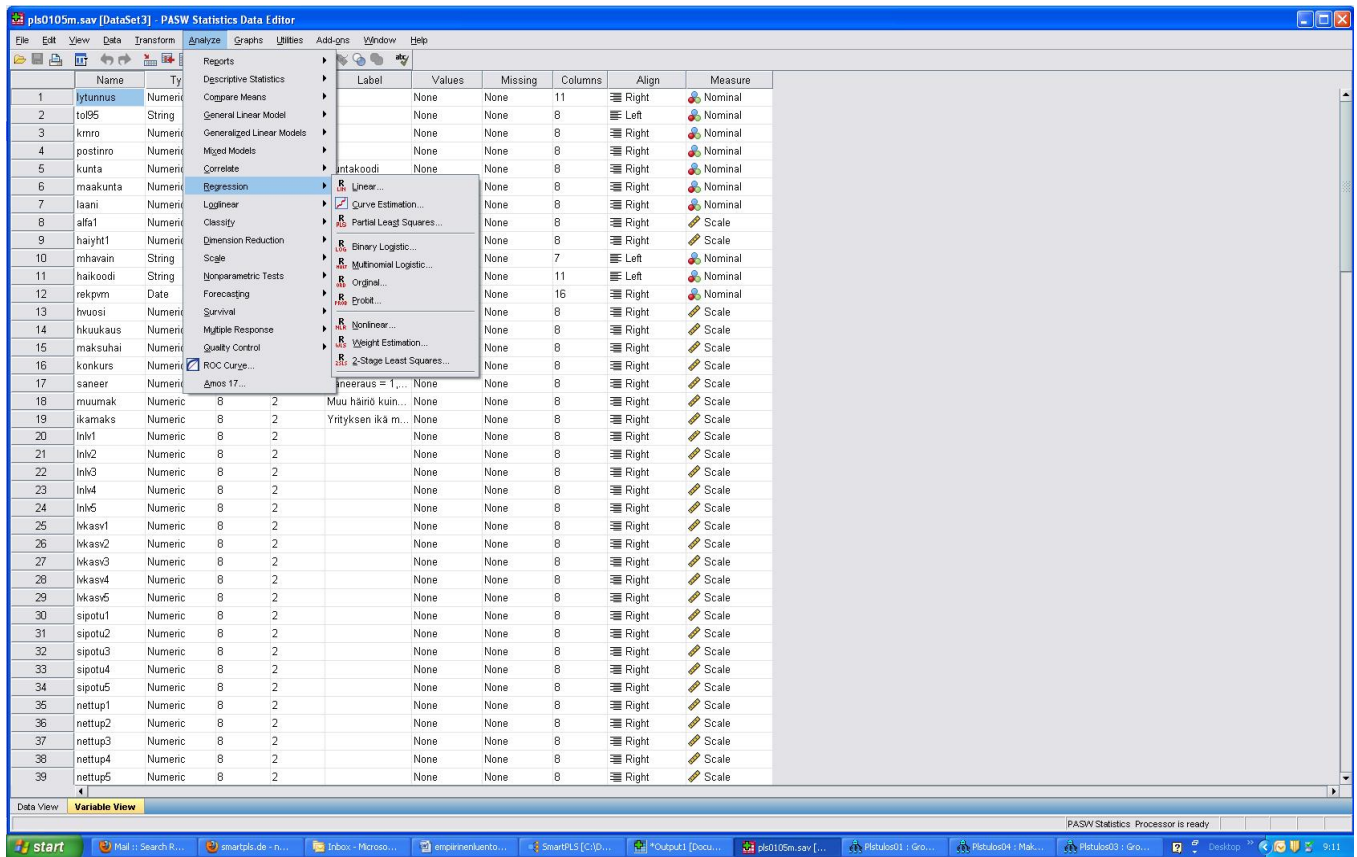
1. PLS

- osittaisen pienimmän neliösumman regressio (PLS, partial least squares regression) muistuttaa faktorianalyysiä, mutta siihen yhdistyy myös regressioanalyysi
- menetelmälle on olennaista se, että siinä etsitään sellainen faktoriratkaisu, joka selittää tehokkaimmin ennustettavaa muuttujaa
- esimerkiksi selitetään binääristä 0/1-muuttujalla mitattua maksuhäiriötunnusta tunnusluvuilla: tuloksena saadaan faktorit, jotka selittävät tehokkaimmin maksuhäiriötä
- pienet otokset: havaintoja voi olla vähemmän kuin muuttujia

Tunnuslukujen ja PLS-faktoreiden korrelaatio:

	PLS-faktori 1	PLS-faktori 2
Maksuhäiriötunnus	0.4081	0.2225
	<.0001	<.0001
Logaritminen liikevaihto	-0.2017	-0.3082
	<.0001	<.0001
Liikevaihdon kasvu-%	0.0961	0.4240
	<.0001	<.0001
Pääoman tuotto-%	-0.5716	0.4590
	<.0001	<.0001
Nettotulos-%	-0.7365	0.5081
	<.0001	<.0001
Quick ratio	-0.4731	-0.1622
	<.0001	<.0001
Rahoitustulos-%	-0.7072	0.5062
	<.0001	<.0001
Omavaraisuusaste	-0.7843	-0.4253
	<.0001	<.0001
Takaisinmaksukyky	-0.8099	0.1980
	<.0001	<.0001

- PLS löytyy SPSS:sta (extended version) regressioanalyysistä:

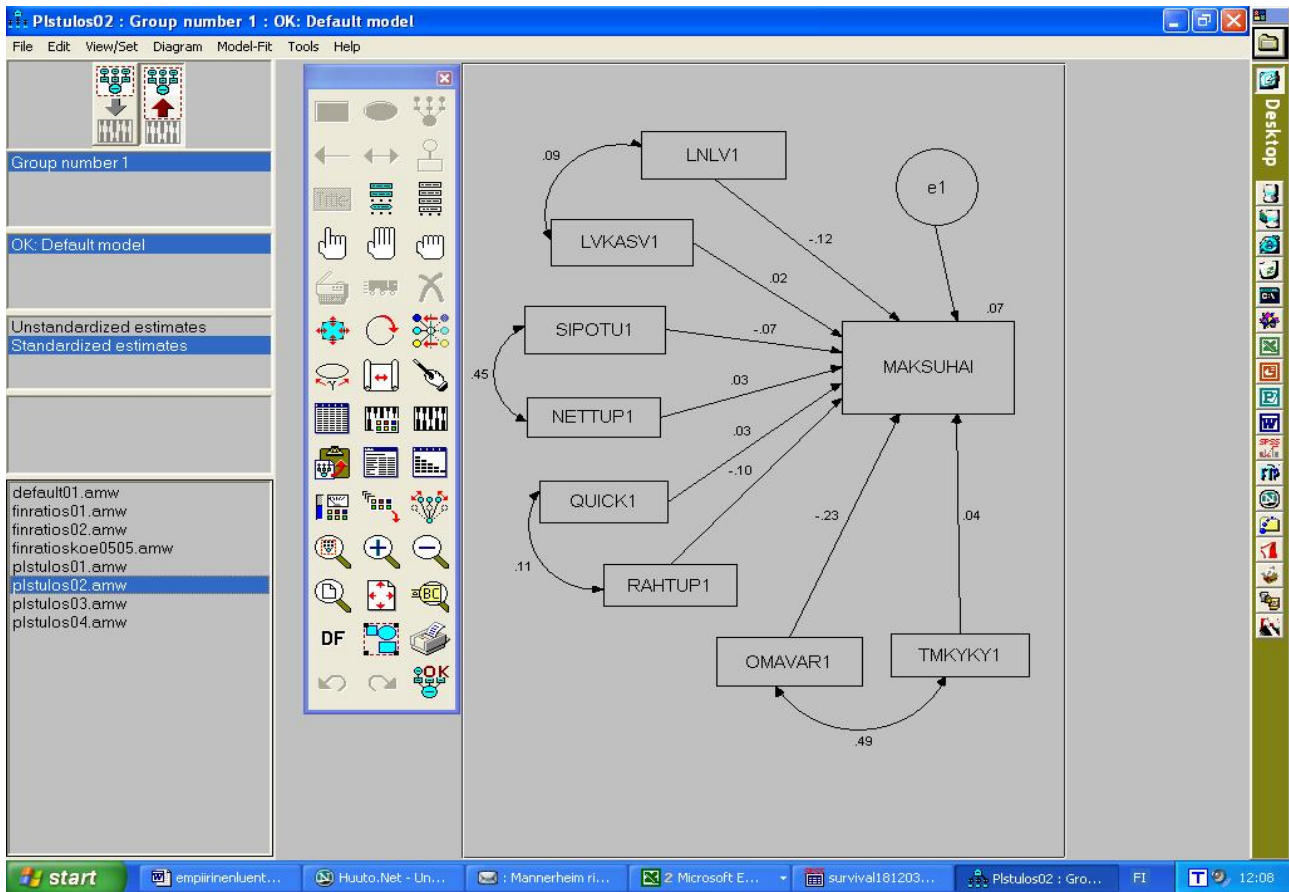


- PLS on varsinaisesti polkumalli, mutta sitä voidaan pitää myös SEM-mallina. Siihen löytyy hyvä ilmaisohjelma netistä, SMARTPLS

2. SEM

- moniyhtälömalli tai rakenneyhtälömalli (structural equation model, SEM) koostuu muuttujista, faktoreista, riippuvuuksista
- kuvataan kausaalisesti eri muuttujien tai niiden yhdistelmien (faktoreiden, konstruktioiden) väliset suhteet
- SPSS sisältää AMOS-nimisen SEM-mallin (lisävarusteena)

SEM: Esimerkki AMOS-mallista: yksinkertainen selitysmalli



Entä logistinen regressionanalyysi?

The screenshot shows the SPSS Data Editor interface. The main window displays a list of variables with their names, types, and measurement scales. The 'Analyze' menu is open, and the 'Regression' submenu is also open, showing options like Linear, Logistic, and Nonlinear regression.

Name	Type	Label	Values	Missing	Columns	Align	Measure	
1	lytunnus	Numeric		None	11	Right	Nominal	
2	ymimi	String		None	72	Left	Nominal	
3	tol	Numeric		None	8	Right	Nominal	
4	henkluok	Numeric		None	8	Right	Nominal	
5	postinro	Numeric		None	8	Right	Nominal	
6	kunta	Numeric		None	8	Right	Nominal	
7	maakunta	Numeric		None	8	Right	Nominal	
8	laani	Numeric		None	8	Right	Nominal	
9	haikoodi	String		None	11	Left	Nominal	
10	rekprm	Date		None	16	Right	Nominal	
11	hvuosi	Numeric		None	8	Right	Scale	
12	hkuukaus	Numeric		None	8	Right	Scale	
13	maksuhai	Numeric		None	8	Right	Scale	
14	konkurs	Numeric		None	8	Right	Scale	
15	saneer	Numeric	8	2	Saneeraus = 1, häiriötön = 0	8	Right	Scale
16	muumak	Numeric	8	2	Muu häiriö kuin konkurssi, san	8	Right	Scale
17	ikamaks	Numeric	8	2	Yrityksen ikä maksuhäiriövuon	8	Right	Scale
18	selvit	Numeric	8	2	Selvitystila = 1, 0 = häiriötön	8	Right	Scale
19	groups	Numeric	8	2	1=häir+ei puutu+ei sel, 2=terve	8	Right	Scale
20	bigroup	Numeric	8	2	1 =häir+ei puutu+ei sel, 2=terv	8	Right	Scale
21	rakent	Numeric	8	2		8	Right	Scale
22	valmist	Numeric	8	2		8	Right	Scale
23	kauppa	Numeric	8	2		8	Right	Scale
24	hotrav	Numeric	8	2		8	Right	Scale
25	liikevar	Numeric	8	2		8	Right	Scale
26	palvelu	Numeric	8	2		8	Right	Scale
27	toimiala	Numeric	8	2	tol: 1=val, 2=rak, 3=kaup, 4=m	8	Right	Scale
28	year	Numeric	4	0		8	Right	Scale
29	vuosi	Numeric	8	2		8	Right	Scale
30	liikev	Numeric	8	2		8	Right	Scale
31	Inliikev	Numeric	8	2		8	Right	Scale
32	sintu	Numeric	8	2		8	Right	Scale

survival181203.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

Logistic Regression

Dependent: Maksuhairio = 1, Ei hair

Block 1 of 1

Covariates: Inliikev, sipotu, nettup, quick, rahtup

Method: Enter

Selection Variable:

OK, Paste, Reset, Cancel, Help, Categorical..., Save..., Options...

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	
1					one	None	11	Right	Nominal	
2					one	None	72	Left	Nominal	
3					one	None	8	Right	Nominal	
4					one	None	8	Right	Nominal	
5					one	None	8	Right	Nominal	
6					one	None	8	Right	Nominal	
7					one	None	8	Right	Nominal	
8					one	None	8	Right	Nominal	
9					one	None	8	Right	Nominal	
10					one	None	8	Right	Nominal	
11					one	None	11	Left	Nominal	
12					one	None	16	Right	Nominal	
13					one	None	8	Right	Scale	
14					one	None	8	Right	Scale	
15					one	None	8	Right	Scale	
16					one	None	8	Right	Scale	
17					one	None	8	Right	Scale	
18	semit	Numeric	0	2	semitystyria = 1, 0 = hamaton	None	None	8	Right	Scale
19	groups	Numeric	8	2	1=hair+ei puutu+ei sel, 2=terve	None	None	8	Right	Scale
20	bigroup	Numeric	8	2	1=hair+ei puutu+ei sel, 2=terv	None	None	8	Right	Scale
21	rakent	Numeric	8	2		None	None	8	Right	Scale
22	valmist	Numeric	8	2		None	None	8	Right	Scale
23	kauppa	Numeric	8	2		None	None	8	Right	Scale
24	hotrav	Numeric	8	2		None	None	8	Right	Scale
25	liikevar	Numeric	8	2		None	None	8	Right	Scale
26	palvelu	Numeric	8	2		None	None	8	Right	Scale
27	toimiala	Numeric	8	2	tol: 1=val, 2=rak, 3=kaup, 4=m	None	None	8	Right	Scale
28	year	Numeric	4	0		None	None	8	Right	Scale
29	vuosi	Numeric	8	2		None	None	8	Right	Scale
30	liikev	Numeric	8	2		None	None	8	Right	Scale
31	Inliikev	Numeric	8	2		None	None	8	Right	Scale
32	sipotu	Numeric	8	2		Nnne	Nnne	8	Right	Scale

Data View Variable View

SPSS Processor is ready

start | empiirinenlu... | kyselysuomi... | auditqualific... | 2 Netscp | 2 Microsof... | survival181... | Output2 - S... | FI | 12:03

2. LOGISTINEN REGRESSIOANALYYSI

- esimerkkinä maksuhäiriön ennustaminen

Luennon sisältö:

- 2.1. Ennustettava muuttuja**
- 2.2. Selittävät muuttujat**
- 2.3. Otanta**
- 2.4. Menetelmä**
- 2.5. Testaus**
- 2.6. Esimerkki**

Menetelmän soveltamisen idea:

- pyritään ymmärtämään maksuhäiriöprosessi
- selitetään prosessin tulosta
- prosessia kuvaavilla tekijöillä
- ja mallinnetaan se
- testataan malli luotettavuuden selvittämiseksi

2.1. Ennustettava muuttuja

- maksuhäiriöprosessin tulos
- mahdollisimman yksikäsitteinen ja selkeä
- tavallisesti: "maksusitoumus jätetty hoitamatta sen tullessa maksuun"
- vakavuudelta olennainen (riski rahoittajalle)
- suhteellisen homogeeninen
 - konkurssi, saneeraus
 - maksuhäiriö, vakava
 - maksuviive

2.2. Selittävät muuttujat

- ymmärretään maksuhäiriöprosessi:
"käytettävissä oleva rahoitus ei riitä maksuvelvoitteiden hoitamiseen"
- maksukyvyttömyys

- ilmiö:

Velvoitteet: Rahoitus:

- tulot

- ostot
- palkat
- vuokrat
- korot
- verot

= rahoitustulos

- rahoitusomaisuuden
käyttö

- lyhytaikaisen velan
lisäys

- pitkäaikaisen velan
lisäys

- oman pääoman lisäys

- (sisäsyntyisen) ilmiön selittäjät:

- rahoitustulos

- rahoitustulos / liikevaihto

- pääoman tuottosuhde

- kasvunopeus

- omavaraisuusaste

- vuokra/osto

- rahoitusomaisuus ja

- lyhytaikaiset velat
 - quick ratio
 - omavaraisuusaste
- pitkäaikaiset velat
 - omavaraisuusaste
 - koko
- oma pääoma
 - omistajan varallisuus
 - yhtiömuoto
 - koko

- muut tekijät (ulkoiset tekijät): default correlation
 - riippuvuus muista yrityksistä
 - riskikeskittymät, ketjureaktio
 - arvoketju
 - omistus

- (vakava) ilmiö prosessina:
 - pohjana heikko omavaraisuusaste
 - < 10-15 %
 - tuottoprosentti heikkenee
 - < 5-7 %
 - tai kasvu liian nopea
 - > 30-40 %
 - rahoitustulos heikkenee
 - < 0-1 %
 - quick ratio heikkenee
 - < 0.5
 - rahoitus ei riitä

- prosessi vaikeasti havaittavissa
 - jos ilmiö ei vakava
 - sattumat
 - nopeus
 - ulkoiset tekijät keskeisiä
 - henkilöyhtiö

2.3. Otanta

- vastinpareittainen otanta, jos halutaan eliminoida
 - koko
 - toimiala
 - suhdanteet
- satunnainen otanta, jos halutaan pitää ne mallissa
- riittävä otos, noin 100 häiriöllistä
 - ei mukaan epäpuhtaita tapauksia
 - tahalliset
 - rikolliset
 - arvoitukset
 - terveitäkin karsitaan
 - poikkeuksellisen heikot
 - keinotekoiset
- ajankohta (esimerkiksi suhdannevaihe)
 - ei poikkeava ajankohta: edustava
 - ei ratkaisevaa vaikutusta
- aikaviive ilmiön (eventin) ja selittävien mittauksen välillä (tilinpäätös ---- häiriö)
 - kiinnitetään
 - suhteellisen vakio

2.4. Menetelmä

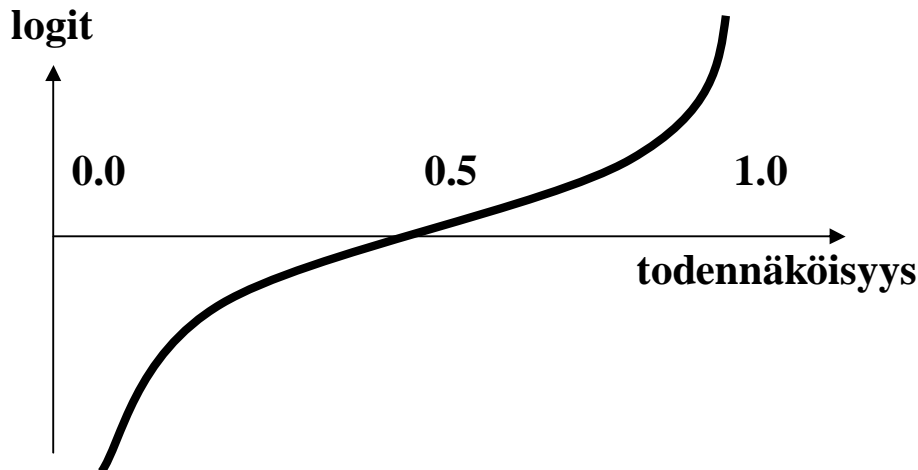
- logistinen regressioanalyysi

$$\text{logit} = a + b_1X_1 + b_2X_2 + \dots$$

- lineaarinen riskin mittari: logit
 - kompensatio
- a, b_i vakio ja regressiokertoimet
- X_i selittävät tekijät
- vertaa regressioanalyysi!
- mutta: logistinen muunnos

- muunnetaan ehdolliseksi todennäköisyydeksi kuulua johonkin luokkaan

$$\text{todennäköisyys} = 1/[1+\exp(-\text{logit})]$$



- esimerkiksi todennäköisyys kuulua maksuhäiriöisiin yrityksiin
- muuttujien ei tarvitse noudattaa normaalijakaumaa
- keskeisten muuttujien pakottaminen
- eteenpäin askeltava
 - ensin paras
 - muiden selityskyky
 - Khii toiseen
- muuttujia 4-7 kpl
 - osa dummy-muuttujia
 - toimiala
 - ikä
 - muuttaa riskitasoa
- muuttujien etumerkit
- muuttujien merkitys:
 - Waldin testi, Khii toiseen (t-testi approximaatiot)
- mallin hyvyys
 - 2 Log likelihood (goodness of fit)
 - Schwartzin kriteeri (goodness of fit)
 - Cox & Snell R-Square (selitysaste)
 - Nagelkerke R-Square (selitysaste)
 - Hosmer & Lemeshowin test (luokittelu)
 - Coefficient of concordance (yhteensopivuus)

2.5. Testaus

- estimointiaineisto ja testiaineisto
- testiaineisto tärkeä
 - isoja, pieniä
 - eri toimialoja
- Lachenbruchin ristiinvalidointi:
 - ei testiaineistoa
 - havainto otetaan pois estimoinnista
 - tämä havainto luokitellaan estimoidulla mallilla
 - jne

2.6. Esimerkki: saneeraukseen hakemisen ennustaminen (maksukyvyttömyys)

- Voitto plus –tietokanta (SAT)
- 47 saneeraukseen hakenutta yritystä
- 457 tervettä yritystä
 - ei vastinparimenettelyä

2.6.1. Tiedoston muuntaminen

- syntax –muoto: helppo tehdä muutoksia ja monistaa

```
DATA SANEERAL;
  INFILE 'sandata1.dat' LRECL=1000;

  INPUT X1 TILIKP YH X3 $ VUOSI X5 X6 X7 X8 X9 X10 X11 X12 X13 X14
        X15 X16 X17 X18 X19 X20 X21 X22 X23 X24 X25 X26 X27
        X28 X29 X30 X31 X32 X33 X34 X35 X36 X37
        X38 X39 X40 X41 X42 X43 X44
        X45 X46 X47 X48 X49 X50 X51 X52 X53 X54
        X55 X56 X57 X58 X59 X60 X61 X62 X63 X64
        X65 X66 X67 X68 X69 X70 X71 X72 X73 X74
        X75 X76 X77 X78 X79 X80 X81 X82 X83 X84
        X85 X86 X87 X88 X89 X90 X91 X92 X93 X94
        X95 X96 X97 X98 X99 X100 X101 X102 X103
        X104 X105 X106;

M = 12/TILIKP;

/* TULOSLASKELMA */
/*LISÄTTY MUUT LIIKET TUOTOT*/
LVAIHTO = (X6+X9)*M;
ULKPAL = -X13*M;
HLOMKUL = -X14*M;
```

```

MKATE = X16*M;
HLOKKUL = -X18*M;
VUOKRAT = -X19*M;
KKATE = X21*M;

/* TASE */
/* OMAISUUS */
PERMENOT = X43;
TUTMENOT = X44;
AINOIK = X45;
LIJARVO = X46;
MUPIME=X47;

/* KANNATTAVUUS */

TUOTTOPR = 100*(TULEVV + KORRKKUL)/TASELSU;

/* MAKSUVALMIUS */

/* PERINTEINEN */

RAHOM = X65+X66+X67+X68+X69+X70+X71+X72+X73+X74;
VOM = X61+X62+X63+X64;
LYVELAT = X99+X100+X101+X102+X103+X104+X105;
QURRAT = (RAHOM+VOM)/(LYVELAT);
QUICKRAT = RAHOM/(LYVELAT-LVENNAK);

/* DYNAAMINEN */

RAHTULOS = TULEVV+POISTOT-VVEROT;
RAHTULPR = 100*RAHTULOS/LVAIHTO;

```

2.6.2. Aineiston arviointi

```

PROC SORT;
  BY TUNNUS;
PROC MEANS;
  VAR VUOSI LVAIHTO TASELSU LVKASVU TASKASVU RAHTULPR QUICKRAT
  OVELKN TULOSPR TUOTTOPR OMAVARAS VPTAKA ZLUKU;
  BY TUNNUS;

```

TUNNUS=0

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
VUOSI	47	98.0000000	0	98.0000000	98.0000000
LVAIHTO	47	3541.10	3314.08	214.0000000	17460.00
TASELSU	47	2547.40	3540.80	116.0000000	22087.00
LVKASVU	39	48.2842821	98.6200935	-69.3320000	418.7500000
TASKASVU	39	59.4927179	120.6378014	-47.9170000	495.7450000
RAHTULPR	45	0.7226000	13.7836434	-42.5180000	24.6260000
QUICKRAT	47	0.7576723	1.2118977	0.0132700	6.5154600
OVELKN	42	160.4019048	207.8543859	0	806.8100000
TULOSPR	45	-3.4903556	13.9908718	-50.1940000	31.4730000
TUOTTOPR	45	-5.3321778	27.0906538	-79.9540000	46.0740000
OMAVARAS	44	-1.7954773	40.2674463	-138.8160000	73.8720000
VPTAKA	46	0.1438696	26.5848160	-65.9910000	68.5610000
ZLUKU	42	13.0723251	51.1751657	-101.9586214	158.6473644

```
-----
```

TUNNUS=1

Variable	N	Mean	Std Dev	Minimum	Maximum
LVAIHTO	457	42412.36	129637.05	1.0000000	1963050.00
TASELSU	457	40156.74	162412.70	14.0000000	3000000.00
LVKASVU	328	112.0759165	743.4059898	-99.7800000	11140.66
TASKASVU	329	25.2920596	84.8061691	-57.0370000	1224.40
RAHTULPR	454	9.7242817	12.6503970	-65.8800000	85.0607000
QUICKRAT	454	1.3450586	1.0615812	0.0029000	13.5595000
OVELKN	407	85.5592555	97.3617602	0	725.8000000
TULOSPR	451	5.5578051	11.9102998	-58.6679000	50.1425000
TUOTTOPR	455	12.9969637	16.1036400	-70.7726000	77.7360000
OMAVARAS	457	25.7505263	24.3809097	-110.8790000	94.2460000
VPTAKA	438	22.5236820	24.0491372	-74.8175000	100.0000000
ZLUKU	452	49.6938193	36.6109041	-101.9051018	194.6177080

```
-----
```

2.6.3. Logistinen regressioanalyysi

```
PROC LOGISTIC;
  MODEL TUNNUS2 = LVAIHTO TASELSU LVKASVU TASKASVU RAHTULPR QUICKRAT
  OVELKN TULOSPR TUOTTOPR OMAVARAS VPTAKA
  / SELECTION=STEPWISE SLENTY=0.15 SLSTAY=0.15 DETAILS CTABLE MAXSTEP=10;
```

The LOGISTIC Procedure

Model Information

Data Set	WORK.KASVIAA
Response Variable	TUNNUS2
Number of Response Levels	2
Number of Observations	306
Link Function	Logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	TUNNUS2	Total Frequency
1	0	274
2	1	32

NOTE: 198 observations were deleted due to missing values for the response or explanatory variables.

Stepwise Selection Procedure

Step 0. Intercept entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	2.1474	0.1868	132.1301	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
49.3829	11	<.0001

Analysis of Effects Not in the Model

Effect	DF	Score Chi-Square	Pr > ChiSq
LVAIHTO	1	3.3072	0.0690
TASELSU	1	1.6330	0.2013
LVKASVU	1	0.4002	0.5270
TASKASVU	1	11.4391	0.0007
RAHTULPR	1	12.9782	0.0003
QUICKRAT	1	4.6533	0.0310
OVELKN	1	3.0913	0.0787
TULOSPR	1	11.0063	0.0009
TUOTTOPR	1	13.7410	0.0002
OMAVARAS	1	22.1633	<.0001
VPTAKA	1	12.1353	0.0005

Step 1. Effect OMAVARAS entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	207.033	192.048
SC	210.756	199.495
-2 Log L	205.033	188.048

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	16.9846	1	<.0001
Score	22.1633	1	<.0001
Wald	16.1713	1	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	1.7334	0.2091	68.6983	<.0001
OMAVARAS	1	0.0246	0.00611	16.1713	<.0001

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses

Percent Concordant	77.0	Somers' D	0.549
Percent Discordant	22.2	Gamma	0.553
Percent Tied	0.8	Tau-a	0.103
Pairs	8768	c	0.774

Analysis of Effects in Model

Effect	DF	Wald Chi-Square	Pr > ChiSq
OMAVARAS	1	16.1713	<.0001

Analysis of Effects Not in the Model

Effect	DF	Score Chi-Square	Pr > ChiSq
LVAIHTO	1	2.7292	0.0985
TASELSU	1	1.1589	0.2817
LVKASVU	1	0.4474	0.5036
TASKASVU	1	12.8700	0.0003
RAHTULPR	1	3.6168	0.0572
QUICKRAT	1	1.0623	0.3027
OVELKN	1	2.3615	0.1244
TULOSPR	1	2.1856	0.1393
TUOTTOPR	1	2.2123	0.1369
VPTAKA	1	4.3013	0.0381

Step 2. Effect TASKASVU entered: jne muuttujia valitaan ja pudotetaan

Lopullinen tulos:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	207.033	132.871
SC	210.756	151.489
-2 Log L	205.033	122.871

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	82.1621	4	<.0001
Score	32.8003	4	<.0001
Wald	27.1545	4	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-0.8972	0.5079	3.1203	0.0773
LVAIHTO	1	0.000300	0.000074	16.3901	<.0001
TASKASVU	1	-0.00557	0.00282	3.9127	0.0479
QUICKRAT	1	0.5487	0.2937	3.4907	0.0617
VPTAKA	1	0.0189	0.00976	3.7386	0.0532

Association of Predicted Probabilities and Observed Responses

Percent Concordant	92.4	Somers' D	0.850
Percent Discordant	7.4	Gamma	0.851
Percent Tied	0.2	Tau-a	0.160
Pairs	8768	c	0.925

- huom! ei omavaraisuusastetta mukana

- askellus etsii parhaimman mallin: yksittäisten muuttujien merkitys riippuu siitä, mitä muuttujia mallissa on

Testaus: Lachenbruch crossvalidation

32/306 = ~11 % = vertailutodennäköisyys

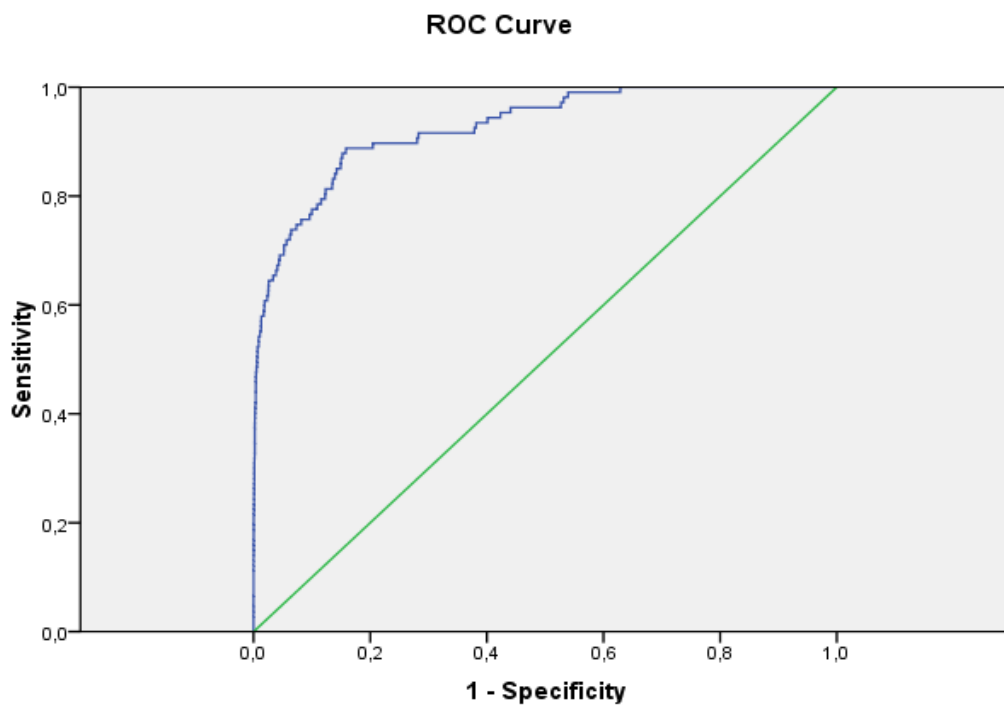
Classification Table

Prob Level	Correct		Incorrect		Correct	Percentages			False POS	False NEG
	Event	Non-Event	Event	Non-Event		Sensitivity	Specificity			
0.440	270	8	24	4	90.8	98.5	25.0	8.2	33.3	
0.460	270	8	24	4	90.8	98.5	25.0	8.2	33.3	
0.480	270	8	24	4	90.8	98.5	25.0	8.2	33.3	
0.500	269	9	23	5	90.8	98.2	28.1	7.9	35.7	
0.520	268	12	20	6	91.5	97.8	37.5	6.9	33.3	
0.540	268	12	20	6	91.5	97.8	37.5	6.9	33.3	
0.560	268	13	19	6	91.8	97.8	40.6	6.6	31.6	
0.580	266	13	19	8	91.2	97.1	40.6	6.7	38.1	
0.600	266	15	17	8	91.8	97.1	46.9	6.0	34.8	
0.620	262	18	14	12	91.5	95.6	56.3	5.1	40.0	
0.640	261	19	13	13	91.5	95.3	59.4	4.7	40.6	
0.660	260	20	12	14	91.5	94.9	62.5	4.4	41.2	
0.680	257	20	12	17	90.5	93.8	62.5	4.5	45.9	
0.700	257	21	11	17	90.8	93.8	65.6	4.1	44.7	
0.720	253	23	9	21	90.2	92.3	71.9	3.4	47.7	
0.740	249	23	9	25	88.9	90.9	71.9	3.5	52.1	
0.760	244	23	9	30	87.3	89.1	71.9	3.6	56.6	
0.780	239	24	8	35	85.9	87.2	75.0	3.2	59.3	
0.800	235	24	8	39	84.6	85.8	75.0	3.3	61.9	
0.820	232	26	6	42	84.3	84.7	81.3	2.5	61.8	
0.840	229	26	6	45	83.3	83.6	81.3	2.6	63.4	
0.860	223	26	6	51	81.4	81.4	81.3	2.6	66.2	
0.880	215	27	5	59	79.1	78.5	84.4	2.3	68.6	
0.900	211	27	5	63	77.8	77.0	84.4	2.3	70.0	
0.920	204	28	4	70	75.8	74.5	87.5	1.9	71.4	
0.940	199	28	4	75	74.2	72.6	87.5	2.0	72.8	
0.960	190	29	3	84	71.6	69.3	90.6	1.6	74.3	
0.980	170	29	3	104	65.0	62.0	90.6	1.7	78.2	
1.000	0	32	0	274	10.5	0.0	100.0	.	89.5	

- Luokittelun tarkkuutta voidaan mitata ROC-käyrällä

The screenshot shows the PASW Statistics Data Editor interface. The 'Analyze' menu is open, and 'ROC Curve...' is selected. The main window displays a list of variables with their labels, values, missing data, columns, alignment, and measures. The 'ROC Curve...' option is highlighted in the menu.

Name	Type	Label	Values	Missing	Columns	Align	Measure		
1	lytunnus	Numeri	None	None	11	Right	Nominal		
2	to95	String	None	None	8	Left	Nominal		
3	kmro	Numeri	None	None	8	Right	Nominal		
4	postintro	Numeri	None	None	8	Right	Nominal		
5	kunta	Numeri	None	None	8	Right	Nominal		
6	maakunta	Numeri	None	None	8	Right	Nominal		
7	laani	Numeri	None	None	8	Right	Nominal		
8	alfa1	Numeri	None	None	8	Right	Nominal		
9	hayht1	Numeri	None	None	8	Right	Scale		
10	mhavain	String	None	None	7	Left	Nominal		
11	haikoodi	String	None	None	11	Left	Nominal		
12	rekpm	Date	None	None	16	Right	Nominal		
13	hvuosi	Numeri	None	None	8	Right	Scale		
14	hkuukaus	Numeri	None	None	8	Right	Scale		
15	maksuhai	Numeri	None	None	8	Right	Scale		
16	konkurs	Numeri	None	None	8	Right	Scale		
17	saneer	Numeri	None	None	8	Right	Scale		
18	muumak	Numeri	8	2	None	None	8	Right	Scale
19	akamaks	Numeri	8	2	None	None	8	Right	Scale
20	ln1	Numeri	8	2	None	None	8	Right	Scale
21	ln2	Numeri	8	2	None	None	8	Right	Scale
22	ln3	Numeri	8	2	None	None	8	Right	Scale
23	ln4	Numeri	8	2	None	None	8	Right	Scale
24	ln5	Numeri	8	2	None	None	8	Right	Scale
25	Mkasv1	Numeri	8	2	None	None	8	Right	Scale
26	Mkasv2	Numeri	8	2	None	None	8	Right	Scale
27	Mkasv3	Numeri	8	2	None	None	8	Right	Scale
28	Mkasv4	Numeri	8	2	None	None	8	Right	Scale
29	Mkasv5	Numeri	8	2	None	None	8	Right	Scale
30	sipotu1	Numeri	8	2	None	None	8	Right	Scale
31	sipotu2	Numeri	8	2	None	None	8	Right	Scale
32	sipotu3	Numeri	8	2	None	None	8	Right	Scale
33	sipotu4	Numeri	8	2	None	None	8	Right	Scale
34	sipotu5	Numeri	8	2	None	None	8	Right	Scale
35	nettup1	Numeri	8	2	None	None	8	Right	Scale
36	nettup2	Numeri	8	2	None	None	8	Right	Scale
37	nettup3	Numeri	8	2	None	None	8	Right	Scale
38	nettup4	Numeri	8	2	None	None	8	Right	Scale
39	nettup5	Numeri	8	2	None	None	8	Right	Scale



- mittari: alueen pinta-ala: AUC = 0 ----- 1 (täydellinen malli)

Area Under the Curve (AUC)					
Test Result Variable(s):	Predicted reorganization probability				
Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	Upper Bound	
0.926967	0.013179	9.55E-53	0.901136	0.952798	

- a. Under the nonparametric assumption
- b. Null hypothesis: true area = 0.5

2.6.4. Multi-logit analyysi

- selitettävällä muuttujalla useita (>2) luokkia (ei siis binäärinen)
- esimerkiksi: maksuhäiriö, selvitystila, saneeraus, konkurssi

- multilogit-mallissa on kaksi perustyyppiä:

1) a generalized multinomial logit model for nominal outcomes

- produces separate parameter vectors for each of the generalized logit equations of interest

- oma logit-malli jokaiselle N-1:lle luokalle (1 luokka on referenssiluokka)

SPSS

2) the proportional odds model with cumulative logit link for ordinal responses

- produces a common slope but separate intercepts for each of the cumulative logit equations of interest

- oma vakio jokaiselle N-1 luokalle, muuttujien kertoimet samat

SAS

Esimerkki multinomial-mallista: SPSS

- selitettävä ei binäärinen, vaan 4-luokkainen:

0 = maksuhäiriötön

1 = muu maksuhäiriö

2 = saneeraus

3 = konkurssi

- vuotta ennen häiriötä

- selittäjinä tunnusluvut

- jokaisella luokalla N-1 oma logistinen funktio:

- SPSS:n kuvaruutu Multinomial logistic regression:

SPSS Data Editor window showing a data file named 'satesitelma230806.sav'. A 'Multinomial Logistic Regression' dialog box is open, with 'Luokka3(First)' selected as the dependent variable. The dialog includes sections for 'Factor(s)' and 'Covariate(s)'. The main data grid is visible in the background, listing variables such as 'lyt', 'yrr', 'tol', 'he', 'ma', 'ku', 'rek', 'hu', 'hkv', 'ma', 'kor', 'sa', 'mu', 'ika', 'se', 'gro', 'bigroup', 'rakent', 'valmist', 'kauppa', 'hotrav', 'liikevar', 'palvelu', 'toimiala', 'year', 'vuosi', 'liikev', 'Inliikev', and 'sintu'.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1 lyt					None	None	11	Right	Nominal
2 yrr					None	None	72	Left	Nominal
3 tol					None	None	8	Right	Nominal
4 he					None	None	8	Right	Nominal
5 ma					None	None	8	Right	Nominal
6 ku					None	None	8	Right	Nominal
7 rek					None	None	8	Right	Nominal
8 hu					None	None	8	Right	Nominal
9 hkv					None	None	8	Right	Nominal
10 ma					None	None	11	Left	Nominal
11 kor					None	None	16	Right	Nominal
12 sa					None	None	8	Right	Scale
13 mu					None	None	8	Right	Scale
14 ika					None	None	8	Right	Scale
15 se					None	None	8	Right	Scale
16 gro					None	None	8	Right	Scale
17 bigroup	Numeric	8	2	1 =häir+ei puutu+ei sel, 2=terv	None	None	8	Right	Scale
18 rakent	Numeric	8	2		None	None	8	Right	Scale
19 valmist	Numeric	8	2		None	None	8	Right	Scale
20 kauppa	Numeric	8	2		None	None	8	Right	Scale
21 hotrav	Numeric	8	2		None	None	8	Right	Scale
22 liikevar	Numeric	8	2		None	None	8	Right	Scale
23 palvelu	Numeric	8	2		None	None	8	Right	Scale
24 toimiala	Numeric	8	2	tol: 1=val, 2=rak, 3=kaup, 4=m	None	None	8	Right	Scale
25 year	Numeric	4	0		None	None	8	Right	Scale
26 vuosi	Numeric	8	2		None	None	8	Right	Scale
27 liikev	Numeric	8	2		None	None	8	Right	Scale
28 Inliikev	Numeric	8	2		None	None	8	Right	Scale
29 sintu	Numeric	8	2		None	None	8	Right	Scale

Tulokset:

Case Processing Summary

		N	Marginal Percentage
Luokka	.00	31375	93.1%
3			
	1.00	2112	6.3%
	2.00	45	.1%
	3.00	183	.5%
Valid		33715	100.0%
Missing		52314	
Total		86029	
Subpopulation		33711(a)	

a The dependent variable has only one value observed in 33711 (100.0%) subpopulations.

Model Fitting Information

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	18726.324	18751.601	18720.324			
Final	16668.628	16921.399	16608.628	2111.696	27	.000

Pseudo R-Square

Cox and Snell	.061
Nagelkerke	.142
McFadden	.113

Parameter Estimates

Luokka3(a)		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
1.00	Intercept	-.628	.089	49.480	1	.000			
	ika	-.050	.003	217.621	1	.000	.951	.945	.958
	lnliikev	-.149	.013	123.738	1	.000	.861	.839	.884
	sipotu	-.002	.001	12.986	1	.000	.998	.997	.999
	nettup	.015	.003	27.370	1	.000	1.015	1.010	1.021
	quick	.018	.004	17.684	1	.000	1.018	1.009	1.026
	rahtup	-.018	.003	44.076	1	.000	.982	.977	.987
	omavar	-.015	.001	626.326	1	.000	.985	.984	.987
	tmkyky	-.002	.001	13.628	1	.000	.998	.997	.999
	lvkasvu	.002	.000	27.099	1	.000	1.002	1.001	1.003
2.00	Intercept	-8.280	.694	142.447	1	.000			
	ika	-.018	.017	1.054	1	.305	.982	.949	1.016
	lnliikev	.400	.081	24.518	1	.000	1.492	1.273	1.748
	sipotu	-.009	.004	3.896	1	.048	.991	.983	1.000
	nettup	-.025	.012	4.493	1	.034	.975	.952	.998
	quick	-.621	.351	3.140	1	.076	.537	.270	1.068
	rahtup	.005	.012	.183	1	.669	1.005	.981	1.030
	omavar	-.026	.004	43.863	1	.000	.975	.967	.982
	tmkyky	.001	.006	.051	1	.822	1.001	.989	1.014
	lvkasvu	.005	.002	4.777	1	.029	1.005	1.001	1.009
3.00	Intercept	-3.777	.308	150.149	1	.000			
	ika	-.031	.010	10.626	1	.001	.969	.951	.988
	lnliikev	-.008	.044	.031	1	.860	.992	.911	1.081
	sipotu	-.003	.002	2.984	1	.084	.997	.993	1.000
	nettup	.028	.008	10.932	1	.001	1.028	1.011	1.046
	quick	-.162	.091	3.165	1	.075	.851	.712	1.017
	rahtup	-.031	.008	14.446	1	.000	.969	.954	.985
	omavar	-.021	.002	153.463	1	.000	.979	.976	.983
	tmkyky	-.007	.002	8.500	1	.004	.993	.988	.998
	lvkasvu	.001	.001	.493	1	.483	1.001	.998	1.004

a The reference category is: .00.

Classification

Observed	Predicted				Percent Correct
	.00	1.00	2.00	3.00	
.00	31324	51	0	0	99.8%
1.00	2056	56	0	0	2.7%
2.00	44	1	0	0	.0%
3.00	171	12	0	0	.0%
Overall Percentage	99.6%	.4%	.0%	.0%	93.1%

- ei juuri kykene erottelemaan pois referenssiluokasta
- muokattavissa kriittisellä todennäköisyydellä