SUOMEN VAKUUTUSTIETEEN KESKUSLIITTO
VAKUUTUSKIRJASTO

WORKING PAPERS ISSN 0781-4410

SUOMEN AKTUAARIYHDISTYS
The Actuarial Society of Finland

30

Yrjö Romppainen

KONKURSSIN ENNUSTAMINEN EROTTELU-ANALYYSIN AVULLA (1990)

FORECASTING THE BANKRUPTCY BY USING THE DISCRIMINANT ANALYSIS

The research contained Finnish industrial enterprises, 93 bankruptcies and 843 healthy conserns. Using official balance sheets 20 variables xi, i=1,...,20 , were formed. These single variables were used again to build new variables in form square xi*xj and cubic xi*xj*xk, where i,j,k=1,...,20.

First was used only the set of single variables to choose the most important ones by stepwise discriminant procedure and the linear canonical discriminant function, h1(x), was calculated.

The second stage was to expand the set of variables so that it contained 20 single and 210 square variables; again the stepwise disciminant procedure was used and 14 most important variables were chosen and then the linearized square canonical discriminant function, h2(x), was calculated.

The third stage was similar to the second but the set of cubic variables was only 92 so there were 322 variables altohether. Again 14 best variables were selected and the linearized cubic canonical discriminant function, h3(x), was obtained.

The cubic function, h3(x), gave only some improvement to discriminate between bankruptcies and healthy enterprises, so the square function, h2(x), was useful to base the tariff of credit insurance.

Yrjö Romppainen
Helsinki 16.01.1990

S I S Ä L L Y S L U E T T E L O

1.0 Tiivistelmä....1
2.0 Tutkimusmenetelmä....1
2.1 Perusoletuksia....1
2.2 Luokitteluongelma....2
3.0 Hyvän luokittelun kriteerit....3
3.1 Merkintöjä....3
3.2.1 Minimoidaan väärinluokittelun kokonaistodennäköisyys....5
3.2.2 Baeysiläinen lähestymistapa....5
3.2.3 Minimoidaan maksimaalinen väärinluokittelutodennäköisyys6
3.2.4 Minimoidaan väärinluokittelun seurauksena syntyneet kustannukset....7
3.2.5 Yhteenveto....8
4.0 Kaksi multinormaalisti jakautunutta populaatiota, kun jakaumaparametrit tunnetaan....8
4.1 Erottelufunktio U:n jakauma....10
5.0 Kaksi multinormaalisti jakautunutta populaatiota, kun jakaumaparametreja ei tunneta....12
5.1 Erottelufunktio W:n jakauma....12
5.2 W:n asymptoottinen jakauma....14
6.0 Muita lähestymistapoja....15
6.1 Regressioanalyyttinen lähestymistapa....15
6.2 Suurin uskottavuus....17
6.3 Väärinluokittelutodennäköisyydet....19
7.0 Kaksi multinormaalisti jakautunutta populaatiota, kun kovarianssimatriisit erisuuret....27
7.1 Uskottavuusmenetelmä....27
7.2 Sijoitusmenetelmä....28
7.3 Lineaariset menetelmät....29
7.3.1 Lineaaristen menetelmien käyttö....33
8.0 Etäisyysmittoihin perustuva erottelu....35
8.1 B/W:n maksimointi....38
8.2 !W!/!B!:n minimointi....38
8.3 Kanonisen korrelaatioanalyysin ja erotteluanalyysin välinen yhteys....39
8.4 Ryhmien välisten erojen testaamisesta....41
8.5 Kanoninen erottelufunktio....42
9.0 Muuttujien valinta....43
9.1 Perusteellinen etsintä....44
9.2 Kiihdytetty etsintä....45
9.3 Osaoptimaalinen etsintä....45
9.3.1 Etenevä valinta....45
9.3.2 Laskeva valinta....46
9.3.3 Askeltava valinta....46
9.4 Valinta kriteeri....46
9.4.1 Osittais F-suhde....46
9.4.2 Mahalanobisetäisyys....47
9.5 Muuttujien valinnalle asetettavat minimivaatimukset....48
10.0 Perusoletusten paikkansa pitävyys....48
11.0 Tutkimussunnitelma....50
11.1 Lähtöaineisto....51
11.2 Alkuehdot....51
11.3 Tutkimusaineisto....51
11.4 Tariffitekijät....53
11.5 Mahalanobisetäisyys....60
11.6 Havaintoaineiston karsinta....60
11.7 Multinormaalisuus....62
11.8 Kovarianssimatriisit....73
11.9 Tariffitekijöiden valinta....74
11.10 Erottelufunktiot....76
11.11 Kanoninen erotteluanalyysi....77
11.12 Kanoninen erottelufunktio....79
11.13 Tariffin muodostus....97
12.0 ETK:n luottovakuutustariffi....97

TAULUKOT

1.0 Yritysten tilinpäätösajankohdat.....52
2.0 Tilastotietoja konkursseista Suomessa
2.1 Yhtiömuoto.....54
2.2 Toimiala.....54
3.0 Havainnot teollisuusryhmittäin.....55
4.0 Konkurssiyritysten erottelumuuttujien tunnuslukuja.....59
5.0 Toimivien yritysten erottelumuuttujien tunnuslukuja.....59
6.0 Tunnuslukujen keskiarvot.....61
7.0 Tunnuslukujen keskihajonnat.....61
8.0 Valitut muuttujat, kun vain ensimmäisen asteen termit mukana.....75
9.0 Valitut muuttujat, kun toisen asteen termit mukana.....77
10.0 Sisäisten rakennekerrointen itseisarvot.....78
11.0 Kanonisen erottelufunktion muuttujat.....79
12.0 Kanonisen erottelufunktion kertoimet.....79
13.0 Mahalanobisetäisyys.....80
14.1 Korvatut 1.7.87-31.12.89 välisenä aikana, ryhmät I - III
4 vuotta ennen korvausta 100
14.2 3 ' ' ' 101
14.3 2 ' ' ' 101
14.4 1 ' ' ' 102

KUVAT

1 Multinormaalisuus konkurssit alfa=0.0.....63
2 ' toimivat ' 64
3 ' konkurssit alfa=0.001.....65
4 ' toimivat ' 66
5 ' konkurssit alfa=0.01.....67
6 ' toimivat ' 68
7 ' konkurssit alfa=0.05.....69
8 ' toimivat ' 70
9 ' konkurssit alfa=0.10.....71
10 ' toimivat ' 72
11 Neliömuotoisen erottelufunktion
pistemäärien kertymäfunktio alfa=0.0.....81
12 ' ' alfa=0.001.....82
13 ' ' alfa=0.01.....83
14 ' ' alfa=0.05.....84
15 ' ' alfa=0.10.....85
16 Ensimmäisen asteen erottelufunktion
pistemäärien kertymäfunktio alfa=0.0.....86
17 ' ' alfa=0.001.....87
18 ' ' alfa=0.01.....88
19 ' ' alfa=0.05.....89
20 ' ' alfa=0.10.....90
21 Kolmannen asteen erottelufunktion
pistemäärien kertymäfunktio alfa=0.0.....92
22 ' ' alfa=0.001.....93
23 ' ' alfa=0.01.....94
24 ' ' alfa=0.05.....95
25 ' ' alfa=0.10.....96
26 U% ja Z% 4 vuotta ennen korvausta.....103
27 ' 3 ' 104
28 ' 2 ' 105
29 ' 1 ' 106
30 Oikeinluokittelukyky.....107
31 Mahalanobiksen etäisyys.....109

LIITTEET

1 Mahdolliset tariffitekijät
2 Mahdolliset tariffitekijät tase- ja tuloslaskelmassa
3 Tilinpäätöksen tärkeimmät joustokohdat
4 Ensimmäisen asteen muuttujien kovarianssimatriisit
5 Toisen asteen muuttujien valinta
6 Ensimmäisen asteen muuttujien testitunnuslukuja, rakennekertoimet, erottelufunktion muuttujat ja kertoimet sekä Mahalanobisetäisyys
7 Kolmannen asteen muuttujien valinta
8 Kolmannen asteen muuttujien testitunnuslukuja, rakennekertoimet, erottelufunktion muuttujat ja kertoimet sekä Mahalanobisetäisyys
9 Lähdeluettelo

KONKURSSIN ENNUSTAMINEN EROTTELUANALYYSIN AVULLA

1.0 TIIVISTELMÄ

Muodostetaan erotteluanalyysin avulla matemaattinen malli, jonka perusteella ennustetaan virallisten tilinpäätöstietojen avulla yrityksen konkurssi.

Mallittamisessa käytetään monimuuttujamenetelmiä päämääränä olevan erottelufunktion ja päätössäännön luomisessa luottovakuutustoiminnan tarpeita silmällä pitäen.

Aineisto on Eläketurvakeskuksen luottovakuutusosaston asiakaskannasta, tietokonelaitteistona on käytetty IBM 3090/200 -järjestelmää sekä itse tilastolliset analyysit on tehty SAS-ohjelmistolla.
(SAS Institute Inc., Cary, NC, USA)

2.0 TUTKIMUSMENETELMÄ

Erotteluanalyysi auttaa analysoimaan ryhmien välisiä eroja ja/tai antaa keinot luokitella minkä tahansa havainnon kuulumaan ryhmään, jota se eniten muistuttaa.
Erotteluanalyysi on tekniikka, joka liittää yhden luokittelumuuttujan useaan erottelumuuttujaan. Erotteluanalyysillä on siten dualistinen tehtävä eli toisaalta selittää ryhmien välisiä eroja ja toisaalta luokitella alkuperältään tuntematon havainto johonkin ryhmään kuuluvaksi.

2.1 PERUSOLETUKSIA

Eroteltavia ryhmiä on kaksi tai enemmän ja ryhmiä kuvataan nominaalitason muuttujalla. Merkitään ryhmien lukumäärää g:llä, jolloin ryhmiä merkitään G(1), G(2),..., G(g).

Tunnuslukuja, joita käytetään havaitsemaan eroja ryhmien välillä, kutsutaan erottelumuuttujiksi. Näiden muuttujien on oltava mitattavissa intervalli- tai suhdeasteikolla, jotta keskiarvot ja kovarianssit voitaisiin laskea ja siten niitä voidaan oikeutetusti käyttää matemaattisissa yhtälöissä. Merkitään erottelumuuttujien lukumäärää p:llä, jolloin erottelumuuttujia merkitään x1,...,xp:llä sekä n*p-ulotteista havaintomatriisia X=(X1,...,Xp):llä, missä Xi=(X1i,...,Xni)'.

Havaintojen kokonaislukumäärän on oltava suurempi kuin erottelumuuttujien lukumäärä. Kun havaintojen lukumäärää ryhmässä i merkitään ni:llä, niin kaikkien havaintojen lukumäärää merkitään n:llä.

Mikään erottelumuuttuja ei saa olla muiden erottelumuuttujien lineaarikombinaatio. Samoin ei voida käyttää erottelumuuttujia, jotka ovat täysin korreloituneita.

Lisäksi oletus, jota tarvitaan monissa sovelluksissa, on, että kullakin ryhmällä on yhtäsuuret kovarianssimatriisit. Mikäli kovarianssimatriisit ovat yhtä suuret, saadaan lineaarinen erottelufunktio eli erottelumuuttujien suhteen ensimmäistä astetta oleva funktio. Mutta mikäli oletus yhtäsuurista kovarianssimatriiseista ei päde, saadaan neliömuotoinen erottelufunktio eli erottelumuuttujien suhteen toista astetta oleva funktio.

Myöskin tarvitaan oletusta, että kukin ryhmä on peräisin populaatiosta, jolla on multinormaalijakauma. Tämä sallii tarkan merkitsevyystestin suorittamisen sekä ryhmiin kuulumistodennäköisyyksien laskemisen. Kun oletus ei päde, lasketut todennäköisyydet eivät ole tarkkoja, mutta ne voivat silti olla hyödyllisiä, mikäli niitä tulkitaan varovasti. On kuitenkin osoittautunut, ettei erotteluanalyysi ole erityisen herkkä vähäisille poikkeamille multinormaalisuusoletuksesta. Seurauksena on johdetun erottelufunktion vähentynyt tehokkuus ja tarkkuus.

Edellä mainitut oletukset muodostavat sen matemaattisen mallin puitteet, josta yleisimmät erotteluanalyysin lähestymistavat ovat peräisin. Mikäli aineisto johonkin tiettyyn ongelmaan ei täytä näitä oletuksia, tilastolliset tulokset eivät ole tarkalleen todellisuuden heijastuksia.

Erotteluanalyysi on siis menetelmä, joka erottelee kaksi tai useampaa tunnettua perusjoukkoa mahdollisemman etäälle toisistaan. Tässä tapauksessa nämä perusjoukot ovat konkurssiyritykset ja toimivat yritykset.

Ominaisuudet, joiden perusteella erottelu suoritetaan, ovat yrityksen tilinpäätöstiedoista laskettuja tunnuslukuja.

Käsitteellä määrittelyjoukko tarkoitetaan joukkoa, josta päätössääntö on muodostettu. Väärinluokitteluasteella tai virheasteella tarkoitetaan väärinluokiteltujen osuutta. Ilmeinen virheaste on tulos siitä, kun määrittelyjoukkoon sovelletaan muodostettua erottelufunktiota. Todellinen virheaste on puolestaan tulos siitä, kun riippumattomaan joukkoon sovelletaan muodostettua erottelufunktiota.

2.2 LUOKITTELUONGELMA

Luokitteluongelma tulee siitä, kun havainnosta tehdään monia mittauksia ja tarkoituksena on luokitella havainto johonkin ryhmään kuuluvaksi mittaustulostensa perusteella. Monissa tapauksissa voidaan olettaa, että on olemassa äärellinen määrä ryhmiä tai populaatioita, joista havainto saattaa olla peräisin ja kutakin populaatiota kuvataan mittaustulosten todennäköisyysjakaumilla. Täten havaintoa voidaan pitää satunnaisena havaintona tästä populaatiosta. Kysymys on tilanteesta, kun havainnosta on käytettävissä tietyt mittaustulokset ja halutaan tietää, mistä populaatiosta havainto on peräisin.

Luokitteluongelmaa voidaan pitää tilastollisena päätösongelmana. Päätösteoriassa päätössääntö jakaa avaruuden osa-avaruuksiksi Ri, i=1,... ,g. Havainnon sanotaan tulevan ryhmästä tai populaatiosta G(i), mikäli sitä vastaava vektoriesitys on Ri:ssä. Osa-avaruuksien väliset rajat ovat päätöspintoja. Mikäli havainnosta ei ole käytettävissä muuta informaatiota kuin a'prioritodennäköisyydet, qi, niin silloin havainnon sanotaan tulevan ryhmästä G(i), jos qi>qk, kaikille k=1,...,g ja k≠i. Mikäli havainnosta on käytettävissä mittaustulosvektori x, niin havainvainto tulee ryhmästä G(i), jos P(G(i)!x) > P(G(k)!x) kaikille k=1,...,g ja k≠i. Tätä kutsutaan myös Bayesin minimivirheen säännöksi.

Käytettävissä on useita hypoteeseja eli kukin hypoteesi on, että alkuperältään tuntemattoman havainnon jakauma on annettua tiettyä muotoa. Tällöin täytyy hyväksyä yksi näistä hypoteeseistä ja hylätä muut.

Joissakin yhteyksissä ryhmät voidaan tunnistaa etukäteen siinä mielessä, että mittaustulosten todennäköisyysjakaumat oletetaan täydellisesti tunnetuiksi. Toisissa tapauksissa kunkin jakauman muoto saattaa olla tunnettu, mutta jakauman parametrit täytyy estimoida tuon populaation otoksesta.

3.0 HYVÄN LUOKITTELUN KRITEERIT

Rakennettaessa luokittelumenetelmää on toivottavaa minimoida tiettyjä luokittelutodennäköisyyksiä tai luokittelusta johtuvia kustannuksia tai maksimoida ryhmien välisiä eroja. Luokittelun perusteeksi asetetaan kriteeri, jonka mukaan luokittelu- eli päätössääntö muodostetaan.

3.1 MERKINTÖJÄ

Tarkastellaan etupäässä tilannetta, kun luokiteltavia ryhmiä on vain kaksi ja laajennetaan myöhemmin soveltuvin osin useampien ryhmien tapaukseen.

Oletetaan, että havainto on peräisin joko G(1):stä tai G(2):sta. Havainnon luokittelu riippuu mittaustulosvektorista $X'=(X1,\ldots,Xp)$. Asetetaan sääntö, että mikäli havaintoa luonnehditaan tiettyjen x1:n,..., xp:n arvojen joukon kautta, niin silloin havainto on peräisin G(1):stä ja mikäli toiset arvot, niin silloin G(2):sta. Voidaan siten ajatella, että havainto on piste p-ulotteisessa avaruudessa.

Tämä avaruus jaetaan kahteen osa-avaruuteen, jolloin mikäli havainnon vektoriesitys on R1:ssä, niin silloin havainto on peräisin ryhmästä G(1) ja jos vektoriesitys on R2:ssa, niin silloin se on G(2):sta.

Seuraavassa on luokittelumenetelmä, jossa voidaan tehdä kahdenlaisia luokitteluvirheitä: havainto on todellisuudessa peräisin G(1):stä, mutta se luokitellaan G(2):teen tai päivastoin. Myöhemmin havaitaan, että tarvitsee vain tietää näiden kahdenlaisten väärinluokittelun aiheuttamat suhteelliset ei-toivottavat tulokset. Näitä ei-toivottuja tuloksia voidaan monissa tapauksissa kuvata kustannuksina. Olkoon siten C(2!1) sellaisen tapauksen kustannus, joka syntyy, kun alkuperältään G(1):stä peräisin oleva luokitellaan G(2):teen kuuluvaksi. Oletetaan jatkossa, että C(1!1)=C(2!2)=0.

Oletetaan, että on käytettävissä myöskin a'prioritodennäköisyydet g1 ja g2 eli todennäköisyys sille, että havainto on populaatiosta G(1), on q1 ja vastaavasti populaatiosta G(2), on q2 ,(q1+q2=1). Populaation G(1) todennäköisyysominaisuudet määräytyvät jakaumafunktion mukaan. Mukavuuden vuoksi käsitellään tapausta, missä jakaumalla on tiheysfunktio, vaikkakin diskreettitapaus johtaa melkein samanlaiseen käsittelyyn.

Olkoon siis populaation G(1) tiheysfunktio f1(x) ja populaation G(2) f2(x). Mikäli on olemassa osa-avaruus R1 luokitteluperusteeksi sille, että havainto on peräisin populaatiosta G(1), silloin oikeinluokittelutodennäköisyys tapaukselle, että havainto todella on populaatiosta G(1), on

$$P(1!1) = \int_{R1} f1(x)\, dx, \qquad (3.1.1)$$

missä dx = dx1...dxp, sekä populaatiosta G(1) peräisin olevan havainnon väärinluokittelutodennäköisyys on

$$P(2!1) = \int_{R2} f1(x)\, dx. \qquad (3.1.2)$$

Vastaavasti populaatiosta G(2) olevan havainnon oikeinluokittelutodennäköisyys on

$$P(2!2) = \int_{R2} f2(x)\, dx, \qquad (3.1.3)$$

ja väärinluokittelutodennäköisyys

$$P(1!2) = \int_{R1} f2(x)\, dx. \qquad (3.1.4)$$

Koska todennäköisyys sille, että havainto on populaatiosta G(1), on q1, niin todennäköisyys tapaukselle, että havainto on populaatiosta G(1) ja luokitellaan oikein, on

$$q1*P(1!1) = q1 \int_{R1} f1(x)\, dx, \qquad (3.1.5)$$

sekä vastaavasti, kun se väärinluokitellaan, on

$$q1*P(2!1) = q1 \int_{R1} f1(x)\, dx. \qquad (3.1.6)$$

Tällöin myös vastaavat todennäköisyydet populaatiosta G(2) olevalle havainnolle, että se oikeinluokitellaan

$$q2*P(2!2) = q2 \int_{R2} f2(x)\, dx, \qquad (3.1.7)$$

sekä vastaavasti, kun se väärinluokitellaan, on

$$q2*P(1!2) = q2 \int_{R1} f2(x)\, dx. \qquad (3.1.8)$$

3.2.1 MINIMOIDAAN VÄÄRINLUOKITTELUN KOKONAISTODENNÄKÖISYYS ((1))*

Väärinluokittelun kokonaistodennäköisyys on

$$P(R,f) = q1 \int_{R2} f1(x)\,dx + q2 \int_{R1} f2(x)\,dx$$

$$= q1*\Big(1 - \int_{R1} f1(x)\,dx\Big) + q2 \int_{R1} f2(x)\,dx$$

$$= q1 + \int_{R1} (q2*f2(x) - q1*f1(x))\,dx \qquad (3.2.1.1)$$

Tämä minimoituu, kun osa-avaruus R1 valitaan siten, että

$$q2*f2(x) - q1*f1(x) \leq 0 \text{ , kaikille x kuuluu R1:teen} \qquad (3.2.1.2)$$

Päätössäännöksi saadaan, että valitse alueet R1 ja R2 siten, että

$$\begin{aligned} R1 &: \quad q1*f1(x) \geq q2*f2(x) \\ R2 &: \quad q1*f1(x) < q2*f2(x) \end{aligned} \qquad (3.2.1.3)$$

Mikäli q1*f1(x)=q2*f2(x), niin piste voidaan luokitella joko G(1):teen tai G(2):teen kuuluvaksi, yllä se on laitettu R1:teen.

3.2.2 BAYESILÄINEN LÄHESTYMISTAPA

Määritetään päätössäännöksi, että havainto kuuluu siihen ryhmään, jolla on suurin posterioritodennäköisyys.

Posterioritodennäköisyydestä

$$P(G(i)!x) = \frac{qi*fi(x)}{\sum_{k=1}^{g} qk*fk(x)} \qquad (3.2.2.1)$$

seuraa, että havainto tulee ryhmästä G(k), mikäli

$$P(G(i)!x) \geq P(G(k)!x)$$

$$\Longrightarrow \quad qi*fi(x) \geq qk*fk(x), \text{ kaikille } k=1,\ldots,g \text{ ja } k \neq i. \qquad (3.2.2.2)$$

Päätössääntö on siten ekvivalenttia säännön kanssa, että minimoidaan väärinluokittelun kokonaistodenäköisyys.

* Merkintä ((1)) on kirjallisuusviittaus. Luettelo käytetystä kirjallisuudesta on liitteenä.

Yhtä hyvin oltaisiin voitu määrittää painot

$$w1 = \frac{C(2!1)*q1}{C(2!1)*q1+C(1!2)*q2}$$

$$w2 = \frac{C(1!2)*q2}{C(2!1)*q1+C(1!2)*q2} \qquad (3.2.4.5)$$

ja minimoitu T = w1*P(2!1)+w2*P(1!2) eli tämä on jälleen ekvivalenttia säännön kanssa, että minimoidaan väärinluokittelun kokonaistodennäköisyys.

3.2.5 YHTEENVETO

Edellä olleissa luokittelukriteereissä ollaan kiinnostuneita suhteellisista arvoista eli onko f1(x)/f2(x) > vakio, eikä absoluuttisista arvoista. Yleissäännöt ovat siten, että luokittele x ryhmään G(k), mikäli

$$gk(x) > gj(x) \text{ kaikilla } j=1,\ldots,g \text{ ja } j \neq k \qquad (3.2.5.1)$$

tai kun g=2, x kuuluu ryhmään G(1), mikäli

$$h(x) > \text{vakio, muulloin x kuuluu G(2):teen.} \qquad (3.2.5.2)$$

Näitä yleisfunktioita g ja h kutsutaan erottelufunktioiksi.

4.0 KAKSI MULTINORMAALISTI JAKAUTUNUTTA POPULAATIOTA,

KUN JAKAUMAPARAMETRIT TUNNETAAN

Tarkastellaan tilannetta, kun on kaksi multinormaalisti jakautunutta populaatiota ja tarkoituksena on luokitella havainto jompaan kumpaan populaatioon kuuluvaksi.

Olkoon G(i) $N(\mu_i, \Sigma_i)$-jakautunut, kun i=1,2. Tällöin i:nnen populaation tiheysfunktio on

$$fi(x) = (2*\pi)^{-0.5p} * |\Sigma_i|^{-0.5} * \exp\left\{-0.5*(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)\right\}$$

$$= (2*\pi)^{-0.5p} * \exp\left\{-0.5*\left[(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)+\ln|\Sigma_i|\right]\right\}. \qquad (4.0.1)$$

Päätössääntö on siten:

havainto on G(1):stä, jos

$$H(x) = \frac{f1(x)}{f2(x)} = \frac{\exp\left\{-0.5*\left[(x-\mu_1)'\,\Sigma_1^{-1}(x-\mu_1) + \ln|\Sigma_1|\right]\right\}}{\exp\left\{-0.5*\left[(x-\mu_2)'\,\Sigma_2^{-1}(x-\mu_2) + \ln|\Sigma_2|\right]\right\}} > C \qquad (4.0.2)$$

Tämä saadaan muotoon, kun h(x) = ln H(x)

$$h(x) = x'(\Sigma_1^{-1} - \Sigma_2^{-1})x + 2*(\mu_2'\,\Sigma_2^{-1} - \mu_1'\,\Sigma_1^{-1})x + C1 \qquad (4.0.3)$$

missä vakio C1 on

$$C1 = 2*\ln\left[q1*C(2!1)/q2*C(1!2)\right] + \ln(|\Sigma_1|/|\Sigma_2|) + \mu_1'\,\Sigma_1^{-1}\mu_1 - \mu_2'\,\Sigma_2^{-1}\mu_2$$

Päätössäännöksi saadaan

havainto on G(1):stä, jos h(x) < 0 , muuten G(2):sta.

Erottelufunktiota h(x) kutsutaan neliömuotoiseksi.

Tästä funktiosta saadaan erikoistapauksina, kun kovarianssimatriisit ovat yhtä suuret eli $\Sigma_1 = \Sigma_2 = \Sigma$, niin havainto on G(1):stä, mikäli

$$h(x) = x'\,\Sigma^{-1}(\mu_1 - \mu_2) - C2 > 0 \qquad (4.0.4)$$

missä vakio C2 on

$$C2 = \ln\left[q2*C(1!2)/q1*C(2!1)\right] + 0.5(\mu_1 + \mu_2)'\,\Sigma^{-1}(\mu_1 - \mu_2)$$

Erottelufunktiota h(x) kutsutaan nyt lineaariseksi.

Usein (4.0.4):n epäyhtälön ensimmäistä termiä yksinään kutsutaan erottelufunktioksi.

Mikäli a'prioritodennäköisyyksiä q1:tä ja q2:ta ei tunneta, voidaan katkaisukohta C valita siten, että asetetaan luokitteluvirheistä johtuvien kustannusten odotusarvot yhtäsuuriksi.

Mikäli luokitteluvirhekustannuksia ei tunneta, on katkaisukohdan C valinta vaikeata. Minimax-sääntöä voidaan käyttää, mikäli C(1!2)=C(2!1), mutta muuten sen valinta perustuu lähinnä tarkasteltavaan tilanteeseen.

Kun C(1!2)=C(2!1), niin Bayesin kaavasta (3.2.2.1) saadaan todennäköisyydet logististen erottelufunktioiden avulla

$$P(G(1)!x) = P(G(2)!x)*\exp(h(x)+C)$$

$$P(G(2)!x) = (1 + \exp(h(x)+C))^{-1} \qquad (4.0.5)$$

missä, kun $\Sigma_1 \neq \Sigma_2$, niin

$$h(x) = x'(\Sigma_1^{-1} - \Sigma_2^{-1})x + 2*(\mu_2'\Sigma_2^{-1} - \mu_1'\Sigma_1^{-1})x$$

ja

$$C = 2*\ln(q1/q2) + \ln(|\Sigma_1|/|\Sigma_2|) + \mu_1'\Sigma_1^{-1}\mu_1 - \mu_2'\Sigma_2^{-1}\mu_2$$

sekä, kun $\Sigma_1 = \Sigma_2$, niin

$$h(x) = x'\Sigma^{-1}(\mu_1 - \mu_2)$$

ja

$$C = \ln(q2/q1) + 0.5(\mu_1 + \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$$

4.1 EROTTELUFUNKTIO U:N JAKAUMA

Mikäli populaatioilla on samat a'prioritodennäköisyydet q1=q2=1/2 ja C(1!2)=C(2!1) sekä $\Sigma_1 = \Sigma_2 = \Sigma$, niin päätössäännöksi tulee:

havainto tulee ryhmästä G(1), mikäli

$$R1: \quad x'\Sigma^{-1}(\mu_1 - \mu_2) > 0.5*(\mu_1 + \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2) \qquad (4.1.5)$$

Olkoon X satunnainen havainto. Tarkoituksena on löytää U:n jakauma, kun kovarianssimatriisit ovat yhtä suuret, $\Sigma_1 = \Sigma_2 = \Sigma$.

$$U = X'\Sigma^{-1}(\mu_1 - \mu_2) - 0.5*(\mu_1 + \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2) \qquad (4.1.6)$$

Kun oletetaan ensin, että X on $N(\mu_1, \Sigma)$ jakautunut, niin

$$E(U) = \mu_1'\Sigma^{-1}(\mu_1 - \mu_2) - 0.5*(\mu_1 + \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$$

$$= 0.5*(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2) \qquad (4.1.7)$$

$$VAR(U) = E\left[(\mu_1 - \mu_2)'\Sigma^{-1}(X - \mu_1)(X - \mu_1)'\Sigma^{-1}(\mu_1 - \mu_2)\right]$$

$$= (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2) \qquad (4.1.8)$$

Ja sitten, että X on $N(\mu_2, \Sigma)$ jakautunut, niin

$$E(U) = \mu_2'\Sigma^{-1}(\mu_1 - \mu_2) - 0.5*(\mu_1 + \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$$

$$= -0.5*(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2) \qquad (4.1.9)$$

$$VAR(U) = E\left[(\mu_1-\mu_2)'\Sigma^{-1}(\mu_2 - X)(\mu_2 - X)'\Sigma^{-1}(\mu_1-\mu_2)\right]$$

$$= (\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2) \qquad (4.1.10)$$

Jakaumien $N(\mu_1,\Sigma)$ ja $N(\mu_2,\Sigma)$ välinen Mahalanobisetäisyys on

$$D^2(1!2) = (\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2) = m^2 \qquad (4.1.11)$$

Joten U on $N((-1)^i * 0.5 * m^2, m^2)$ jakautunut, kun X on $N(\mu_i,\Sigma)$ jakautunut, kun i=1,2.

Tapahtuman, että havainto on peräisin populaatiosta G(1), mutta se luokitellaan G(2):teen, todennäköisyys on

$$P(2!1) = \int_{-\infty}^{C} (2*\pi)^{-0.5} * m^{-1} * \exp(-0.5(z-0.5m^2)/m^2)\, dz$$

$$= \int_{-\infty}^{(C-0.5m^2)/m} (2*\pi)^{-0.5} * \exp(-0.5y^2)\, dy \qquad (4.1.12)$$

sekä vastaavasti havainto on peräisin populaatiosta G(2), mutta se luokitellaan G(1):teen, tapahtuman todennäköisyys on

$$P(1!2) = \int_{C}^{\infty} (2*\pi)^{-0.5} * m^{-1} * \exp(-0.5(z+0.5m^2)/m^2)\, dz$$

$$= \int_{(C+0.5m^2)/m}^{\infty} (2*\pi)^{-0.5} * \exp(-0.5y^2)\, dy \qquad (4.1.13)$$

Minimax-ratkaisuksi saadaan ratkaisemalla katkaisukohta C yhtälöstä

$$C(1!2)*P(1!2) = C(2!1)*P(2!1)$$

eli

$$C(1|2) \int_{(C+0.5m^2)/m}^{\infty} (2*\pi)^{-0.5} *\exp(-0.5y^2)dy$$

$$= C(2|1) \int_{-\infty}^{(C-0.5m^2)/m} (2*\pi)^{-0.5} *\exp(-0.5y^2)dy \qquad (4.1.14)$$

Päätössäännöksi saadaan siten

mikäli G(i):llä on tiheysfunktio, i=1,2

$$fi(x) = (2*\pi)^{-0.5p} *|\Sigma|^{-0.5} * \exp(-0.5*(x-\mu_i)'\Sigma^{-1}(x-\mu_i)) \qquad (4.1.15)$$

niin luokittelun minimax-alueet ovat

$$R1: \quad x'\Sigma^{-1}(\mu_1-\mu_2) - 0.5*(\mu_1+\mu_2)'\Sigma^{-1}(\mu_1-\mu_2) \geq C$$

$$R2: \quad x'\Sigma^{-1}(\mu_1-\mu_2) - 0.5*(\mu_1+\mu_2)'\Sigma^{-1}(\mu_1-\mu_2) < C$$

missä C on valittu siten, että (4.1.14) toteutuu.

Tästä saadaan erikoistapaukseksi, kun väärinluokittelukustannukset ovat yhtäsuuret eli C(1|2) = C(2|1), silloin C=0 ja väärinluokittelutodennäköisyydet ovat

$$P(1|2) = P(2|1) = \int_{0.5m}^{\infty} (2*\pi)^{-0.5} *\exp(-0.5y^2)dy = \Phi(-m/2)$$

5.0 KAKSI MULTINORMAALISTI JAKAUTUNUTTA POPULAATIOTA,

KUN JAKAUMAPARAMETREJA EI TUNNETA

Useimmissa tapauksissa multinormaalisti jakautuneiden populaatioiden parametreja ei tunneta, vaan ne joudutaan estimoimaan kunkin ryhmän otoksesta. Tarkastellaan tapausta, jossa on otos kummastakin kahdesta multinormaalisti jakautuneesta populaatiosta ja pyritään käyttämään saatua informaatiota hyväksi luokittelemaan uusi havainto jompaan kumpaan populaation kuuluvaksi.

Oletaan , että on otos $x_1^{(1)},\ldots,x_{n1}^{(1)}$ $N(\mu_1,\Sigma)$:stä ja otos $x_1^{(2)},\ldots,x_{n2}^{(2)}$ $N(\mu_2,\Sigma)$:stä, käytetään estimaatteja

$$\mu_i\text{:lle} \qquad \bar{x}^{(i)} = \sum_{j=1}^{ni} x_j^{(i)} / ni \qquad i=1,2 \tag{5.0.1}$$

$$\Sigma\text{:lle} \qquad S = \frac{1}{n1+n2-2} * \left[\sum_{i=1}^{2} \sum_{j=1}^{ni} (x_j^{(i)} - \bar{x}^{(i)})(x_j^{(i)} - \bar{x}^{(i)})' \right]$$

Näillä estimaateilla korvataan (4.1.5):n parametrit, niin saadaan

$$W(x) = x'S^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)}) - 0.5(\bar{x}^{(1)} + \bar{x}^{(2)})'S^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)}) \tag{5.0.2}$$

Pyritään käyttää tätä funktiota luokittelukriteerinä samalla tavalla kuin (4.1.5) käytettiin. Kaavan (5.0.2) ensimmäistä termiä kutsutaan kahteen otokseen perustuvaksi erottelufunktioksi. Se on havainnon x suhteen lineaarinen funktio, jolla on suurin varianssi joukkojen välillä suhteessa varianssiin joukkojen sisällä. ((4))

Kun a'prioritodennäköisyydet tunnetaan, voidaan väittää, että luokittelukriteeri on paras siinä mielessä, että se minimoi tappion odotusarvon. Kuitenkin W:tä ei voida käyttää samalla tavalla kuin teoreettista U:ta käytettiin, tämä johtuu W:n jakaumasta.

5.1 EROTTELUFUNKTIO W:N JAKAUMA

Olkoon

$$W = X'S^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) - 0.5(\bar{X}^{(1)} + \bar{X}^{(2)})'S^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)})$$

$$= \left[X - 0.5(\bar{X}^{(1)} + \bar{X}^{(2)}) \right]'S^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) \tag{5.1.1}$$

missä on nyt satunnaismuuttujat X, $\bar{X}^{(1)}$, $\bar{X}^{(2)}$ ja S. Tällöin W:n jakauma on äärimmäisen monimutkainen riippuen otoskoi'ista, n1:stä ja n2:sta sekä tuntemattomasta Mahalanobisetäisyydestä.

Olkoon uudet satunnaismuuttujat Y1 ja Y2, ((3))

$$Y1 = C1*\left[X - (n1*\bar{X}^{(1)} + n2*\bar{X}^{(2)})/(n1+n2) \right] \tag{5.1.2}$$

$$Y2 = C2*(\bar{X}^{(1)} - \bar{X}^{(2)}) \tag{5.1.3}$$

missä

$$C1 = \left[(n1+n2)/(n1+n2+1)\right]^{0.5}$$

$$C2 = \left[(n1*n2)/(n1+n2)\right]^{0.5}$$

Tällöin

$$EY1 = C1*n2*(\mu_1-\mu_2)/(n1+n2) \text{ , kun X on G(1):stä}$$

$$EY1 = -C1*n1*(\mu_1-\mu_2)/(n1+n2) \text{ , kun X on G(2):stä} \qquad (5.1.4)$$

$$EY2 = C2*(\mu_1-\mu_2) \text{ , kun X on G(1):stä tai G(2):sta}$$

$$VAR(Y2) = VAR(Y1) = \Sigma .$$

Silloin Y1 ja Y2 ovat riippumattomia normaalijakautuneita muuttujia. Olkoon Y = (Y1,Y2) ja

$$M = Y'S^{-1}Y = \begin{bmatrix} m11 & m12 \\ m21 & m22 \end{bmatrix}$$

Jolloin saadaan

$$W = m12*\left[(n1+n2+1)/n1*n2\right]^{0.5} + m22*(n1-n2)/(2*n1*n2) \qquad (5.1.5)$$

Jos n1=n2, niin W:n jakauma, kun X tulee G(1):stä, on sama kuin -W:n, kun X tulee G(2):sta. Joten jos W > 0 on alue, johon luokitellaan G(1):stä tulevat, niin väärinluokittelutodennäköisyydet ovat yhtäsuuret, P(1!2)=P(2!1).

5.2 W:N ASYMPTOOTTINEN JAKAUMA

Siinä tapauksessa, että otokset ovat suuria $N(\mu_1,\Sigma)$:stä ja $N(\mu_2,\Sigma)$:sta, voidaan soveltaa jakaumien raja-arvoteoriaa.

Koska $\bar{x}^{(1)}$ on n1:n riippumattoman havainnon otoskeskiarvo $N(\mu_1,\Sigma)$:stä, niin tiedetään, että

$$\text{plim } \bar{x}^{(1)} = \mu_1 \text{ , kun } n1 \longrightarrow \infty \qquad (5.2.1)$$

tarkalleen ilmaistuna tämä tarkoittaa, että mielivaltaisia positiivisia lukuja δ ja ε kohti on olemassa luku N siten, että

$$\text{kun } n1 > N \text{ , niin } P(|\bar{x}^{(1)} - \mu_1| < \delta) > 1 - \varepsilon .$$

Vastaavasti saadaan, kun n1,n2—>∞,

$$\text{plim } \bar{x}^{(2)} = \mu_2$$

$$\text{plim } S = \Sigma$$

$$\Longrightarrow \text{plim } S^{-1} = \Sigma^{-1}$$

$$\Longrightarrow \text{plim } S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)}) = \Sigma^{-1} (\mu_1 - \mu_2) \qquad (5.2.2)$$

$$\Longrightarrow \text{plim } (\bar{x}^{(1)} + \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)}) = (\mu_1 + \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)$$

Tällöin W:n jakauman raja-arvona on U:n jakauma eli W:n rajajakauma on $N((-1)^i * 0.5 * m^2, m^2)$, kun x on $N(\mu_i, \Sigma)$, i=1,2.

6.0 MUITA LÄHESTYMISTAPOJA

6.1 REGRESSIOANALYYTTINEN LÄHESTYMISTAPA ((5))

Olkoon

$$y_j^{(1)} = \frac{n2}{n1+n2} \quad ,\text{kun } j=1,\ldots,n1$$

ja

$$y_j^{(2)} = \frac{-n1}{n1+n2} \quad ,\text{kun } j=1,\ldots,n2. \qquad (6.1.1)$$

tällöin

$$\bar{y} = 0$$

ja

$$\bar{x} = \frac{n1 * \bar{x}^{(1)} + n2 * \bar{x}^{(2)}}{n1 + n2} \quad , \text{ missä } \bar{x}^{(i)} = \frac{\sum_{j=1}^{ni} x_j^{(i)}}{ni}$$

Minimoidaan lauseke

$$\sum_{i=1}^{2} \sum_{j=1}^{ni} \left[y_j^{(i)} - b'(x_j^{(i)} - \bar{x}) \right]^2$$

eli

$$\frac{d \left\{ \sum_{i=1}^{2} \sum_{j=1}^{ni} \left[y_j^{(i)} - b'(x_j^{(i)} - \bar{x}) \right]^2 \right\}}{db} \qquad (6.1.2)$$

$$= \sum_{i=1}^{2} \sum_{j=1}^{ni} \left[-2y_j^{(i)} (x_j^{(i)} - \bar{x}) + 2(x_j^{(i)} - \bar{x})(x_j^{(i)} - \bar{x})'b \right] = 0$$

josta saadaan

$$\sum_{i=1}^{2}\sum_{j=1}^{ni} y_j^{(i)}\,(x_j^{(i)} - \bar{x}) = \sum_{i=1}^{2}\sum_{j=1}^{ni} (x_j^{(i)} - \bar{x})(x_j^{(i)} - \bar{x})'b$$

Tällöin vasenpuoli on

$$X'y = \sum_{i=1}^{2}\sum_{j=1}^{ni} y_j^{(i)} * (x_j^{(i)} - \bar{x}) = \frac{n1*n2}{n1+n2}(\bar{x}^{(1)} - \bar{x}^{(2)}) \qquad (6.1.3)$$

ja oikea puoli ilman b:tä on

$$X'X = \sum_{i=1}^{2}\sum_{j=1}^{ni} (x_j^{(i)} - \bar{x})(x_j^{(i)} - \bar{x})'$$

$$= \sum_{j=1}^{n1} (x_j^{(1)} - \bar{x})(x_j^{(1)} - \bar{x})' + \sum_{j=1}^{n2} (x_j^{(2)} - \bar{x})(x_j^{(2)} - \bar{x})'$$

$$= \sum_{j=1}^{n1} (x_j^{(1)} - \bar{x}^{(1)})(x_j^{(1)} - \bar{x}^{(1)})' + \sum_{j=1}^{n2} (x_j^{(2)} - \bar{x}^{(2)})(x_j^{(2)} - \bar{x}^{(2)})'$$

$$+ n1*(\bar{x}^{(1)} - \bar{x})(\bar{x}^{(1)} - \bar{x})' + n2*(\bar{x}^{(2)} - \bar{x})(\bar{x}^{(2)} - \bar{x})'$$

$$= (n1+n2-2)S + \frac{n1*n2}{n1+n2}(\bar{x}^{(1)} - \bar{x}^{(2)})(\bar{x}^{(1)} - \bar{x}^{(2)})' \qquad (6.1.4)$$

Merkitään $a = (\bar{x}^{(1)} - \bar{x}^{(2)})'b$ (skalaari), saadaan koska

$$X'Xb = X'y$$

niin

$$\left[(n1+n2-2)S + \frac{n1*n2}{n1+n2}*(\bar{x}^{(1)} - \bar{x}^{(2)})(\bar{x}^{(1)} - \bar{x}^{(2)})'\right]b = \frac{n1*n2}{n1+n2}*(\bar{x}^{(1)} - \bar{x}^{(2)})$$

ja ottamalla merkintä huomioon

$$(n1+n2-2)Sb = \frac{n1*n2}{n1+n2}*(\bar{x}^{(1)} - \bar{x}^{(2)})(1 - a) \qquad (6.1.5)$$

siis vektori b on verrannollinen $S^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)})$:teen.

6.2 SUURIN USKOTTAVUUS

Olkoon testattavana nollahypoteesina, että

$x, x_1^{(1)},\ldots,x_{n1}^{(1)}$ ovat peräisin $N(\mu_1,\Sigma)$:stä ja $x_1^{(2)},\ldots,x_{n2}^{(2)}$ $N(\mu_2,\Sigma)$:sta

kun vastahypoteesina on

$x_1^{(1)},\ldots,x_{n1}^{(1)}$ ovat peräisin $N(\mu_1,\Sigma)$:stä ja $x, x_1^{(2)},\ldots,x_{n2}^{(2)}$ $N(\mu_2,\Sigma)$:sta

Ho:n vallitessa suurimman uskottavuuden estimaatit μ_1:lle, μ_2:lle ja Σ:lle, ovat

$$\underset{0}{\bar{x}}^{(1)} = \frac{n1*\bar{x}^{(1)} + x}{n1 + 1}$$

$$\underset{0}{\bar{x}}^{(2)} = \bar{x}^{(2)}$$

$$S_0 = \frac{1}{n1+n2+1}\left[\sum_{j=1}^{n1}(x_j^{(1)} - \underset{0}{\bar{x}}^{(1)})(x_j^{(1)} - \underset{0}{\bar{x}}^{(1)})' + (x-\underset{0}{\bar{x}}^{(1)})(x-\underset{0}{\bar{x}}^{(1)})' + \sum_{j=1}^{n2}(x_j^{(2)} - \underset{0}{\bar{x}}^{(2)})(x_j^{(2)} - \underset{0}{\bar{x}}^{(2)})'\right]$$

missä S_0 voidaan kirjoittaa muotoon

$$S_0 = \frac{1}{n1+n2+1}\left[A + \frac{n1}{n1+1}(x - \underset{0}{\bar{x}}^{(1)})(x - \underset{0}{\bar{x}}^{(1)})'\right]$$

kun $A = \sum_{i=1}^{2}\sum_{j=1}^{ni}(x_j^{(i)} - \bar{x}^{(i)})(x_j^{(i)} - \bar{x}^{(i)})'$ ja siis $S^{-1} = A^{-1}/(n1+n2-2)$

Hypoteesi Ho on siten

Ho:

$$\underset{0}{\bar{x}}^{(1)} = \frac{n1*\bar{x}^{(1)} + x}{n1 + 1}$$

$$\underset{0}{\bar{x}}^{(2)} = \bar{x}^{(2)} \qquad (6.2.1)$$

$$S_0 = \frac{1}{n1+n2+1}\left[A + \frac{n1}{n1+1}(x - \underset{0}{\bar{x}}^{(1)})(x - \underset{0}{\bar{x}}^{(1)})'\right]$$

Vaihtoehtoinen hypoteesi on

H1:

$$\underset{1}{\bar{x}}^{(1)} = \bar{x}^{(1)}$$

$$\underset{1}{\bar{x}}^{(2)} = \frac{n2*\bar{x}^{(2)} + x}{n2 + 1} \qquad (6.2.2)$$

$$S_1 = \frac{1}{n1+n2+1}\left[A + \frac{n2}{n2+1}(x - \bar{x}^{(2)})(x - \bar{x}^{(2)})'\right]$$

Suurimman uskottavuuden säännöksi saadaan

$$\frac{\left| S_1 \right|^{(n1+n2+1)/2}}{\left| S_0 \right|^{(n1+n2+1)/2}} = \frac{\left| A + \frac{n2}{n2+1}(x - \bar{x}^{(2)})(x - \bar{x}^{(2)})' \right|^{(n1+n2+1)/2}}{\left| A + \frac{n1}{n1+1}(x - \bar{x}^{(1)})(x - \bar{x}^{(1)})' \right|^{(n1+n2+1)/2}}$$

tämä saadaan muotoon

$$= \frac{\left[n + \frac{n2}{n2+1}(x - \bar{x}^{(2)})'S^{-1}(x - \bar{x}^{(2)}) \right]^{(n1+n2+1)/2}}{\left[n + \frac{n1}{n1+1}(x - \bar{x}^{(1)})'S^{-1}(x - \bar{x}^{(1)}) \right]^{(n1+n2+1)/2}} \overset{>}{=} 1 \qquad (6.2.3)$$

missä n = n1+n2-2.

Päätössäännöksi saadaan: havainto on G(1):stä, jos alue R1 valitaan siten, että

$$\text{R1:} \quad n + \frac{n2}{n2+1}(x - \bar{x}^{(2)})'S^{-1}(x - \bar{x}^{(2)})$$

$$\overset{>}{=} n + \frac{n1}{n1+1}(x - \bar{x}^{(1)})'S^{-1}(x - \bar{x}^{(1)}) \qquad (6.2.4)$$

Olkoon

$$Z = \frac{1}{2}\left[\frac{n2}{n2+1} (x - \bar{x}^{(2)})'S^{-1} (x - \bar{x}^{(2)}) - \frac{n1}{n1+1} (x - \bar{x}^{(1)})'S^{-1} (x - \bar{x}^{(1)}) \right] \qquad (6.2.5)$$

Suurimman uskottavuuden sääntö on tällöin:

havainto on G(1):stä, mikäli $Z > 0$ ja G(2):sta, mikäli $Z < 0$.

Koska

$$\lim (W - Z) = 0, \text{ kun } n1,n2 \longrightarrow \infty \qquad (6.2.6)$$

niin väärinluokittelun asymptoottiset ominaisuudet ovat samat W:lle ja Z:lle.

6.3 VÄÄRINLUOKITTELUTODENNÄKÖISYYDET

Tarkastellaan tilannetta, kun käytetään erottelufunktiona W:tä tai Z:taa. Kuten kohdissa 5.1 ja 6.2 jo esitettiin, W:n ja Z:n jakaumat ovat äärimmäisen hankalia laskea. Siksi käsitellään niiden asymptoottisia laajennuksia, kun n1 ja n2 kasvavat.

Perusteena on, että W ja Z ovat $N((-1)^{-i} *0.5*m^2 , m^2)$, kun havainto on G(i):stä, i=1,2.

Esitetään tulokset lauseiden muodossa, itse todistukset löytyvät kirjallisuusviittausten avulla.

Olkoon $\Phi(.)$ ja $\phi(.)$ N(0,1)-jakauman kertymä- ja tiheysfunktio ja $o(.)$ on funktio, jolle $\lim o(n)=0$, kun $n \longrightarrow 0$.

Kohdassa (5.1.1):ssä W määriteltiin

$$W = x'S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)}) - 0.5(\bar{x}^{(1)} + \bar{x}^{(2)})'S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})$$

ja kohdassa (6.2.5):ssä Z vuorostaan

$$Z = \frac{1}{2}\left[\frac{n2}{n2+1} (x - \bar{x}^{(2)})'S^{-1} (x - \bar{x}^{(2)}) - \frac{n1}{n1+1} (x - \bar{x}^{(1)})'S^{-1} (x - \bar{x}^{(1)}) \right]$$

LAUSE 1A. Kun n1—>∞, n2—>∞ ja n1/n2—>a (>0), n=n1+n2-2, niin ((6))

$$P\left[\frac{W-0.5*m^2}{m} < u \;\middle|\; G(1)\right] =$$

$$\Phi(u) - \phi(u)*\left\{\frac{1}{2*n1*m^2}*\left[u^3 + (p-3)u - pm\right]\right.$$

$$+ \frac{1}{2*n2*m^2}*\left[u^3 + 2mu^2 + (p-3+m^2)u + (p-2)m\right]$$

$$\left. + \frac{1}{4*n}*\left[4u^3 + 4mu^2 + (6p-6+m^2)u + 2(p-1)m\right]\right\}$$

$$+ o(n^{-2}) \qquad (6.3.1)$$

ja

$$P\left[-\frac{W-0.5*m^2}{m} < u \;\middle|\; G(2)\right]$$ on (6.3.1), kun n1 ja n2 vaihdetaan.

Lauseesta saadaan väärinluokittelutodennäköisyydet, kun $u=(C-0.5m^2)/m$ ja $u=-(C+0.5m^2)/m$. Kun C = 0, silloin on u = -0.5*m. Jos n1=n2, niin lause määrittää tarkan minimax-menetelmän. ((7))

Lausetta 1a vastaava lause Z:lle on seuraava.

LAUSE 1B. Kun n1—>∞, n2—>∞ ja n1/n2—>a (>0), n=n1+n2-2, niin ((8))

$$P\left[\frac{Z-0.5*m^2}{m} < u \;\middle|\; G(1)\right] =$$

$$\Phi(u) - \phi(u)*\left\{\frac{1}{2*n1*m^2}*\left[u^3 + mu^2 + (p-3)u - m\right]\right.$$

$$\left. + \frac{1}{2*n2*m^2}*\left[u^3 + mu^2 + (p-3-m^2)u - m^3 - m\right]\right.$$

$$+ \frac{1}{4*n} * \left[4u^3 + 4mu^2 + (6p-6+m^2)u + 2(p-1)m \right] \Bigg\}$$

$$+ o(n^{-2}) \qquad (6.3.2)$$

ja

$$P\left[-\frac{W+0.5*m^2}{m} \underset{=}{<} u \;\middle|\; G(2) \right]$$ on (6.3.2), kun n1 ja n2 vaihdetaan.

Kun C=0, silloin on u = -0.5m. Kun n1 = n2, niin Z:n päätössääntö on identtinen W:n kanssa ja väärinluokittelutodennäköisyys on (6.3.3).

KOROLLAARI 1A. Kun n1—>∞, n2—>∞ ja n1/n2—>a (>0), niin ((4))

$$P\left[W < 0 \;\middle|\; G(1), \lim_{n\to\infty} n1/n2 = 1 \right]$$

$$= \Phi(-0.5m) + \frac{\phi(0.5m)}{n} * \left[\frac{p-1}{m} + \frac{p}{4} * m \right] + o(n^{-1}) \qquad (6.3.3)$$

$$= P\left[W > 0 \;\middle|\; G(2), \lim_{n\to\infty} n1/n2 = 1 \right]$$

Koska m on yleensä tuntematon, niin käytetään sen otosestimaattia d eli Mahalanobiksen kahden otoksen välistä etäisyyden neliöjuurta

$$D^2(1!2) = (\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)}) = d^2, \qquad (6.3.4)$$

jonka odotusarvo on, ((9))

$$E\left[D^2(1!2) \right] = \frac{n1+n2-2}{n1+n2-p-3} \left[m^2 + p * \frac{n1+n2}{n1*n2} \right] \qquad (6.3.5)$$

LAUSE 2A. Kun n1/n2—>a (>0), n—> ∞, niin
((10))

$$P\left[\frac{W-0.5*d^2}{d} \leq u \,\middle|\, G(1)\right] =$$

$$\Phi(u) - \phi(u)*\left\{\frac{1}{n1}\left[\frac{u}{2} - \frac{p-1}{m}\right] + \frac{1}{n}\left[\frac{u^3}{4} + (p-0.75)u\right]\right\} \tag{6}$$

$$+ o(n^{-2}) \tag{6.3.6}$$

ja

$$P\left[-\frac{W+0.5*d^2}{d} \leq u \,\middle|\, G(2)\right]$$ on (6.3.6) kun n1 ja n2 vaihdetaan.

LAUSE 2B. Kun n1/n2—>a (>0), n—> ∞, niin
((11))

$$P\left[\frac{Z-0.5*d^2}{d} \leq u \,\middle|\, G(1)\right] =$$

$$\Phi(u) - \phi(u)*\left\{\frac{1}{2*n1*m}\left[u^2 + mu - p + 1\right]\right.$$

$$- \frac{1}{2*n2*m}\left[u^2 + 2mu + p - 1 + m^2\right]$$

$$\left. + \frac{1}{4*n}\left[u^3 + (4p-3)u\right]\right\}$$

$$+ o(n^{-2}) \tag{6.3.7}$$

ja

$$P\left[-\frac{Z+0.5*d^2}{d} \leq u \,\middle|\, G(2)\right]$$ on (6.3.7) kun n1 ja n2 vaihdetaan.

Tavallisesti ollaan kiinnostuneita tilanteesta u < 0 eli pienistä virhetodennäköisyyksistä. Tällöin korjaustermi on positiivinen ja normaaliapproksimaatio aliestimoi väärinluokittelutodennäköisyyden.

Voidaan asettaa katkaisupiste C siten, että väärinluokittelu on kontroloitu. Olkoon siis $P(W < C \mid G(1)) = \alpha$.

LAUSE 3A. Olkoon u0 sellainen, että $\Phi(u0) = \alpha$ ja olkoon ((10))

$$u = u0 - \frac{1}{n1}*\left[\frac{p-1}{d} - 0.5*u0\right] + \frac{1}{n}*\left[(p-0.75)*u0 + 0.25*u0^{3}\right] \qquad (6.3.8)$$

Silloin kun $n1 \to \infty$, $n2 \to \infty$ ja $n1/n2 \to a$ (>0), niin

$$P\left[\frac{W-0.5*d^{2}}{d} \leq u \mid G(1)\right] = \alpha + o(n^{-2}) \qquad (6.3.9)$$

Kun katkaisukohta $C = d*u + 0.5*d^{2}$, niin väärinluokittelutodennäköisyys on likipitäen α .

LAUSE 3B. Olkoon u0 sellainen, että $\Phi(u0) = \alpha$ ja olkoon ((12))

$$u = u0 + \frac{1}{2*n1*d}\left(u0^{2} + d*u0 - p + 1\right)$$

$$- \frac{1}{2*n2*d}\left(u0^{2} + d*u0 + p - 1 - d^{2}\right)$$

$$+ \frac{1}{4*n}\left(u0^{3} + (4p-5)u0\right) \qquad (6.3.10)$$

Silloin kun $n1 \to \infty$, $n2 \to \infty$ ja $n1/n2 \to a$ (>0), niin

$$P\left[\frac{W-0.5*d^{2}}{d} \leq u \mid G(1)\right] = \alpha + o(n^{-2}) \qquad (6.3.11)$$

Koska W on ehdollisesti normaalijakautunut, ehtoina $\bar{x}^{(1)}$, $\bar{x}^{(2)}$ ja S, niin ehdollinen odotusarvo on, kun x on G(i):stä, i=1,2

$$E\left[W \mid \bar{x}^{(1)}, \bar{x}^{(2)}, S\right] = \left[\mu^{(i)} - 0.5*(\bar{x}^{(1)} + \bar{x}^{(2)})\right]' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})$$

$$= \mu^{(i)}(\bar{x}^{(1)}, \bar{x}^{(2)}, S) \qquad (6.3.12)$$

ja ehdollinen varianssi

$$\mathrm{VAR}\left[W \,\middle|\, \bar{x}^{(1)}, \bar{x}^{(2)}, S \right] = (\bar{x}^{(1)} - \bar{x}^{(2)})' \Sigma^{-1} S \Sigma^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})$$

$$= m^2 (\bar{x}^{(1)}, \bar{x}^{(2)}, S) \qquad (6.3.13)$$

Näillä on todennäköisyysraja-arvot

$$\operatorname*{plim}_{n1,n2 \to \infty} \mu^{(i)} (\bar{x}^{(1)}, \bar{x}^{(2)}, S) = (-1)^{i+1} 0.5*m^2$$

$$\operatorname*{plim}_{n1,n2 \to \infty} m^2 (\bar{x}^{(1)}, \bar{x}^{(2)}, S) = m^2 \qquad (6.3.14)$$

Kun C on katkaisupiste, niin väärinluokittelutodennäköisyydet ovat

$$P\left[2 \,\middle|\, 1;C,\bar{x}^{(1)},\bar{x}^{(2)}, S \right] = \Phi\left\{ \frac{C - \mu^{(1)} (\bar{x}^{(1)}, \bar{x}^{(2)}, S)}{m^2 (\bar{x}^{(1)}, \bar{x}^{(2)}, S)} \right\} \longrightarrow \Phi(u1) \qquad (6.3.15)$$

$$P\left[1 \,\middle|\, 2;C,\bar{x}^{(1)},\bar{x}^{(2)}, S \right] = 1-\Phi\left\{ \frac{C - \mu^{(2)} (\bar{x}^{(1)}, \bar{x}^{(2)}, S)}{m^2 (\bar{x}^{(1)}, \bar{x}^{(2)}, S)} \right\} \longrightarrow 1-\Phi(u2) \qquad (6.3.16)$$

$n1,n2 \to \infty$

missä (6.3.15):ssa on $c = d*u1 + 0.5*d^2$, jolloin $\Phi(y)$:n argumentti y

$$y = \frac{u1*d+(\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})}{m} \longrightarrow u1 \quad , \text{ kun } n1,n2 \to \infty$$

sekä (6.3.16):ssa on $c = d*u2 + 0.5*d^2$ sekä vastaavasti

$$y = \frac{u2*d+(\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})}{m} \longrightarrow u2 \quad , \text{ kun } n1,n2 \to \infty$$

Tarkastellaan (6.3.15):n tilannetta , kun $C = d*u1 + 0.5*d^2$ ja u1 valitaan siten, että kontrolloidaan

$$P(2!1;C,\bar{x}^{(1)},\bar{x}^{(2)}, S):ää.$$

Tämä ehdollinen todennäköisyys $\bar{x}^{(1)}$:n, $\bar{x}^{(2)}$:n ja S:n funktiona on satunnaismuuttuja, jonka jakauma voidaan approksimoida.

LAUSE 4A. Kun n1—>oo, n2—>oo ja n1/n2—>a (>0), niin ((14))

$$P\left\{ n^{0.5} * \frac{P[2|1;d*u1 + 0.5*d^2, \bar{x}^{(1)}, \bar{x}^{(2)}, S] - \Phi(u1)}{\phi(u1)(0.5*u1^2 + n/n1)^{0.5}} < x \right\}$$

$$= \Phi\left\{ x - \frac{(p-1)*n/n1 - (p-0.75+n/n1)*u1 - 0.25*u1^3}{(n*(0.5*u1^2 + n/n1))^{0.5}} \right\} + o(n^{-2}) \qquad (6.3.17)$$

Z:n ehdollinen jakauma ei ole normaalinen, koska Z on x:n suhteen toista-astetta, ellei n1=n2. Joten (6.3.12):ta ja (6.3.13):ta vastaavia esitysmuotoja ei saada Z:lle.

LAUSE 4B. Kun n1—>oo, n2—>oo ja n1/n2—>a (>0), niin ((15))

$$P\left\{ k* \frac{P[2|1; 0, \bar{x}^{(1)}, \bar{x}^{(2)}, S] - \Phi(-0.5*m)}{\phi(0.5*m)} < x \right\}$$

$$= \Phi\left\{ x - k*\left[\frac{1}{16*n1*m}*(4p - 4 - m^2) + \frac{1}{16*n2}*(4p - 4 + 3m^2) - \frac{1}{4*n}*(p - 1)*m \right] \right\} + o(n^{-2}) \qquad (6.3.19)$$

missä $k = 2 * \sqrt{\frac{n1*n2}{n1+n2}}$

Peter A. Lachenbruch ((9)) on johtanut seuraavat approksimaatiot

$$E\left[W \mid G(1)\right] = 0.5*\frac{n1+n2-2}{n1+n2-p-3}*\left[m^2 - \frac{p(n2-n1)}{n1*n2}\right]$$

$$E\left[W \mid G(2)\right] = -0.5*\frac{n1+n2-2}{n1+n2-p-3}*\left[m^2 + \frac{p(n2-n1)}{n1*n2}\right] \qquad (6.3.20)$$

$$E\left[VAR(W)\right] = \frac{(n1+n2-3)(n1+n2-2)^2}{(n1+n2-p-2)(n1+n2-p-3)(n1+n2-p-5)}\left[m^2 + \frac{p(n1+n2)}{n1*n2}\right]$$

Näitä tuloksia voidaan käyttää arvoitaessa otoskoon n (=n1+n2) riittävyyttä. Sillä havaintojen lukumäärä riippuu erottelumuuttujien lukumäärästä p, ryhmien välisestä Mahalanobisetäisyydestä m sekä väärinluokittelutodennäköisyydestä P(2!1).

Tiedetään, että jos jokaista δ , ε paria kohti on olemassa N siten, että kun n > N, niin $P(|U - W| < \delta) > 1 - \varepsilon$, joten W on U:n konsistenssi estimaatti.

Kun ryhmät G(i) ovat multinormaalisti jakautuneita, i=1,2 sekä $\Sigma_1 = \Sigma_2$, n1=n2 ja q1=q2, niin saadaan tarvittavien havaintojen lukumäärille seuraavat arviot.((16))

Mikäli väärinluokittelutodennäköisyyden P(2!1) pitäisi olla < 0.05 ja jos m > 2, niin havaintoja tarvitaan noin 2.5*p kappaletta, mutta jos m < 2, niin havaintoja noin 3.5*p kappaletta.

Mutta jos väärinluokittelutodennäköisyyden P(2!1) pitäisi olla < 0.01 ja jos m > 3, niin havaintoja tarvitaan jo 10*p kappaletta, mutta jos m < 1, niin havaintoja tarvitaankin peräti 20*p kappaletta.

7.0 KAKSI MULTINORMAALISTI JAKAUTUNUTTA POPULAATIOTA, KUN KOVARIANSSIMATRIISIT OVAT ERISUURET

7.1 USKOTTAVUUSMENETELMÄ

Olkoon G(i) $N(\mu_i, \Sigma_i)$ jakautunut , i=1,2 ja $\mu_1 \neq \mu_2$ sekä $\Sigma_1 \neq \Sigma_2$. Kun parametrit tunnetaan, uskottavuussuhde on

$$\frac{f1(x)}{f2(x)} = \frac{|\Sigma_2|^{0.5} \exp\left[-0.5(x-\mu_1)' \Sigma_1^{-1} (x-\mu_1)\right]}{|\Sigma_1|^{0.5} \exp\left[-0.5(x-\mu_2)' \Sigma_2^{-1} (x-\mu_2)\right]} \qquad (7.1.1)$$

$$= \exp\left[0.5\left\{(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2) - (x-\mu_1)'\Sigma_1^{-1}(x-\mu_2) + \ln\frac{|\Sigma_2|}{|\Sigma_1|}\right\}\right]$$

Kun parametrit ovat tuntemattomia, testataan nollahypoteesia

$x, x_1^{(1)}, \ldots, x_{n1}^{(1)}$ ovat peräisin $N(\mu_1, \Sigma_1)$:stä ja $x_1^{(2)}, \ldots, x_{n2}^{(2)}$ $N(\mu_2, \Sigma_2)$:sta

kun vastahypoteesina on

$x_1^{(1)}, \ldots, x_{n1}^{(1)}$ ovat peräisin $N(\mu_1, \Sigma_1)$:stä ja $x, x_1^{(2)}, \ldots, x_{n2}^{(2)}$ $N(\mu_2, \Sigma_2)$:sta

Ho:n vallitessa suurimman uskottavuuden estimaatit μ_i:lle ja Σ_i:lle, kun i=1,2 , ovat

H0:

$$\bar{x}^1(0) = \frac{n1 * \bar{x}^{(1)} + x}{n1 + 1}$$

$$\bar{x}^2(0) = \bar{x}^{(2)}$$

$$S^1(0) = \frac{1}{n1+n2+1}\left[A1 + \frac{n1}{n1+1}(x - \bar{x}^{(1)})(x - \bar{x}^{(1)})'\right] \qquad (7.1.2)$$

$$S^2(0) = \frac{1}{n2} A2$$

kun $Ai = \sum_{j=1}^{ni} (x_j^{(i)} - \bar{x}^{(i)})(x_j^{(i)} - \bar{x}^{(i)})'$, kun i=1,2

Vaihtoehtoinen hypoteesi on

H1:

$$\bar{x}^{1}(1) = \bar{x}^{(1)}$$

$$\bar{x}^{2}(1) = \frac{n2*\bar{x}^{(2)} + x}{n2 + 1}$$

$$S^{1}(1) = \frac{1}{n1} A1 \qquad (7.1.3)$$

$$S^{2}(1) = \frac{1}{n1+n2+1} \left[A2 + \frac{n2}{n2+1} (x - \bar{x}^{(2)})(x - \bar{x}^{(2)})' \right]$$

Uskottavuussuhde on,

$$\frac{\left| S^{1}(1) \right|^{n1/2} \left| S^{2}(1) \right|^{(n2+1)/2}}{\left| S^{1}(0) \right|^{(n1+1)/2} \left| S^{2}(0) \right|^{n2/2}}$$

$$= \frac{\left\{ \left[1+(x-\bar{x}^{(2)})'A2^{-1}(x-\bar{x}^{(2)}) \right]^{n2+1} \frac{(n1+1)p}{(n1+1)} \frac{n2*p}{n2} |A2| \right\}^{0.5}}{\left\{ \left[1+(x-\bar{x}^{(1)})'A1^{-1}(x-\bar{x}^{(1)}) \right]^{n1+1} \frac{(n2+1)p}{(n2+1)} \frac{n1*p}{n1} |A1| \right\}^{0.5}} \qquad (7.1.4)$$

Päätössäännöksi saadaan:

havainto x kuuluu G(1):teen, mikäli (7.1.4) on > 1 ja
G(2):teen, mikäli (7.1.4) on < 1 .

7.2 SIJOITUSMENETELMÄ

Yhtä hyvin oltaisi voitu sijoittaa estimaatit logaritmilausekkeeseen ja käyttää erottelufunktiona

$$h(x) = 0.5* \left[(x-\bar{x}^{(2)})'S2^{-1}(x-\bar{x}^{(2)}) - (x-\bar{x}^{(1)})'S1^{-1}(x-\bar{x}^{(1)}) \right] \qquad (7.2.1)$$

Päätössäännöksi saadaan tällöin:

havainto x kuuluu G(1):teen, mikäli (7.2.1) on suuri ja
G(2):teen, mikäli (7.2.1) on pieni.

Tästä saadaan erikoistapauksina:

1. Kun a'prioritodennäköisyydet ja väärinluokittelukustannukset tunnetaan, päätössäännöksi saadaan tällöin:

   havainto x kuuluu G(1):teen, mikäli (7.2.2) on pienempi kuin 0 ja
   G(2):teen, mikäli (7.2.2) on suurempi kuin 0.

$$h(x) = x'(S1^{-1} - S2^{-1})x + 2*(\bar{x}^{(2)}{}'S2^{-1} - \bar{x}^{(1)}{}'S1^{-1})x + C1 \qquad (7.2.2)$$

   missä vakio C1 on

$$C1 = 2*\ln\left[\frac{q1*C(2|1)}{q2*C(1|2)}\right] + \ln\frac{|S1|}{|S2|} + \bar{x}^{(1)}{}'S1^{-1}\bar{x}^{(1)} - \bar{x}^{(2)}{}'S2^{-1}\bar{x}^{(2)}$$

2. Kun a'prioritodennäköisyyksiä ja väärinluokittelukustannuksia ei tunneta, niin

   A) voidaan yrittää lineaarisia menetelmiä kohta 7.3.

   B) etsitään katkaisukohta siten, että lasketaan h(x):n arvot sekä ryhmästä G(1) ja G(2) ja muodostetaan näistä arvoista kertymäfunktiot. Katkaisukohta eli vakio C määrätään joko minimax-menetelmällä tai valitsemalla se sopivasti.

   C) otetaan mukaan vain ne termit, joiden kerroin on itseisarvoltaan suuri eli ne, joilla on merkitystä h(x):n arvoon. Katkaisukohta etsitään samalla tavalla kuin kohdassa B.

7.3 LINEAARISET MENETELMÄT ((9))

Kun $S1 \neq S2$, parhaat menetelmät eivät ole lineaarisia. Jos parametrit tunnetaan, parhaat menetelmät perustuvat havaintovektorin neliömuotoiseen funktioon. Menetelmän hyvyys riippuu hyvin paljon oletetusta normaalisuudesta.
Luokittelumenetelmää voidaan kuitenkin käyttää tilanteessa, jossa kaksi populaatiota on keskittynyt eri pisteiden ympärille, populaatioilla on erilaiset hajontamallit sekä populaatioiden multinormaalisuuksia voidaan pitää tarpeeksi järkevinä hyvään aproksimaatioon, vaikkakin ei kovin kaukana populaatioiden keskipisteistä. Tälläisissä tapauksissa avaruus voidaan jakaa kahteen luokittelualueeseen käyttäen jotakin yksinkertaista käyrää tai pintaa. Yksinkertaisin on käyttää suoraa tai hypertasoa, jolloin erottelufunktio saattaa muodostua lineaariseksi.

Olkoon siis b p-komponenttinen vektori ja c skalaari. Havainto x luokitellaan G(1):teen, mikäli $b'x>c$ ja muutoin G(2):teen. Ollaan ensisijaisesti kiinnostuneita kahden populaation keskusten välisestä alueesta. Oletetaan, että $\mu_1 \neq \mu_2$ ja $\Sigma_1 \neq \Sigma_2$ sekä Σ_i on ei-singulaarinen eli käänteismatriisi on olemassa, i=1,2.

Kun G(i) on multinormaalisti jakautunut, $N(\mu_i, \Sigma_i)$, niin b'x on normaalijakautunut , $N(b'\mu_i, b'\Sigma_i b)$ eli

$$\begin{aligned} &\text{odotusarvo on} && E(\, b'x \mid G(i)\,) = b'\mu_i \\ &\text{ja varianssi on} && VAR(\, b'x \mid G(i)\,) = b'\Sigma_i b. \end{aligned} \qquad (7.3.1)$$

Pyrkimyksenä on määrittää kerroinvektori b ja katkaisukohta c samaan tapaan kuin lineaarisessa erottelufunktiossa oli

$$h(x) = x'S^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)}) > c \qquad (7.3.2)$$

missä $b = S^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)})$.

ja

$$S = \frac{1}{n1+n2-2}\left[\sum_{j=1}^{n1}(x_j^{(1)} - \bar{x}^{(1)})(x_j^{(1)} - \bar{x}^{(1)})' + \sum_{j=1}^{n2}(x_j^{(2)} - \bar{x}^{(2)})(x_j^{(2)} - \bar{x}^{(2)})'\right]$$

eli

$$S = \frac{n1-1}{n1+n2-2} * S1 + \frac{n2-1}{n1+n2-2} * S2 \qquad (7.3.3)$$

Kun havainto on G(1):stä, niin väärinluokittelutodennäköisyys on

$$\begin{aligned} P(2|1) &= P\left[b'x < c \mid G(1)\right] \\ &= P\left[\frac{b'x - b'\mu_1}{\sqrt{b'\Sigma_1 b}} < \frac{c - b'\mu_1}{\sqrt{b'\Sigma_1 b}} \mid G(1)\right] \\ &= \Phi\left[\frac{c - b'\mu_1}{\sqrt{b'\Sigma_1 b}}\right] \qquad (7.3.4) \\ &= 1 - \Phi\left[\frac{b'\mu_1 - c}{\sqrt{b'\Sigma_1 b}}\right] \end{aligned}$$

Vastaavasti kun havainto on G(2):stä, niin väärinluokittelutodennäköisyys on

$$P(1|2) = P\left[b'x > c \mid G(2)\right]$$

$$= P\left[\frac{b'x - b'\mu_2}{\sqrt{b'\Sigma_2 b}} > \frac{c - b'\mu_2}{\sqrt{b'\Sigma_2 b}} \;\middle|\; G(2)\right]$$

$$= 1 - P\left[\frac{b'x - b'\mu_2}{\sqrt{b'\Sigma_2 b}} < \frac{c - b'\mu_2}{\sqrt{b'\Sigma_2 b}} \;\middle|\; G(2)\right] \qquad (7.3.5)$$

$$= 1 - \Phi\left[\frac{c - b'\mu_2}{\sqrt{b'\Sigma_2 b}}\right]$$

Luokittelukriteerinä on minimoida väärinluokittelutodennäköisyydet. Kun pyritään saamaan P(1|2) ja P(2|1) pieniksi, niin se on sama kuin normaalijakauman argumentit saatetaan suuriksi. Siis maksimoidaan y1 ja y2

$$y1 = \frac{b'\mu_1 - c}{\sqrt{b'\Sigma_1 b}}$$

$$y2 = \frac{c - b'\mu_2}{\sqrt{b'\Sigma_2 b}} \qquad (7.3.6)$$

Asetetaan P(1|2) vakioksi eli y2 on vakio.

Eliminoidaan c (7.3.6):stä ja sijoitetaan y1:n lausekkeeseen, jolloin y1 saadaan muotoon

$$y1 = \left[b'v - y2(b'\Sigma_2 b)^{0.5}\right](b'\Sigma_1 b)^{-0.5} \qquad (7.3.7)$$

missä $v = \mu_1 - \mu_2$

Maksimoidaan y1 b:n suhteen ja saadaan

$$\frac{dy1}{db} = \left[v - y2\sqrt{b'\Sigma_2 b}\,\Sigma_2 b\right](b'\Sigma_1 b)^{-0.5} - \left[b'v - y2\sqrt{b'\Sigma_2 b}\right]\sqrt{b'\Sigma_1 b}\;\Sigma_1 b \qquad (7.3.8)$$

Merkitään

$$t1 = \frac{b'v - y2\sqrt{b'\Sigma_2 b}}{b'\Sigma_1 b}$$

$$t2 = \frac{y2}{\sqrt{b'\Sigma_2 b}} \qquad (7.3.9)$$

ja saadaan kertoimet,

$$\frac{dy1}{db} = 0 \quad ===> \quad b = \left[t1*\Sigma_1 + t2*\Sigma_2 \right]^{-1} (\mu_1 - \mu_2) \qquad (7.3.10)$$

Jos on olemassa pari t1,t2 ja vektori b, jolle pätee (7.3.10), silloin

$$c = t2*b'\Sigma_2 b + b'\mu_2 \qquad (7.3.11)$$

$$P(2|1) = 1 - \Phi(t1\sqrt{b'\Sigma_1 b}) \qquad (7.3.12)$$

Lineaariseen malliin verrattuna S:n paikalla on nyt $t1*\Sigma_1+t2*\Sigma_2$, jossa oli

$$t1 = \frac{n1-1}{n1+n2-2}$$
$$t2 = \frac{n2-1}{n1+n2-2} \qquad (7.3.13)$$

Tarkastellaan yhtälöä (7.3.10) t:n funktiona eli olkoon t1=t ja t2=1-t (0 < t < 1). Silloin

$$b = \left[t1*\Sigma_1+t2*\Sigma_2\right]^{-1}(\mu_1 - \mu_2) = \left[t*\Sigma_1 + (1-t)*\Sigma_2\right]^{-1} v \qquad (7.3.14)$$

Merkitään

$$v1 = t1\sqrt{b'\Sigma_1 b}$$
$$v2 = t2\sqrt{b'\Sigma_2 b} \qquad (7.3.15)$$

jolloin

$$v1^2 = t1^2 (b'\Sigma_1 b) = t^2 \left\{ v' \left[t*\Sigma_1+(1-t)*\Sigma_2 \right]^{-1} \Sigma_1 \left[t*\Sigma_1+(1-t)*\Sigma_2 \right]^{-1} v \right\} \qquad (7.3.16)$$

$v1^2$:n derivaatta t:n suhteen

$$\frac{dv1^2}{dt} = 2tv' \left[t\Sigma_1+(1-t)\Sigma_2 \right]^{-1} \Sigma_1 \left[t\Sigma_1+(1-t)\Sigma_2 \right]^{-1} v$$
$$- t^2 v' \left[t\Sigma_1+(1-t)\Sigma_2 \right]^{-1} (\Sigma_1-\Sigma_2) \left[t\Sigma_1+(1-t)\Sigma_2 \right]^{-1} \Sigma_1 \left[t\Sigma_1+(1-t)\Sigma_2 \right]^{-1} v$$
$$- t^2 v^2 \left[t\Sigma_1+(1-t)\Sigma_2 \right]^{-1} \Sigma_1 \left[t\Sigma_1+(1-t)\Sigma_2 \right]^{-1} (\Sigma_1-\Sigma_2) \left[t\Sigma_1+(1-t)\Sigma_2 \right]^{-1} v$$

$$= tv' \left[t\Sigma_1+(1-t)\Sigma_2\right]^{-1}\left\{ \Sigma_2 \left[t\Sigma_1+(1-t)\Sigma_2\right]^{-1}\Sigma_1 + \Sigma_1\left[t\Sigma_1+(1-t)\Sigma_2\right]^{-1}\Sigma_2\right\} * \left[t\Sigma_1+(1-t)\Sigma_2\right]^{-1} v$$

$$> 0 \qquad (7.3.17)$$

LEMMA 1 Jos Σ_1 ja Σ_2 ovat positiivisesti definiittejä sekä t1, t2 > 0 niin

$$\Sigma_2\left[t\,\Sigma_1+(1-t)\,\Sigma_2\right]^{-1}\Sigma_1 \qquad (7.3.18)$$

on positiivisesti definiitti.

Siis $dv1^2/dt > 0$ ja vastaavalla tavalla voidaan näyttää, että $dv2^2/dt<0$ Koska v1>0 ja v2>0, saadaan, että v1(t) kasvaa 0:sta $t\cdot\sqrt{b'\,\Sigma_1\,b}$:teen ja v2(t) laskee $t\cdot\sqrt{b'\,\Sigma_1\,b}$:sta ja 0:aan.

Tällöin siis tunnettua $0 < y2 < (v'\,\Sigma_2^{-1}\,v)^{0.5}$ kohti on olemassa t siten, että $y2 = v2 = t2\cdot\sqrt{b'\,\Sigma_2 b}$ ja b toteuttaa $(t1\cdot\Sigma_1+t2\cdot\Sigma_2)b = v$, kun t1 = t ja t2 = (1-t). Silloin tuolla y2:n arvolla maksimoituu $y1 = v1 = t1\cdot\sqrt{b'\,\Sigma_1 b}$. Vastaavasti jos y1 on tunnettu, saadaan y2 maksimoitua.

7.3.1 LINEAARISTEN PROSEDUURIEN KÄYTTÖ

Kun t1 ja t2 tunnetaan, jolloin siis $t1*\Sigma_1+t2*\Sigma_2$ on positiivisesti definiitti, voitaisiin laskea optimaalinen b ratkaisemalla lineaariset yhtälöt

$$c = t2b'\,\Sigma_2 b + b'\mu_2 \qquad (7.3.1.1)$$

ja sitten laskea c joko yhtälöstä

$$y1 = \frac{b'\mu_1 - c}{\sqrt{b'\,\Sigma_1 b}}$$

tai

$$y2 = \frac{c - b'\mu_2}{\sqrt{b'\,\Sigma_2 b}} \qquad (7.3.1.2)$$

Yleensä t1 ja t2 ovat tuntemattomia, tällöin on lähinnä kolme tapaa

1) Minimoimalla toinen väärinluokittelutodennäköisyys pitämällä toista kiinnitettynä.

2) Minimax-menetelmä. Asetetaan y1=y2 ja etsitään t = t1 = 1-t2 siten, että

$$0 = y1^2 - y2^2 = b'\left[t^2\Sigma_1 - (1-t)^2\Sigma_2\right]b \qquad (7.3.1.3)$$

tai yhtä hyvin yhtälöstä

$$y1 = \frac{b'\mu_1 - c}{\sqrt{b'\Sigma_1 b}} = \frac{c - b'\mu_2}{\sqrt{b'\Sigma_2 b}} = y2 \qquad (7.3.1.4)$$

ja ratkaisemalla c:n

$$c = y1\sqrt{b'\Sigma_1 b} - b'\mu_1 = y2\sqrt{b'\Sigma_2 b} + b'\mu_2 \qquad (7.3.1.5)$$

Sillloin y1=y2:n yhteinen arvo on

$$\frac{b'(\mu_1+\mu_2)}{\sqrt{b'\Sigma_1 b} - \sqrt{b'\Sigma_2 b}} \qquad (7.3.1.6)$$

ja etsitään se b:n arvo, joka maksimoi tämän, kun b on muotoa

$$b = \left[t\Sigma_1 + (1-t)\Sigma_2\right]^{-1} v \qquad 0 < t < 1 \qquad (7.3.1.7)$$

3) Kun a'prioritodennäköisyydet tunnetaan, silloin väärinluokitteluto-dennäköisyys

$$q1*\left[1 - \Phi(y1)\right] + q2*\left[1 - \Phi(y2)\right] = 1 - \left[q1*\Phi(y1) + q2*\Phi(y2)\right] \qquad (7.3.1.8)$$

minimoidaan ja ratkaisu t saadaan yhtälöstä

$$q1*\Phi(y1)*\frac{dy1}{dt} + q2*\Phi(y2)*\frac{dy2}{dt} = 0 \qquad (7.3.1.9)$$

missä $\Phi(u) = (2*\pi)^{-0.5}\exp(-0.5*u^2)$.

On huomattava, että yhtälön (7.3.1.9):n vasen puoli ei ole välttämät-tä ole monotoninen, jolloin yhtälöllä saattaa olla useita ratkaisuja.

8.0 ETÄISYYSMITTOIHIN PERUSTUVA EROTTELU

Tarkoituksena on osittaa havaintojen vektoriesityksiä sisältävä avaruus alueisiin, joista kukin sisältäisi ideaalisesti vain yhden ryhmän edustavia havaintoja. Tällöin havaintojen ja jakaumien väliset etäisyydet on tärkeä käsite.((17))

Havaintojen p-komponenttisten vektoreiden x ja y välisen etäisyyden

1. Euklidinen mitta

$$d1(x,y) = \left[\sum_{i=1}^{p} (xi - yi)^2 \right]^{1/2} \qquad (8.0.1)$$

2. City-block mitta

$$d2(x,y) = \sum_{i=1}^{p} |xi - yi| \qquad (8.0.2)$$

3. Yleinen Minkowski-mitta

$$d3(x,y) = \left[\sum_{i=1}^{p} (xi - yi)^r \right]^{1/r} \qquad (8.0.3)$$

Kun r=2, niin d3(x,y) = d1(x,y) ja kun r=1, niin d3(x,y) = d2(x,y), mikäli jokaisella i xi > yi.

4. Kun r=∞, niin

$$d4(x,y) = \underset{i=1}{\overset{p}{\text{maksimi}}} |xi - yi| \qquad (8.0.4)$$

5. Kun A on p*p-matriisi, niin

$$d5(x,y) = (x-y)'A(x-y) \qquad (8.0.5)$$

Kun A = I, niin $d5(x,y) = d1(x,y)^2$.

Kun x ja y ovat kahden jakauman odotusarvot ja A on yhteisen kovarianssimatriisin käänteismatriisi, niin tällöin d5 on jakaumien välinen Mahalanobisetäisyys.

Edellä olevat mitat ovat metrisiä mittoja, koska ne toteuttavat metriselle mitalle asetettavat vaatimukset

a) $d(x,y) \geq 0$ ja $d(x,x) = 0$

b) $d(x,y) = d(y,x)$

c) $d(x,y) < d(x,z) + d(z,y)$

Jakaumien f1(x) ja f2(x) väliselle etäisyydelle voidaan määrittää myöskin erilaisia mittoja. Tälläiselle mitalle asetettavia toivottavia ominaisuuksia ovat

d) $J(G(1),G(2)) = 0$, kun $f1(x) = f2(x)$

e) $J(G(1),G(2)) \geq 0$ (8.0.7)

f) $J(G(1),G(2))$ saavuttaa maksimin, kun ryhmät ovat erillisiä

6. Hajaannusmitta

$$J1(G(1),G(2)) = \int \left[f1(x) - f2(x) \right] \ln \left[f1(x)/f2(x) \right] dx \qquad (8.0.8)$$

Kun x on multinormaalinen $N(\mu_i, \Sigma_i)$ ja $\Sigma_1 \neq \Sigma_2$, niin

$$J1(G(1),G(2)) = 0.5*(\mu_1-\mu_2)' \left[\Sigma_1 + \Sigma_2 \right]^{-1} (\mu_1-\mu_2) + 0.5*\mathrm{tr}\left[\Sigma_1^{-1}\Sigma_2 + \Sigma_2^{-1}\Sigma_1 - 2I \right]$$

Kun lisäksi $\Sigma_1 = \Sigma_2$, niin

$$J1(G(1),G(2)) = (\mu_1-\mu_2)' \Sigma^{-1} (\mu_1-\mu_2) = m^2 \text{, Mahalanobisetäisyys.}$$

7. Chernoff-etäisyys

$$J2(G(1),G(2)) = -\ln \left[\int f1(x)^{1-s} f2(x)^{s} dx \right], \quad 0 \leq s \leq 1 \qquad (8.0.9)$$

Kun x on multinormaalinen $N(\mu_i, \Sigma_i)$ ja $\Sigma_1 \neq \Sigma_2$, niin

$$J2(G(1),G(2)) = 0.5*\left\{ s(1-s)(\mu_1-\mu_2)' \left[(1-s)\Sigma_1 + s\Sigma_2 \right]^{-1} (\mu_1-\mu_2) + \ln\left[\left| (1-s)\Sigma_1 + s\Sigma_2 \right| / \left(|\Sigma_1|^{1-s} |\Sigma_2|^{s} \right) \right] \right.$$

Kun lisäksi $\Sigma_1 = \Sigma_2$, niin

$$J2(G(1),G(2)) = (\mu_1-\mu_2)' \Sigma^{-1} (\mu_1-\mu_2) = m^2 \text{, Mahalanobisetäisyys.}$$

Väärinluokittelun kokonaistodennäköisyydelle saadaan yläraja

$$P(R,f) \leq g1^{1-s} * q2^{s} * \exp(-J2)$$

8. Bhattacharya etäisyys saadaan, kun s=0.5

$$J3(G(1),G(2)) = -\ln\left[\int \sqrt{f1(x)\ f2(x)}\ dx\right].$$

(8.0.10)

Kun x on multinormaalinen $N(\mu_i, \Sigma_i)$ ja $\Sigma_1 \neq \Sigma_2$, niin

$$J3(G(1),G(2)) = 1/8 * (\mu_1-\mu_2)' \left[0.5(\Sigma_1+\Sigma_2)\right]^{-1}(\mu_1-\mu_2)$$
$$+1/2 * \ln\left[\left|0.5(\Sigma_1+\Sigma_2)\right| / \left(\left|\Sigma_1\right|^{0.5}\left|\Sigma_2\right|^{0.5}\right)\right]$$

Kun lisäksi $\Sigma_1 = \Sigma_2$, niin

$$J3(G(1),G(2)) = (\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2) = m^2,$$ Mahalanobisetäisyys.

Väärinluokittelun kokonaistodennäköisyydelle saadaan rajat

$$0.5\left\{1-\left[1-4*q1*q2*\exp(-2J3)\right]^{0.5}\right\} \underset{=}{<} P(R,f) \underset{=}{<} \left\{q1*q2 * \exp(-J3)\right\}^{0.5}$$

Alussa esitettiin jo ajatus, että pyritään saamaan eroteltavat ryhmät mahdollisimman etäälle toisistaan. Pyritään tähän päämäärään hajontaan perustuvien mittojen avulla.

Määritellään aluksi,

$$T = \sum_{j=1}^{q}\sum_{i=1}^{nj}(X_{ij} - \bar{X})(X_{ij} - \bar{X})'$$ kokonaishajontamatriisi (8.0.11)

$$W = \sum_{j=1}^{q}\sum_{i=1}^{nj}(X_{ij} - \bar{X}_j)(X_{ij} - \bar{X}_j)'$$ ryhmien sisäinen hajontamatriisi (8.0.12)

$$B = \sum_{j=1}^{q}\sum_{i=1}^{nj}(\bar{X}_j - \bar{X})(\bar{X}_j - \bar{X})'$$ ryhmien välinen hajontamatriisi (8.0.13)

missä

$$\bar{X}_j = \frac{1}{nj}\sum_{i=1}^{nj} X_{ij}$$ on j:nnen ryhmän keskiarvo

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{q} nj*\bar{X}_j$$ on kokonaiskeskiarvo

Tällöin

$$T = W + B$$ (8.0.14)

8.1 B/W:N MAKSIMOINTI

Pyritään maksimoimaan ryhmien välinen etäisyys suhteessa ryhmien sisäisiin etäisyyksiin eli halutaan maksimoida $W^{-1}B$. Etsitään erottelumuuttuja $Z = a'X$, joka maksimaalisesti erottelee ryhmät siinä mielessä, että $J = a'Ba/a'Wa$ maksimoituu Jos voidaan olettaa, että W on ei-singulaarinen, niin

$$\frac{dJ}{da} = \frac{2Ba\ a'Wa - 2Wa\ a'Ba}{(a'Wa)^2} = 0 \qquad (8.1.1)$$

$$==> \quad 2W^{-1}(W^{-1}B - JI)a = 0 \qquad (8.1.2)$$

Tämä on ekvivalenttia ratkaista yleistetty ominaisarvotehtävä

$$Ba = JWa \qquad (8.1.3)$$

Tällöin on itseasiassa $t = \min(g-1,p)$ kappaletta erottelumuuttujaa eli $Z1, Z2, \ldots, Zt$ vastaten nollasta poikeavia ominaisarvoja $J1 \geq J2 \geq \ldots \geq Jt$.

On selvää, että ominaisvektori da, missä d on vakio, toteuttaa myös edellä olevan yhtälön. Monasti asetetaan am*am=1 tai suurin ominaisarvo J1 normeerataan ykköseksi.

Ominaisarvon suuruus osoittaa erottelukykyä ja siten Z1 on maksimaalierottelija. Erottelumuuttuja Z2 on toiseksi paras ja niin edelleen. Kanonisilla erottelumuuttujilla on lisäksi ominaisuus

$$CORR(Zi,Zj)=0, \text{ kun } i \neq j. \qquad (8.1.4)$$

Kirjallisuudessa käytetään erottelumuuttujasta Zt usein nimitystä erottelufunktio, mutta jatkossa näin johdetuista muuttujista käytetään nimitystä kanoninen erottelufunktio.

8.2 !W!/!T!:N MINIMOINTI

$$\text{Wilksin lambda} \quad L = \frac{|W|}{|T|} = \frac{|W|}{|W+B|} = \frac{|W|}{|W||I+W^{-1}B|} = \frac{1}{|I+W^{-1}B|} \qquad (8.2.1)$$

Siis $|W|/|T|$:n minimointi, on sama kuin $1/|I+W^{-1}B|$:n maksimointi.

Toisaalta jos J on $W^{-1}B$:n ominaisarvo ja a on sitä vastaava ominaisvektori eli

$$W^{-1}Ba - JIa = 0 \text{ tällöin } (W^{-1}B + I)a - (J+1)Ia = 0 \qquad (8.2.2)$$

joten J+1 on $(I + W^{-1}B)$:n on ominaisarvo ja a sitä vastaava ominaisvektori.

Havaintaan, että B/W:n maksimointi on ekvivalenttia $|W|/|T|$:n minimoinnin kanssa.

Tämän Fisherin lähestymistavan yleistystä kutsutaan kanoniseksi muuttuja-analyysiksi ja on erotteluanalyysimenetelmä, jota käytetään yleensä kuvaamaan ryhmien välisiä eroja.

8.3 KANONISEN KORRELAATIOANALYYSIN JA EROTTELUANALYYSIN VÄLINEN YHTEYS ((18))

Kanonisen korrelaatioanalyysin tyypillinen käyttö on seuraava: halutaan tutkia yhtäläisyyksiä muuttujajoukon x1,...,xp ja toisen muuttujajoukon y1,...,ys muuttujien välillä. Muodostetaan ensimmäisestä muuttujajoukosta lineaarikombinaatio

$$U = u1*x1 + u2*x2 + \ldots + up*xp$$

ja toisesta (8.3.1)

$$V = v1*y1 + v2*y2 + \ldots + vs*ys.$$

Tehtävänä on määrätä kertoimet u' = (u1,...,up) ja v' = (v1,...,vs) siten, että lineaarikombinaatioiden välinen korrelaatio maksimoituu.

Tarkastellaan korrelaatiota

$$CORR(U,V) = \frac{\sum_{i=1}^{n} Ui*Vi}{\sqrt{\left[\sum_{i=1}^{n} Ui^2\right]\left[\sum_{i=1}^{n} Vi^2\right]}} \qquad (8.3.2)$$

missä n on havaintojen lukumäärä. Koska

$$\sum_{i=1}^{n} Ui^2 = u'S(x,x)u$$

$$\sum_{i=1}^{n} Vi^2 = v'S(y,y)v \qquad (8.3.3)$$

$$\sum_{i=1}^{n} Ui*Vi = u'S(x,y)v$$

missä S(x,x) on p*p , S(y,y) on s*s ja S(x,y) on p*s ulotteisia hajontamatriiseja. Joten

$$CORR(U,V) = \frac{u'S(x,y)v}{\sqrt{u'S(x,x)u\ v'S(y,y)v}} \qquad (8.3.4)$$

Vastaavasti kuin erotteluanalyysissä painot u ja v ovat suhteellisia vakioita eli jos a ja b ovat samanmerkkisiä vakioita, niin CORR(U,V) on sama käytettäessä au:ta ja bv:tä tai u:ta ja v:tä.

Asetetaan side-ehdoksi

$$u'S(x,x)u = v'S(y,y)v = 1 \qquad (8.3.5)$$

Käytetään Lagrangen kerroinmenetelmää, (h/2 ja k/2 on valittu vain laskennallista syistä), jolloin maksimoitava funktio on

$$F(u,v) = u'S(x,y)v - h/2*\left[u'S(x,x)u - 1\right] - k/2*\left[v'S(y,y)v - 1\right] \qquad (8.3.6)$$

joten

$$\frac{d\,F(u,v)}{d\,u} = S(x,y)v - h*S(x,x)u = 0$$

$$\frac{d\,F(u,v)}{d\,v'} = u'S(x,y) - k*v'S(x,x) = 0 \qquad (8.3.7)$$

näistä saadaan ottamalla side-ehto huomioon, että

$$u'S(x,y)v = h = k \qquad (8.3.8)$$

Jos S(x,x) ja S(y,y) ovat ei-singulaarisia, niin saadaan

$$\left[S(x,x)^{-1}\,S(x,y)S(y,y)^{-1}\,S(y,x) - h^2 I\right]u = 0 \qquad (8.3.9)$$

Tämän ominaisarvotehtävän suurin ominaisarvo $h1^2$ antaa CORR(U,V):n neliömaksimin ja siihen liittyvän ominaisvektorin u1. Toiseksi suurin h2 antaa toiseksi suurimman korrelation ja niin edelleen. Tällöin CORR(U1,U2) = CORR(V1,V2) = 0, mutta myös CORR(U1,V2) = CORR(U2,V1) = 0

Näitä muuttujapareja U1, V1 ja U2, V2 ja...ja Ur, Vr kutsutaan kanonisiksi muuttujiksi ja niitä on

$$\text{rank}\left[S(x,x)^{-1}\,S(x,y)S(y,y)^{-1}\,S(y,x)\right] = \min(p,s) = r \text{ kappaletta.}$$

Kun J on $W^{-1}B$:n ominaisarvo ja h^2 on $S(x,x)^{-1}\,S(x,y)S(y,y)^{-1}\,S(y,x)$:n ominaisarvo, niin tällöin

$$h^2 = \frac{J}{1+J} \qquad (8.3.10)$$

8.4 RYHMIEN VÄLISTEN EROJEN TESTAMISESTA ((19))-((21))

Olkoon muodostettu erottelufunktiot Y = b'X , missä X on p*g-matriisi ja b on p*g-kerroinmatriisi. Testattavana hypoteesina on

$$\text{Ho:} \qquad LbM = d \tag{8.4.1}$$

Tämä voidaan kirjoitaa muotoon

$$(Lb - rj)M = 0 \tag{8.4.2}$$

missä L on g*p matriisi, M on g*g-matriisi liittyen luokittelumuuttujiin G(i), d on g*g-vakiomatriisi, r on g*1-vakiovektori ja j on 1*g ykkösiä sisältävä rivivektori.

Konstruoidaan kaksi matriisia B ja W vastaten F-testin osoittajaa ja nimittäjää.

$$\begin{aligned} B &= M'(Lb-rj)'\left[L(X'X)^{-1}L'\right]^{-1}(Lb-rj)M \\ W &= M'\left[Y'Y-b'(X'X)b\right]M \end{aligned} \tag{8.4.3}$$

Olkoon rank(W+B) = p, joka on vähemmän tai yhtäsuuri kuin M:n sarakkeiden lukumäärä ja v olkoon W:n vapausaste sekä rank $\left[L(X'X)^{-1}L\right]$ = q. Merkitään lisäksi s=min(p,q), m=0.5(|p-q|-1) ja n = 0.5*(v-p-1).

Neljä tärkeätä testitunnuslukua, jotka perustuvat $W^{-1}B$:n tai $T^{-1}B$:n ominaisarvoihin Ji:hin ja hi^2:seen, ovat likipitäen F-jakautuneita.

A) Wilksin lambda

$$L = \frac{|W|}{|T|} = \prod_i \frac{1}{1+Ji} = \prod_i (1 - hi^2) \tag{8.4.4}$$

testiluku

$$F = \frac{1 - L^{1/t}}{L^{1/t}} * \frac{rt - 2u}{pq} \quad \text{on likipitäen } F(pq, rt-2u) \tag{8.4.5}$$

missä $r = v-(p-q+1)/2$

$u = (pq-2)/4$

ja

$$t = \sqrt{\frac{p^2 * q^2 - 4}{p^2 + q^2 - 5}} \quad \text{kun } p^2 + q^2 - 5 > 0 \text{ ja 1 muuten}$$

Tämä approksimaatio on tarkka, mikäli min(p,q) < 3 eli kun ryhmiä on vain kaksi, on testiluku F F-jakautunut vapausastein p,n1+n2-p-1.

B) Pillain trace

$$V = tr\,(BT^{-1}) = \sum_i \frac{Ji}{1+Ji} = \sum_i hi^2 \qquad (8.4.6)$$

testiluku

$$F = \frac{(2n + s + 1)\ V}{(2m + s + 1)(s-V)} \quad \text{on likipitäen } F[s(2m+s+1), s(2n+s+1)] \qquad (8.4.7)$$

C) Hotelling-Lawley trace

$$U = tr\,(W^{-1}B) = \sum_i Ji = \sum_i \frac{hi^2}{1 - hi^2} \qquad (8.4.8)$$

testiluku

$$F = \frac{2(sn+1)\ U}{s^2\,(2m + s + 1)} \quad \text{on likipitäen } F[s(2m+s+1), 2(sn+1)] \qquad (8.4.9)$$

D) Royn maksimijuuri

$$J1 = \max Ji \qquad (8.4.10)$$

testiluku

$$T = \frac{v-r-1}{r}\ J1 \quad \text{on likipitäen } F(r, v-r-1) \qquad (8.4.11)$$

missä r = max (p,q).

8.5 KANONINEN EROTTELUFUNKTIO ((18))

Usein tarkastellaan kanonisten erottelufunktioiden avulla alkuperäistä aineistoa. Tämä tapahtuu tutkimalla ryhmien sisäistä rakennetta eli korrelaatioita kanonisen erottelijafunktion ja alkuperäisten erottelumuuttujien välillä eli

$$CORR(U,X) = A'WD(1/\sqrt{wii}\,) \qquad (8.5.1)$$

missä A:lla on sarakkeittain skaalatut ominaisarvot aiWaj = a(i,j)' ja D on lävistäjämatriisi i:nä diagonaalielementtinä $1/\sqrt{wii}$ ja wii on W:n i:nnes diagonaalielementti. Nämä korrelaatiot kuvaavat alkuperäisten muuttujien tärkeyttä, tällöin itseisarvoltaan suuri kerroin

tarkoittaa, että kyseinen muuttuja saa aikaan tärkeätä myötävaikutusta ryhmien väliseen erotteluun. Tätä kutsutaan erotteluvoimaksi.

Wilksin lambdalla voidaan tutkia myös ovatko kaikki r=min(g-1,p) erottelufunktiota tarpeellisia. Tämä tapahtuu tarkastelemalla jäännöserottelua

$$Lk = \prod_{i=k+1}^{r} 1/(1+Ji) \qquad (8.5.2)$$

missä k on tarkasteltavien kanonisten erottelufunktioiden lukumäärä. Testimuuttuja

$$-[n-1-(p+g)/2]\ln Lk \qquad (8.5.3)$$

on χ^2-jakautunut vapausastein (r-k)(r-k-1).

Tarkastelemalla erottelukriteeriä J = a'Ba/a'Wa havaitaan, että pyrittäessä saamaan ryhmät mahdollisimman etäälle toisistaan, niin sellaisen muuttujan ottaminen malliin, jonka ryhmien sisäiset korrelaatiot ovat suuria positiivisia lukuja, mutta ryhmien väliset korrelaatiot tulevat negatiivisiksi, antaa lisää erotteluvoimaa. Vastaavasti mikäli ryhmien sisäiset korrelaatiot ovat negatiivisia ja ryhmien väliset positiivisia, niin silloin erotteluvoima lisääntyy. Jos ryhmien sisäiset ja väliset korrelaatiot ovat saman merkkisiä, on erotteluvoiman lisäys vähäisempi. Mikäli sisäiset korrelaatiot ovat pieniä, mutta väliset korrelaatiot suuria, niin tällöin erottelufunktion kertoimista tulee epästabiileja. ((26))

9.0 MUUTTUJIEN VALINTA

Tarkastellaan algoritmejä, miten alkuperäisistä p:stä kappaleesta muuttujia valitaan p' kappaletta. Ollaan erityisesti kiinnostuneita siitä, antaako valitsematta olevat (p-p') muuttujaa erotteluun merkittävästi parannusta. Tämä parannus voi olla, että virheellinen luokittelu pienenee, joukkojen etäisyys kasvaa tai jokin muu erottelukriteerin ominaisuus paranee. ((17)) ((22)) ((24))

Tarkastellaan suhdetta d-muuttujaa/virheaste erottelumitoille, joille pätee

$$J(X_1) = J(X_2) = \ldots = J(X_m) \qquad (9.0.1)$$

kun

$$X_1 \subset X_2 \subset \ldots \subset X_m \qquad (9.0.2)$$

missä X_i on joukko, joka käsittää i kappaletta muuttujia xi.

Tutkitaan tilannetta Mahalanobiksen etäisyyden kannalta. Suoritetaan satunnaishavainnolle x muunnos

$$y = P^{-0.5} A'x \qquad (9.0.3)$$

missä A' kääntää akselit siten, että ne tulevat korreloimattomiksi A'x:n kanssa ja diagonaalimatriisi P skaalaa uudet akselit siten, että

$$COV(y) = P^{-0.5} A'SAP^{-0.5} = I \qquad (9.0.4)$$

Tällöin satunnaishavainnon y Mahalanodisetäisyys ryhmän keskuksesta u

$$\begin{aligned} D^2(u|y) = y'y &= (Rx)'(RSR')^{-1}(Rx) \\ &= x'R'R'^{-1}S^{-1}R^{-1}Rx \\ &= x'S^{-1}x \\ &= D^2(u|x), \end{aligned} \qquad (9.0.5)$$

$$\text{missä } R = P^{-0.5}A'.$$

Joten Mahalanobisetäisyys ei muutu transformaatiossa. Lisätään dimensiota tässä uudessa koordinaatistossa

$$D^2(u|y,y_p) = y'y + y_p^2 > D^2(u|y) \qquad (9.0.6)$$

Joten etäisyys yleensä kasvaa lisättäessä dimensiota, mutta samanaikaisesti toinen erottelukriteeri voi huonontua esimerkiksi väärinluokittelu saattaa kasvaa. Onkin havaittu, että muuttujien lisäämisen seurauksena erottelukyky väärinluokittelu mielessä ensin paranee, mutta sitten huononee. Tätä ilmiötä kutsutaan dimension kiroukseksi.

9.1 PERUSTEELLINEN ETSINTÄ

Kun p on riittävän pieni, tämä menetelmä on käyttökelpoinen.
Lähtötilanne on, että mallissa on kaikki p kappaletta muuttujia ja hylätään hypoteesi, että kaksi ryhmää ovat samoja, jos

$$T^2(p) > \frac{(n1+n2-2)p}{n1+n2-1-p} * F^v(p,n1+n2-1-p) \qquad (9.1.1)$$

missä $T^2(p)$ on Hotellingin T^2 p-dimenssiolla ja $F^v(p,n1+n2-1-p)$ on tasoa v vastaava F-jakauman arvo vapausastein p,n1+n2-1-p.

Jos tätä lähtöhypoteesiä ei hylätä, silloin täyden mallin muuttujat eivät erottele ryhmiä ja analyysi päättyy. Jos hypoteesi hylätään, voidaan tarkastella p' kappaleen osajoukkoja tutkimalla erottelevatko ne ryhmät toisistaan.

Yleisesti mikäli testitunnusluvulle pätee

$$T(p')^2 < \frac{(n1+n2-1-p)*T(p)^2 - (n1+n2-2)*p*F^v(p,n1+n2-1-p)}{n1+n2-1-p+p*F^v(p,n1+n2-1-p)} \qquad (9.1.2)$$

niin p' kappaleen muuttujajoukko ei erottele joukkoja tarpeeksi tehokkaasti.

9.2 KIIHDYTETTY ETSINTÄ

Tätä menetelmää kutsutaan myös haarautumis- ja rajamenetelmäksi. Etsitään paras p'-dimensionaalinen osajoukko p-dimensionaalisesta täydestä joukosta.

Ehtona on, että luokittelukriteeri J kasvaa monotonisesti, kun muuttujia lisätään malliin.

Lähtötilanteessa kaikki muuttujat ovat mukana ja konstruioidaan puurakenne, jossa muuttujia tullaan hylkäämään. Aluksi on p-muuttujan joukko, seuraavaksi muuttujia on p-1:n kappaleen joukko ja niin edelleen kunnes on p' muuttujaa jäljellä.

Oletetaan, että jotakin haaraa pitkin ollaan tultu jonkin kriteerin mukaiseen kriittiseen arvoon J1. Jos jonkin haaraan J2 < J1, niin tällöin kyseistä haaraa pitkin ei kannata edetä alemmaksi, koska oletuksena oli J:n monotoninen kasvaminen.

9.3 OSAOPTIMAALINEN ETSINTÄ

Eräs ilmeinen tapa valita p' parasta erottelumuuttujaa p:stä mahdollisesta, on ottaa p' yksittäisesti parasta. Tämä ei välttämättä johda hyvään tariffitekijöiden joukkoon, sillä tässä menetelmässä hylätään monimuuttujien välinen yhteys. Voi käydä niin, että kaksi tai useampi erottelumuuttuja jakaa saman erotteluinformaation, vaikka ne yksinään ovat hyviä erottelijoita. Kun jotkut näistä sisällytetään malliin, ovat jäljelle jääneet muuttujat redundantteja. Vaikka ne olisivat yksinään hyviä, eivät nämä redundattimuuttujat anna mallille mitään lisää. Mikäli niiden läsnäololle ei ole mitään teoreettista pakkoa, on heikko tai redundantti muuttuja poistettava, ne vain lisäävät esimerkiksi väärinluokitteluosuutta. Vastaavasti yksinään heikko muuttuja, voi yhdessä toisen muuttujan kanssa tuoda malliin lisää erotteluvoimaa.

9.3.1 ETENEVÄ VALINTA ((23))

Oletetaan, että p1 muuttujaa on jo valittu muodostaen joukon X(p1). Nyt kutakin vielä valitsematta olevaa muuttujaa x(r) kohti lasketaan kriteeriarvo J(X(p1) U x(r)). Näitä kriteeriarvoja on siten p-p1 kappaletta. Se x(r), joka tuottaa maksimi kriteeriarvon, valitaan ja lisätään X(p1):teen muodostaen joukon X(p2). Täten siis kussakin vaiheessa lisätään malliin se muuttuja, joka maksimoi J-kriteerin. Proseduuri päättyy, kun ei ole muuttujia, jotka täyttäisivät J-kriteerin.
Menetelmällä on ainakin kaksi huonoa puolta eli mitään lukua ei pidetä vielä valitsemattomien muuttujien keskinäisistä suhteista ja toisaalta

kun muuttuja on lisätty malliin, ei ole tapaa, millä se poistetaan, vaikka se olisi jo menettänyt merkityksensä.

9.3.2 LASKEVA VALINTA

Menetelmä on tavallaan edellisen peilikuva, nimittäin nyt lähdetään täydestä p-muuttujan mallista ja pudotetaan yksi kerrallaan pois. Poispudotettu on se muuttuja x(r), joka on vähiten pienentänyt J-kriteeriä. Siis poistetaan se x(r), jonka J(X(p)-x(r)) on suurin. Seuraavaksi poistetaan se muuttuja x(s), jonka J(X(p)-x(r)-x(s)) on suurin ja niin edelleen. Haittapuolena on se, että poistettua muuttujaa ei voida enää lisätä malliin.

9.3.3 ASKELTAVA VALINTA

Tämä valintamenetelmä on edellisten yhdistelmä. Tyypillisesti tämä toimii etenevästi muuttujaa valittaessa ja mikäli jokin muuttuja on menettänyt erottelukykynsä J-kriteeri mielessä, niin se poistetaan. Proseduuri päättyy, kun aikaisemmin poistettu muuttuja otetaan uudestaan malliin tai kun ei ole muuttujaa, joka täyttäisi J-valintakriteeriä. On huomattava, että valinta- ja poistakriteerin J-arvot voivat olla eri.

Tämän menetelmän eräs muunnos on 'ota l-kappaletta ja poista r-kappaletta muuttujia'-menetelmä.

9.4 VALINTAKRITEERI

Valintaproseduurin täytyy käyttää jotakin mittaa valintakriteerinä. Wilksin lambda voi olla sellainen, mutta on olemassa vaihtoehtoja, joita voidaan käyttää maksimoimaan erilaisia joukkojen välisiä erilaisuuskäsitteitä. Mikä seuraavista on paras, riippuu tutkimustilanteesta.

9.4.1 OSITTAIS F-SUHDE

Edellä oli neljä testitunnuslukua, jotka perustuivat Wilksin lambdaan, Pillain traceihin, Hotelling-Lawley traceihin ja Royn maksimijuureen, jotka kaikki ovat likipitäen F-jakautuneita.

Esimerkiksi Wilksin lambda on tunnusluku, joka ottaa huomioon sekä ryhmien välisen erottelun että samanlaisuuden ryhmien sisällä. Samanlaisuudella tarkoitetaan tilannetta, kun havainnot keskittyvät ryhmien keskusten läheisyyteen. Täten muuttuja, joka lisää samanlaisuutta muuttamatta keskusten välistä erottelua, valitaan paremminkin kuin muuttuja, joka lisää erottelua muuttamatta samanlaisuutta.

Koska Wilksin lambda on ominaisarvojen käänteisluku

$$L = \prod_i \frac{1}{1+Ji} \qquad (9.4.1.1)$$

valitaan se muuttuja, joka tuottaa pienimmän lambdan kyseisellä valintakerralla tai suurimman F-arvon.

9.4.2 MAHALANOBISETÄISYYS ((24))

Sen sijaan, että keskitytään ryhmien väliseen kokonaiserotteluun, voidaankin ottaa muuttuja, joka tuottaa suurimman erottelun toisiaan lähimpänä olevien ryhmäparien välillä. Täten pyritään pakottaa kaikki ryhmät erilleen. On olemassa ainakin kolme tunnuslukua, joilla mitataan tätä ominaisuutta ja ne perustuvat Mahalanobiksen etäisyyteen ryhmäparien keskusten välillä.
Yksi näistä on Mahalanobiksen etäisyys itse antaen saman painon kullemin ryhmälle.

1) Testiluku T

$$T = \frac{ni+nj-p-1}{p-p'} * C * \frac{D^2(p) - D^2(p')}{1 + C*D^2(p')} \qquad (9.4.2.1)$$

missä

$$C = \frac{ni*nj}{(ni+nj)*(ni+nj-2)}$$

$D^2(p)$ on Mahalanobiksen etäisyys, kun muuttujia p kappaletta

$D^2(p')$ on Mahalanobiksen etäisyys, kun muuttujia p' kappaletta

T on F-jakautunut vapausastein p-p', ni+nj-p-1. Erikoisesti kun p'= p-1 eli tuoko yhden muuttujan lisääminen erotteluvoimaa. ((25))

2) Tunnusluku, jolla myös voidaan tutkia kahden ryhmän välistä eroa, on

$$\frac{(n-g-p'+1)ni*nj}{p'(n-g)(ni+nj)} D^2\left[G(i)\,|\,G(j)\right], \quad \text{missä } n = \sum_{i=1}^{g} ni \qquad (9.4.2.2)$$

Tämä tunnusluku on F-jakautunut vapausastein n-g, n-g-p'+1.

Tällä tunnusluvulla on taipumus maksimoida ne ryhmäparit, joilla havaintojen lukumäärät ovat suuret.

3) Tunnusluvulla R minimoidaan ryhmien välinen jäännösvarianssi

$$R = \sum_{i=1}^{g-1} \sum_{j=i+1}^{g} 4/\left\{4+D^2\left[G(i)\,|\,G(j)\right]\right\} \qquad (9.4.2.3)$$

Kukin termi summalausekkeessa on $1-R^2(X,(i,j))$:n estimaatti, missä X on joukko mukana olevista muuttujista x1,...,xp' ja (i,j) on dummymuuttu-

ja, joka identifioi vastaavat ryhmäparit. Tämä on jäännösvarianssi, koska kukin termeistä on dummymuuttujan varianssin osuus, jota muuttujat eivät selitä. Joskus R jaetaan vielä g(g-1)/2:lla, jolloin se johtaa keskimääräiseen jäännösvarianssiin ryhmien välillä. Tämä tunnusluku eroaa osittais F-testeistä siinä mielessä, että osittais F-testeillä on taipumus jättää kaksi tai useampaa ryhmää lähelle toisiaan, mikäli saavutetaan kokonaiserottelun kannalta suotuisampi tulos.

9.5 MUUTTUJAN VALINNALLE ASETETTAVAT MINIMIVAATIMUKSET

Useimmat valintaproseduurit asettavat muuttujan hyväksymiselle malliin tietyt minimivaatimukset ennenkuin se testataan valintakriteerillä. Näitä ehtoja ovat sallitun mittapoikkeaman testi, jolla varmistetaan laskennallinen tarkkuus, ja osittais F-tunnuslukutesti, jolla tarkistetaan, että kasvanut erottelu ylittää käyttäjän määräämän tason.

Mittapoikkeamatesti on määritelty varjelemaan laskennallista tarkkuutta. Valitsemattoman muuttujan mittapoikkeama on

$$1 - R^2(x;1,\ldots,p') \qquad (9.5.1)$$

missä 1,...,p' on jo valitut muuttujat ja x on tarkasteltava muuttuja. Korrelaatiot lasketaan kunkin ryhmän sisäisestä korrelaatimatriisista. Jos tarkasteltava muuttuja on lähes linearinen kombinaatio yhdestä tai useammasta jo valitusta muuttujasta, niin silloin kyseisen muuttujan mittapoikkeama on lähes nolla. Muuttuja, jolla on pieni mittapoikkeama tulee aiheuttamaan epätarkkuutta lopulliseen ryhmien väliseen matriisiin, koska pyöristysvirheet kasvavat nopeasti.

F-sisään on tunnusluku, jolla testataan tarkasteltavan muuttujan tuomaa lisäystä erotteluun ottamalla huomioon mallissa mukana olevien muuttujien aiheutama erottelu. Jos tarkasteltavan muuttujan F-luku on pieni, ei muuttujaa oteta malliin, koska se ei tuo malliin tarpeeksi erottelua. F:ää käytetään siis testaamaan, onko muuttujan tuoma erottelun lisäys tilastollisesti merkitsevä. Muuttujan on läpäistävä mittapoikkeamatesti ja F-sisääntesti ennenkuin se testataan valintakriteerillä.

F-poista on tunnusluku, jolla testataan erottelun pienentymistä muuttujaa poistettaessa. Testi suoritetaan kunkin valintakierroksen alussa, jotta nähtäisiin onko jokin muuttuja menettänyt erottelukykynsä. Muuttuja, joka oli arvokas erottelija aikaisemmilla valintakierroksilla, ei ehkä enää myöhemmin olekkaan hyvä, koska myöhemmin malliin otetut muuttujat ovat tuoneet lähes saman erotteluinformaation.

10. PERUSOLETUSTEN PAIKKANSA PITÄVYYS

Kun aineisto ei täytä multinormaalijakauma ja yhtäsuuret kovarianssimatriisit oletusta, useat tutkijat ovat osoittaneet, että erotteluanalyysi on silti melko vahva tekniikka, joka sietää joitakin poikkeamia näistä oletuksista.

Kun oletus, että muuttujat ovat peräisin populaatiosta, joka noudattaa multinormaalijakaumaa, ei päde, niin todellinen otosjakauma kyseiselle muuttujalle on kutakuinkin eri kuin teoreettisesti johdettu jakauma. Erot näiden kahden jakauman välillä riippuuu tutkimusaineistosta.

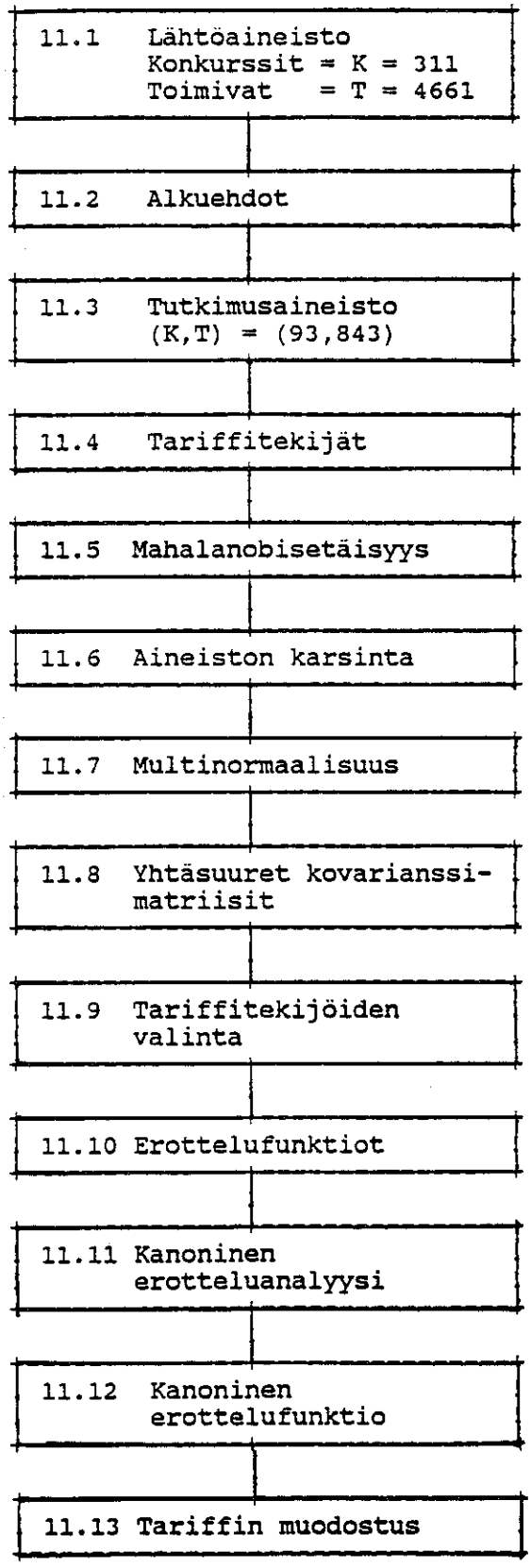
Lachenbruch ((24)) osoitti, ettei erotteluanalyysi ole erityisen herkkä vähäisille poikkeamille normaalijakaumasta. Seurauksena on vähentynyt tehokkuus ja tarkkuus. Normaalisuusoletus on tärkeä ryhmiin kuulumistodennäköisyyksiin perustuvissa luokitteluissa. Nämä todennäköisyydet lasketaan F -jakaumasta, joka on sopiva vain, kun erottelumuuttujilla on multinormaalijakauma. Kun normaalisuusoletus ei päde, lasketut todennäköisyydet ovat epätarkkoja. Lisäksi kun kriteerinä on minimoida väärinluokiteltujen tapausten lukumäärä, niin tulos ei ole optimaalinen.

Kun yhtäsuuret kovarianssimatriisit oletus ei päde ja käytetään yhdistettyä sisäistä kovarianssimatriisia, niin taipumuksena on saada vääristynyt kanoninen erottelufunktio ja luokittelusääntö.

Eräs virhelähde on edellä mainittu ryhmien sisäinen kovarianssimatriisi tai W:n matriisin sukulainen. Ryhmien sisäisen kovarianssimatriisin tarkoitus on olla yhteisen ryhmän kovarianssimatriisin estimaatti populaatiossa, joka koostuu useista ryhmistä. Jos ryhmien sisäiset kovarianssimatriisit ovat hyvin erilaiset, voidaan yhdistetty matriisi kyllä kin laskea, mutta se ei täysin täytä sitä roolia, joka sillä on useimmissa kaavoissa. Kanoniset erottelufunktiot eivät tuota maksimaalista erottelua ryhmien välillä ja ryhmiin kuulumistodennäköisyydet ovat vääristyneitä. Kun ei näytä olevan mitään proseduuria parantamaan kanonisia erottelufunktioita, niin usein suositellaan käytettäväksi yksittäisiä ryhmien sisäisiä kovarianssimatriiseja laskettaessa ryhmiin kuulumistodennäköisyyksiä.

Erotteluanalyysi voidaan suorittaa, vaikka normaalisuus- ja yhtäsuuret kovarianssimatriisit-oletukset eivät päde. Ongelma syntyy vasta kun käytetään johdettuja erottelufunktioita. Eli ongelmana on: mitä saatu funktio tarkoittaa tai kuinka paljon virheitä siedetään?

11. TUTKIMUSSUUNNITELMA



11.1 LÄHTÖAINEISTO

Tutkimusaineisto on Eläketurvakeskuksen (ETK) luottovakuutuustoiminnan tietokannasta. Kaikki konkurssiyritykset ja toimivat yritykset on otettu tutkimukseen mukaan myöhemmin esitettävien rajoitusten puitteissa. Konkurssilla tarkoitetaan tilannetta, kun yritys ei ole pystynyt huolehtimaan vieraan pääoman takaisinmaksustaan.

11.2 ALKUEHDOT

Tarkastelun kohteena on 1.1.1980-31.12.1986 välisenä aikana sattuneet konkurssit. Konkurssiyrityksiltä vaaditaan, että viralliset tilinpäätöstiedot ovat käytettävissä kolmelta konkurssia edeltävältä vuodelta.

Toisena määrittelyjoukkona on toimivat yritykset, joilta vaaditaan, että viralliset tilinpäätöstiedot ovat käytettävissä viideltä viimeiseltä vuodelta.

Konkurssiyrityksistä lasketaan mahdolliset tariffitekijät kahdesta konkurssia edeltävän vuoden tilinpäätöstiedoista.

Toimivien yritysten tilinpäätöstietoja tarkastellaan kolmea vuotta vanhempia tietoja eli tilinpäätösajankohtia vuosilta 1984 ja 1985. Toimivaksi yritykseksi, jota kutsutaan myös terveeksi yritykseksi, luokitellaan sellainen yritys, joka tarkasteluhetkestä katsoen on toiminut ainakin kolme vuotta. Näin varmistetaan se, ettei terveisiin yrityksiin tule tapauksia, jotka jo tarkasteluhetkeä seuraavana vuonna menevät konkurssiin.

Sekä konkurssien että toimivien yrityksien tilikauden tulee olla 12 kuukauden pituinen. (Taulukko 1) ((27))

11.3 TUTKIMUSAINEISTO

Tilinpäätösanalyysin avulla pyritään selvittämään yrityksen taloudelliset toimintaedellytykset osana laajempaa yritysanalyysiä. Päätavoitteena on antaa mahdollisimman oikea kuva yrityksen kannattavuudesta ja taloudellisesta asemasta sekä samalla saada yrityksen eri tilikaudet ja eri yritykset vertailukelpoisiksi sekä kannattavuuden että rahoitusrakenteen osalta. ((28))

Yritysanalyysillä tarkoitetaan yrityksen analysointia kokonaisuudessaan, jolloin suuri paino asetetaan yrityksen reaalipuolen tarkastelulle ja pyritään löytämään syyt hyvälle tai huonolle menestymiselle. Tällöin reaaliprosessi sisältää yrityksen menestymisen syyt ja rahaprosessi kuvaa seurauksia, jotka näkyvät tilinpäätöslaskelmissa. Yritysanalyysissä pyritään siten selvittämään yrityksessä vallitsevat kausaalisuhteet, jolloin taloudellinen tilanne selitetään reaalipuolen tapahtumien perusteella.

Tunnuslukuanalyysi on se osa tilinpäätösanalyysiä, jossa taloudellinen tilanne tiivistetään avainluvuiksi erilaisten suhdelukujen avulla. Mutta tunnuslukuanalyysin teoreettisen perustan puuttuminen aiheuttaa kaksi perusongelmaa: verrattavuus- ja johdonmukaisuusongelman. ((29))

TAULUKKO 1. Yritysten tilinpäätösajankohdat

Tilin-päätös	Toimivat Lukumäärä	%	Konkurssit Lukumäärä		%
310182			1		
310183			1		
310184			1		
310185	6	0.7%	1	4	4.3%
280283			2		
280285	25	3.0%	2	4	4.3%
310380			1		
310381			1		
310382			2		
310385	82	9.7%	1	5	5.4%
300483			1		
300485	43	5.1%	-	1	1.1%
310580			1		
310584			1		
310585	14	1.7%	-	2	2.2%
300680			2		
300681			2		
300682			4		
300683			3		
300684			3		
300685	67	7.9%	1	15	16.1%
310784			1		
310785	15	1.8%	1	2	2.2%
310883			1		
310884			2		
310885	31	3.7%	1	4	4.3%
300984			3		
300985	43	5.1%	-	3	3.2%
311084			1		
311085	32	3.8%	-	1	1.1%
301182			1		
301185	16	1.9%	-	1	1.1%
311280			6		
311281			11		
311282			7		
311283			11		
311284			11		
311285	469	55.6%	5	51	54.8%

Verrattavuusongelmalla tarkoitetaan sitä, että tunnuslukuanalyysiä tulisi soveltaa vain keskenään vertailukelpoisiin yrityksiin. Tällöin huomioitavia tekijöitä ovat muun muassa yrityksen koko, toimiala, työvoimavaltaisuus, käytetäänkö omia vai vuokrattuja koneita, laitteita tai tiloja. Ongelman muodostaa se, ettei tiedetä mitkä tekijät vertailukelpoisuuteen vaikuttavat ja miten ne ilmenevät tunnusluvuissa.

Johdonmukaisuusongelmalla tarkoitetaan puolestaan sitä, että tietyt tunnusluvut johtavat erilaisiin yritysjärjestyksiin, vaikka niiden tusi mitata samaa asiaa. Tunnusluvut voivat antaa eri tuloksen, koska tilinpäätöksen joustokohdat ja ulkopuoliset tekijät vaikuttavat niihin eri tavoin.

Pyritään välttämään verrattavuusongelmaa siten, että tarkastellaan vain teollisuusyrityksiä, joten vastaavanlainen tutkimus tulisi suorittaa kullekin toimialalle erikseen.

Perinteisesti toimialat on jaettu kuuteen pääryhmään eli maa- ja metsätalouden, teollisuuden, rakennustoiminnan, kaupan, liikenteen sekä palvelualan yrityksiin. (Taulukko 2) Tarkasteltaessa kehitystä 80-luvulla havaitaan konkurssien lukumäärien kasvu kussakin pääryhmässä. Erityisesti sekä teollisuuden että kaupan konkurssien lukumäärät ovat vaihdelleet 20% - 30% välillä ollen suurimmat ryhmät kunakin vuonna.

Joudutaan tekemään lisäksi oletus, että havaintoaineistossa olevat tapaukset ovat toisistaan riippumattomia edustaen läpileikkausta 80-luvun alkupuolen suomalaista yrityspopulaatiota. Tarkalleen ottaen tämä ei päde, koska aineistossa on varmasti tapauksia, joissa yhden yrityksen konkurssiin meno on myöskin voinut aiheuttaa toisen yrityksen kaatumisen. Toisaalta ETK:n kannan muodostumiseen on vaikuttanut aikaisemmin riskin perusteena ollut Prihtin ((31)) teollisuusyritysten aineistosta johdettu meno-tulovirtalaskelma.

Tässä tutkimuksessa käytetään 93 konkurssiin menneen ja 843 toimivan teollisuusyrityksen aineistoa.(Taulukko 3) Tekstiiliteollisuuden suuri lukumäärä, 29 kappaletta eli melkein joka kolmas teollisuuskonkurssi lienee 80-luvun alkupuolelle tunnusomaista. Tilanne kehittyi 80-luvun loppua kohden vielä dramaattisemmaksi, sillä vuosikymmenen alun konkurssit olivat tyypiltään pieniä ja keskisuuria, kun taas vuosikymmenen loppupuolella entistä suuremmat tekstiiliteollisuuden yritykset kohtasivat määränpäänsä.

11.4 TARIFFITEKIJÄT

Luonnollisin tapa valita tariffitekijät, olisi tietää, mitkä syyt ovat aiheuttaneet konkurssin ja käyttää näitä muuttujia mittaamaan konkurssiriskiä. Vaikka lukuisissa tutkimuksissa on laadittu luetteloita syistä, jotka ovat saattaneet vaikuttaa yrityksen epäonnistumiseen, on niillä tietyt omat rajoitteensa. ((31))

Ensinnäkin esitetyt syyt ovat olleet hyvin erilaisia eri tutkimuksissa. Toiseksi ei tiedetä tarkalleen, mitkä syyt erottavat parhaiten menestymisen ja epäonnistumisen eikä tiedetä monta syytä tarvitaan, ennen kuin yritys joutuu konkurssiin. Lopulta ei ole varmaa tietoa edes siitä, kuinka näitä syitä pitäisi käyttää ennustamiseen tai konkurssin välttäseen. ((30))

TAULUKKO 2. Tilastotietoja konkursseista Suomessa
(Konkurssitiedote 8/1989, Tilastokeskus)

2.1 Yhtiömuoto

Vuosi	Konkurssit yhteensä	Yhtiöt ja yhteisöt		Yrittäjät		Muut	
1980	856	529	62%	185	22%	142	17%
1981	1219	767	63%	260	21%	192	16%
1982	1447	906	63%	330	23%	211	15%
1983	1322	896	68%	312	23%	114	9%
1984	1704	1140	67%	399	23%	165	10%
1985	2122	1503	70%	438	20%	181	9%
1986	2463	1772	72%	627	25%	64	3%
1987	2816	2005	71%	746	26%	65	2%
1988	2547	1812	71%	623	24%	112	4%

Kohta 'Muut' on sekä muut yksityishenkilöt että yhtiömuoto tuntematon.

2.2 Toimiala

Vuosi	Maa-, metsä-talous		Teolli-suus		Raken-nus		Kauppa		Lii-kenne		Pal-velut		Tunte-maton		Yht.
1980	11	1%	156	18%	119	14%	216	25%	17	2%	79	9%	258	30%	856
1981	5	0%	229	19%	197	16%	229	19%	30	2%	97	8%	432	35%	1219
1982	15	1%	336	23%	172	12%	267	18%	43	3%	150	10%	464	32%	1447
1983	31	2%	345	27%	152	11%	282	21%	54	4%	112	8%	390	30%	1322
1984	15	1%	421	25%	207	12%	335	20%	63	4%	149	9%	514	30%	1704
1985	18	1%	580	27%	313	15%	489	23%	82	4%	216	10%	424	20%	2122
1986	29	1%	615	25%	406	16%	703	28%	99	4%	345	14%	266	10%	2463
1987	31	1%	661	23%	477	17%	804	29%	126	4%	380	13%	337	12%	2816
1988	45	2%	552	22%	396	16%	737	29%	147	6%	371	15%	290	11%	2547

TAULUKKO 3. Havainnot teollisuusryhmittäin

Toimiala	Toimivia Lkm	%	Konkursseja Lkm	%
Elintarvikkeiden, juomien ja tupakan valmistus	64	7.6	2	2.2
Tekstiilien, vaatteiden ja nahkatuotteiden valmistus	107	12.7	29	31.2
Puutavaran valmistus	92	10.9	15	16.1
Paperiteollisuustuotteiden valmistus, graafinen tuotanto	146	17.3	6	6.5
Kemiallisten, maaöljy-, kumi- ja muovituotteiden valmistus	68	8.1	2	2.2
Savi-, lasi- ja kivituotteiden valmistus	34	4.0	5	5.4
Metallien valmistus	19	2.3	5	5.4
Metalli- ja konepajatuotteiden valmistus	283	33.6	28	30.1
Muu valmistus	30	3.6	1	1.1
Yhteensä	843	100.0	93	100.0

Osa syistä on ollut yrityksen sisäisiä eli syitä, johon yritysjohdon olisi pitänyt voida vaikuttaa, ja osa ulkoisia eli syitä, johon yritysjohdolla on vähäiset vaikuttamismahdollisuudet. Esimerkiksi Dupontin ((32)) tutkimus osoitti, että yritysjohdon tehottomuus ja kokemattomuus aiheutti 90% konkursseista. Kun taas eräät ranskalaiset tutkimukset osoittavat, että vain 20% konkursseista voidaan lukea yritysjohdon tilille. Luetaanko yritysjohdon virheeksi esimerkiksi kova kilpailu, valtiovallan rajoitustoimenpiteet, suhdannemuutokset, korkokannan muutokset tai muut ulkoiset syyt, vai onko kysessä yritysjohdon kyvyttömyys reagoida strategisten tekijöiden muutoksiin. Ratia ((33)) päätyi näkemykseen, että Suomessa on niin epäonnen kuin menestyksenkin todelliseksi takuumieheksi katsottava toimitusjohtaja kahdeksassa tapauksessa kymmenestä.

Argentin mukaan ((34)) konkurssiin on syynä yritysjohdon huonot liikkeenjohdon taidot, lähestyvää konkurssia ei nähdä, mutta tähän saattaa olla perimmäisenä syynä laskentainformaation riittämättömyys.

Dewing ((35)) ja Hartigan ((36)) mainitsevat syiksi kannattamattoman kasvun, mikä taas on peräisin hinnoitteluvirheistä eli sisäisen laskentatoimen puutteellisuudesta.

Liiketaloudellinen tutkimuslaitos on tutkimuksessaan ((37)) todennut pienyritysten epäonnistumisen syiksi Suomessa: rahoituksen, tuotekehittelyn, kustannustason, investointien suunnittelun, perheyritysten jatkuvuuden, yrityksen johdon ja organisaation sekä laskentatoimen. Nämä luetellut syyt olivat puutteellisia tai virheellisiä.
Prihti ((31)) lisäsi luetteloon vielä suhdannevaihtelut, yrityksen koon ja kasvuvauhdin.

Sharman ja Mahajanin ((35)) tutkimuksen mukaan yrityksen lähestyvä konkurssi näkyy selvästi heikosta kannattavuudesta, maksuvalmiudesta ja raskaasta velkaantumisessa, kuten myös Prihti ((31)) osoitti, että vastaavat tekijät ennustavat lähestyvän konkurssin myöskin Suomessa.

Yrityksen konkurssin lähestymisen näkee myös muusta informaatiosta kuin tilinpäätöksestä. Peel, Peel ja Pope ((35)) osoittivat, että yrityksen konkurssia voidaan erilaisten tunnuslukujen lisäksi ennustaa myös esimerkiksi yritysjohdossa tapahtuvien muutosten ja tilinpäätösten julkaimisen lykkäämisen perusteella. ETK:ssa on vastaavien havaintojen lisäksi mainittava myöskin tilintarkastajien vaihtuvuus, tilinpäätösajankohdan muuttaminen, tilinpäätöstietojen toimittamattomuus ETK:lle sekä omana riskiryhmänään yrityskaupat. On osoittautunut, että noin 50% konkurssiyrityksistä jättää konkurssia edeltävän vuoden tietonsa toimittamisen ETK:lle, noin 30%:lta kaksi vuotta ja noin 20%:lta kolme vuotta ennen konkurssia olevat tilinpäätöstiedot puuttuvat. Neljä vuotta ennen konkurssia ja toimivilta yrityksiltä puuttuu suunnilleen saman verran eli noin 5% - 10% tapauksista. Syynä tietojen toimittamattomuuteen saattaa olla välinpitämättömyys tai yksinkertaisesti yritys ei ole laatinut tilinpäätöstietojansa. Puuttuvista tilinpäätöstiedoista johtuen erityisesti konkurssiaineisto supistui merkittävästi, sillä lähtöaineiston 197 konkurssiin menneen teollisuusyrityksen joukosta on peräkkäiset, kahdet viimeiset tilinpäätöstiedot käytettävissä ainoastaan 93 tapauksella, kun taas toimivien yritysten lähtöpopulaatio supistui 889:stä 843:een yritykseen.

Yrityksen ikä mainittiin jo eräänä konkurssiriskiin mahdollisesti vaikuttavana tekijänä, sillä Tilastokeskuksen mukaan ((35)) noin 50% 80-luvun alkupuolella konkurssiin menneistä yrityksistä oli iältään alle 7 vuotta, noin 40% oli iältään 7-12 vuotta, noin 10% iältään 12-15 vuotta ja vain 1% oli vanhempia kuin 15 vuotta. Vastaavat ikäjakaumat käyvät yksiin myös ETK:n aineiston kanssa. Ikäväliä 3 - 6 vuotta kutsutaankin usein sopivasti yrityksen kehityshistorian kannalta kuoleman laaksoksi.

Koska erityisesti konkurssiaineisto pirstoutuisi liian pieniksi ryhmiksi jakamalla tutkimusaineisto boniteettilukujen mukaisiin osiin, käytetään toimialaa ainoana jakoperusteena. Boniteettiluvuilla pyritään tyypittelemään yritys entistä hienojakaisempiin luokkiin ja ottamalla mukaan muun muassa yrityksen ikä-, kasvu- ja yhtiömuotoseikkoja sekä alueellisia tekijöitä. ((38))

Osaan edellä olevista konkurssin ominaisuuksista toivotaan voitavan kuvata jollakin tase- tai tuloslaskelmasuureen avulla. Osalle konkurssisyitä on itseasiassa hyvin vaikeata keksiä mitattavissa olevaa yksikköä

Yleensä konkurssiennusteet ((39)) laaditaan taseista tunnuslukujen avulla. Tarkasteltaessa epäonnistumisprosessia rahavirroista koostuvilla tunnusluvuilla tausta-ajatuksena on, että tunnusluvut kuvaavat yritystoimintaa. Yritystoiminta jaetaan kahteen osa-alueeseen, reaali- ja rahaprosesseihin. Joten rahaprosessin tunnusluvuilla pyritään kuvaamaan reaaliprosessin onnistumista tai epäonnistumista. Täten tunnusluku ei anna itse niitä tekijöitä, jotka ovat vaikuttaneet epäonnistumiseen.

Tunnusluvut voidaan jakaa kahteen ryhmään sen mukaan onko niitä käytetty yhdestä tai useasta tunnusluvusta koostunutta mallia epäonnistumisen ennustamiseen. Esimerkiksi Beaver ((40)) käytti yhden tunnusluvun mallia, kun taas Altman ((41)) ja Prihti ((31)) käyttivät usean tunnusluvun lineaarista mallia.

Perinteisen taseanalyysin tietoperustana on yrityksen viralliset tilinpäätökset, jolloin tuloslaskelmat ja taseet muokataan mahdollisimman luotettavaksi tietopohjaksi joustokohdat huomioiden yrityksen taloudellisen menestyksen mittaamista varten. Tärkeimmät joustokohdat eli yrityksen johdon harkinnan varassa olevien kohtien käytöstä on tarkemmin Liite 3:ssä. On huomattava, että joustokohtien enimmäismäärät vaihtelevat vuosittain. Koska ETK:n tietokantaan ei ole tallennettu tarvittavia tietoja puhdistamisen suorittamiseksi, joudutaan turvautumaan puhdistamattomiin tunnuslukuihin eikä esimerkiksi Yritystutkimusneuvottelukunnan ((28)) suosittamia tunnuslukuja voida muodostaa. (Liite 2)

Hyvälle tunnusluvulle voidaan asettaa tiettyjä ominaisuuksia. Näitä ovat ensinäkin, että tunnusluku on validi eli se mittaa sitä ominaisuutta, jota sen on tarkoituskin mitata. Toisena ominaisuutena on, että tunnusluku on reliaabeli eli antaa luotettavia tuloksia ja johon yritysjohdon käyttämät harkinnanvaraisuudet eivät vaikuta. ((29)) ((42))

Mahdolliset tariffitekijät eli erottelumuuttujat muodostetaan tase- tai tuloslaskelmasuure jaettuna liikevaihdolla. Kokeiltiin yksittäinen tilinpäätössuure liikevaihdolla jaettuna ja havaittiin, ettei kunnollisia tuloksia saada puuttuvien tietojen vuoksi. Siksi tunnusluku x3 muodostetaan summana käyttöomaisuuseristä, x6 oman pääoman luonteisten erien summana sekä käyttökate x7 muutetaan varastovarauksen vapaaehtoisuusluonteesta johtuen vertailukelpoiseksi kaikkien yritysten kesken. Näillä toimenpiteillä pyrittiin huomioimaan tunnusluvun validisuuskysymys.

Reliabilisuuden vastakohtana muodostettiin tunnusluku x8, jonka arvoihin voidaan olettaa yritysjohdon nimenomaisesti haluavan vaikuttaa. Nimittäin hyvin menestyvän yrityksen voidaan olettaa tekevän mahdollisimman suuret poistot, kasvattavan varauksia ja näyttävän vielä kohtuullista tulosta. Kun taas heikosti menestyvä yritys saattaa tinkiä poispoistoista, purkaa ehkä varauksia, mutta saattaa silti näyttää kohtuullista tulosta. ((39)) ((40))

Taseen vastaavien puolelta muodostetaan 3 erottelumuuttujaa, kuten myös vastattavien puolelta sekä tuloslaskelmasta saadaan 4 erottelumuuttujaa. Lisäksi muodostetaan vastaavien suureiden peräkkäisten vuosien erotuksen suhde viimeisen vuoden liikevaihtoon. (Liite 1)

Kun erottelumuuttujista tutkitaan yksimuuttujallisia graafisia esityksiä eli varsi-lehti-kuvio, laatikko-kuvio sekä normaalisuustodennäköisyys-kuvio, sekä numeerisia yhteenvetoja eli keskiarvoja, keskihajontoja, vinouksia ja huipukkuuksia, havaitaan, ettei mikään erottelumuuttuja yksinään noudata Kolomogorovin D-tunnusluku mielessä normaalijakaumaa. (Taulukot 4 ja 5)

Vaikka rajajakaumien normaalisuudesta ei seuraa multinormaalisuutta, on aineistoa kuitenkin seulottava ja poistettava nämä erikoistapaukset eli outlierit, jotka poikkeavien ominaisuuksiensa vuoksi eivät oikeastaan kuulu kyseiseen perusjoukkoon, vaan ovat joutuneet sinne satunnaisten, ennalta arvaamattomien tekijöiden vuoksi. ((43))

Tiedetään, että todelliset outlierit piileskelevät korkempi ulotteisissa avaruuksissa ja niitä voidaan etsiä myös tutkimalla kaksiulotteisia hajontakuvioita, mutta kuvioiden suuri määrä $g*p*(p+1)/2 = 420$ kappaletta, ei anna kovin tyydyttävää tulosta hallitsemaan erottelumuuttujien välisiä yhteyksiä ryhmien sisällä.

TAULUKKO 4. Konkurssiyritysten erottelumuuttujien tunnuslukuja, havaintoja 93 kappaletta.

Muuttuja	Keski-arvo	Keski-hajonta	Vinous	Huipukkuus	Normaalisuus testi: D	Tod. > D
x1	27.67	22.31	2.77	10.58	0.1930	<.01
x2	25.40	19.98	1.84	5.91	0.1251	<.01
x3	46.58	37.04	1.04	0.58	0.1331	<.01
x4	49.48	24.21	1.02	1.37	0.1192	<.01
x5	38.17	28.36	1.30	1.92	0.1287	<.01
x6	9.52	21.92	0.63	3.58	0.1725	<.01
x7	0.73	13.03	-1.07	1.75	0.1172	<.01
x8	-4.75	13.00	-1.13	1.88	0.1469	<.01
x9	7.65	4.71	1.31	2.85	0.1336	<.01
x10	0.60	1.00	3.20	15.00	0.2596	<.01
ex1	0.31	12.74	-0.62	2.61	0.1247	<.01
ex2	3.44	9.21	0.94	2.37	0.1296	<.01
ex3	8.01	17.38	0.34	3.41	0.1820	<.01
ex4	7.96	17.26	-0.94	4.44	0.0604	>.15
ex5	5.27	14.02	-0.57	5.31	0.1449	<.01
ex6	-2.88	11.76	-0.58	3.05	0.1742	<.01
ex7	-4.48	12.76	-1.47	5.68	0.1192	<.01
ex8	-5.81	15.30	-1.98	8.16	0.1610	<.01
ex9	1.40	3.02	-0.47	1.89	0.1080	<.01
ex10	0.07	1.17	1.45	13.53	0.2130	<.01

TAULUKKO 5. Toimivien yritysten erottelumuuttujien tunnuslukuja, havaintoja 483 kappaletta.

Muuttuja	Keski-arvo	Keski-hajonta	Vinous	Huipukkuus	Normaalisuus testi: D	Tod. > D
x1	28.14	27.38	7.78	94.41	0.1946	<.01
x2	15.36	13.66	1.84	2.24	0.1303	<.01
x3	32.13	30.54	3.85	27.11	0.1515	<.01
x4	27.45	21.15	8.61	135.27	0.1681	<.01
x5	26.32	25.10	5.13	47.10	0.1587	<.01
x6	20.51	23.93	3.07	17.91	0.1507	<.01
x7	9.33	12.43	-3.68	58.33	0.1516	<.01
x8	7.83	12.52	6.24	94.56	0.1673	<.01
x9	3.92	3.53	3.90	29.48	0.1442	<.01
x10	0.76	1.33	3.31	27.99	0.2242	<.01
ex1	4.25	18.64	4.13	93.53	0.2165	<.01
ex2	1.08	8.07	-5.06	86.52	0.1819	<.01
ex3	3.96	18.22	2.94	116.33	0.2248	<.01
ex4	2.85	11.29	0.16	10.49	0.1142	<.01
ex5	3.60	14.95	9.54	163.08	0.2192	<.01
ex6	2.81	13.56	-3.08	105.37	0.2279	<.01
ex7	0.39	8.85	-1.58	15.52	0.1126	<.01
ex8	0.15	15.39	-2.16	163.18	0.2199	<.01
ex9	0.58	1.77	2.89	42.68	0.1782	<.01
ex10	-0.03	1.32	-3.24	63.00	0.1887	<.01

11.5. MAHALANOBISETÄISYYS

Multinormaalijakaumaa ja monimuuttujaoutleireiden vaikutusta arvioidaan tutkimalla Mahalanobiksen neliöetäisyyksiä

$$D^2(i) = (X(i) - \bar{X})' S^{-1} (X(i) - \bar{X})$$

missä X(i) on tarkasteltava havaintovektori, $\bar{X}$ on keskiarvovektori ja S on kovarianssimatriisi. Sekä keskiarvovektori että kovarianssimatriisi lasketaan kummallekin ryhmälle erikseen.

Neliöetäisyydet ovat likipitäen χ^2-jakautuneita vapausastein p. Joten jos $D^2(i) > \chi^2(\alpha,p)$, niin havainto tulisi luetella mahdolliseksi outlieriksi, mikäli α on tarpeeksi pieni.
$\chi^2(\alpha,p)$ on χ^2-jakauman $1-\alpha$ kvartiili vapausastein p.

10.6 HAVAINTOAINEISTON KARSINTA

Annetaan α:lle viisi eri arvoa eli 0, 0.001, 0.01, 0.05 ja 0.10. Tutkitaan, miten poistetut havainnot vaikuttavat valittaviin tariffitekijöihin sekä itse erottelufunktioihin, sillä kuten myöhemmin tulemme havaitsemaan päädymme neliömuotoisiin erottelufunktioihin.

α	χ^2-arvo	Poistetaan Konkursseja	Poistetaan Toimivia	Havaintoja jäljellä Konkursseja	Havaintoja jäljellä Toimivia
0.0	-	0	0	93	843
0.001	45.315	6	63	87	780
0.01	37.566	11	76	82	767
0.05	31.410	17	92	76	751
0.10	28.412	18	106	75	737

Taulukoissa 6-7 on keskiarvojen ja -hajontojen muuttumista tarkasteltu poistettaessa havaintoja. Havaitaan, että sekä konkurssi että toimivien yritysten kohdalla keskiarvot tunnusluvuissa x1-x6, x9 ja ex9 pienenevät havaintoja poistettaessa, kun taas tunnuslukujen ex7 ja ex8 keskiarvot kasvavat. Tunnuslukujen x7, x8, ex1, ex2, ex3, ex4, ex5 ja ex6 kohdalla konkurssien keskiarvojen kasvaessa toimien pienenee tai päinvastoin. Hämmästyttävää on tunnusluku x10 (verot/liikevaihto), sen keskiarvo on melkein stabiili ja on samansuuruinen sekä konkursseille että toimiville yrityksille. Verojen muutosta kuvaavan tunnusluvun (ex10) keskiarvot eivät tunnu kovin järkeviltä eli verojen kasvu suhteessa liikevaihtoon on konkurssiyrityksille luonteenomaisempaa kuin toimiville yrityksille.
Keskihajonnat pienenevät kaikkien tunnuslukujen osalta, mutta eroja konkurssien ja toimivien yritysten välillä esiintyy runsaasti.

Taulukko 6. Tunnuslukujen keskiarvot

	Konkurssiyritykset					Toimivat yritykset				
α	.0	.001	.01	.05	.10	.0	.001	.01	.05	.10
n	93	87	82	76	75	843	780	767	751	737
x1	27.67	25.78	25.22	23.61	23.66	28.14	25.21	24.98	24.68	24.50
x2	25.40	23.79	23.43	23.70	23.56	15.36	14.53	14.41	14.16	14.15
x3	46.58	45.85	41.90	41.91	41.46	32.13	28.98	28.67	28.48	28.35
x4	49.48	49.00	47.05	46.86	46.63	27.45	25.72	25.51	25.14	25.11
x5	38.17	36.96	34.94	33.58	33.73	26.32	23.80	23.54	23.37	23.15
x6	9.52	7.73	6.80	7.87	7.13	20.51	18.22	18.06	17.90	17.87
x7	0.73	1.45	1.78	2.16	2.19	9.33	9.49	9.42	9.35	9.26
x8	-4.75	-4.47	-4.06	-3.65	-3.59	7.83	7.03	6.99	6.92	6.87
x9	7.65	7.29	7.04	6.91	6.89	3.92	3.57	3.52	3.50	3.46
x10	0.60	0.60	0.63	0.63	0.63	0.76	0.69	0.67	0.68	0.67
ex1	0.31	1.11	1.21	1.00	0.94	4.25	3.19	3.12	2.98	2.96
ex2	3.44	3.25	3.19	3.35	3.36	1.08	1.38	1.38	1.27	1.26
ex3	8.01	8.85	7.53	8.09	8.35	3.96	3.58	3.56	3.45	3.48
ex4	7.96	9.24	8.29	8.10	8.17	2.85	2.71	2.67	2.52	2.56
ex5	5.27	6.50	6.26	6.60	6.65	3.60	3.14	3.17	3.13	3.04
ex6	-2.88	-3.31	-3.46	-2.67	-2.71	2.81	2.32	2.23	2.06	2.13
ex7	-4.48	-4.58	-4.12	-3.74	-3.70	0.39	0.56	0.59	0.54	0.54
ex8	-5.81	-5.90	-5.41	-5.01	-4.99	0.15	0.19	0.21	0.15	0.16
ex9	1.40	1.66	1.45	1.30	1.31	0.58	0.58	0.56	0.56	0.54
ex10	0.07	0.08	0.10	0.13	0.13	-0.03	-0.05	-0.05	-0.03	-0.02

Taulukko 7. Tunnuslukujen keskihajonnat

	Konkurssiyritykset					Toimivat yritykset				
α	.0	.001	.01	.05	.10	.0	.001	.01	.05	.10
n	93	87	82	76	75	843	780	767	751	737
x1	22.31	17.38	16.28	13.38	13.46	27.38	14.54	14.07	13.60	13.15
x2	19.98	17.42	16.65	16.49	16.55	13.66	11.58	11.36	10.78	10.80
x3	37.04	37.66	34.38	34.03	34.04	30.54	21.40	21.21	20.83	20.73
x4	24.21	24.21	22.15	21.51	21.56	21.15	13.00	12.58	11.83	11.74
x5	28.36	26.98	24.38	23.42	23.54	25.10	15.81	15.48	15.40	15.12
x6	21.92	20.04	19.36	17.66	16.54	23.93	16.93	16.56	16.18	15.91
x7	13.03	12.64	11.72	10.80	10.87	12.43	8.23	8.02	7.93	7.77
x8	13.00	12.48	11.78	10.23	10.28	12.52	7.63	7.43	7.06	6.79
x9	4.71	4.28	4.09	3.54	3.55	3.53	2.47	2.39	2.36	2.30
x10	1.00	1.02	0.96	0.99	0.99	1.33	0.98	0.93	0.94	0.92
ex1	12.74	11.51	9.50	9.23	9.27	18.64	9.17	8.71	8.39	8.13
ex2	9.21	8.89	9.03	9.14	9.20	8.07	5.19	5.01	4.72	4.69
ex3	17.38	14.08	12.79	12.81	12.70	18.22	9.50	9.02	8.86	8.90
ex4	17.26	14.82	14.52	14.12	14.20	11.29	8.74	8.53	8.13	7.95
ex5	14.02	11.59	11.17	11.41	11.48	14.95	8.38	8.32	8.26	7.94
ex6	11.76	11.31	11.49	10.19	10.26	13.56	7.39	6.74	6.27	6.08
ex7	12.76	12.81	10.59	10.36	10.43	8.85	7.29	7.08	6.88	6.72
ex8	15.30	12.58	10.10	9.51	9.57	15.39	6.93	6.74	6.22	6.00
ex9	3.02	2.80	2.66	2.44	2.46	1.77	1.13	1.09	1.10	1.06
ex10	1.17	1.21	1.03	0.94	0.95	1.32	0.99	0.95	0.91	0.88

11.7. MULTINORMAALISUUS ((44))

Multinormaalisuutta on tarkasteltu myöskin χ^2-jakauman avulla, sillä χ^2-jakauma vapausastein p on gammajakauman erikoistapaus asteikkoparametrina 2 ja muotoparametrina p/2. Joten mikäli perimmäisenä jakaumana on multinormaalijakauma, pitäisi gammatodennäköisyyskuvion olla suora. Kuvat 1 - 10.

Kun $\alpha = 0$ (kuvat 1 ja 2) eli kun aineistosta ei ole poistettu havaintoja, niin sekä konkurssien että toimivien yritysten kuviot ovat vähemmän suoraa muistuttavia. Erityisesti toimivissa yrityksissä näyttäisi olevan outliereita.

Kun $\alpha = 0.001$ (kuvat 3 ja 4) eli kun kuusi konkurssia ja 63 toimivaa yritystä on poistettu, niin konkurssiyritysten kuvio on kohtuullisen suora, mutta toimivissa saattaa esiintyä vielä outliereita.

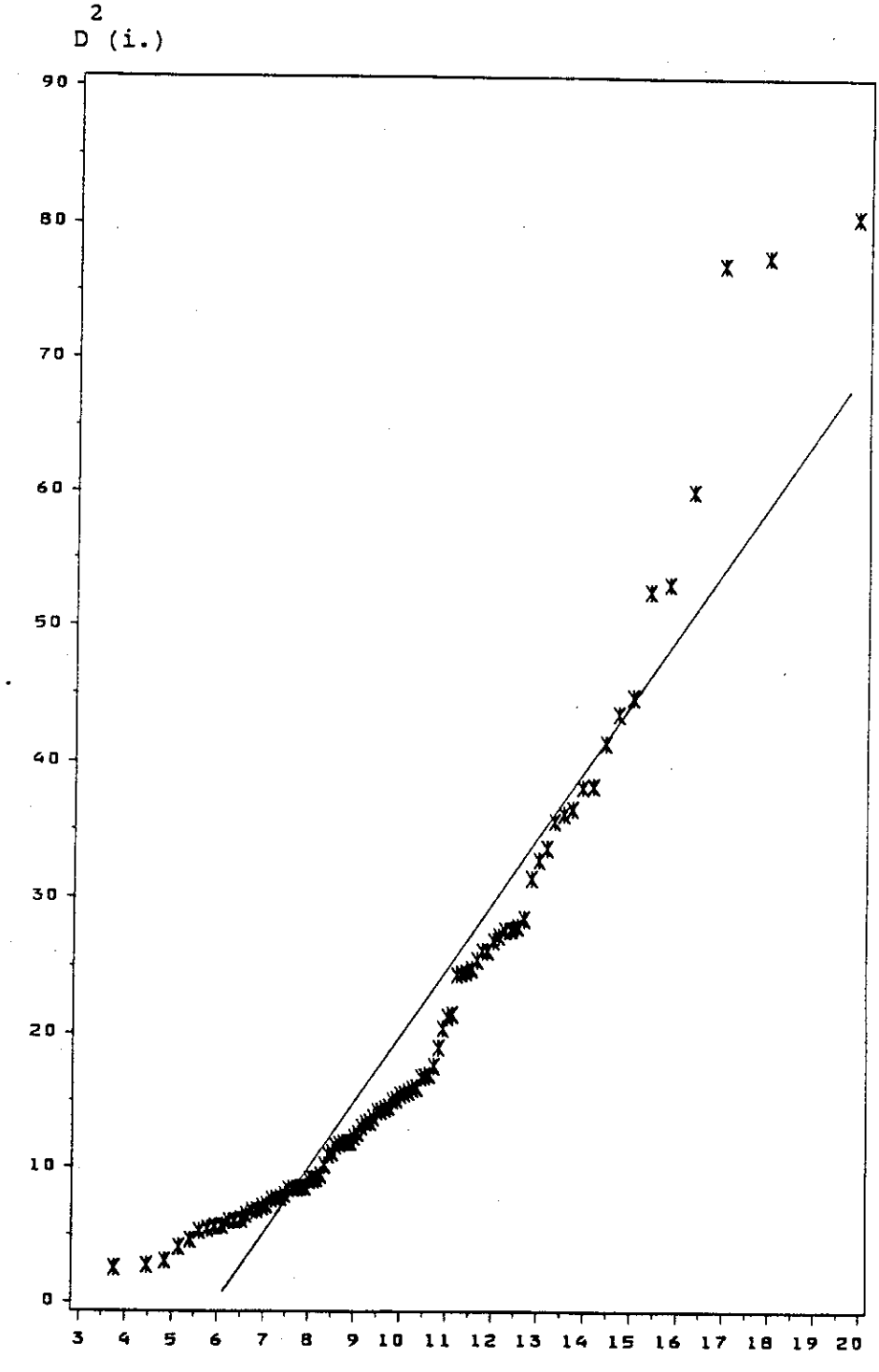
Kun $\alpha = 0.01$, 0.05 tai 0.10 (kuvat 5 - 10), silloin sekä konkurssi- että toimivien yritysten kuviot ovat suorahkoja. Taustajakaumana saattaa olla jokin muu jakauma kuin multinormaalijakauma, mutta voitaneen olettaa, että kumpikin populaatio noudattaa multinormaalijakaumaa ainakin kun aineistosta on poistettu ne tapaukset, joiden Mahalanobisetäisyyden neliö on suurempi kuin 37.566 eli kun $\alpha > 0.01$.

Tämä ajatus, että populaatiot eivät itseasiassa noudata multinormaalijakaumaa, johtaa mallin kehittelyn kannalta katsoen mielenkiintoiseen kokeiluun, mutta tästä hiukan myöhemmin.

KUVA 1

Multinormaalisuus konkurssit

$\alpha = 0.0$

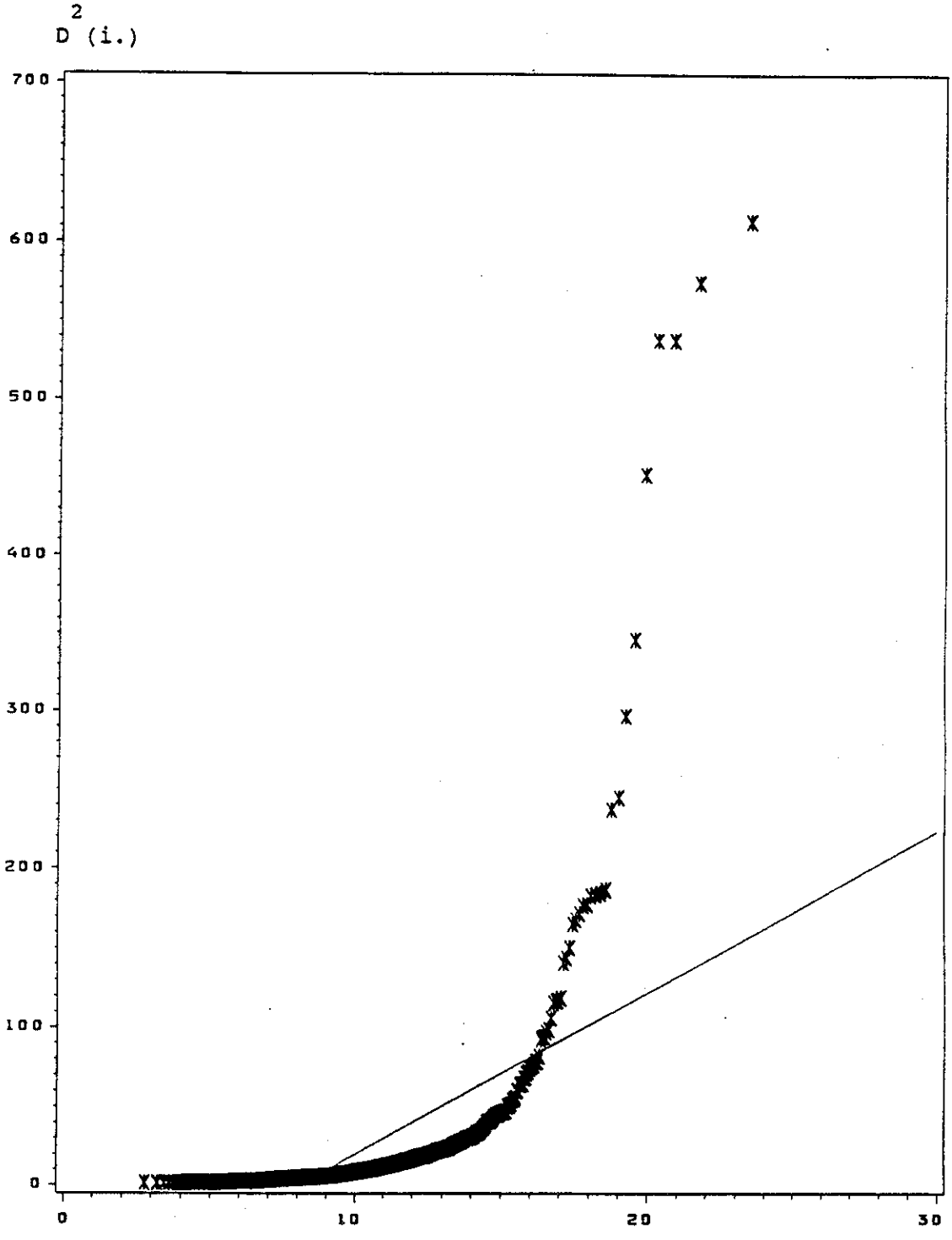


$$\text{Gamma}^{-1}\left(\frac{D^2(i.) - 0.5}{93}, 10\right)$$

KUVA 2

Multinormaalisuus toimivat

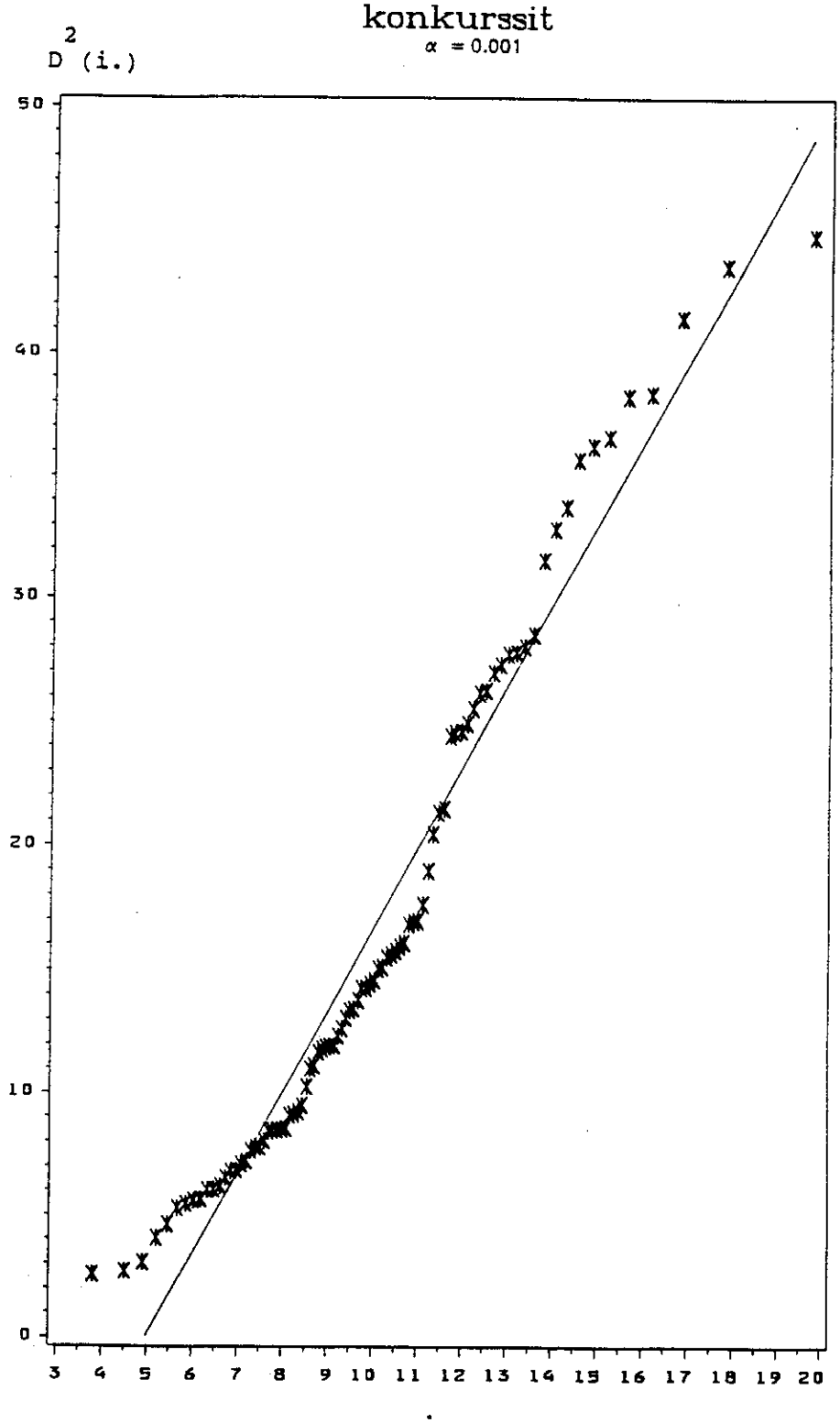
$\alpha = 0.0$



KUVA 3

Multinormaalisuus konkurssit

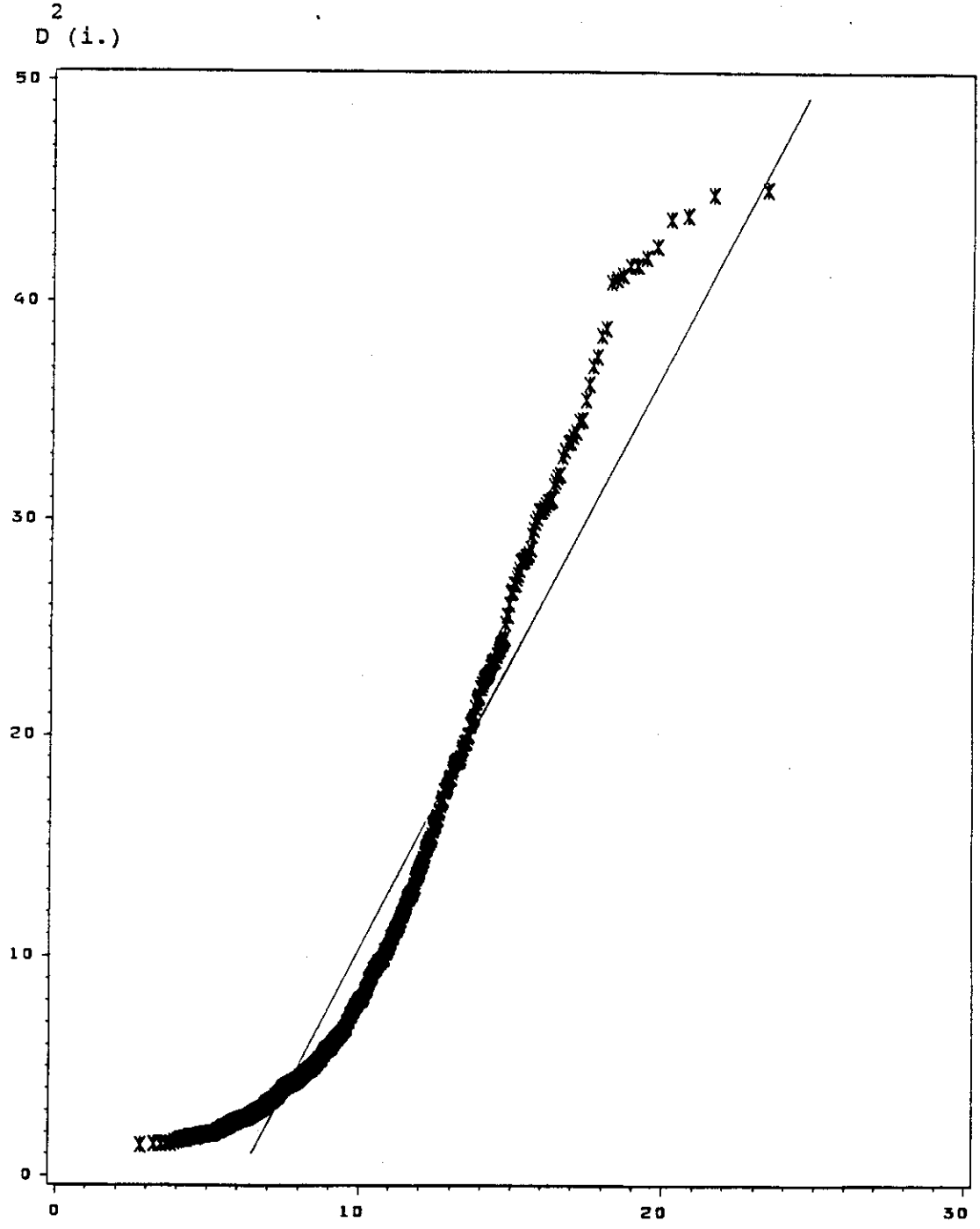
$\alpha = 0.001$



KUVA 4

Multinormaalisuus toimivat

$\alpha = 0.001$

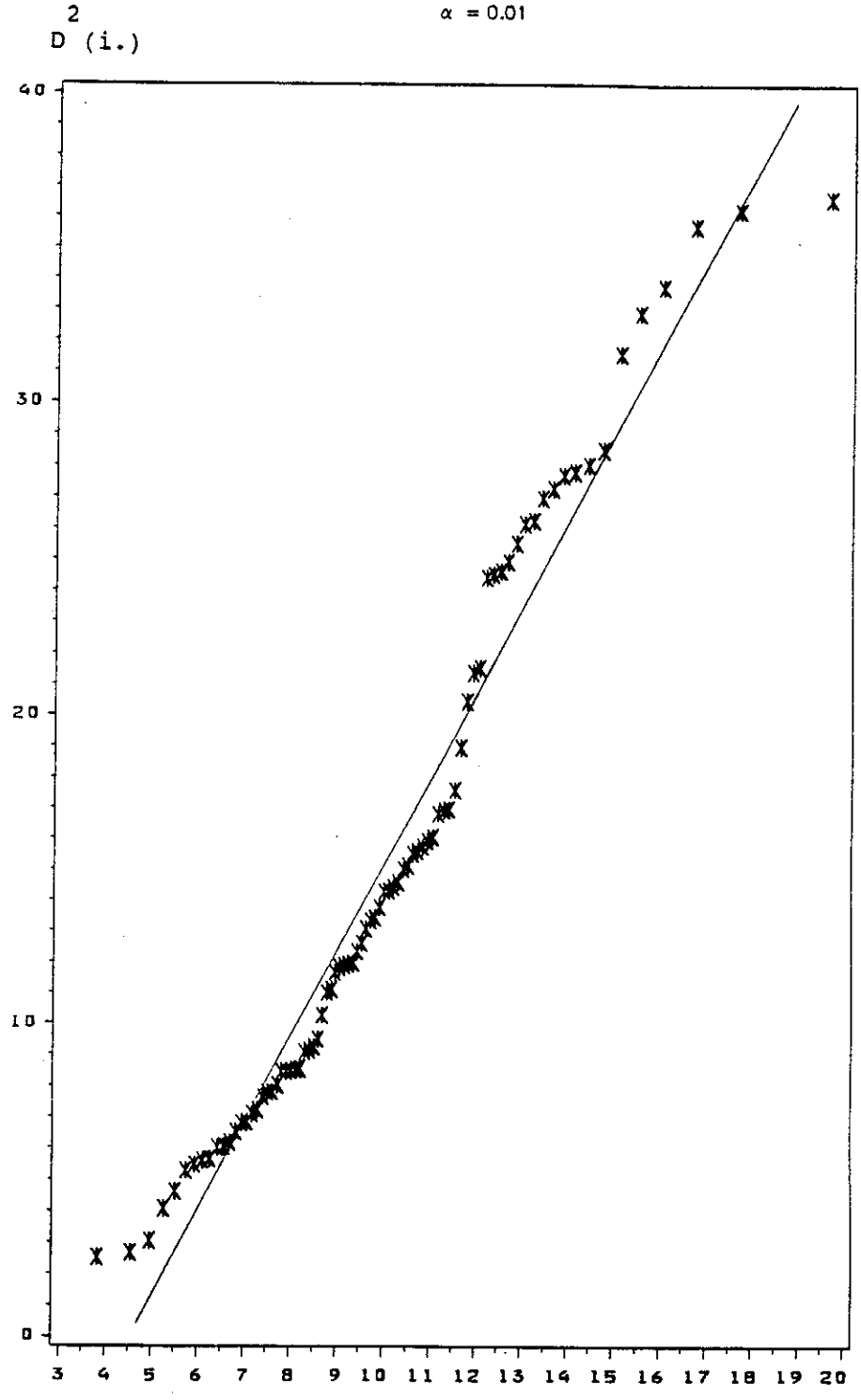


$$\text{Gamma}^{-1}\left(\frac{D^2(i.) - 0.5}{780}, 10\right)$$

KUVA 5

Multinormaalisuus konkurssit

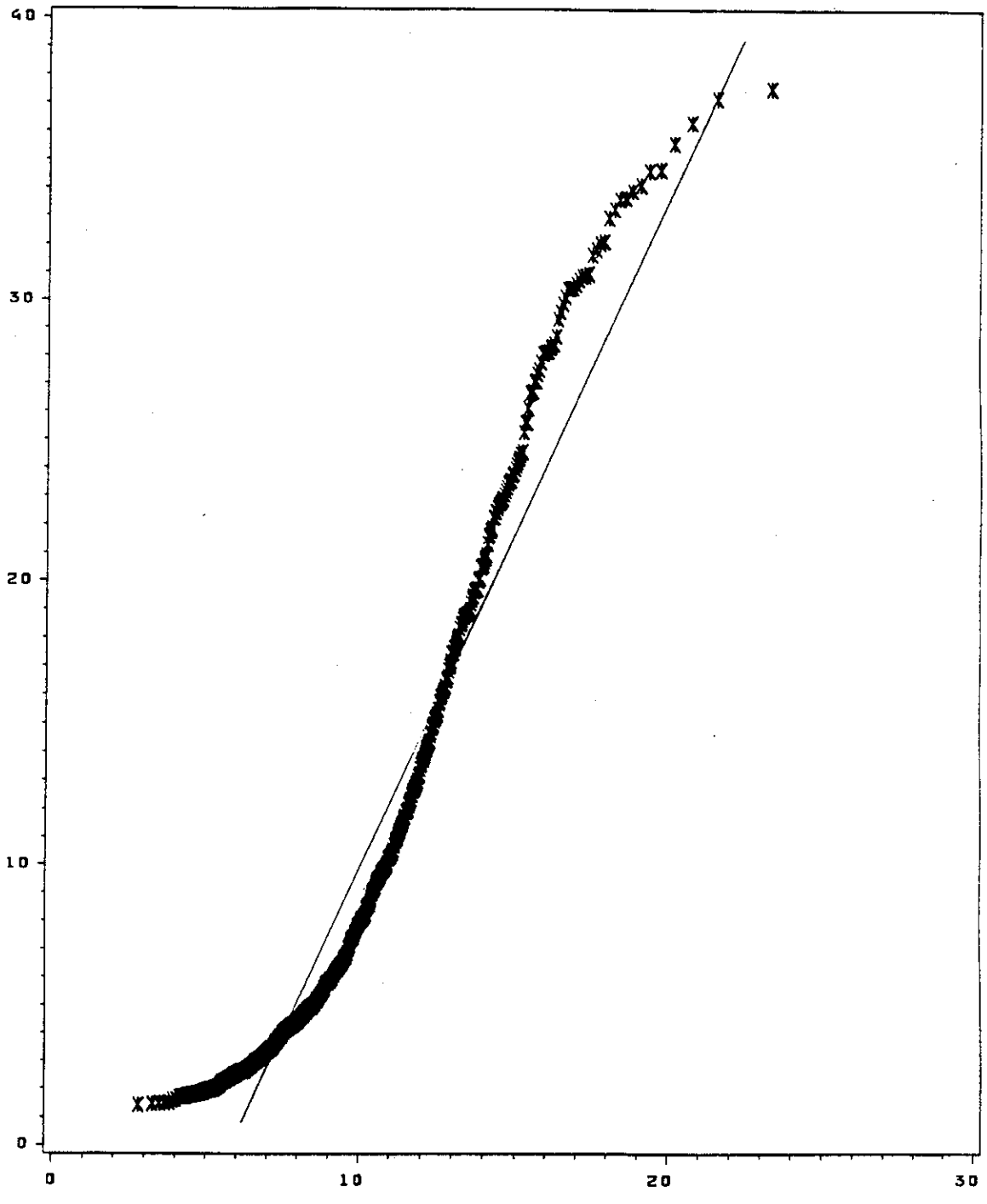
$\alpha = 0.01$



KUVA 6

Multinormaalisuus
toimivat

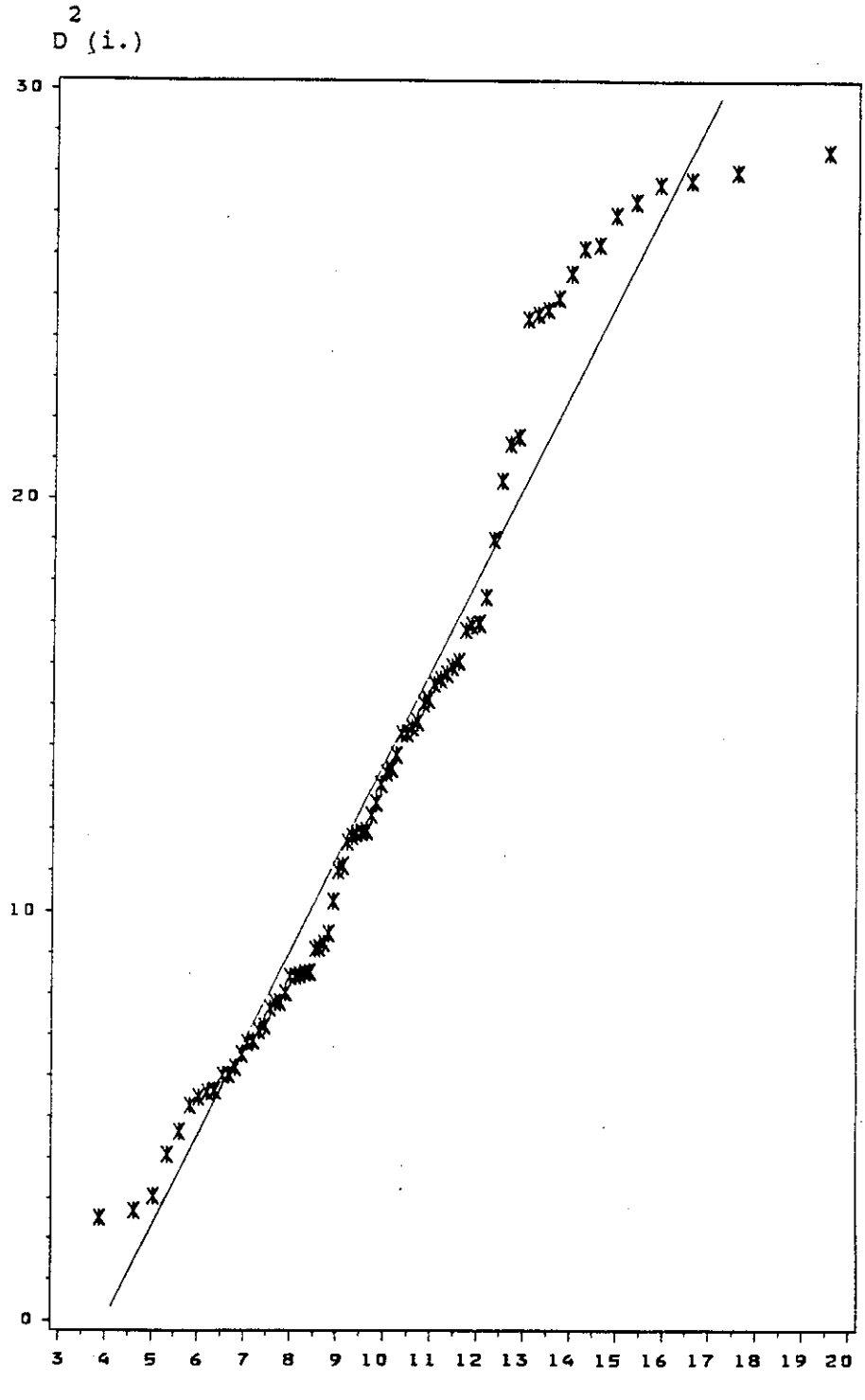
$\alpha = 0.01$



KUVA 7

Multinormaalisuus konkurssit

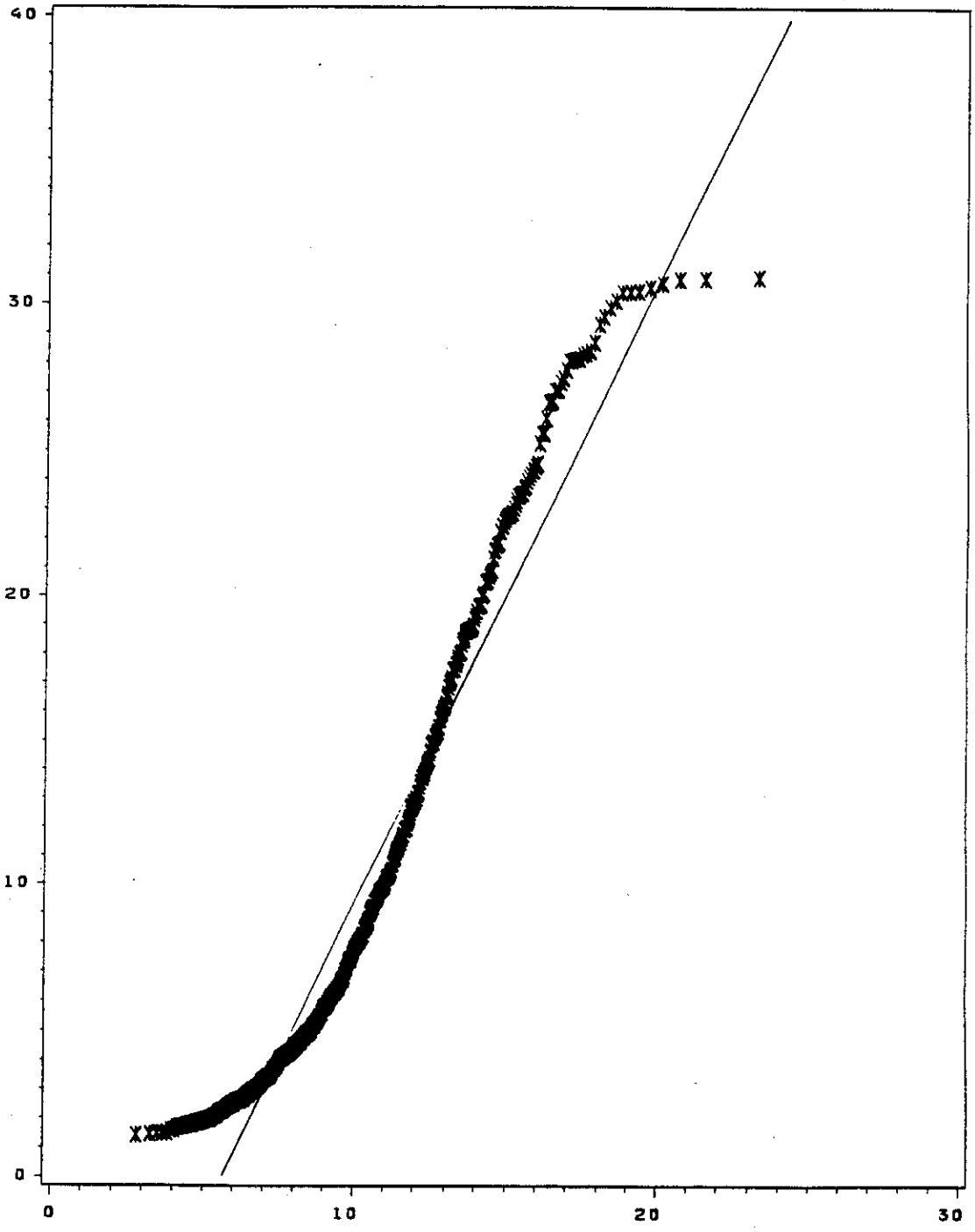
$\alpha = 0.05$



KUVA 8

Multinormaalisuus toimivat

$\alpha = 0.05$

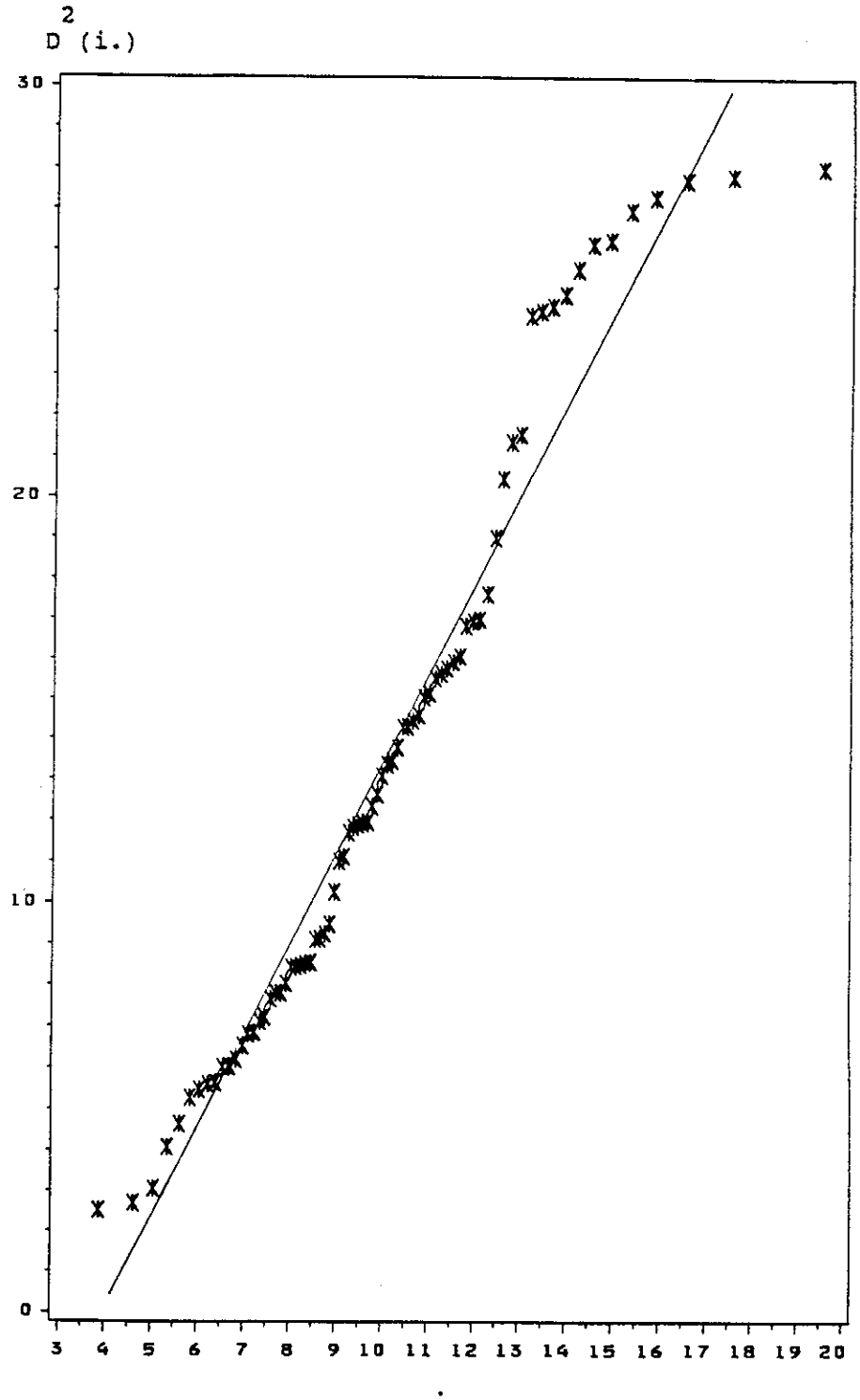


$$\text{Gamma}^{-1}\left(\frac{D^2(i.) - 0.5}{751}, 10\right)$$

KUVA 9

Multinormaalisuus konkurssit

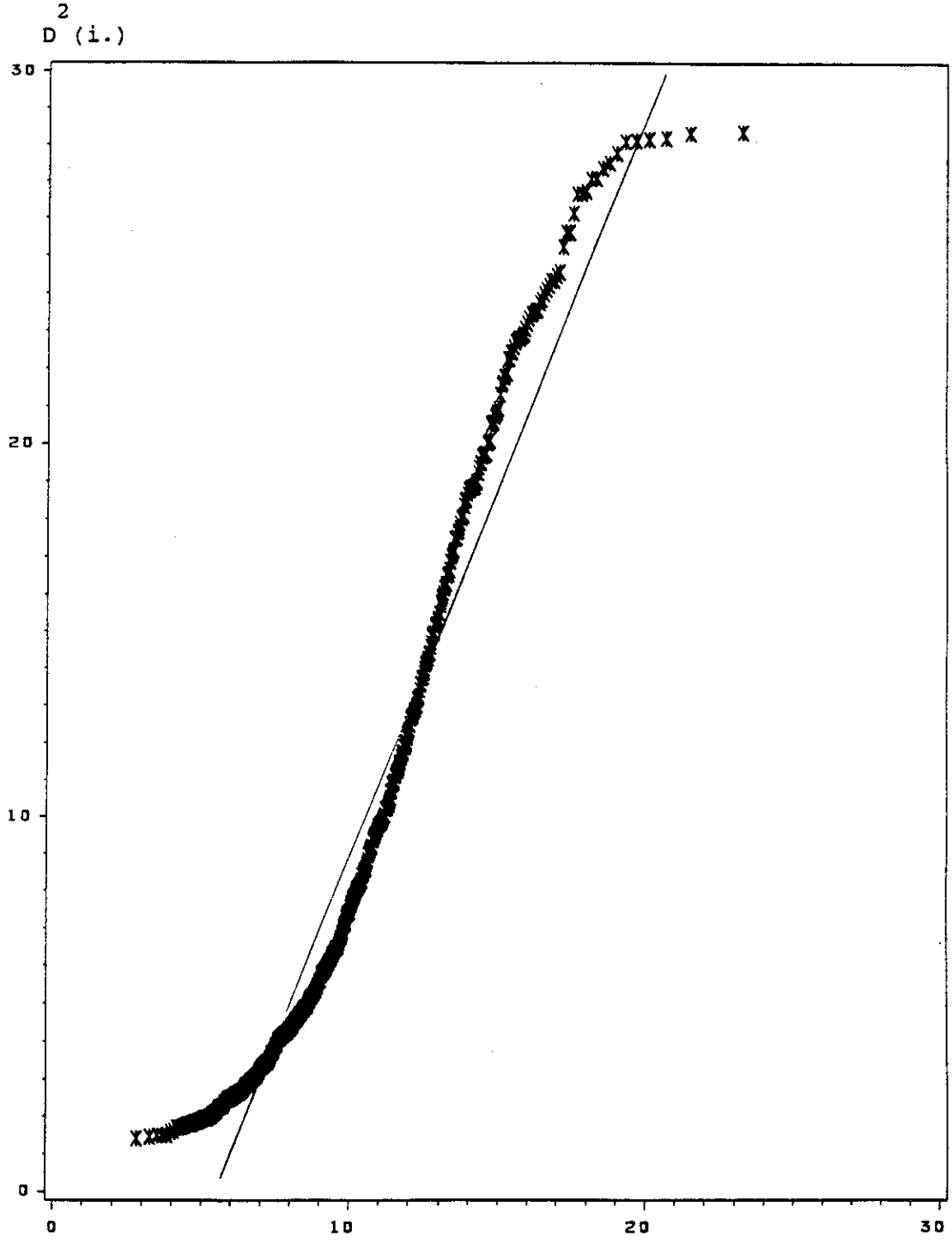
$\alpha = 0.10$



KUVA 10

Multinormaalisuus toimivat

$\alpha = 0.10$



10.8 KOVARIANSSIMATRIISIT

Toinen tärkeä erotteluanalyysin oletuksista on populaatioiden kovarianssimatriisien yhtäsuuruus. Suoritetaan se Bartlettin testin monimuuttuja yleistyksellä. ((45))

Merkitään
- g = ryhmien lukumäärä
- p = muuttujien lukumäärä
- n = havaintojen kokonaislukumäärä
- ni = havaintojen lukumäärä i:nnessä ryhmässä
- Wi = ryhmän i sisäinen hajontamatriisi (7.1.2:n Ai)
- W = ryhmien yhdistetty sisäinen hajontamatriisi

$$V = \frac{\prod |Wi|^{ni/2}}{|W|^{n/2}}$$

$$\rho = 1.0 - \left[\sum_{i=1}^{g} \frac{1}{ni - 1} - \frac{1}{n-g} \right] \frac{2p^2 + 3p - 1}{6(p+1)(g-1)}$$

$$f = (g-1)*p*(p+1)/2$$

sekä testitunnuslukua

$$T = -2*\rho*Ln\left[\frac{n^{p*n/2} \, V}{\prod_{i=1}^{g} ni^{p*ni/2}} \right]$$

Hypoteesina Ho on, että kovarianssimatriisit ovat yhtäsuuret eli tarkasteltavassa tapauksessa

Ho: W1 = W2

Tällöin Ho:n vallitessa T on likipitäen χ^2-jakautunut vapausastein f.

α	T	Tod. $>\chi^2(f)$
0.0	1107.62	0.0001
0.001	1199.74	0.0001
0.01	1107.66	0.0001
0.05	869.16	0.0001
0.10	887.69	0.0001

Käytettäessä χ^2-approksimaatiota vapausastein 210 johtaa todennäköisyysarvoon < 0.0001, joten oletus yhtäsuurista kovarianssimatriiseista on hylättävä 0.10:n merkitsevyystasolla. Tämä antaa viitteen siitä, että erottelufunktio on neliömuotoinen.

10.9 TARIFFITEKIJÖIDEN VALINTA

Käytetään askeltavaa valintamenetelmää (9.3.3) tärkeiden tariffitekijöiden löytämiseksi. Sisääntulokriteerinä käytetään 0.15 ja poistakriteerinä 0.30, nämä F-testille olevat arvot ovat Afifin ja Clarkin suosittelemia. ((46))

Yhteenvetona muuttujien valinnasta saadaan.

I) Kun α =0.0 , jolloin n1=843 ja n2=93.

Askel	Muuttuja Sisään	Poista	Mukana	Osittais R**2	F-luku	Tod>F	Wilksin lambda	Tod < lambda
1	x4		1	.086	88.205	.0001	.9137	.0001
2	x8		2	.079	79.858	.0001	.8417	.0001
3	x9		3	.020	19.087	.0001	.8248	.0001
4	ex8		4	.017	16.390	.0001	.8105	.0001
5	ex10		5	.004	4.125	.0425	.8069	.0001
6	x1		6	.004	3.438	.0640	.8040	.0001
7	ex9		7	.004	3.534	.0604	.8009	.0001
8	ex7		8	.003	2.536	.1116	.7987	.0001
9	ex2		9	.004	3.515	.0611	.7957	.0001

II) Kun α =0.001 , jolloin n1=780 ja n2=87.

Askel	Muuttuja Sisään	Poista	Mukana	Osittais R**2	F-luku	Tod>F	Wilksin lambda	Tod < lambda
1	x4		1	.189	201.467	.0001	.8111	.0001
2	x8		2	.078	73.004	.0001	.7479	.0001
3	x9		3	.015	13.076	.0003	.7367	.0001
4	ex5		4	.004	3.372	.0667	.7339	.0001
5	x10		5	.003	2.509	.1136	.7317	.0001

III) Kun α =0.01 , jolloin n1=767 ja n2=82.

Askel	Muuttuja Sisään	Poista	Mukana	Osittais R**2	F-luku	Tod>F	Wilksin lambda	Tod < lambda
1	x4		1	.176	180.789	.0001	.8241	.0001
2	x8		2	.074	67.625	.0001	.7631	.0001
3	x9		3	.016	13.801	.0002	.7508	.0001
4	x10		4	.004	2.963	.0856	.7482	.0001
5	x6		5	.004	3.117	.0778	.7455	.0001

IV) Kun α =0.05 , jolloin n1=751 ja n2=76.

Askel	Muuttuja Sisään	Poista	Mukana	Osittais R**2	F-luku	Tod>F	Wilksin lambda	Tod < lambda
1	x4		1	.189	192.462	.0001	.8198	.0001
2	x8		2	.080	71.696	.0001	.7459	.0001
3	x9		3	.015	12.175	.0005	.7351	.0001
4	ex10		4	.005	3.889	.0489	.7316	.0001
5	ex5		5	.004	3.007	.0833	.7289	.0001
6	x5		6	.005	4.092	.0434	.7253	.0001
7	ex9		7	.009	7.004	.0083	.7192	.0001

V) Kun α =0.10 , jolloin n1=737 ja n2=75.

Askel	Muuttuja Sisään	Poista	Mukana	Osittais R**2	F-luku	Tod>F	Wilksin lambda	Tod < lambda
1	x4		1	.188	188.063	.0001	.8116	.0001
2	x8		2	.088	78.365	.0001	.7399	.0001
3	x9		3	.017	14.235	.0002	.7271	.0001
4	ex10		4	.005	4.395	.0363	.7232	.0001
5	x6		5	.003	2.724	.0992	.7207	.0001
6	ex5		6	.003	2.173	.1409	.7188	.0001
7	ex9		7	.004	3.018	.0827	.7161	.0001
8	x5		8	.006	4.594	.0324	.7120	.0001

Tarkastellaan vielä muuttujia rinnakkain vertailun helpottamiseksi.

TAULUKKO 8 Valitut muuttujat, kun vain ensimmäisen asteen termit mukana

Askel	Valitut muuttujat sekä F-todennäköisyysarvo									
	I		II		III		IV		V	
1	x4	.0001	x4	.0001	x4	.0001	x4	.0001	x4	.0001
2	x8	.0001	x8	.0001	x8	.0001	x8	.0001	x8	.0001
3	x9	.0001	x9	.0003	x9	.0002	x9	.0005	x9	.0002
4	ex8	.0001	ex5	.0667	x10	.0856	ex10	.0049	ex10	.0363
5	ex10	.0425	x10	.1136	x6	.0778	ex5	.0833	x6	.0992
6	x1	.0640					x5	.0434	ex5	.1409
7	ex9	.0604					ex9	.0083	ex9	.0827
8	ex7	.1116							x5	.0324
9	ex2	.0611								

Muuttujat x4, x8 ja x9 ovat outliereistä riippumattomia ja ne tulevat valituiksi joka kerta samassa tärkeysjärjestyksessä, muut muuttujat ovat enemmän ja vähemmän tarkasteltavasta aineistosta riippuvaisia. On muistettava, että aikaisemmin todettiin kovarianssimatriisien eri suuruus, mikä ilmeisestikin häiritsee muuttujien valintaa, koska jokaisella valinta-askeleella lasketaan yhdistetty sisäinen hajontamatriisi.

11.10 EROTTELUFUNKTIOT

Erottelufunktioiksi saadaan 7.2.2:sta, kun ryhmä G(1) on toimivat yritykset ja G(2) on konkurssiyritykset. Keskiarvovektorin voi muodostaa käyttäen taulukon 6 arvoja sekä liitteessä 4 on matriisit S1 ja S2, kun lisäksi oletetaan, että väärinluokittelukustannukset ja a'prioritodennäköisyydet ovat yhtäsuuret, saadaan erottelufunktio h(x) ratkaistua.

Päätössäännöksi saadaan silloin:
havainto x kuuluu G(1):teen, mikäli (11.10.1) on pienempi kuin 0 ja
G(2):teen, mikäli (11.10.1) on suurempi kuin 0.

$$h(x) = x'(S1^{-1} - S2^{-1})x + 2*(\bar{x}^{(2)}{}'S2^{-1} - \bar{x}^{(1)}{}'S1^{-1})x + C \qquad (11.10.1)$$

missä vakio C on

$$C = \ln \frac{|S1|}{|S2|} + \bar{x}^{(1)}{}'S1^{-1}\bar{x}^{(1)} - \bar{x}^{(2)}{}'S2^{-1}\bar{x}^{(2)}$$

Erottelufunktio saadaan muotoon

$$h(x) = \sum_{i=1}^{p}\sum_{j=i}^{p} a(i,j)*xi*xj + \sum_{i=1}^{p} b(i)*xi + c$$

Ongelman muodostaa nyt se, että erottelufunktiossa h(x) tulee olemaan runsaasti muuttujia. Jos matriisi S on p*p-ulotteinen, niin erottelufunktiossa on toisen asteen termejä p(p+1)/2, ensimmäisen asteen termejä p kappaletta sekä vakio termi eli yhteensä termejä (p**2+3p+2)/2.

	p	Termejä
0.0	9	55
0.001	5	21
0.01	5	21
0.05	7	36
0.10	8	48

Koska muuttujien määrä on näinkin suuri, pyritään kaavoja linearisoimaan valitsemalla erottelun kannalta merkitykselliset muuttujat. Muodostetaan uusia muuttujia kertomalla muuttujat keskenään eli esimerkiksi muuttuja x1x2 on muuttujien x1:n ja x2:n tulo. Valitaan näin muodostetuista muuttujista sekä alkuperäisistä muuttujista askeltavaa valintamenetelmää käyttäen tariffitekijät. Muuttujajoukko käsittää nyt yhteensä 230 mahdollista tariffitekijää. Askelten maksimimääränä käytetään 30. Liitteessä 5 on yhteenveto merkityksellisistä muuttujista.

Havaitaan, että valituiksi tulisi lukuisa määrä tulomuotoisia termejä. Mallin kannalta kaikki 30 termiä ei varmastikaan ole tarpeellisia, joten poispudottamisen voi aloittaa viimeisestä askeleesta 30. Tarkastellaan sitten Mahalanobiksen neliöetäisyyden muuttumista, kun muuttuja jätetään mallista pois. ETK:n kaavaan valittiin siten 14 ensimmäistä muuttujaa, joten tarkastellaan jatkossa myöskin 14 muuttujan malleja.

TAULUKKO 9 Valitut muuttujat, kun toisen asteen termit mukana

Lkm	I	II	III	IV	V
1	x4	x4x9	x4x9	x4x9	x4x9
2	x4x8	x8	x2x8	x8x9	x8x9
3	x1ex4	x8x8	x4	ex9ex9	ex9ex9
4	x1x8	ex5ex9	ex9ex9	x4ex9	x4ex9
5	ex5ex9	x4x10	x4ex9	ex4ex4	ex4ex4
6	ex2ex10	x10ex9	x5ex10	x1x3	ex4ex6
7	x3ex2	x10	x1x3	x3ex10	x3ex10
8	x9ex1	x1x9	x5ex3	ex4ex6	x1x3
9	ex9ex10	ex2ex10	x8ex3	x5ex3	ex1ex7
10	ex3ex4	ex8ex9	x6x6	x7ex1	ex2ex2
11	ex5	ex5ex5	ex7ex8	ex2ex2	ex5ex9
12	ex2ex4	x9ex7	ex2ex2	ex8ex9	ex5ex6
13	x1x4	ex3ex3	ex10	x2ex8	x5x8
14	x8	x9x10	ex2ex10	x10ex5	x6ex5

Ensimmäisen asteeen muuttujia esiintyy I:ssä erottelufunktiossa kolme, II:ssa kaksi ja III:ssa yksi muuttuja, kaikki muut ovat toisen asteen termejä. I:llä ja II:llä erottelufunktiolla on kolme samaa muuttujaa, nimittäin muuttujat ex5ex9, ex2ex10 ja x8, II:lla ja III:lla kaavalla kaksi, x4x9 ja ex2ex10, III:lla ja IV:lla jo kuusi samaa muuttujaa, x4x9, x2x8, ex9ex9, x4ex9, x5ex10 ja x1x3 sekä IV:llä ja V:llä peräti seitsemän samaa, joista viisi ensimmäistä on otettu vieläpä samassa järjestyksessä, x4x9, x8x9, ex9ex9, x4ex9 ja ex4ex4 sekä vielä x1x3, x3ex10 ja ex4ex6.

Verrattaessa taulukon 8 muuttujia ja taulukon 9 muuttujia keskenään havaitaan, että muuttujat x4, x8 ja x9 esiintyvät nytkin, mutta useimmiten tulomuotoisissa termeissä. Taulukon 8 I-muuttujaryhmän ex8 ja ex7 eivät esiinny enää taulukon 9 I-ryhmän muuttujissa, ne tulisivat valituiksi vasta 19. ja 27. askeleella (liite 5). Vastaavasti muuttuja x10 III-ryhmän muuttujista tulisi valituksi vasta 15. askeleella.

11.11 KANONINEN EROTTELUANALYYSI

Lasketaan konkurssien ja toimivien yritysten välisen korrelaation maksimi käytettäessä muuttujia I - V.

Oheisessa taulukossa on lisäksi laskettu ryhmien välisen eron testaamista tarvittavia lukuja. (8.4.4, 8.4.6 ja 8.4.8)

Muuttujat	Kanoninen korrelaatio	Ominaisarvo	Wilksin lambda	Pillain trace	Hot.Law. trace
I	.552	.4388	.695	.305	.439
II	.605	.5781	.634	.366	.578
III	.595	.5489	.646	.354	.549
IV	.618	.6172	.618	.382	.617
V	.630	.6593	.603	.397	.659

Suoritettaessa monimuuttujatestit (8.4.5, 8.4.7, 8.4.9 ja 8.4.11) saadaan testiluvun F arvoksi 28.869, jonka todennäköisyysarvo < 0.0001, koska tunnusluku F on F-jakutunut vapausastein 14 ja 921, kun käytetään I-muuttujia. Vastaavasti käytettäessä II-muuttujia saadaan F:n arvoksi 35.183, joka on myös <0.0001 (vapausasteet 14 ja 852). III-muuttujajoukon osalta F-luku on 32.700 ja edelleen todennäköisyysarvo on < 0.0001 (vapausasteet 14 ja 834). IV-muuttujien F-luku on 35.495 ja < 0.0001 (vapausasteet 14 ja 812) sekä vielä V-muuttujajoukon F-luvun arvo 37.533 ja <0.0001 (vapausastein 14 ja 797).

Joten muuttujajoukot erottelevat konkurssit ja toimivat yritykset toisistaan käytettiinpä mitä tahansa I - V muuttujia.

Tarkastellaan seuraavaksi erottelumuuttujien ja kanonisen erottelufunktion välisiä korrelaatioita eli sisäistä rakennetta. (8.5)

I-muuttujajoukosta korkeimmat arvot ovat muuttujilla x4 ja x8 sekä alle 0.10 arvoja esiintyy peräti seitsemällä muuttujalla. II-muuttujajoukossa suurimmat arvot x4x9:llä ja x8:lla sekä alle 0.10 arvoja kahdella. III-muuttujajoukossa korkeimmat x4x9:llä ja x4:llä ja pienet myöskin kahdella. Muuttujoukoissa IV-V korkeimmat arvot x4x9:llä ja x8x9:llä sekä alhaiset kahdella.

Alhainen arvo tarkoittaa, ettei kyseisellä muuttujalla ole itseasiassa paljoakaan erotteluvoimaa, mutta ei ole mitään sääntöä siitä, miten pieni sisäinen rakennekerroin saa olla muuttujan poisjättämiselle.

TAULUKKO 10 Sisäisten rakennekerrointen itseisarvot

I-muuttujat n1=843 n2=93		II-muutt. n1=780 n2=87		III-muutt. n1=767 n2=82		IV-muuttujat n1=751 n2=76		V-muuutt. n1=737 n2=75	
x4	.46	x4x9	.67	x4x9	.69	x4x9	.69	x4x9	.67
x4x8	.24	x8	.55	x2x8	.56	x8x9	.55	x8x9	.54
x1ex4	.07	x8x8	.10	x4	.62	ex9ex9	.51	ex9ex9	.52
x1x8	.10	ex5ex9	.32	ex9ex9	.53	x4ex9	.34	x4ex9	.34
ex5ex9	.18	x4x10	.18	x4ex9	.40	ex4ex4	.39	ex4ex4	.40
ex2ex10	.06	x10ex9	.08	x5ex10	.15	x1x3	.12	ex4ex6	.29
x3ex2	.14	x10	.04	x1x3	.13	x3ex10	.14	x3ex10	.13
x9ex1	.05	x1x9	.34	x5ex3	.24	ex4ex6	.29	x1x3	.12
ex9ex10	.00	ex2ex10	.15	x8ex3	.21	x5ex3	.25	ex1ex7	.04
ex3ex4	.23	ex8ex9	.23	x6x6	.07	x7ex1	.07	ex2ex2	.31
ex5	.05	ex5ex5	.16	ex7ex8	.26	ex2ex2	.31	ex5ex9	.32
ex2ex4	.06	x9ex7	.28	ex2ex2	.26	ex8ex9	.23	ex5ex6	.16
x1x4	.03	ex3ex3	.20	ex10	.06	x2ex8	.40	x5x8	.48
x8	.45	x9x10	.20	ex2ex10	.10	x10ex5	.04	x6ex5	.01

11.12 KANONINEN EROTTELUFUNKTIO

Merkitään kanonista erottelufunktiota h(y):llä

$$h(y) = \sum_{i=1}^{14} a(i)*y(i) + C$$

missä muuttujavektori y(i) saadaan taulukosta 11 ja vastaava kerroinvektori taulukosta 12.

TAULUKKO 11 Kanonisen erottelufunktion muuttujat

y(i)	I	II	III	IV	V
y(1)	x4	x4x9	x4x9	x4x9	x4x9
y(2)	x4x8	x8	x2x8	x8x9	x8x9
y(3)	x1ex4	x8x8	x4	ex9ex9	ex9ex9
y(4)	x1x8	ex5ex9	ex9ex9	x4ex9	x4ex9
y(5)	ex5ex9	x4x10	x4ex9	ex4ex4	ex4ex4
y(6)	ex2ex10	x10ex9	x5ex10	x1x3	ex4ex6
y(7)	x3ex2	x10	x1x3	x3ex10	x3ex10
y(8)	x9ex1	x1x9	x5ex3	ex4ex6	x1x3
y(9)	ex9ex10	ex2ex10	x8ex3	x5ex3	ex1ex7
y(10)	ex3ex4	ex8ex9	x6x6	x7ex1	ex2ex2
y(11)	ex5	ex5ex5	ex7ex8	ex2ex2	ex5ex9
y(12)	ex2ex4	x9ex7	ex2ex2	ex8ex9	ex5ex6
y(13)	x1x4	ex3ex3	ex10	x2ex8	x5x8
y(14)	x8	x9x10	ex2ex10	x10ex5	x6ex5

TAULUKKO 12 Kanonisen erottelufunktion kertoimet

y(i)	I	II	III	IV	V
y(1)	-5.00	-0.352	-0.237	0.461	-0.423
y(2)	0.121	8.22	0.232	1.03	0.556
y(3)	0.0779	-0.161	-2.99	-9.37	-8.06
y(4)	-0.0244	-2.69	-7.60	0.526	0.612
y(5)	-0.849	-1.49	0.560	-0.0888	-0.123
y(6)	-4.44	18.6	-1.76	0.0425	0.383
y(7)	-0.0547	40.0	0.0551	-0.672	-0.699
y(8)	-0.183	0.288	-0.0693	0.248	0.0125
y(9)	3.48	-3.55	0.180	-0.0509	-0.175
y(10)	-0.0785	-2.74	-0.0162	-0.145	-0.187
y(11)	1.282	0.156	-0.124	-0.173	-1.18
y(12)	0.0967	0.412	-0.142	-1.55	-0.332
y(13)	0.00871	-0.0474	27.7	0.118	0.117
y(14)	2.30	-5.99	-3.18	1.43	0.0766

Kertoimet on mukavuussyistä kerrottu 100:lla, jolloin erottelupistemäärä tulee myös 100:lla kerrotuksi. Käytetään jatkossa erottelupistemäärien kokonaislukuarvoja.

Kanonisen erottelufunktion kaavassa vakio c saadaan määritettyä laskemalla kokonaiskeskiarvovektori Y ja sitten c = a'Y. Näin menetellen saadaan lasketuksi konkurssien ja toimivien yritysten erottelupistemäärin keskiarvot.

TAULUKKO 13 Mahalanobisetäisyys

	Erottelupistemäärien keskiarvot Konkurssit	Toimivat	Mahalanobis-etäisyys
I	-199	22	1.49
II	-227	25	1.59
III	-225	24	1.58
IV	-247	25	1.65
V	-254	26	1.67

Kuvissa 11 - 15 on neliömuotoisten erottelufunktioiden pistemäärien kertymäfunktiot. Vertailun vuoksi on kuvissa 16 -20 ensimmäisen asteen erottelufunktioiden pistemäärien kertymäfunktiot.

Ensimmäisenä havaintona on toimivien ja konkurssiyritysten kertymäfunktioiden välisen etäisyyden kasvaminen, kun määrittelyjoukosta poistetaan outliereita.

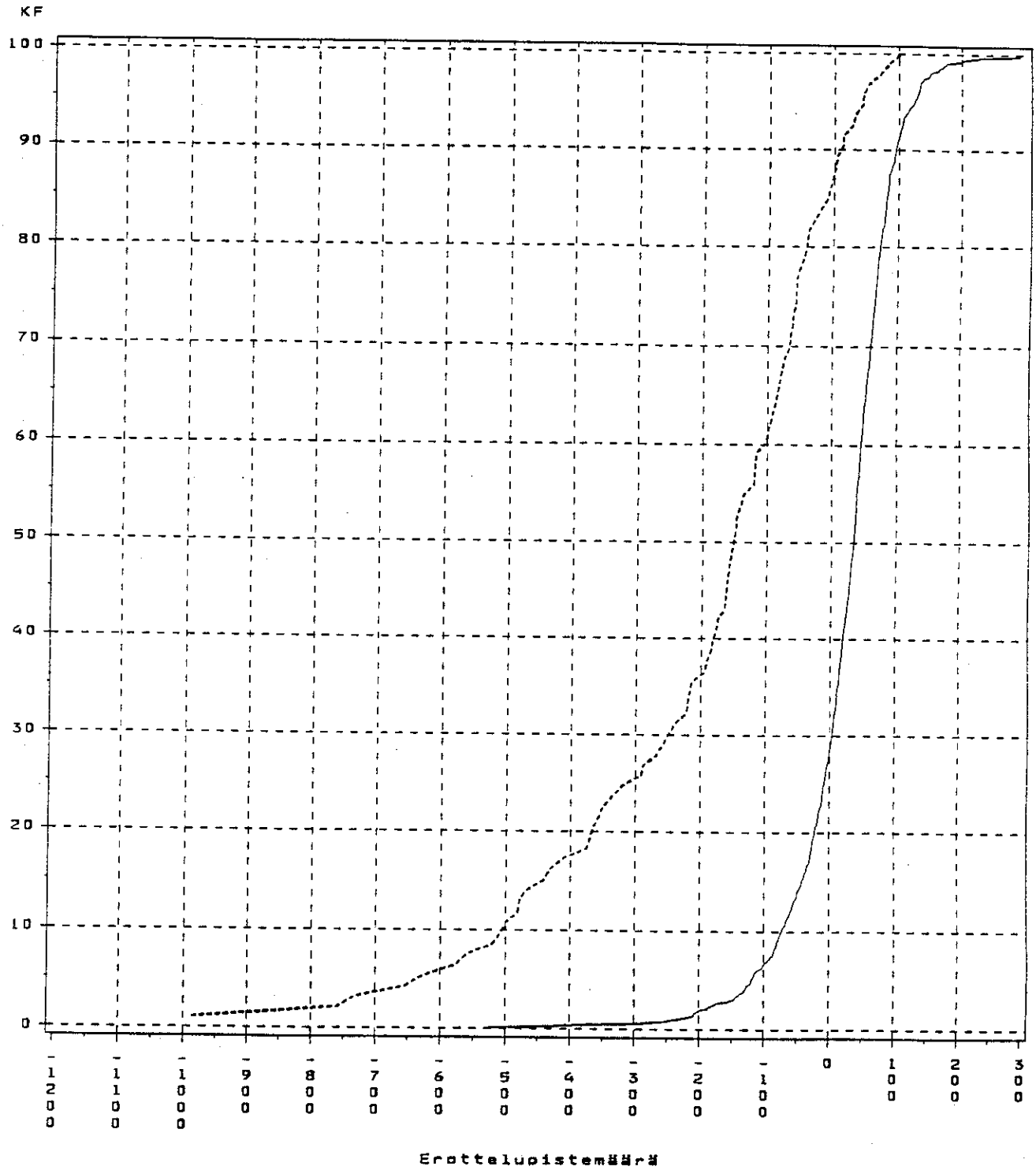
Toinen havainto on neliömuotoisen ja ensimmäisen asteen kertymäfunktioiden välinen ero verrattaessa kuvia keskenään samalla alfan arvolla. (Liitteessä 6 on ensimmäisen asteen kanonisen erottelufunktion tiedot)

Luokitteluvirheitä ovat ne tapaukset, joissa kertymäfunktiot ovat päällekkäin. Esimerkiksi kuvasta 11 nähdään, mikäli katkaisukohtana pidetään erottelupistemäärää 0, niin silloin noin 30% toimivista yrityksistä luokiteltaisiin konkursseiksi ja konkursseista luokiteltaisiin vastaavasti toimiviksi noin 10%.

KUVA 11

Neliömuotoisen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.0$

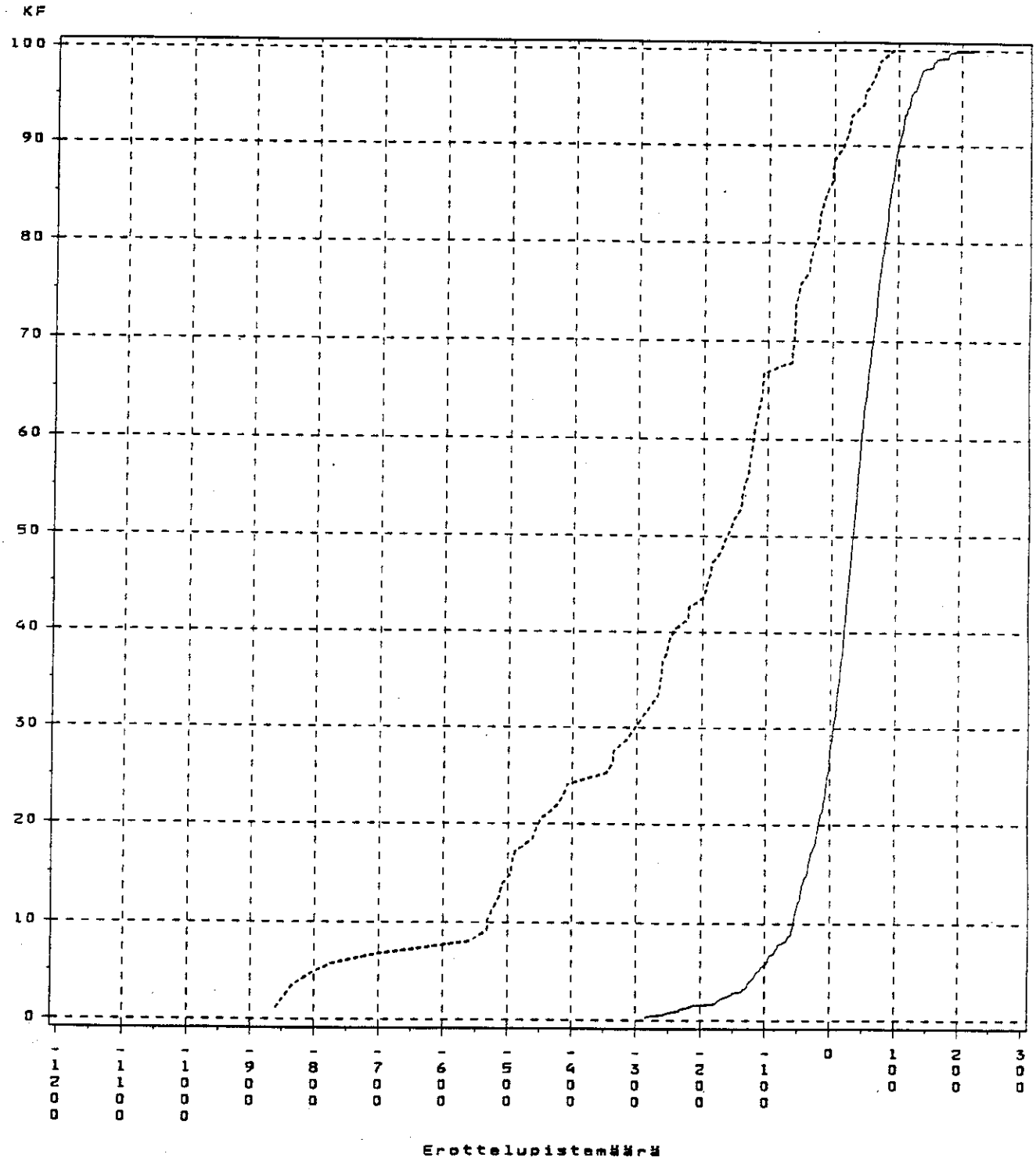


Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 12

Neliömuotoisen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.001$

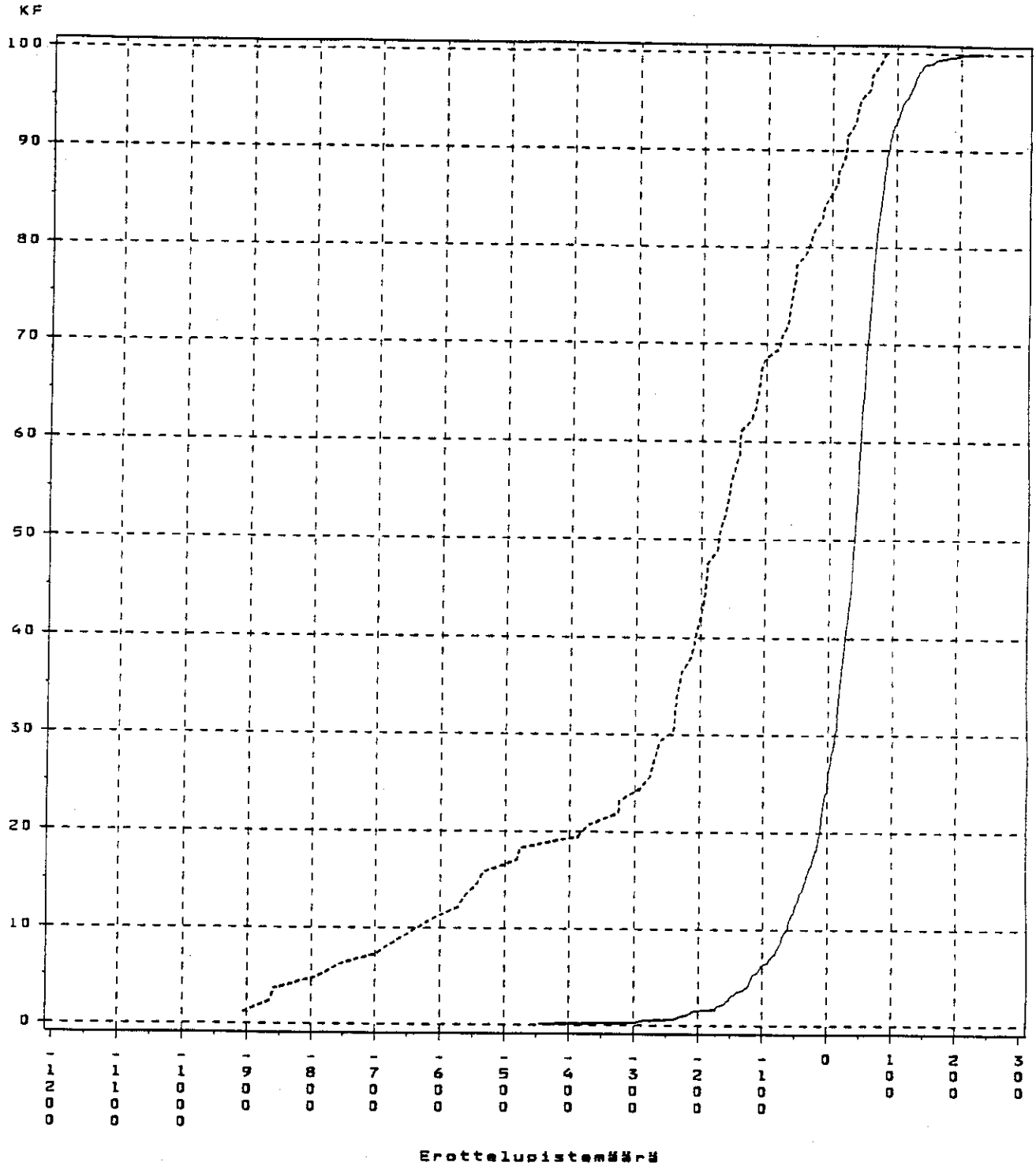


Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 13

Neliömuotoisen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.01$

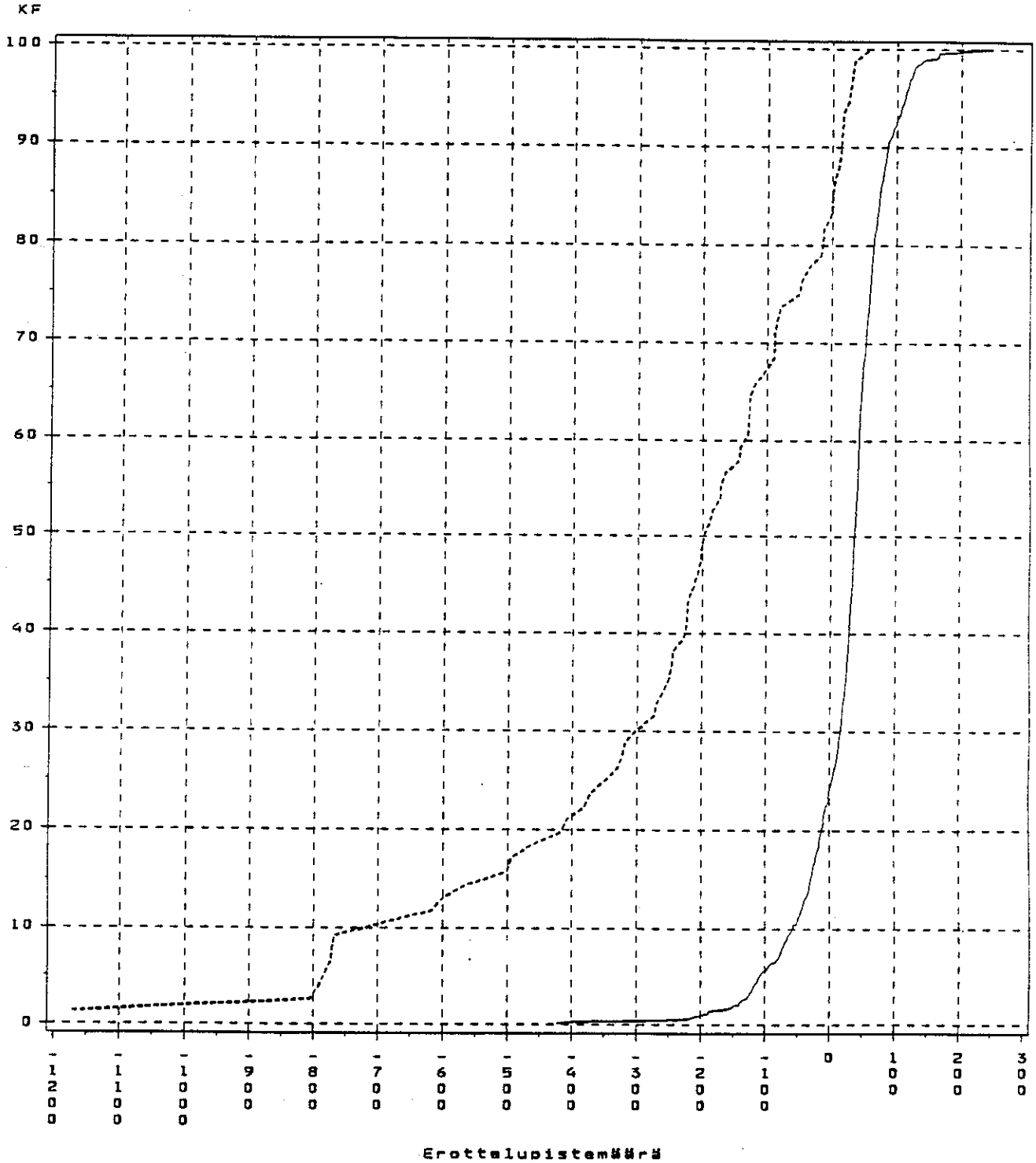


Yhtenäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 14

Neliömuotoisen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.05$

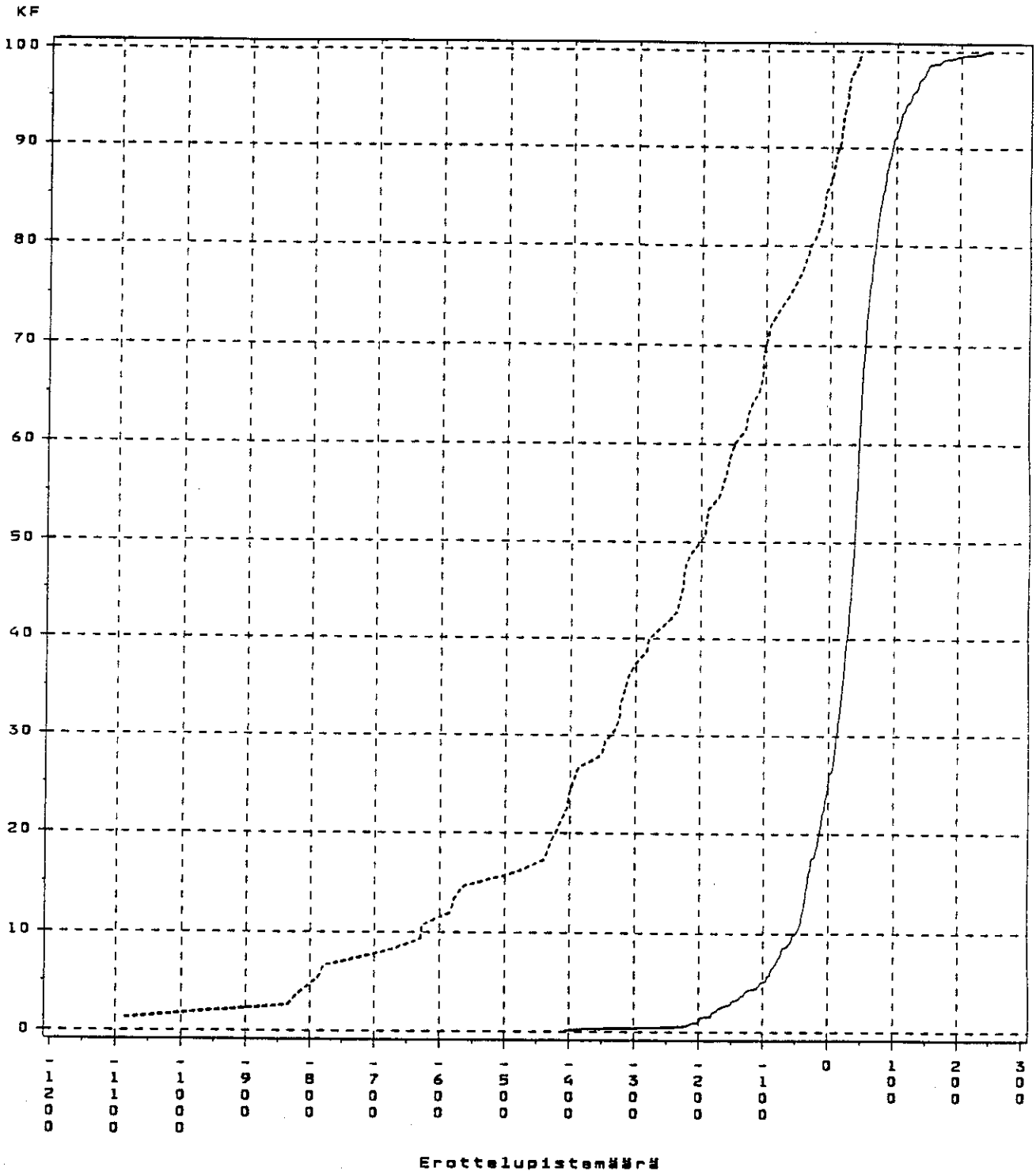


Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 15

Neliömuotoisen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.10$

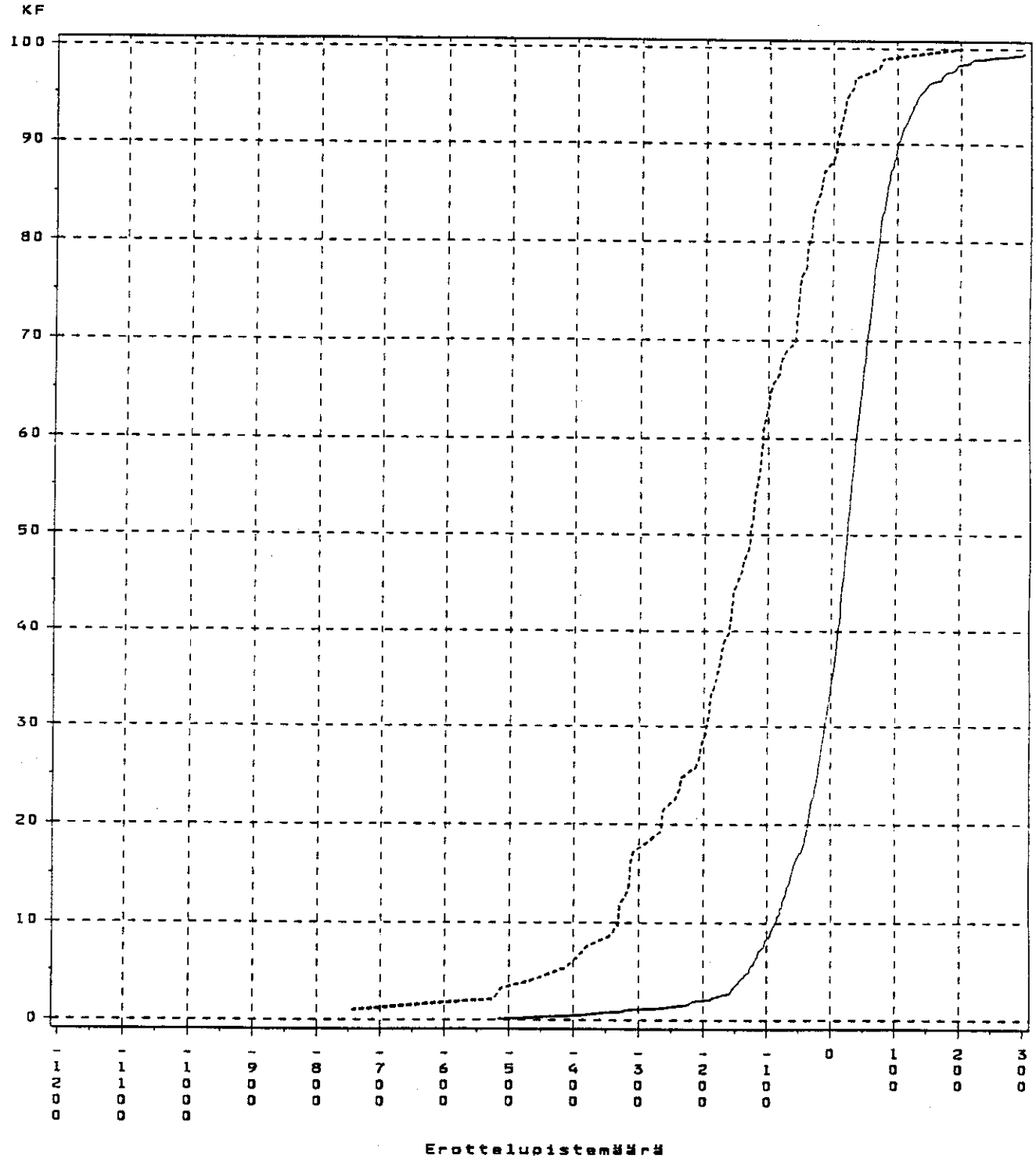


Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 18

Ensimmäisen asteen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.0$

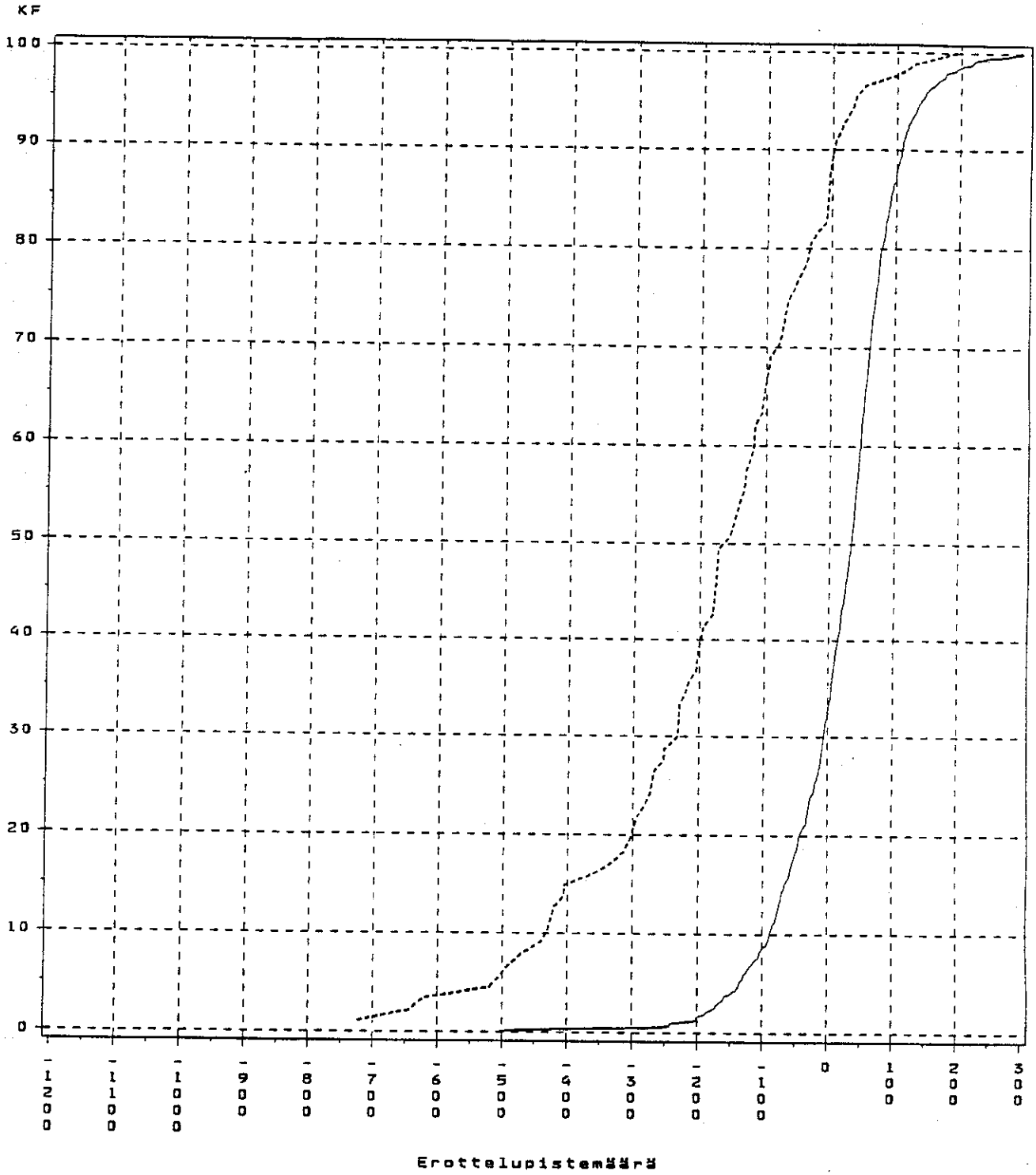


Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 17

Ensimmäisen asteen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.001$

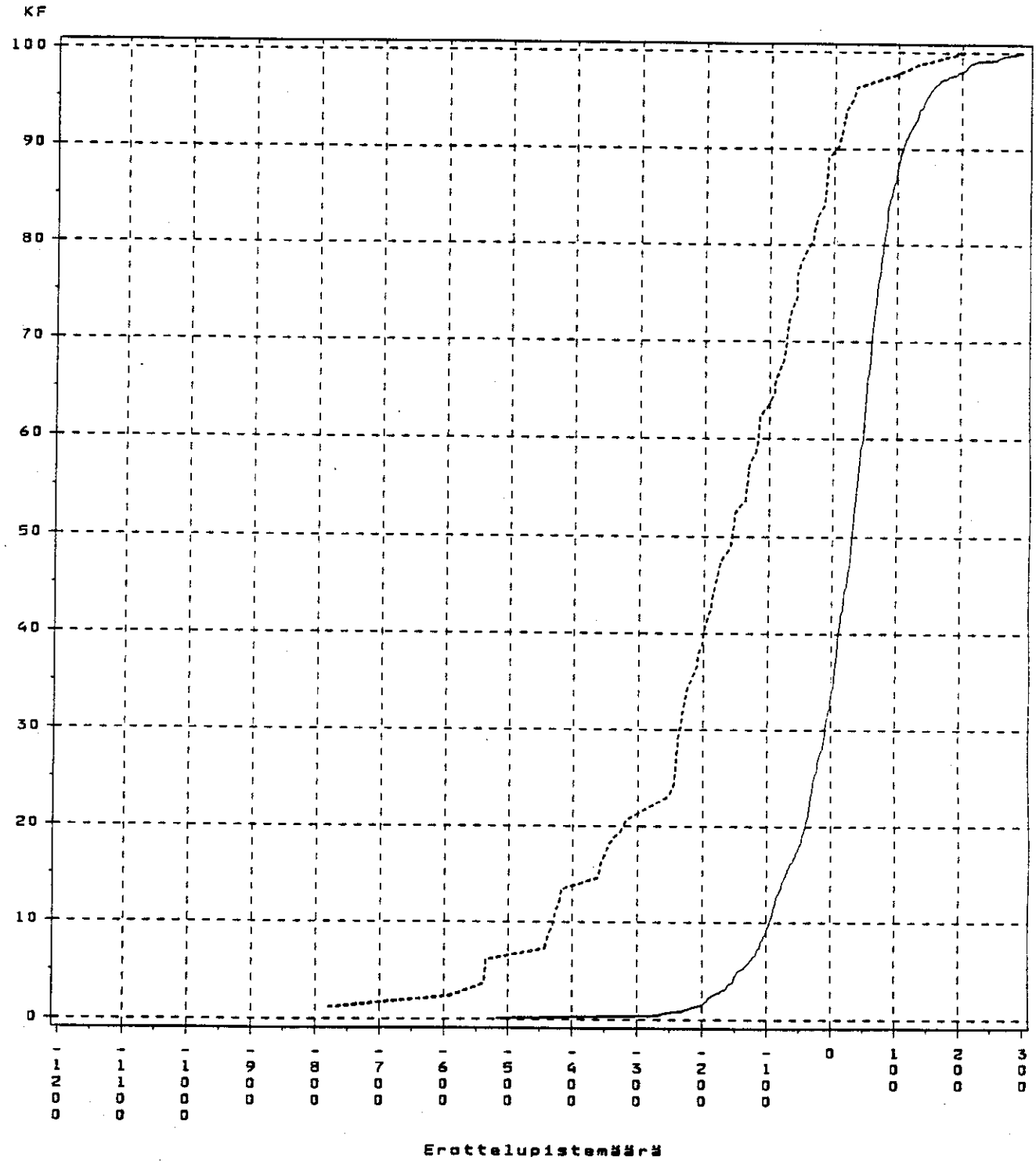


Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 18

Ensimmäisen asteen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.01$

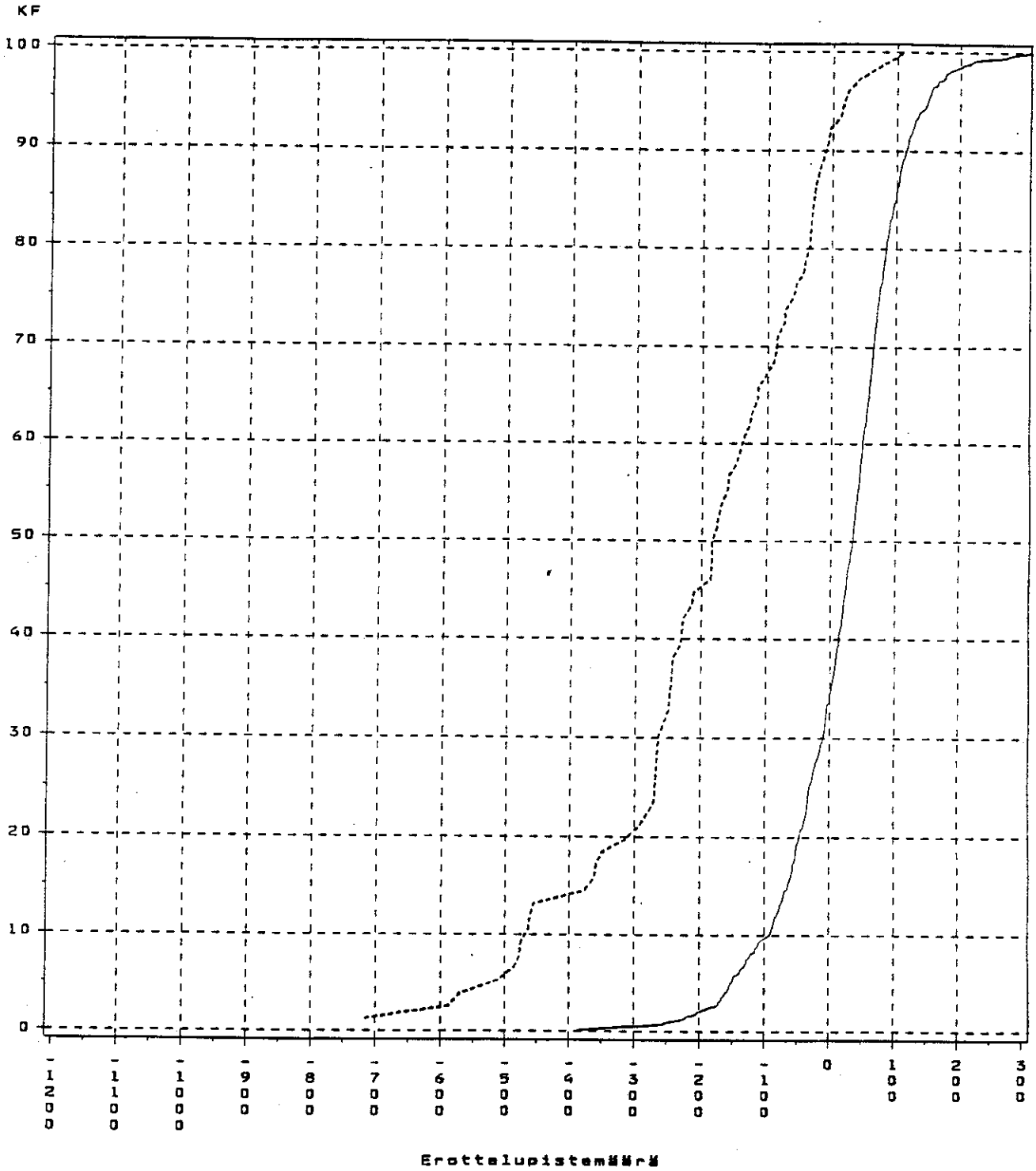


Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 19

Ensimmäisen asteen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.05$

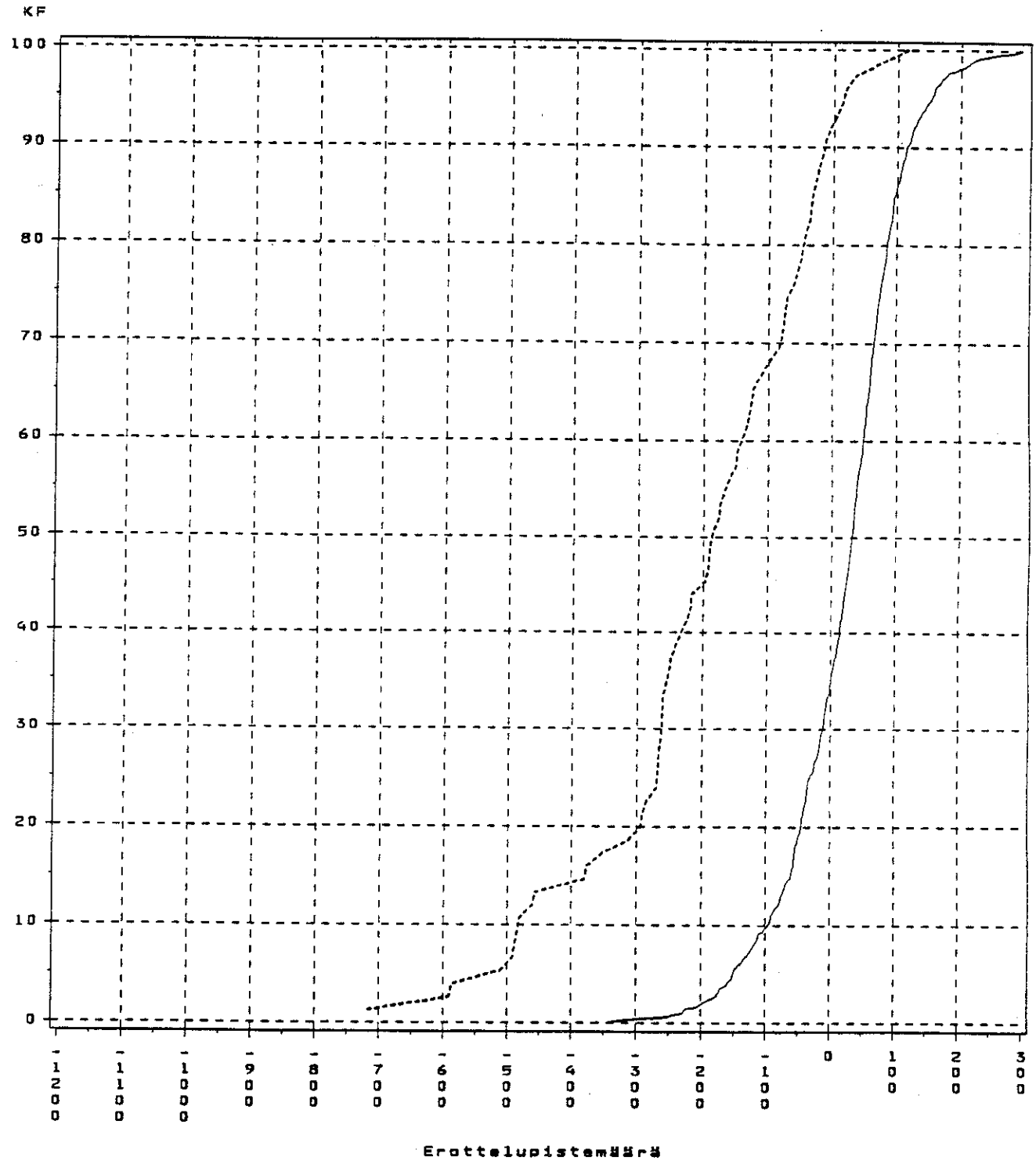


Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 20

Ensimmäisen asteen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.10$



Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

Edellä oli jo maininta siitä, että mitä tapahtuu jos multinormaalisuus-oletus ei pätisikään. Voisiko erottelufunktion kaavaan ilmestyä kolmannen asteen termejä eli voisiko erottelufunktio olla muotoa

$$h(x) = \sum_{i=1}^{p} \sum_{j=i}^{p} \sum_{k=j}^{p} a(i,j,k)*xi*xj*xk + \sum_{i=1}^{p} \sum_{j=i}^{p} b(i,j)*xi*xj$$

$$+ \sum_{i=1}^{p} c(i)*xi + d$$

Muodostetaan siis uusia muuttujia, jotka ovat kolmen alkuperäisen muuttujan tuloja. Kun otetaan ensimmäisen, toisen ja kolmennen asteen muuttujat mahdollisiksi tariffitekijöiksi, kohdataan kaksi ongelmaa, ensinnäkin tutkimusaineisto ei ole riittävä, mikäli kaikki mahdolliset muuttuja kombinaatiot ovat mukana ja toiseksi kokeilu olisi suhteettoman kallista tietokoneajan käytön kannalta. Joten muodostetaan seuraavanlaisia termejä

x1x1x1, x2x2x2, ... , ex10ex10ex10,
x1x4x7, x1x4x8, ... , x1x4x10,
x1x5x7, x1x5x8, ... , x1x5x10,
x1x6x7, x1x6x8, ... , x1x6x10,
x2x4x7, x2x4x7, ... , x2x4x10,

...

ex3ex6ex7, ex3ex6ex8, ... , ex3ex6ex10.

Muuttujat x1-x3 ja ex1-ex3 ovat taseen vastaavien puolelta,
muuttujat x4-x6 ja ex4-ex6 taseen vastattavien puolelta sekä
muuttujat x7-x10 ja ex7-ex10 ovat tuloslaskelmasta.

Uudet muuttujat on muodostettu siis siten, että vastaavien puolen muuttuja kertaa vastattavien muuttuja kertaa tuloslaskelman muuttuja sekä lisäksi kolmannet potenssit, mutta kuitenkin vielä sillä rajoituksella, että x:t ja ex:t eivät esiinny samassa tulossa.

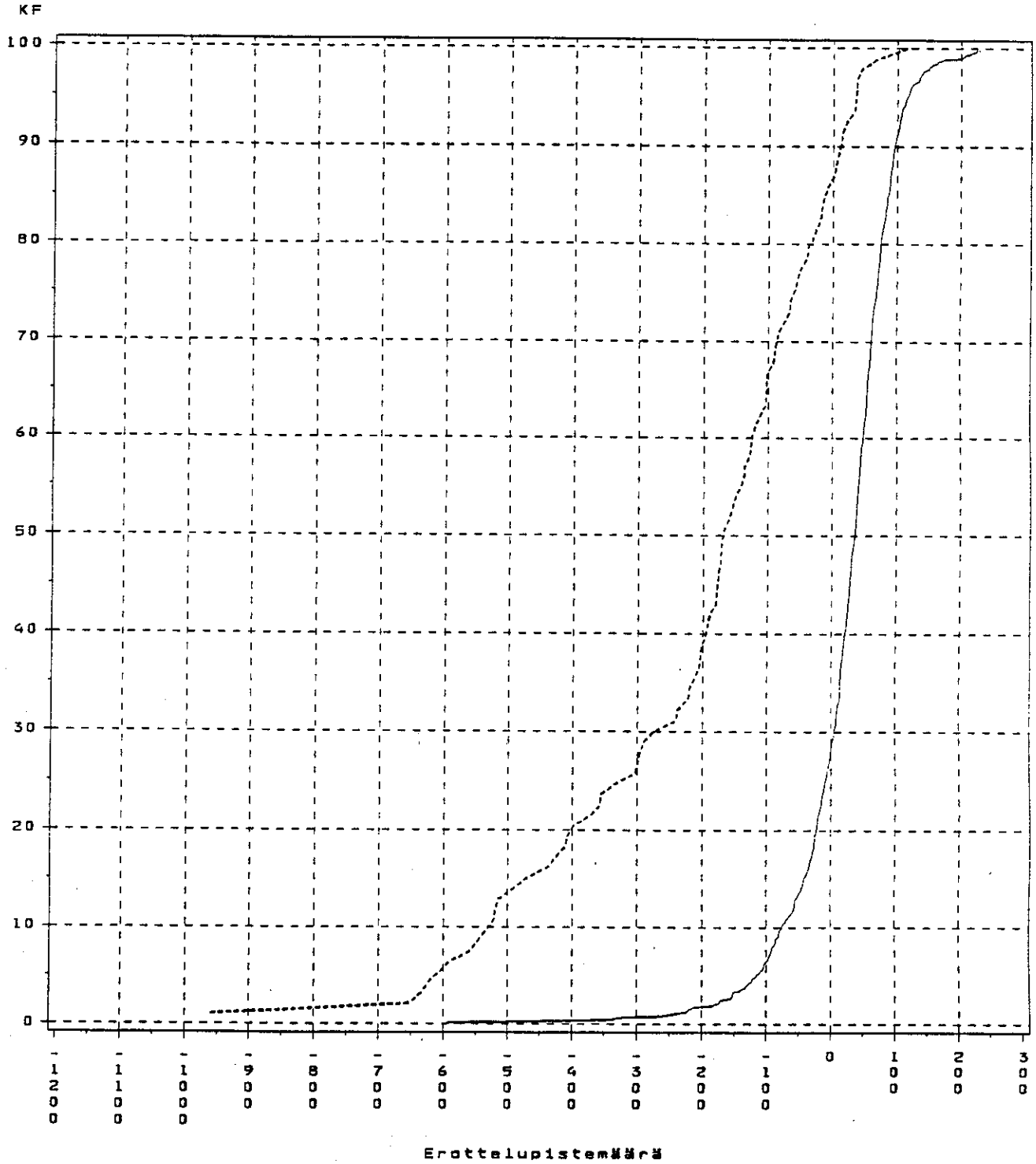
Muuttujajoukko käsittää siten ensimmäisen asteen muuttujia 20 kappaletta, toisen asteen muuttujia 210 ja kolmannen asteen muuttujia vain 92 eli yhteensä 322 muuttujaa. Kolmannen asteen muuttujien erilaisten kombinaatioiden joukko olisi ollut 1.540 eli kaiken kaikkiaan olisi ollut 1.770 mahdollista tariffitekijää, joten valituiksi tulleiden muuttujien joukosta puuttuu mahdollisia tärkeitä kolmannen asteen termejä.

Käytetään jälleen askeltavaa valintamenetelmää, mutta maksimaalisena askelten määränä käytetään nyt 20. Liitteessä 7 on saadut muuttujajoukot eri havaintojoukkojen ko'ille laskettuna. Liitteessä 8 testitunnuslukujen tulokset ja kanonisten erottelufunktioiden kertoimet. Kuvissa 21 - 25 on kanonisten erottelufunktioiden pistemäärien kertymäfunktiot. Erottelu ei sanottavammin parane kriittisillä alueilla eli konkurssiyritysten kertymäfunktion yläosassa ja toimivien alaosassa, parannus kohdistuu lähinnä kertymäfunktioiden keskiosiin.

KUVA 21

Kolmannen asteen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.0$

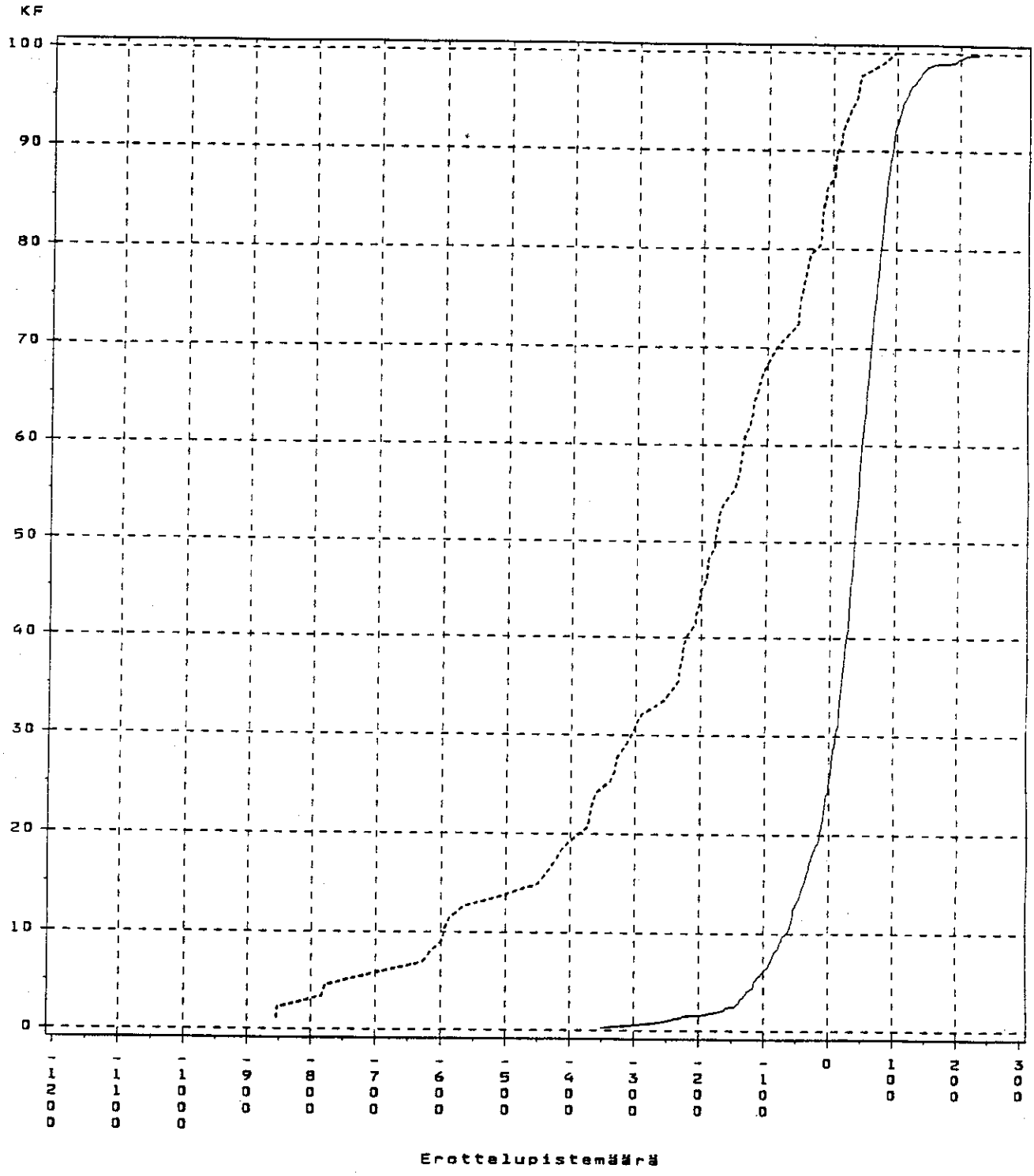


Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 22

Kolmannen asteen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.001$

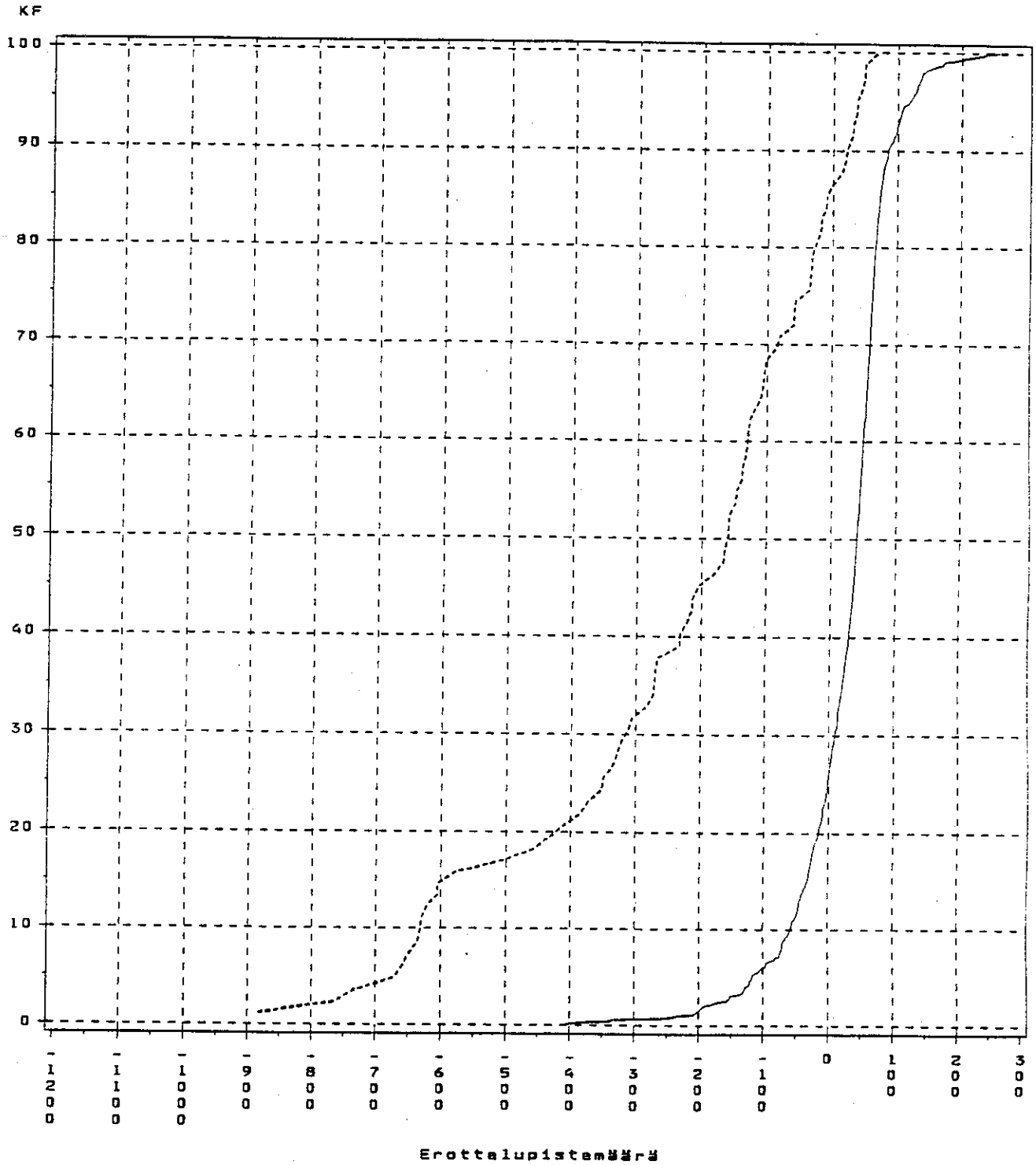


Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 23

Kolmannen asteen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.01$

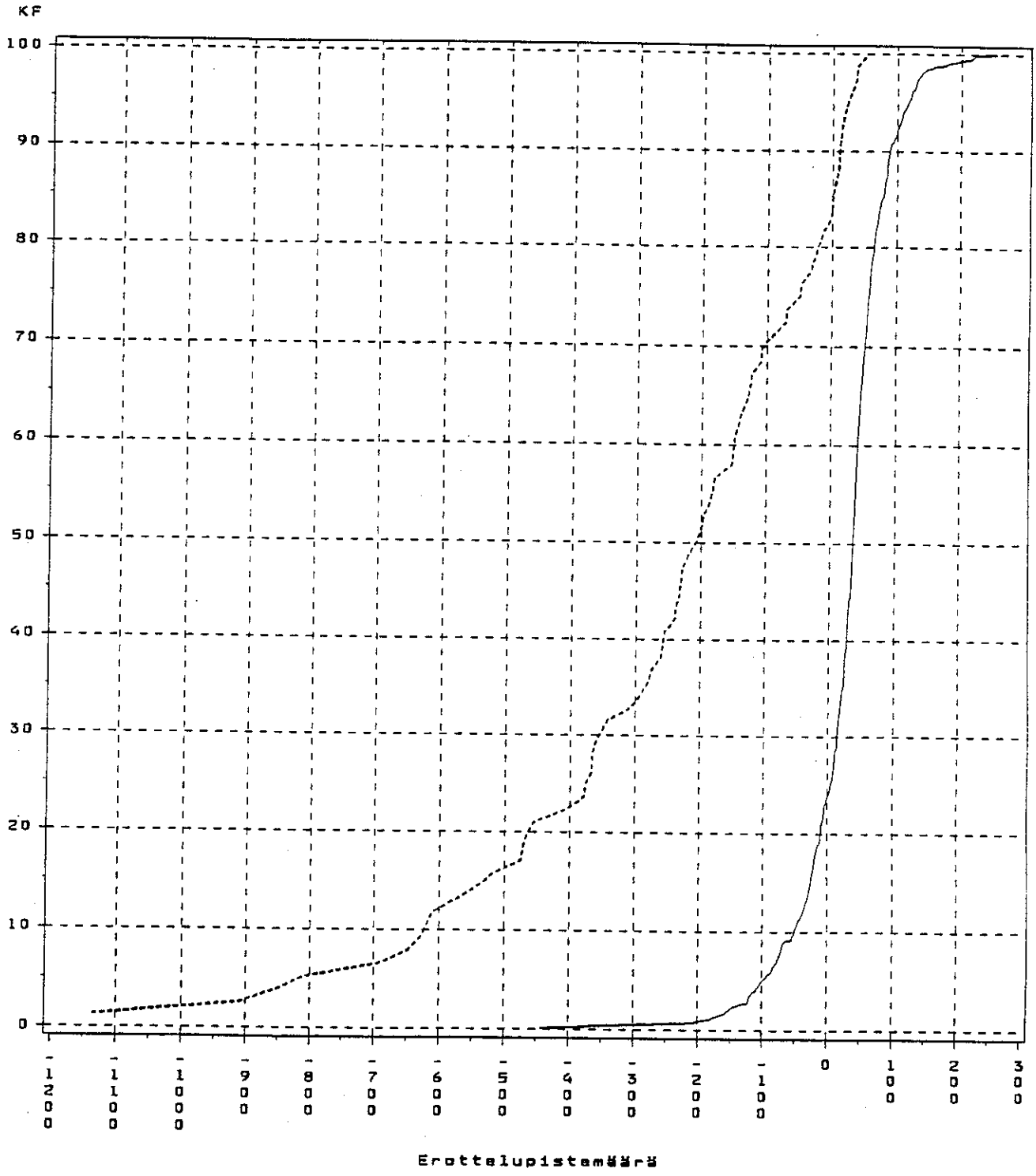


Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 24

Kolmannen asteen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.05$

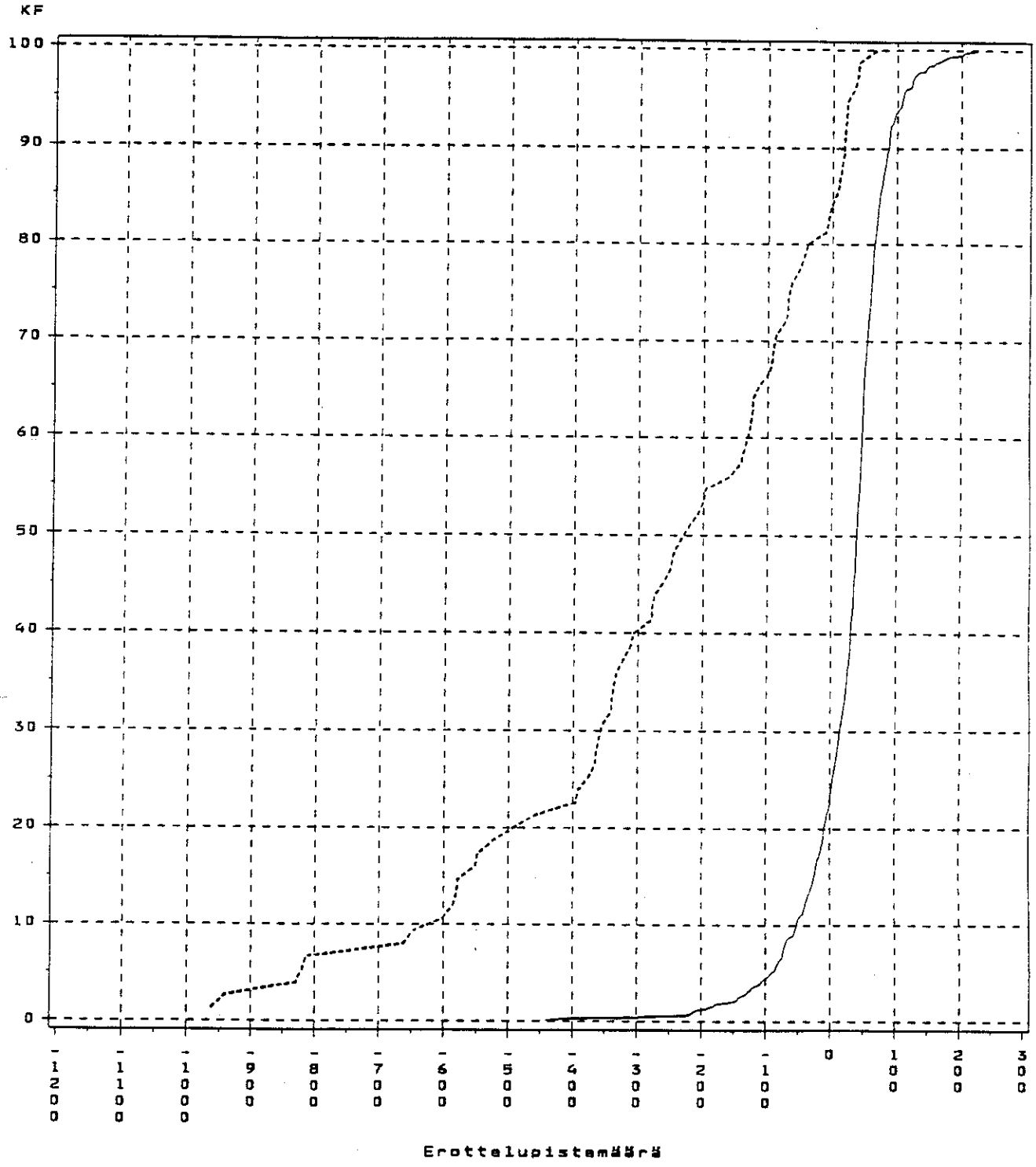


Yhtenäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

KUVA 25

Kolmannen asteen erottelufunktion pistemäärien kertymäfunktio,

$\alpha = 0.10$



Yhteinäinen viiva = toimivat yritykset
Katkoviiva = konkurssiyritykset

11.13 TARIFFIN MUODOSTUS

Erotteluanalyysin avulla voidaan luottovakuutustariffi muodostaa lähinnä kahdella eri tavalla eli joko perustuen Bayesiläiseen lähetymistapaan (3.2.2.1 ja 4.05) tai kanoniseen erottelufunktion tulokseen sovellettuna konkurssiyritysten erottelupistemäärien kertymäfunktioon.

Tarkastellaan tässä jälkimmäistä tapaa.

Koska

$$P2(h(X) \leq h(x)) = 1 - F2(h(x)) \text{ eli}$$

$$P2(h(X) > h(x)) = F2(h(x))$$

missä 1-F2 on konkurssiyritysten erottelupistemäärien kertymäfunktio. Approksimoidaan F2:n jakaumaa F:llä

Olkoon F2 muotoa

$$F(h(x)) = \min (b1*e^{b2*h(x)+b3} + b4 ; b1) \qquad 11.13.1$$

missä vakiot bi, i=1,...,4 , täyttävät käyttäjän asettamat ehdot. Näillä vakioilla huomioidaan ennen kaikkea varmuuslisä, joka pitää asettaa, koska tiedetään, että h(x) on ylioptimaalinen. Vakio b4:llä säädellään minimimaksua, b3 on ainoastaan h(x):n asteikon siirto, b2 nopeuttaa tai hidastaa exponenttilausekkeen kasvua, kun h(x) pienenee, sekä b1 on maksimimaksutason säätelijä. Oleellista on, että F:n lauseke on koko ajan empiirisen F2(h(x)):n yläpuolella.

Vakiot voidaan asettaa myös siten, että asiakasta valittaessa vakioiden arvot ovat eri kuin seuraavina vuosina määrättävän vakuutusmaksun kaavan vakiot.

Riskiä kuvavaan vakuutusmaksun kaava on siten,

$$\text{vakuutusmaksu} = F(h(x)) * \text{vakuutussumma}$$

12 ETK:N LUOTTOVAKUUTUSTARIFFI

Eläketurvakeskuksen luottovakuutustariffin uudistus-projekti käynnistyi syksyllä 1986. Syynä oli korvaustapausten voimakas lisääntyminen ja kun havaittiin lisäksi, ettei silloinen meno-tulovirta-analyysi tuota tarpeeksi aikaisessa vaiheessa hälytystä yrityksen vaikeuksista, oli uudistustarve ilmeinen. Projekti sai työnsä valmiiksi keväällä 1987 ja uudistettu tariffi otettiin käyttöön 1.7.1987.

Tässä tutkimuksessa on käytetty täsmälleen samaa tutkimusaineistoa. Seuraavissa kohdissa poikettiin silloisesta työskentelystä. Ensiksikin outlierit poistettiin vertailemalla kaksiuloitteisia hajontakuvioita. Tariffitekijät olivat samoja, mutta tulomuotoiset termit olivat vain tyyppiä xi*xj tai exi*exj. Mahalanobiksen etäisyystarkastelua ei suoritettu mahdollisten outliereiden paljastamiseksi tai multinormaalisuuden tarkistamiseksi. Yhtäsuurten kovarianssimatriisien tarkistus suori-

tettiin, jonka seurauksena edellä mainitut tulotermit muodostettiin. Tariffitekijöiden valinta noudatti myöskin askeltavaa valintamenetelmää ja tuloksena saatiin 22 merkityksellistä termiä, joista valittiin 14 ensimmäistä termiä tärkeysjärjestyksessä. Muodostettiin kanoninen erottelufunktio. Tariffin muodostamisessa käytettiin toimivien yritysten kanonisen erottelupistemäärien kertymäfunktiota siten, että Prihtin ((31)) Z-luvun toimivien yritysten jakauma oli sama kuin uuden tariffin niin sanottu U-jakauma asettamalla kummankin jakauman ala- ja yläkvartiilit samoiksi.
Tariffitekijät etsittiin teollisuusyritysten aineistosta, mutta kaupan ja kuljetusalan yritysten tariffitekijöiksi valittiin samat kuin teollisuudessa saatiin. Kertoimet tariffitekijöille johdettiin teollisuuden omista aineistoista. Kaupan ja kuljetusalan konkurssit yhdistettiin kaikkien konkurssien kanssa, mutta toimivina yrityksinä oli oman alan yritykset. Tälläisiin yhdistelyihin oli syynä konkurssiaineiston vähäisyys.
Rakennusalan yritysten erottelu ei ollut merkittävästi parempi kuin Prihtin meno-tulovirta-analyysin tulos, joten tälle alalle ei saatu uutta kaavaa muodostettua.

Lopputuloksena oli siten kolme uutta maksuluokkaa:

I) teollisuus
maa-, metsä- ja kalatalous
kaivos- ja muu kaivannaistoiminta
sähkö-, kaasu- ja vesihuolto

II) tukku- ja vähittäiskauppa
ravitsemis- ja majoitustoiminta
liike-elämää palveleva toiminta
yhteiskunnalliset ja henkilökohtaiset palvelukset

III) kuljetus, varastointi ja tietoliikenne

(IV) rakennusala, epäonnistui)

Aineistoa oli seuraavasti, suluissa oman ryhmän konkurssien osuus

Ryhmä		konkursseja	toimivia
I	(87)	87 kpl	826 kpl
II	(26)	129 kpl	728 kpl
III	(7)	129 kpl	241 kpl
IV	(9)	129 kpl	203 kpl

Riskiluku % on muodoltaan (11.13.1), kun

b1 = 100.0
b2 = -0.0177
b3 = -10.755
b4 = 0.32

sekä kanoninen erottelufunktio

$$h(y) = \sum_{i=1}^{14} a(i)*y(i) + C$$

missä

muuttujat y(i)

y(i)	Muuttuja
y(1)	x8
y(2)	x9
y(3)	x4
y(4)	x1*x8
y(5)	x1*x4
y(6)	x8*x9
y(7)	x7*x8
y(8)	x1*x1
y(9)	x1*x9
y(10)	x3
y(11)	x1
y(12)	ex4*ex7
y(13)	ex5*ey9
y(14)	ex2*ex7

sekä näitä vastaavat kertoimet a(i) ja vakio C.

Kerroin a(i)	Ryhmä I	II	III
a(1)	13.0	23.0	17.0
a(2)	-43.0	-53.0	-13.0
a(3)	- 9.3	- 3.70	- 2.70
a(4)	- 0.024	- 0.15	- 0.016
a(5)	0.10	0.0075	- 0.0029
a(6)	1.20	- 0.26	- 0.46
a(7)	- 0.39	- 0.021	- 0.013
a(8)	- 0.084	0.0094	0.035
a(9)	0.39	0.34	- 0.058
a(10)	1.70	0.74	2.10
a(11)	3.30	- 1.40	- 1.90
a(12)	- 0.21	- 0.0062	0.036
a(13)	- 1.40	- 0.016	- 0.79
a(14)	0.38	0.012	- 0.027
C	-160	-220	-370

Uuden tariffin toimivuudesta saanee jonkinlaisen käsityksen, kun vertaa sitä aikaisempaan meno-tulovirtalaskelman tariffiin eli Z-lukuun. Väärinluokitterurajana eli katkaisukohtana on käytetty erottelupistemäärää -405. ETK:lle voi tulla kahdenlaisia korvaustapauksia eli yrityksen

konkurssista johtuvia tai yritys laiminlyö eläkevakuutusyhtiölainansa takaisinmaksuvelvoitteensa, jollaisia tapauksia oheisessa taulukoissa kutsutaan maksuhaluttomiksi. Merkintä 'Ei tp.' tarkoitaa, ettei yritys ole toimittanut tilinpäätöstietojansa. Erot U-tariffin ja Z-tariffin välillä johtuu puuttuvien tietojen osalta siitä, että U-tariffissa tarvitaan kahden peräkkäisen vuoden tietoja, jotta erottelupistemäärä voitaisiin muodostaa.
Havaitaan ensiksin, että korvaustapahtuman lähestyessä tietojen toimittamuus lisääntyy ja toiseksi, että uusi tariffi on reagoinut tuleviin vaikeuksiin paremmin kuin Z-tariffi. Lisäksi on huomattava, että korvaustyypiltään maksuhaluton saattaa ajankuluessa muuttua konkurssiksi.

TAULUKKO 14.1 Korvatut 01.07.87-31.12.89 välisenä aikana, 4 vuotta ennen korvausta, ryhmät I-III

4 Vuotta ennen korvausta		Korvaustyyppi			
		Konkurssi		Maksuhaluton	
Ryhmä	Luokittelukyky	U-Tar. Kpl	Z-Tar. Kpl	U-Tar. Kpl	Z-Tar. Kpl
Teollisuus I	Ei tp.	12	7	13	13
	Väärin	25	32	12	15
	Oikein	26	24	14	11
Kauppa II	Ei tp.	5	5	8	8
	Väärin	15	15	15	19
	Oikein	16	16	9	5
Kuljetus III	Ei tp.	1	1	1	1
	Väärin	1	2	3	4
	Oikein	1	.	2	1
Yht.	Ei tp.	18	13	22	22
	Väärin	41	49	30	38
	Oikein	43	40	25	17
Oikeinluokittelu %		51.2%	44.9%	45.5%	30.9%

TAULUKKO 14.2 Korvatut 01.07.87-31.12.89 välisenä aikana, 3 vuotta ennen korvausta, ryhmät I-III

3 Vuotta ennen korvausta		Korvaustyyppi			
		Konkurssi		Maksuhaluton	
Ryhmä	Luokittelukyky	U-Tar. Kpl	Z-Tar. Kpl	U-Tar. Kpl	Z-Tar. Kpl
Teollisuus I	Ei tp.	8	8	15	11
	Väärin	26	28	8	15
	Oikein	29	27	16	13
Kauppa II	Ei tp.	5	5	10	10
	Väärin	11	17	12	16
	Oikein	20	14	10	6
Kuljetus III	Ei tp.	1	.	2	1
	Väärin	1	2	3	3
	Oikein	1	1	1	2
Yht.	Ei tp.	14	13	27	22
	Väärin	38	47	23	34
	Oikein	50	42	27	21
Oikein-luokittelu %		56.8%	47.2%	54.0%	38.2%

TAULUKKO 14.3 Korvatut 01.07.87-31.12.89 välisenä aikana, 2 vuotta ennen korvausta, ryhmät I-III

2 Vuotta ennen korvausta		Korvaustyyppi			
		Konkurssi		Maksuhaluton	
Ryhmä	Luokittelukyky	U-Tar. Kpl	Z-Tar. Kpl	U-Tar. Kpl	Z-Tar. Kpl
Teollisuus I	Ei tp.	13	11	12	12
	Väärin	17	20	11	11
	Oikein	33	32	16	16
Kauppa II	Ei tp.	8	5	12	12
	Väärin	8	11	4	10
	Oikein	20	20	16	10
Kuljetus III	Ei tp.	2	2	2	2
	Väärin	.	1	2	3
	Oikein	1	.	2	1
Yht.	Ei tp.	23	18	26	26
	Väärin	25	32	17	24
	Oikein	54	52	34	27
Oikein-luokittelu %		68.4%	61.9%	66.7%	52.9%

TAULUKKO 14.4 Korvatut 01.07.87-31.12.89 välisenä aikana, 1 vuosi ennen korvausta, ryhmät I-III

1 Vuosi ennen korvausta		Korvaustyyppi			
		Konkurssi		Maksuhaluton	
Ryhmä	Luokittelukyky	U-Tar. Kpl	Z-Tar. Kpl	U-Tar. Kpl	Z-Tar. Kpl
Teollisuus I	Ei tp.	28	26	23	23
	Väärin	6	9	4	7
	Oikein	29	28	12	9
Kauppa II	Ei tp.	20	20	21	21
	Väärin	3	4	3	5
	Oikein	13	12	8	6
Kuljetus III	Ei tp.	3	3	4	4
	Väärin	.	.	1	1
	Oikein	.	.	1	1
Yht.	Ei tp.	51	49	48	48
	Väärin	9	13	8	13
	Oikein	42	40	21	16
Oikeinluokittelu %		82.4%	75.5%	72.4%	55.2%

Toisena kuvasarjana on korvattujen riskiluku%:n jakaumat.

Merkitään U%, mikäli on käytetty uutta tariffia ja
Z%, mikäli on käytetty edellistä tariffia.

Merkitään lisäksi kirjaimilla K=konkurssi ja M=maksuhaluton.

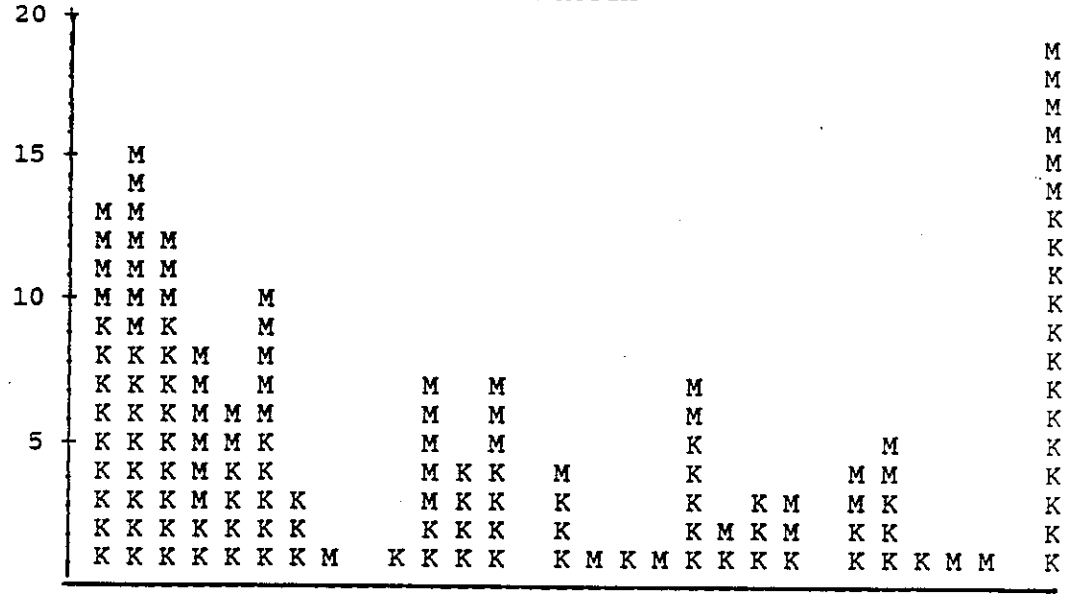
Vaaka-akselina käytetään riski %:a sekä pystyakselina on lukumäärät.

Kuvissa 26 - 29 on riski %:n kehitystä kuvaavat jakaumat korvauksen lähestyessä.

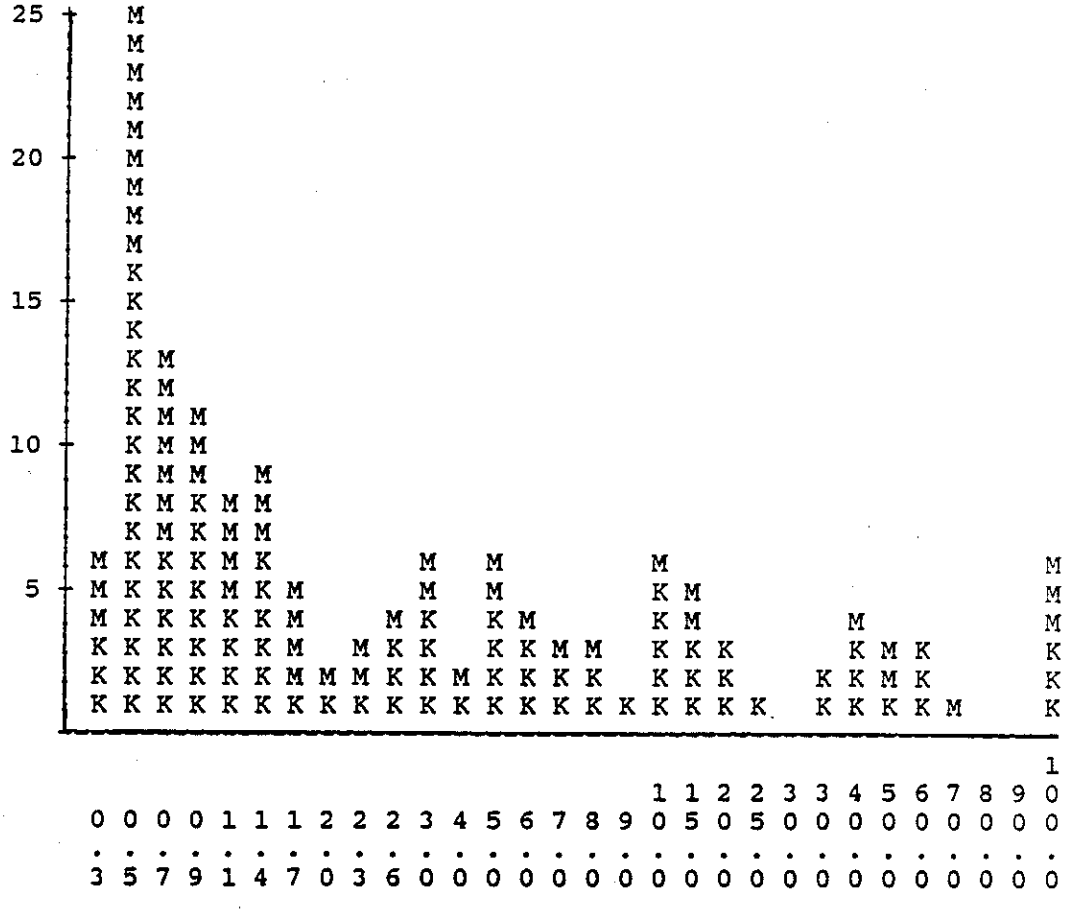
Kuvassa 30 on oikeinluokittelu %:n kehitys U- ja Z-tariffeilla.

KUVA 26

U% 4-VUOTTA ENNEN KORVAUSTA

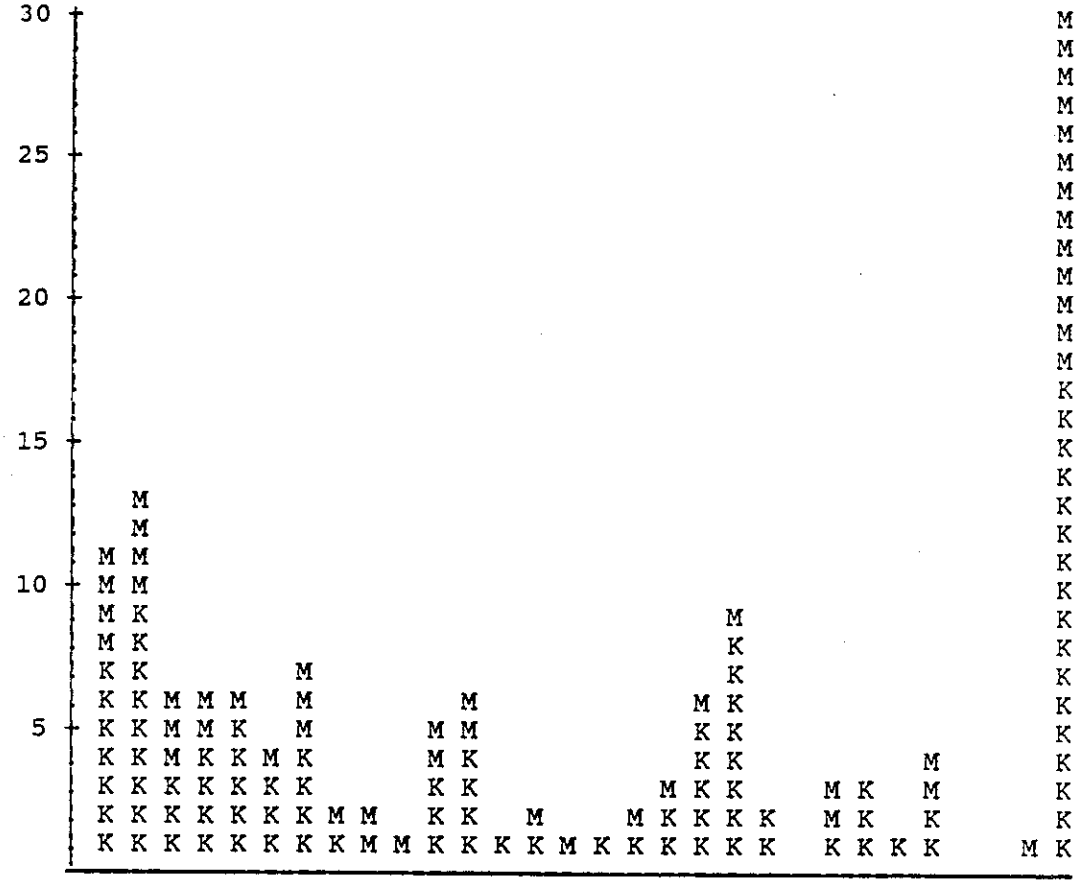


Z% 4-VUOTTA ENNEN KORVAUSTA

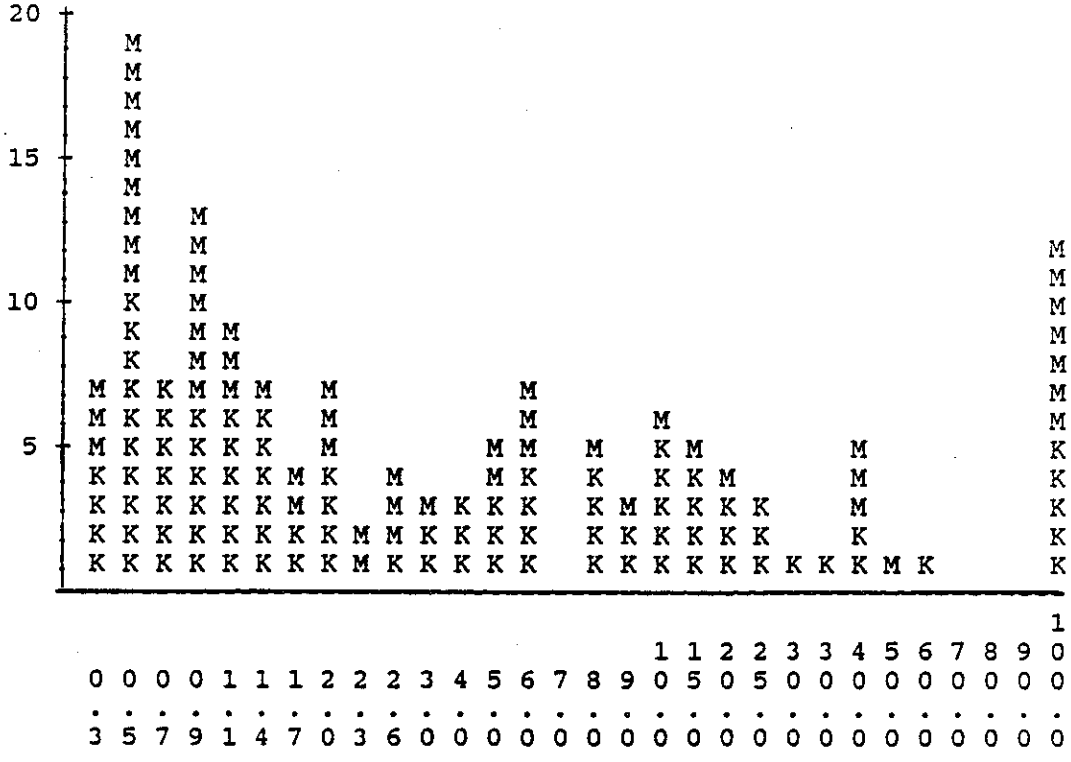


KUVA 27

U% 3-VUOTTA ENNEN KORVAUSTA

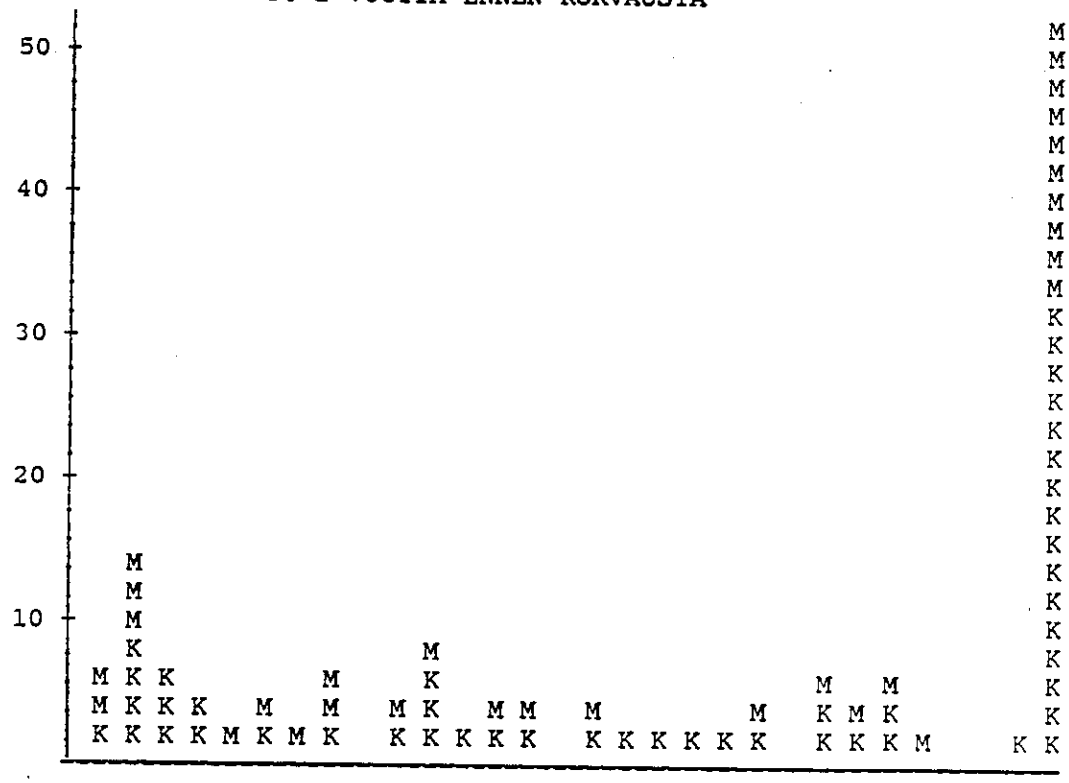


Z% 3-VUOTTA ENNEN KORVAUSTA

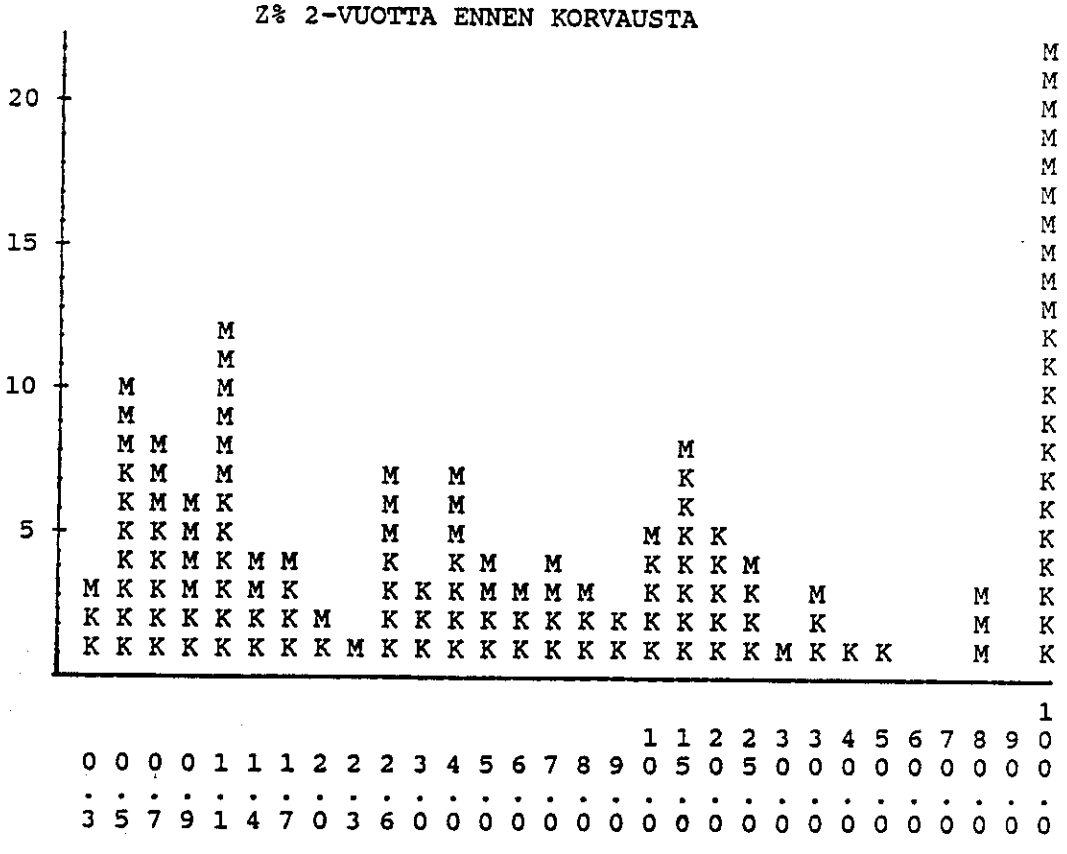


KUVA 28

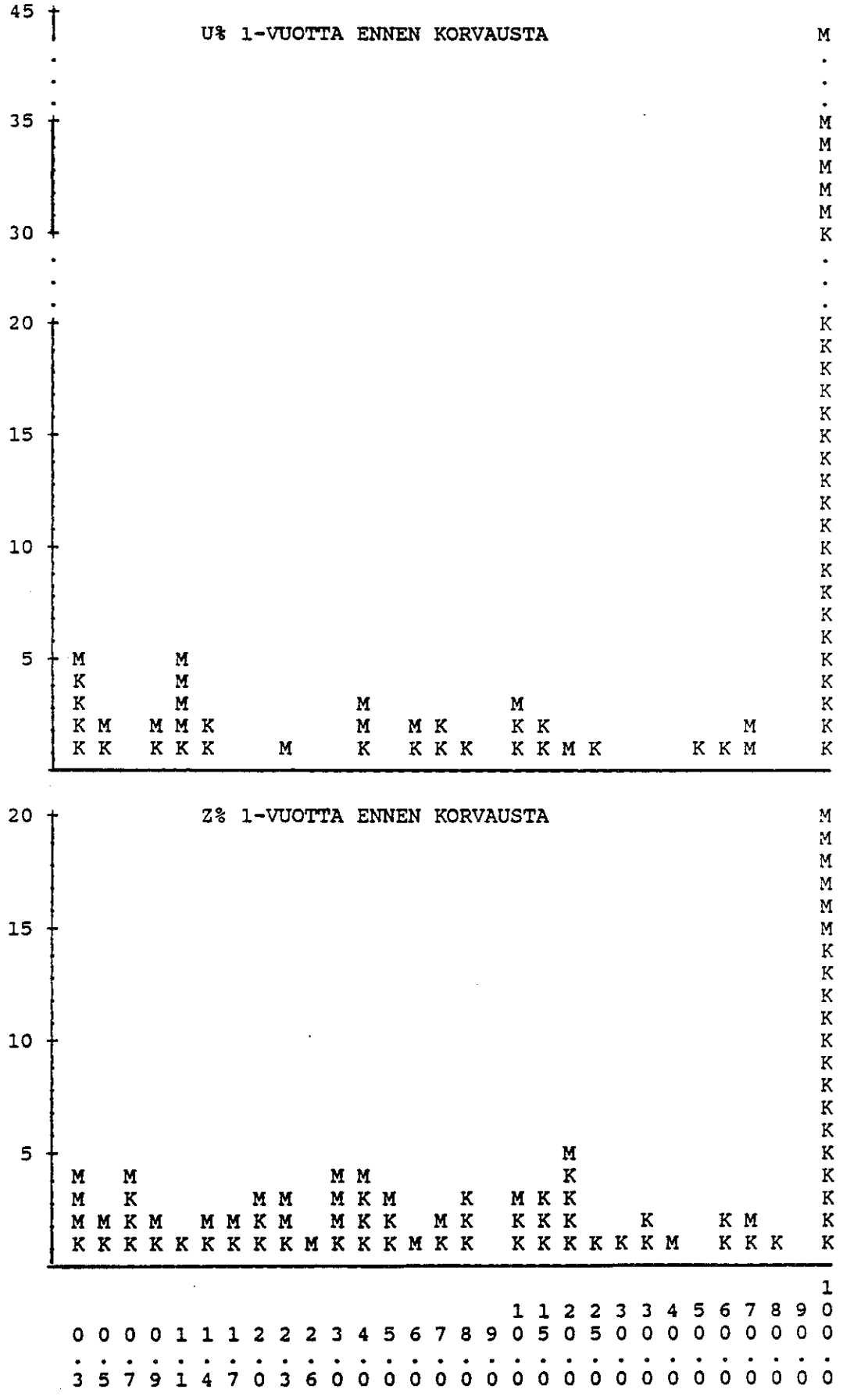
U% 2-VUOTTA ENNEN KORVAUSTA



Z% 2-VUOTTA ENNEN KORVAUSTA

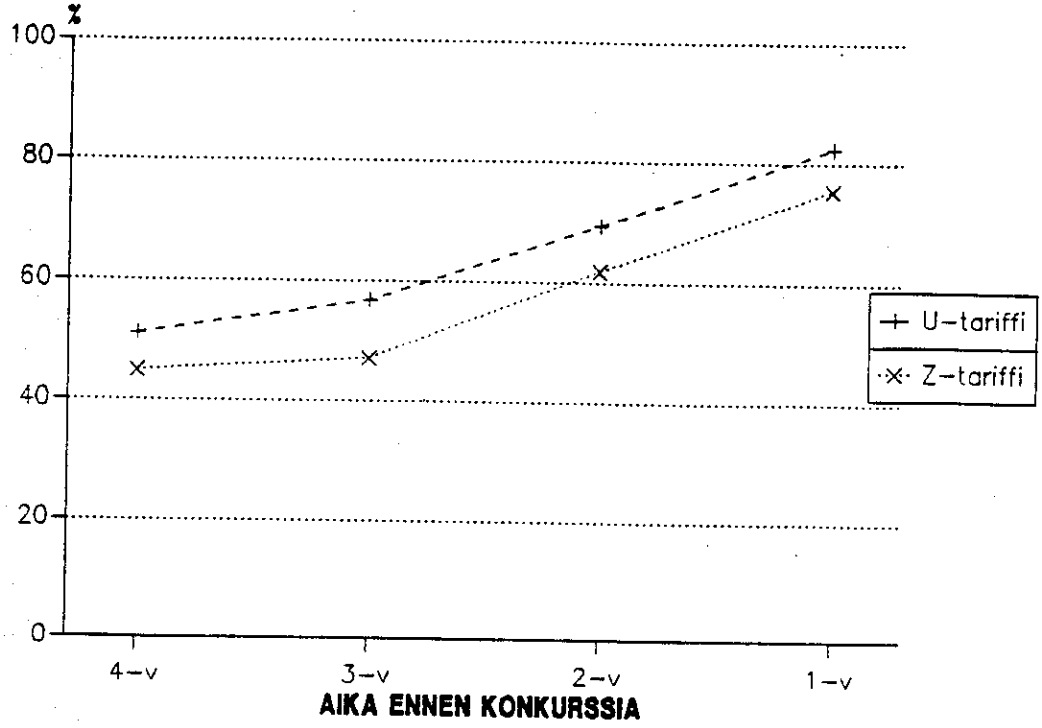


KUVA 29

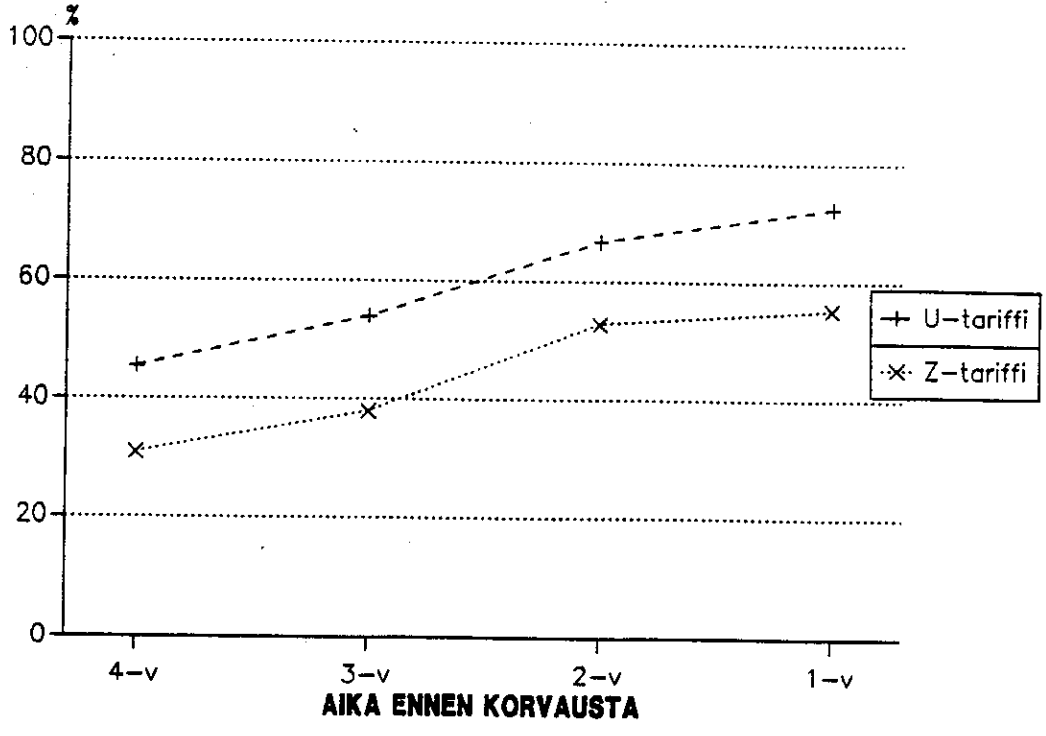

U% 1-VUOTTA ENNEN KORVAUSTA
45
35
30
20
15
10
5
Z% 1-VUOTTA ENNEN KORVAUSTA
20
15
10
5
0.3
0.5
0.7
0.9
1.1
1.4
1.7
2.0
2.3
2.6
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
15.0
20.0
25.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0

KUVA 30

OIKEINLUOKITTELU % : KONKURSSIT



OIKEINLUOKITTELU % : MAKSUHALUTTOMAT



Yhteenvetona voidaan todeta, että mikäli erotteluanalyysin perusoletukset eivät päde, saatiin tällä tutkimusaineistolla ulkonäöltään hyvinkin erilaiset mahdolliset tariffitekijät ja kanoniset erottelufunktiot.

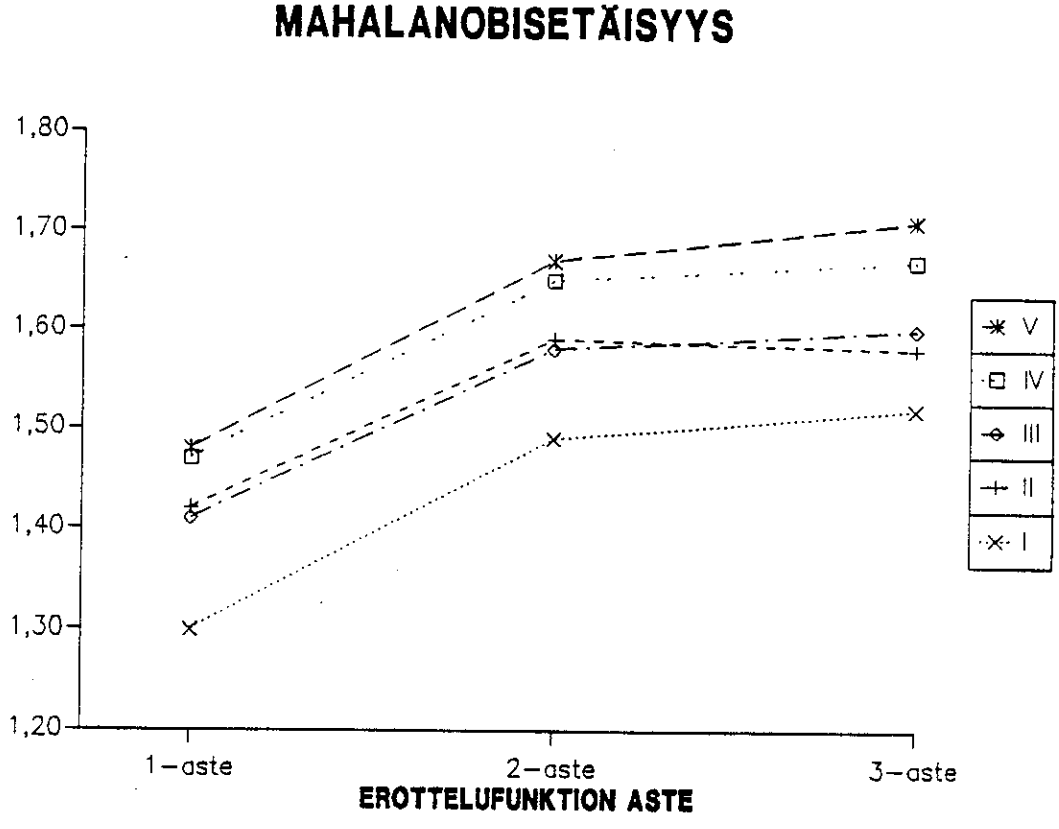
Toisen asteen eli neliömuotoiseen erottelufunktioon päädytään, mikäli kovarianssistruktuuri on erilainen, mutta multinormaalisuus pätee ja toisaalta Desu ((47)) toteaa, että kaikki muut jakaumat kuin multinormaaliset johtavat aina neliömuotoiseen erottelufunktioon. Mutta mikäli multinormaalisuus ja kovarianssioletukset eivät päde, voisi erottelufunktion muoto olla astetta korkeampi eli kolmatta astetta.
Tässä tutkimuksessa saadut kolmannen asteen mallit eivät kylläkään sanottavasti parantaneet erottelua, mutta se ei poissulje sitä mahdollisuutta etteikö niitä esiintyisi. On jäätävä odottamaan erotteluanalyysin teorian kehittymistä tältä osin.

Mahdollisina virhelähteinä saattoi esiintyä tämän tutkimuksen että jo myöskin luottovakuutusprojektin yhteydessä se, että yrityksen toimialakoodi on ollut virheellinen; luvut, joita ei voida koneellisesti tarkistaa esimerkiksi liikevaihto, on ollut virheellinen; aineisto ei edusta perusmuodossaan suomalaisia yrityksiä, koska edellinen tariffi on suosinut tietyn tyyppisiä ja karsinut toisenlaisia. Alussa mainittiin jo se, ettei ETK:n tietokantaan ole tallennettu kaikkia mahdollisesti tarvittavia tietoja joustokohtien huomioonottamiseksi.

Edellä osoitettiin, että määrittelyjoukkoon mukaan otettavilla yrityksillä on merkitystä erottelufunktion kannalta. On osoittautunut, että valittiinpa mikä edellä johdetuista kanonisista erottelufunktiostä tahansa, niin tärkein vaihe on tariffitason asettaminen oikealle paikalleen. Voi käydä niin, että johdettu erottelufunktio erottelee määrittelyjoukkonsa tapaukset erinomaisesti, mutta ei toimi edes kohtuullisesti uusiin tapauksiin. Käytettävän erottelufunktion kertoimia voidaan päivittää, esimerkiksi laskemalla edellä johdetuilla tavoilla vuosittain aina tuoreimmista tiedoista uudet kertoimet.

Lopuksi haluan kiittää Markku Tapperia Tietokonepalvelu Oy:stä antamistaan erinomaisista neuvoista ja ohjeista tietokoneajojen suorittamisissa ja SAS-tilastokielen vaatimasta erityisasiantuntemuksesta. Myöskin esitän mitä suurimmat kiitokseni Eläketurvakeskukselle, esimiehilleni, työtovereilleni ja vaimolleni, sillä ilman heidän neuvojaan, kannustustaan ja ymmärrystään tämä tutkimus olisi valmistunut jo kaksi vuotta sitten.

KUVA 31


MAHALANOBISETÄISYYS
1,80
1,70
1,60
1,50
1,40
1,30
1,20
1–aste
2–aste
3–aste
EROTTELUFUNKTION ASTE
V
IV
III
II
I

LIITE 1

MAHDOLLISET TARIFFITEKIJÄT

x1 =100*rahoitusomaisuus /liikevaihto

x2 =100*(vaihto-omaisuus - varastovaraus)/liikevaihto

x3 =100*(käyttöomaisuus + sijoitusomaisuus + arvostuerät)/lvaihto

x4 =100*(lyhytaikaiset velat - ennakkomaksut)/liikevaihto

x5 =100*pitkäaikaiset velat /liikevaihto

x6 =100*(arvostuserät+varaukset+oma pääoma)/liikevaihto

x7 =100*(käyttökate-/+varastovarauksen muutos)/liikevaihto

x8 =100*(poistot-/+muiden varausten muutos+tilivuoden tulos)/lvaihto

x9 =100*korkokulut /liikevaihto

x10=100*verot /liikevaihto

Varausten vähennys etumerkki - ja lisäys etumerkki +.
Muutosta kuvaavat tunnusluvut, missä viimeinen vuosi=0 ja edellinen=-1

ex1 =100*[rahoitusomaisuus(0) - rahoitusomaisuus(-1)] /liikevaihto(0)

ex2 =100*[vaihto-omaisuus(0) - vaihto-omaisuus(-1)
- varastovaraus(0) + varastovaraus(-1)]/liikevaihto(0)

ex3 =100*[käyttöomaisuus(0) - käyttöomaisuus(-1)
+ sijoitusomaisuus(0) - sijoitusomaisuus(-1)
+ arvostuserät(0) - arvostuserät(-1)] /liikevaihto(0)

ex4 =100*[lyhytaikaiset velat(0) - lyhytaikaiset velat(-1)
- ennakkomaksut(0) + ennakkomaksut(-1)] /liikevaihto(0)

ex5 =100*[pitkäaikaiset velat(0)-pitkäaikaiset velat(-1)] /lvaihto(0)

ex6 =100*[arvotuserät(0) - arvostuserät(-1)
+ varaukset(0) - varaukset(-1)
+ oma pääoma(0) - oma pääoma(-1)] /liikevaihto(0)

ex7 =100*[käyttökate(0) - käyttökate(-1)
-/+ varastovar.muutos(0) +/- varastovar. muutos(-1)]
/liikevaihto(0)

ex8 =100*[poistot(0) - poistot(-1)
-/+ muiden varausten muutos(0) +/- muiden var. muutos(-1)
+ tilivuoden tulos(0) - tilivuoden tulos(-1)]
/liikevaihto(0)

ex9 =100*[korkokulut(0)-korkokulut(-1)]/liikevaihto(0)

ex10=100*[verot(0)-verot(-1)]/liikevaihto(0)

LIITE 3
(3.1)

TILINPÄÄTÖKSEN TÄRKEIMPIÄ JOUSTOKOHTIA

1) Vaihto-omaisuuden jaksotus: KPL § 16 ja EVL § 28.

Verovelvollinen saa niin sanotun epäkuranttivähennyksen jälkeen kirjata verovuoden kuluksi enintään X% vaihto-omaisuuden tilinpäätösarvosta. Tämä X% on vaihdellut ja tullee vaihtelemaan erivuosina siten, että se oli 50% ennen vuotta 1986, vuonna 1989 se oli 35 % ja vuonna 1990 30% ilman toimintavarausta ja mikäli toimintavarausta käytetään vastaavat luvut ovat 30% ja 25%.
Mikäli vaihto-omaisuutena on arvopapereita on tämä niin sanottu varaston aliarvostus vuonna 1989 20% ja 1990 10% edellyttäen, että hallituksen esitys toteutuu.

2) Käyttöomaisuusmenojen jaksotus: KPL § 16 , EVL § 30-45, Verohuojennuslainsäädäntä, Tutkimusmenojen jaksotus EVL § 25 sekä Arvonkorotukset KPL § 18.

Käyttöomaisuuden poisto- ja jaksotussäännöissä on säädetty käyttöomaisuusryhmittäiset enimmäispoistot. Käyttöomaisuutta voidaan jaksottaa valiten poiston väliltä 0% --> kysessä olevan käyttöomaisuusryhmän maksimipoisto. Käyttöomaisuuden jaksotusongelma voidaan nähdä joko poisto- tai arvostusongelmana eli onko käytettävä alkuperäistä vai nykyistä hankintahintaa tai markkinahintaa. Tämä sama ongelma koskee myös vaihto-omaisuuden jaksotusta.
Kirjanpitoasetusta muutettiin vuonna 1985 siten, että sen 7 §:ssä annetaan mahdollisuus esittää tuloslaskelmassa suunnitelman mukaiset poistot. Jos näin menetellään, tuloslaskelmassa voidaan esittää tuotoista vähennettyjen ja suunnitelman mukaisten poistojen erotus omana eränään. Vastaavasti taseeseen voidaan merkitä käyttöomaisuuden hankintamenojen ja muiden pitkävaikutteisten menojen poistetun osan ja suunnitelman mukaan kertyneiden poistojen erotus erikseen käyttöomaisuuteen ja pitkävaikutteisten menojen ryhmään sekä vastattavien varausten ryhmään.

3) Kurssitappioiden jaksotus: KPL § 15 ja EVL § 26.

Liiketoiminnasta johtuneiden velkojen indeksi- ja kurssitappiot voidaan kirjata vaihtoehtoisesti joko sen vuoden kuluksi, jonka aikana ulkomaanrahan kurssi tai hinta- tai kustannusindeksi tai muu vertailuperuste on muuttunut tai tappiota vastaava summa on suoritettu.

4) Arvonkorotukset: KPL § 18.

Yrityksen on mahdollisuus suorittaa käyttöomaisuuden arvonkorotus, jos käyttöomaisuuden todennäköinen luovutushinta on pysyvästi hankitamenoa suurempi. Arvonkorotuksen suuruus voi olla enintään todennäköisen luovutushinnan ja poistamattoman hankintamenon erotuksen suuruinen.

1)-7) Numerointi viittaa tilinpäätöksen tärkeimpiin joustokohtiin

TULOSLASKELMA		aika(0)			aika(-1)		
Myyntituotot			.			.	
Myynninoikaisuerät:							
Alennukset		.			.		
...		.			.		
...		.			.		
Muut oikaisuerät		.	.		.	.	
Liikevaihto			XXX			XXX	
Muuttuvat kulut:	3)						
Aineet ja tarvikkeet		.			.		
...		.			.		
...		.			.		
Varastojen muutos	1)	- .	.		- .	.	
Myyntikate			.			.	
Kiinteät kulut:							
Palkat		.			.		
...		.			.		
...		.			.		
Muut kiinteät kulut	7)	.	.		.	.	
Käyttökate			XXX	x7		XXX	ex7
Poistot	2)						
Rakennuksista ja rakennelmista		YY			YYY		
Koneista ja kalustosta		YY			YY		
...		.			.		
...		.			.		
Muista pitkävaik:sta menoista		.	XX	x8	.	XX	ex8
Liikevoitto			XX			XX	
Muut tuotot ja kulut:	3)						
Korkotuotot		.			.		
...							
...	6)	.			.		
Muut kulut		YY	XX		YY	XX	
Varausten muutos	5)		XX	x8		XX	ex8
Varastovarauksen muutos	5)		YY	x7		YY	ex7
Korkokulut			XX	x9		XX	ex9
Välittömät verot	6)		XX	x10		XX	ex10
Tilikauden tulos			XX	x8		XX	ex8

*

LIITE 2
(2.1)

MAHDOLLISET TARIFFITEKIJÄT TASE- JA TULOSLASKELMASSA

ETK:n tietokantaan tallennettu pakollinen tieto = XX ja tallennttutieto saattaa puuttua = YY. Numerointi 1)-7) viittaa tilinpäätöksen tärkeimpiin joustokohtiin

VASTAAVAA		Tase(0)			Tase(-1)		
Rahoitusomaisuus							
Rahat ja pankkisaamiset		.			.		
...		.			.		
Muut rahoitusvarat		.	XX	x1	.	XX	ex1
Vaihto-omaisuus	1)						
Aineet ja tarvikkeet		.			.		
Valmisteet/tavarat		.	XX	x2	.	XX	ex2
Käyttöomaisuus ja muut pitkävaikutteiset menot	2),4)						
Maa- ja vesialueet		.			.		
Rakennukset ja rakennelmat		YY			YYY		
Koneet ja kalusto		YY			YYY		
Muut aineelliset hyödykkeet		.			.		
...		.			.		
Muut pitkävaikutteiset menot		.	XX	x3	.	XX	ex3
Sijoitusomaisuus	2),3)	.	XX	x3	.	XX	ex3
Arvostuserät	3)	.	XX	x3	.	XX	ex3
			XX			XX	
VASTATTAVAA							
Vieras pääoma							
Lyhytaikainen							
Ostovelat		.			.		
...		.			.		
Ennakkomaksut		YY			YY		
Muut lyhytaiset velat		.	XX	x4	.	XX	ex4
Pitkäainen							
Lainat rahoituslaitoksilta		.			.		
Eläkelainat		YY			XX		
Muut pitkäaikaiset lainat		.	XX	x5	.	XX	ex5
Arvostuserät	4)	.	XX	x6	.	XX	ex6
Varaukset	5)						
...		XX			XX		
Varastovaraus		YY	XX	x6	YY	XX	ex6
Oma pääoma	6)						
Sidottu							
Osakepääoma		YY			YY		
Vapaa							
Voitot aikaisemmilta vuosilta		YY			YY		
Tilikauden tulos		XX	XX	x6	XX	XX	ex6
			XX			XX	

5) Varaukset: erityisvaraukset KPL § 17, EVL § 46,47,49, Jälleenhankintavaraus EVL § 30,43 ja Verohuojennuslaisäädäntö.

Luottotappivarauksena saavat raha-, vakuutus- ja eläkelaitokset vähentää enintään 0.6 % saamisten yhteismäärästä, kuitenkin niin, ettei purkamattomien luottotappiovarausten yhteismäärä saa ylittää 6 % saamisten yhteismäärästä eikä vakuutusmaksusaamisista 3 %. Muiden yritysten osalta on rajoituksena vain luottotappiovarausten yhteismäärä, joka ei saa ylittää 3 % liikesaamisten yhteismäärästä. Varauksen enimmäismäärä on vuonna 1989 5 % raha- ja vastaaville yrityksille sekä vuonna 1990 muille yrityksille se tulee alenemaan 2 % liikesaamisten yhteismäärästä. Takuuvarauksen, jota voi käyttää tuotteistaan ilmenevistä virheistä vastuussa olevat yritykset, esimerkiksi rakennus- , laivanrakennus- ja metalliteollisuutta harjoittava yritys, määrä voi olla 3 % urakkahinnasta. Sitovin kirjallisin sopimuksin kiinteään hintaan tilatun vaihto-omaisuuden osalta voidaan tehdä hinnanlaskuvaraus, mikäli vielä toimittamattomien tavaroiden hinta tilinpäätöspäivänä on vähintään 10 % sopimuksen mukaista hintaa alhaisempi. Toimintavaraus voi olla enintään 30 % 12 kuukauden tilinpäätöstä edeltäneen ajan ennakonpidätyksen alaisista palkoista. Toimintavaraus ja varastovaraus (kohta 1) ovat sidoksissa toisiinsa. Investointivaraus saa päässääntöisesti olla enintään 50 % tilivuoden voitosta verojen jälkeen ennen vuotta 1989 ja vuonna 1989 se on 20 %. Investointivaraksen määrästä puolet talletetaan Suomen Pankkiin investointitalletuksena. Investointivarauksella katettavista investoinneista voidaan tehdä varausten suuruiset poistot jo siinä tilinpäätöksessä, jossa investointivaraus muodostetaan.

6) Ohi tulostilin suoritetut kirjaukset:
Tulojen kirjaaminen ohi tulostilin omaan pääomaan,
Verojen kirjaaminen ohi tulostilin omasta pääomasta.

Nykykäytännön mukaan voi yritys kirjata veroja ja verovapaita tuloja tulostilin ohi suoraan omaan pääomaan.

7) Siirto eläkesäätiön vastuuvajauksen kattamiseksi: EVL § 48.

Tämän joustokohdan käyttö edellyttää olemassa olevaa työnantajan perustamaa työntekijöiden eläkelain tarkoittamaa eläkesäätiötä. Siirrot eläkesäätiölle ovat sen vuoden kulua, jona siirto on tehty tai jona eläkevastuuta on katettu. Eläkesäätiön vastuuvajauksella tarkoitetaan sitä määrää, jolla eläkkeiden ja mahdollisten muiden etuuksien vakuutusteknisesti lasketut vastaiset maksuvelvollisuudet eli eläkevastuu ylittää eläkesäätiön netto-omaisuuden.

LIITE 3
(3.1)

TILINPÄÄTÖKSEN TÄRKEIMPIÄ JOUSTOKOHTIA

1) Vaihto-omaisuuden jaksotus: KPL § 16 ja EVL § 28.

Verovelvollinen saa niin sanotun epäkuranttivähennyksen jälkeen kirjata verovuoden kuluksi enintään X% vaihto-omaisuuden tilinpäätösarvosta. Tämä X% on vaihdellut ja tullee vaihtelemaan erivuosina siten, että se oli 50% ennen vuotta 1986, vuonna 1989 se oli 35 % ja vuonna 1990 30% ilman toimintavarausta ja mikäli toimintavarausta käytetään vastaavat luvut ovat 30% ja 25%.
Mikäli vaihto-omaisuutena on arvopapereita on tämä niin sanottu varaston aliarvostus vuonna 1989 20% ja 1990 10% edellyttäen, että hallituksen esitys toteutuu.

2) Käyttöomaisuusmenojen jaksotus: KPL § 16 , EVL § 30-45, Verohuojennuslainsäädäntä, Tutkimusmenojen jaksotus EVL § 25 sekä Arvonkorotukset KPL § 18.

Käyttöomaisuuden poisto- ja jaksotussäännöissä on säädetty käyttöomaisuusryhmittäiset enimmäispoistot. Käyttöomaisuutta voidaan jaksottaa valiten poiston väliltä 0% --> kysessä olevan käyttöomaisuusryhmän maksimipoisto. Käyttöomaisuuden jaksotusongelma voidaan nähdä joko poisto- tai arvostusongelmana eli onko käytettävä alkuperäistä vai nykyistä hankintahintaa tai markkinahintaa. Tämä sama ongelma koskee myös vaihto-omaisuuden jaksotusta.
Kirjanpitoasetusta muutettiin vuonna 1985 siten, että sen 7 §:ssä annetaan mahdollisuus esittää tuloslaskelmassa suunnitelman mukaiset poistot. Jos näin menetellään, tuloslaskelmassa voidaan esittää tuotoista vähennettyjen ja suunnitelman mukaisten poistojen erotus omana eränään. Vastaavasti taseeseen voidaan merkitä käyttöomaisuuden hankintamenojen ja muiden pitkävaikutteisten menojen poistetun osan ja suunnitelman mukaan kertyneiden poistojen erotus erikseen käyttöomaisuuteen ja pitkävaikutteisten menojen ryhmään sekä vastattavien varausten ryhmään.

3) Kurssitappioiden jaksotus: KPL § 15 ja EVL § 26.

Liiketoiminnasta johtuneiden velkojen indeksi- ja kurssitappiot voidaan kirjata vaihtoehtoisesti joko sen vuoden kuluksi, jonka aikana ulkomaanrahan kurssi tai hinta- tai kustannusindeksi tai muu vertailuperuste on muuttunut tai tappiota vastaava summa on suoritettu.

4) Arvonkorotukset: KPL § 18.

Yrityksen on mahdollisuus suorittaa käyttöomaisuuden arvonkorotus, jos käyttöomaisuuden todennäköinen luovutushinta on pysyvästi hankitamenoa suurempi. Arvonkorotuksen suuruus voi olla enintään todennäköisen luovutushinnan ja poistamattoman hankintamenon erotuksen suuruinen.

Konkurssit

$$S2=\begin{bmatrix} 490.5 & -125.0 & 56.4 & 1.6 & -104.6 \\ -125.0 & 138.8 & -22.6 & 1.9 & 89.8 \\ 56.4 & -22.6 & 16.7 & 0.2 & -9.2 \\ 1.6 & 1.9 & 0.2 & 0.9 & 0.9 \\ -104.6 & 89.8 & -9.2 & 0.9 & 374.9 \end{bmatrix}$$

IV) $\alpha = 0.05$

Toimivat

$$S1=\begin{bmatrix} 139.9 & -11.0 & 14.3 & -0.5 & 0.4 & 40.2 & 3.7 \\ -11.0 & 49.7 & -2.1 & 0.1 & -1.8 & 1.7 & -1.2 \\ 14.3 & -2.1 & 5.5 & -0.0 & 1.9 & 28.9 & 1.1 \\ -0.5 & 0.1 & -0.0 & 0.8 & -0.2 & -0.2 & -0.0 \\ 0.4 & -1.8 & 1.9 & -0.2 & 68.2 & 45.5 & 3.8 \\ 40.2 & 1.7 & 28.9 & -0.2 & 45.5 & 237.1 & 5.1 \\ 3.7 & -1.2 & 1.1 & -0.0 & 3.8 & 5.1 & 1.2 \end{bmatrix}$$

Konkurssit

$$S2=\begin{bmatrix} 462.7 & -91.3 & 45.2 & -0.5 & 36.9 & 79.8 & 10.8 \\ -91.3 & 104.6 & -13.4 & 1.8 & -14.4 & -46.9 & -3.5 \\ 45.2 & -13.4 & 12.5 & 0.1 & 14.3 & 55.3 & 4.1 \\ -0.5 & 1.8 & 0.1 & 0.9 & -0.1 & 1.8 & -0.1 \\ 36.9 & -14.4 & 14.3 & -0.1 & 130.1 & 149.4 & 14.1 \\ 79.8 & -46.9 & 55.3 & 1.8 & 149.4 & 548.4 & 15.1 \\ 10.8 & -3.5 & 4.1 & -0.1 & 14.1 & 15.1 & 6.0 \end{bmatrix}$$

V) $\alpha = 0.10$

Toimivat

$$S1=\begin{bmatrix} 137.8 & -8.2 & 13.8 & -0.7 & 3.7 & 2.2 & 3.5 & 38.8 \\ -8.2 & 46.1 & -1.6 & 0.1 & 40.5 & -2.7 & -1.0 & 3.1 \\ 13.8 & -1.6 & 5.3 & -0.1 & -2.3 & 2.0 & 1.0 & 27.3 \\ -0.7 & 0.1 & -0.1 & 0.8 & -0.1 & -0.2 & -0.0 & -0.4 \\ 3.7 & 40.5 & -2.3 & -0.1 & 253.0 & -6.6 & -2.5 & 7.2 \\ 2.2 & -2.7 & 2.0 & -0.2 & -6.6 & 63.1 & 3.5 & 44.1 \\ 3.5 & -1.0 & 1.0 & -0.0 & -2.5 & 3.5 & 1.1 & 4.6 \\ 38.8 & 3.1 & 27.3 & -0.4 & 7.2 & 44.1 & 4.6 & 228.7 \end{bmatrix}$$

Konkurssit

$$S2=\begin{bmatrix} 464.8 & -91.5 & 45.4 & -0.5 & -66.6 & 38.2 & 11.1 & 83.5 \\ -91.5 & 105.8 & -13.5 & 1.8 & 40.3 & -14.8 & -3.6 & -48.3 \\ 45.4 & -13.5 & 12.6 & 0.1 & 4.5 & 14.6 & 4.2 & 56.3 \\ -0.5 & 1.8 & 0.1 & 0.9 & 3.7 & -0.1 & -0.1 & 1.8 \\ -66.6 & 40.3 & 4.5 & 3.7 & 273.5 & 5.1 & 0.3 & 133.9 \\ 38.2 & -14.8 & 14.6 & -0.1 & 5.1 & 131.7 & 14.2 & 150.8 \\ 11.1 & -3.6 & 4.2 & -0.1 & 0.3 & 14.2 & 6.0 & 15.2 \\ 83.5 & -48.3 & 56.3 & 1.8 & 133.9 & 150.8 & 15.2 & 554.1 \end{bmatrix}$$

I) $\alpha = 0.0$

Toimivat

$$S1 = \begin{bmatrix} 447.4 & 20.5 & 43.6 & 1.5 & -8.8 & 361.3 & 15.0 & -26.6 & 20.0 \\ 20.5 & 156.7 & 3.7 & 126.9 & -2.6 & 160.6 & -0.5 & 24.1 & -2.9 \\ 43.6 & 3.7 & 12.5 & -4.4 & -1.0 & 33.7 & 2.3 & -8.4 & -2.8 \\ 1.5 & 126.9 & -4.4 & 237.0 & -1.0 & 99.5 & 3.2 & 49.2 & -0.6 \\ -8.8 & -2.6 & -1.0 & -1.0 & 1.7 & -8.9 & -0.2 & 1.5 & 0.2 \\ 361.3 & 160.6 & 33.7 & 99.5 & -8.9 & 749.6 & 9.9 & -28.6 & -26.4 \\ 15.0 & -0.5 & 2.3 & 3.2 & -0.2 & 9.9 & 3.1 & -0.2 & 1.6 \\ -26.6 & 24.1 & -8.4 & 49.2 & 1.5 & -28.6 & -0.2 & 78.4 & 21.3 \\ 20.0 & -2.9 & -2.8 & -0.6 & 0.2 & -26.4 & 1.5 & 21.3 & 65.1 \end{bmatrix}$$

Konkurssit

$$S2 = \begin{bmatrix} 586.1 & -132.8 & 72.3 & -65.9 & -1.2 & 127.5 & 20.2 & -48.8 & 62.8 \\ -132.8 & 168.9 & -25.6 & 148.8 & -0.1 & -45.9 & -5.3 & 119.7 & 16.0 \\ 72.3 & -25.6 & 22.2 & -6.2 & -0.6 & 30.7 & 4.6 & -1.9 & 7.1 \\ -65.9 & 148.8 & -6.2 & 234.2 & -3.1 & -82.9 & -3.5 & 153.2 & 23.3 \\ -1.2 & -0.1 & -0.6 & -3.1 & 1.4 & 1.9 & -0.3 & -2.2 & 1.1 \\ 127.5 & -45.9 & 30.7 & -82.9 & 1.9 & 497.8 & -5.8 & -38.7 & 34.9 \\ 20.2 & -5.3 & 4.6 & -3.5 & -0.3 & -5.8 & 9.1 & 4.5 & -2.1 \\ -48.8 & 119.7 & -1.9 & 153.2 & -2.2 & -38.7 & 4.5 & 162.8 & 19.4 \\ 62.8 & 16.0 & 7.1 & 23.3 & 1.1 & 34.9 & -2.1 & 19.4 & 84.8 \end{bmatrix}$$

II) $\alpha = 0.001$

Toimivat

$$S1 = \begin{bmatrix} 169.0 & -15.6 & 17.3 & -1.2 & -1.4 \\ -15.6 & 58.2 & -2.6 & -2.7 & 1.0 \\ 17.3 & -2.6 & 6.1 & 1.7 & -0.4 \\ -1.2 & -2.7 & 1.7 & 70.1 & -0.3 \\ -1.4 & 1.0 & -0.4 & -0.3 & 1.0 \end{bmatrix}$$

Konkurssit

$$S2 = \begin{bmatrix} 586.1 & -120.0 & 67.5 & 30.7 & 0.2 \\ -120.0 & 155.6 & -19.3 & -18.2 & 0.2 \\ 67.5 & -19.3 & 18.3 & 12.3 & -0.2 \\ 30.7 & -18.2 & 12.3 & 134.4 & -1.0 \\ 0.2 & 0.2 & -0.2 & -1.0 & 1.0 \end{bmatrix}$$

III) $\alpha = 0.01$

Toimivat

$$S1 = \begin{bmatrix} 158.3 & -15.2 & 15.6 & -1.2 & 5.1 \\ -15.2 & 55.2 & -2.4 & 0.7 & 45.5 \\ 15.6 & -2.5 & 5.6 & -0.4 & -3.0 \\ -1.2 & 0.7 & -0.4 & 0.9 & 1.1 \\ 5.1 & 45.5 & -3.0 & 1.1 & 274.4 \end{bmatrix}$$

II) Kun α=0.001, jolloin n1=780 ja n2=87

Askel	Muuttuja Sisään	Poista	Mukana	Osittais R**2	F-luku	Tod>F	Wilksin lambda	Tod <lambda
1	X4X9		1	.207	225.997	.0001	.79285	.0001
2	X8		2	.067	61.753	.0001	.73997	.0001
3	X8X8		3	.023	20.619	.0001	.72270	.0001
4	EX7EX9		4	.023	20.267	.0001	.70610	.0001
5	EX5EX9		5	.017	15.282	.0001	.69378	.0001
6	X4X10		6	.013	11.287	.0008	.68479	.0001
7	X10EX9		7	.011	9.689	.0019	.67716	.0001
8	X10		8	.010	8.588	.0035	.67045	.0001
9	X1X9		9	.011	9.161	.0025	.66336	.0001
10	EX2EX10		10	.010	8.227	.0042	.65704	.0001
11	EX8EX9		11	.008	6.453	.0113	.65212	.0001
12		EX7EX9	10	.000	0.011	.9158	.65213	.0001
13	EX5EX5		11	.008	6.791	.0093	.64699	.0001
14	X9EX7		12	.008	6.564	.0106	.64205	.0001
15	EX3EX3		13	.007	5.752	.0167	.63775	.0001
16	X9X10		14	.006	5.498	.0193	.63366	.0001
17	X1X7		15	.007	5.991	.0146	.62923	.0001
18	X4X7		16	.009	7.905	.0050	.62344	.0001
19	X4		17	.009	7.295	.0071	.61812	.0001
20	EX9EX9		18	.008	6.389	.0117	.61350	.0001
21	EX2EX2		19	.006	5.120	.0239	.60982	.0001
22	X4EX9		20	.005	4.376	.0367	.60668	.0001
23	X7EX1		21	.005	3.809	.0513	.60395	.0001
24		X8X8	20	.001	0.567	.4517	.60436	.0001
25		X9EX7	19	.001	0.935	.3338	.60503	.0
26	X6EX4		20	.005	4.466	.0349	.60185	.0
27	X5EX3		21	.005	4.166	.0416	.59890	.0
28	X6EX3		22	.007	5.652	.0177	.59491	.0
29	X10X10		23	.005	4.339	.0376	.59187	.0
30	X5EX10		24	.005	3.848	.0501	.58918	.0

I) Kun α=0.0, jolloin n1=843 ja n2=93.

Askel	Muuttuja Sisään	Poista	Mukana	Osittais R**2	F-luku	Tod>F	Wilksin lambda	Tod <lambda
1	X4		1	.086	88.205	.0001	.9137	.0001
2	X4X8		2	.098	100.911	.0001	.8245	.0001
3	X1EX4		3	.037	36.068	.0001	.7938	.0001
4	X1X8		4	.024	22.707	.0001	.7749	.0001
5	EX5EX9		5	.023	21.383	.0001	.7575	.0001
6	EX2EX10		6	.019	18.095	.0001	.7430	.0001
7	X3EX2		7	.010	9.545	.0021	.7355	.0001
8	X9EX1		8	.012	10.770	.0011	.7270	.0001
9	EX9EX10		9	.010	9.046	.0027	.7200	.0001
10	EX3EX4		10	.008	7.540	.0061	.7142	.0001
11	EX5		11	.009	8.075	.0046	.7080	.0001
12	EX2EX4		12	.008	7.398	.0067	.7023	.0001
13	X1X4		13	.005	4.282	.0388	.6991	.0001
14	X8		14	.006	5.416	.0202	.6950	.0001
15	X10EX3		15	.005	4.385	.0365	.6917	.0001
16	EX9EX9		16	.004	3.792	.0518	.6889	.0001
17	X3EX9		17	.005	4.276	.0389	.6857	.0001
18	X3EX10		18	.010	9.131	.0026	.6789	.0001
19	X5EX8		19	.009	8.697	.0033	.6725	.0001
20	X6EX8		20	.008	7.272	.0071	.6672	.0001
21	X9EX5		21	.006	5.189	.0230	.6635	.0001
22		X9EX1	20	.001	0.659	.4172	.6639	.0001
23	X6EX5		21	.006	5.719	.0170	.6598	.0001
24	X9EX6		22	.009	7.923	.0050	.6541	.0001
25	X4EX6		23	.004	4.069	.0440	.6512	.0001
26	X4EX5		24	.005	4.159	.0417	.6483	.0001
27	EX7		25	.003	2.950	.0862	.6462	.0001
28	X3EX8		26	.002	2.171	.1410	.6446	.0001
29	X1EX5		27	.005	4.178	.0413	.6417	.0001
30		X5EX8	26	.000	0.038	.8451	.6417	.0001

IV) Kun α=0.05, jolloin n1=751 ja n2=76

Askel	Muuttuja Sisään	Poista	Mukana	Osittais R**2	F-luku	Tod>F	Wilksin lambda	Tod <lambda
1	X4X9		1	.2263	241.328	.0001	.7737	.0001
2	X8X9		2	.0749	66.758	.0001	.7157	.0001
3	EX9EX9		3	.0246	20.722	.0001	.6981	.0001
4	X4EX9		4	.0229	19.246	.0001	.6821	.0001
5	EX4EX4		5	.0201	16.865	.0001	.6684	.0001
6	X1X3		6	.0152	12.617	.0004	.6583	.0001
7	X3EX10		7	.0153	12.688	.0004	.6482	.0001
8	EX4EX6		8	.0094	7.742	.0055	.6422	.0001
9	X5EX3		9	.0073	6.012	.0144	.6375	.0001
10	X7EX1		10	.0081	6.699	.0098	.6323	.0001
11	EX2EX2		11	.0052	4.261	.0393	.6290	.0001
12	EX8EX9		12	.0058	4.762	.0294	.6253	.0001
13	X2EX8		13	.0054	4.379	.0367	.6220	.0001
14	X10EX5		14	.0058	4.754	.0295	.6184	.0001
15	X9EX4		15	.0085	6.928	.0086	.6131	.0001
16	X4X4		16	.0053	4.284	.0388	.6099	.0001
17	X9X10		17	.0054	4.385	.0366	.6066	.0001
18	X10		18	.0068	5.541	.0188	.6025	.0001
19	X10X10		19	.0087	7.118	.0078	.5972	.0001
20	EX7EX10		20	.0071	5.731	.0169	.5930	.0001
21	EX5EX5		21	.0061	4.940	.0265	.5894	.0001
22	EX5EX9		22	.0079	6.371	.0118	.5848	.0001
23	X9EX3		23	.0083	6.686	.0099	.5799	.0
24	X10EX9		24	.0073	5.917	.0152	.5757	.0
25	X9EX5		25	.0062	5.034	.0251	.5721	.0
26		X10EX5	24	.0003	0.206	.6502	.5722	.0
27	EX2EX6		25	.0061	4.916	.0269	.5687	.0
28	X10EX4		26	.0047	3.766	.0527	.5661	.0
29	X1X4		27	.0035	2.835	.0926	.5641	.0
30		X4EX9	26	.0012	0.953	.3293	.5647	.0

V) Kun α =0.10, jolloin n1=737 ja n2=75

Askel	Muuttuja Sisään	Poista	Mukana	Osittais R**2	F-luku	Tod>F	Wilksin lambda	Tod <lambda
1	X4X9		1	.2309	243.203	.0001	.7691	.0001
2	X8X9		2	.0821	72.344	.0001	.7060	.0001
3	EX9EX9		3	.0318	26.544	.0001	.6835	.0001
4	X4EX9		4	.0203	16.694	.0001	.6696	.0001
5	EX4EX4		5	.0223	18.421	.0001	.6547	.0001
6	EX4EX6		6	.0189	15.508	.0001	.6423	.0001
7	X3EX10		7	.0184	15.112	.0001	.6305	.0001
8	X1X3		8	.0088	7.090	.0079	.6249	.0001
9	EX1EX7		9	.0079	6.413	.0115	.6200	.0001
10	EX2EX2		10	.0063	5.115	.0240	.6161	.0001
11	EX5EX9		11	.0061	4.903	.0271	.6123	.0001
12	EX5EX6		12	.0064	5.162	.0233	.6084	.0001
13	X5X8		13	.0055	4.416	.0359	.6050	.0001
14	X6EX5		14	.0039	3.113	.0781	.6027	.0001
15	X4		15	.0033	2.597	.1075	.6007	.0001
16	X4X6		16	.0043	3.396	.0657	.5982	.0001
17		X1X3	15	.0001	0.115	.7351	.5982	.0001
18	X5X6		16	.0109	8.789	.0031	.5917	.0
19	X10EX9		17	.0045	3.625	.0573	.5890	.0
20		X8X9	16	.0008	0.607	.4361	.5895	.0
21	X9X10		17	.0050	4.022	.0453	.5865	.0
22	X5X10		18	.0077	6.152	.0133	.5820	.0
23		X4X9	17	.0012	0.929	.3355	.5827	.0
24	EX3EX10		18	.0081	6.495	.0110	.5779	.0
25	X7EX1		19	.0046	3.678	.0555	.5752	.0
26		EX1EX7	18	.0009	0.720	.3964	.5758	.0
27	X5EX1		19	.0061	4.869	.0276	.5723	.0
28	EX2EX9		20	.0035	2.759	.0971	.5703	.0
29		X4EX9	19	.0003	0.252	.6160	.5704	.0
30	EX2EX6		20	.0038	3.013	.0830	.5683	.0

-Kanonisen erottelufunktion muuttujat

y(i)	I	II	III	IV	V
y(1)	x4	x4	x4	x4	x4
y(2)	x8	x8	x8	x8	x8
y(3)	x9	x9	x9	x9	x9
y(4)	ex8	ex5	x10	ex10	ex10
y(5)	ex10	x10	x6	ex5	x6
y(6)	x1			x5	ex5
y(7)	ex9			ex9	ex9
y(8)	ex7				x5
y(9)	ex2				

-Kanonisen erottelufunktion kertoimet

y(i)	I	II	III	IV	V
y(1)	-2.44	-3.65	-3.73	-3.61	-3.64
y(2)	6.13	6.42	6.05	6.70	6.69
y(3)	-12.7	-10.4	-12.0	-25.9	-23.9
y(4)	-3.01	-1.42	-13.1	-14.2	-15.4
y(5)	-12.7	-10.7	0.772	-3.54	0.620
y(6)	0.713			2.27	-3.04
y(7)	7.71			17.6	15.7
y(8)	1.90				1.78
y(9)	-1.83				

-Mahalanobisetäisyys

	Erottelupistemäärien keskiarvot		Mahalanobis-
	Konkurssit	Toimivat	etäisyys
I	-152	17	1.30
II	-181	20	1.42
III	-179	19	1.41
IV	-196	20	1.47
V	-199	20	1.48

ENSIMMÄISEN ASTEEN MUUTTUJIEN

-Testitunnuslukuja

Muut-tujat	Kanoninen korrelaatio	Ominais-arvo	Wilksin lambda	Pillain trace	Hot.Law. trace
I	.452	.2568	.796	.204	.257
II	.518	.3666	.732	.268	.367
III	.505	.3415	.745	.255	.341
IV	.530	.3905	.719	.281	.391
V	.537	.4045	.712	.288	.404

-Sisäisten rakennekerrointen itseisarvot

I-muuttujat n1=843 n2=93		II-muutt. n1=780 n2=87		III-muutt. n1=767 n2=82		IV-muuttujat n1=751 n2=76		V-muuutt. n1=737 n2=75	
x4	.61	x4	.78	x4	.79	x4	.77	x4	.76
x8	.59	x8	.69	x8	.70	x8	.66	x8	.66
x9	.60	x9	.68	x9	.69	x9	.63	x9	.64
ex8	.23	ex5	.19	x10	.02	ex10	.08	ex10	.08
ex10	.04	x10	.05	x6	.34	ex5	.19	x6	.31
x1	.01					x5	.29	ex5	.20
ex9	.25					ex9	.27	ex9	.28
ex7	.31							x5	.30
ex2	.17								

I) Kun α =0.0, jolloin n1=843 ja n2=93.

Kolmannen asteen termit myös mukana

Askel	Muuttuja Sisään	Poista	Mukana	Osittais R**2	F-luku	Tod>F	Wilksin lambda	Tod <lambda
1	x4		1	.0863	88.205	.0001	.91371	.0001
2	x4x8		2	.0976	100.911	.0001	.82453	.0001
3	x1ex4		3	.0373	36.068	.0001	.79381	.0001
4	x1x8		4	.0238	22.707	.0001	.77491	.0001
5	ex5ex9		5	.0225	21.383	.0001	.75749	.0001
6	ex2ex10		6	.0191	18.095	.0001	.74302	.0001
7	ex3ex4ex7		7	.0145	13.607	.0002	.73228	.0001
8	ex9ex10		8	.0132	12.418	.0004	.72260	.0001
9	ex1ex4ex9		9	.0094	8.804	.0031	.71579	.0001
10	x9ex6		10	.0092	8.619	.0034	.70919	.0001
11	x6ex5		11	.0128	12.004	.0006	.70009	.0001
12	ex3ex5ex9		12	.0104	9.728	.0019	.69279	.0001
13	x4ex1		13	.0092	8.584	.0035	.68640	.0001
14	x8		14	.0132	12.285	.0005	.67736	.0001
15	x9x10		15	.0098	9.139	.0026	.67070	.0001
16	ex9ex9		16	.0046	4.292	.0386	.66758	.0001
17	x3ex9		17	.0054	5.021	.0253	.66395	.0001
18	x1x2		18	.0076	6.981	.0084	.65894	.0001
19	x7x9		19	.0074	6.797	.0093	.65408	.0001
20	x8ex4		20	.0058	5.369	.0207	.65027	.0001

II) Kun α =0.001, jolloin n1=780 ja n2=87

Kolmannen asteen termit myös mukana

Askel	Muuttuja Sisään	Poista	Mukana	Osittais R**2	F-luku	Tod>F	Wilksin lambda	Tod <lambda
1	x4x9		1	.2071	225.997	.0001	.79285	.0001
2	x8		2	.0667	61.753	.0001	.73997	.0001
3	x8x8		3	.0233	20.619	.0001	.72270	.0001
4	ex5ex9		4	.0230	20.267	.0001	.70610	.0001
5	ex1ex5ex9		5	.0218	19.144	.0001	.69074	.0001
6	x1x5x9		6	.0174	15.206	.0001	.67874	.0001
7	ex7ex9		7	.0154	13.458	.0003	.66827	.0001
8	x4ex9		8	.0133	11.596	.0007	.65936	.0001
9	ex1ex4ex8		9	.0109	9.446	.0022	.65217	.0001
10	ex5ex5ex5		10	.0079	6.806	.0092	.64702	.0001
11	ex2ex10		11	.0085	7.353	.0068	.64151	.0001
12	ex9ex10		12	.0074	6.379	.0117	.63675	.0001
13	x5ex10		13	.0092	7.886	.0051	.63092	.0001
14	x4ex7		14	.0071	6.132	.0135	.62641	.0001
15	ex1ex4ex9		15	.0064	5.522	.0190	.62237	.0001
16	x1x7		16	.0061	5.221	.0226	.61857	.0001
17	ex1ex6ex9		17	.0087	7.414	.0066	.61322	.0001
18	x4ex8		18	.0089	7.587	.0060	.60778	.0
19	x1x4x7		19	.0067	5.717	.0170	.60370	.0
20	x3x6x8		20	.0060	5.129	.0238	.60007	.0

*

-Kanonisen erottelufunktion muuttujat

y(i)	I	II	III	IV	V
y(1)	x4	x4x9	x4x9	x4x9	x4x9
y(2)	x4x8	x8	x2x8	x8x9	x8x9
y(3)	x1ex4	x8x8	x1x5x9	ex9ex9	ex9ex9
y(4)	x1x8	ex7ex9	x2x5x10	x4ex9	x4ex9
y(5)	ex5ex9	ex1ex5ex9	ex5ex9	ex4ex4	ex4ex4
y(6)	ex2ex10	x1x5x9	x4ex9	ex1ex5ex10	ex4ex6
y(7)	ex3ex4ex7	ex5ex9	ex5ex5ex5	x1x3	x3ex10
y(8)	ex9ex10	x4ex9	x8x9	ex3ex4ex7	ex2ex2ex2
y(9)	ex1ex4ex9	ex1ex4ex8	x1x5	ex4ex6	ex3ex4ex7
y(10)	x9ex6	ex5ex5ex5	ex1ex5ex10	ex2ex2ex2	ex1ex5ex10
y(11)	x6ex5	ex2ex10	ex1ex7	ex6ex9	ex3ex6ex9
y(12)	ex3ex5ex9	ex9ex10	ex2ex2	ex3ex4	ex1ex7
y(13)	x4ex1	x5ex10	x2x4x9	x9ex4	ex1ex6ex7
y(14)	x8	x4ex7	ex8ex9	x7ex1	ex2ex5ex9

-Kanonisen erottelufunktion kertoimet

y(i)	I	II	III	IV	V
y(1)	-4.48	-0.532	-0.769	-0.499	-0.354
y(2)	0.173	6.62	0.159	1.35	1.4
y(3)	0.047	-0.138	0.0131	-9.04	-9.9
y(4)	-0.0635	-2.95	-0.0416	0.519	0.397
y(5)	-1.27	-0.00937	-2.8	-0.0701	-0.132
y(6)	-3.68	0.00587	0.642	-0.39	0.443
y(7)	-0.00505	-1.85	0.0036	0.0338	-0.478
y(8)	3.42	0.483	1.05	-0.0159	-0.00974
y(9)	-0.0291	-0.00741	-0.0534	0.417	-0.0167
y(10)	-0.458	0.00357	-0.275	-0.011	-0.257
y(11)	0.0543	-4.76	-0.171	-1.32	-0.059
y(12)	0.00849	16.9	-0.23	-0.203	-0.341
y(13)	0.0493	-0.962	0.0046	0.475	0.0149
y(14)	3.37	0.0742	-1.63	-0.151	0.122

-Mahalanobisetäisyys

	Erottelupistemäärien keskiarvot Konkurssit	Toimivat	Mahalanobis-etäisyys
I	-208	23	1.52
II	-225	25	1.58
III	-232	25	1.60
IV	-255	26	1.67
V	-263	27	1.71

KOLMANNEN ASTEEN MUUTUUJIEN

-Testitunnuslukuja

Muut-tujat	Kanoninen korrelaatio	Ominais-arvo	Wilksin lambda	Pillain trace	Hot.Law. trace
I	.568	.4763	.677	.323	.476
II	.602	.5674	.638	.362	.567
III	.604	.5746	.635	.365	.575
IV	.630	.6579	.603	.397	.658
V	.643	.7041	.587	.413	.704

-Sisäisten rakennekerrointen itseisarvot

y(i)	I	II	III	IV	V
y(1)	.45	.68	.68	.67	.65
y(2)	.23	.56	.55	.53	.52
y(3)	.06	.11	.32	.49	.51
y(4)	.10	.32	.29	.33	.33
y(5)	.17	.15	.31	.38	.39
y(6)	.06	.36	.39	.18	.28
y(7)	.01	.10	.10	.12	.13
y(8)	.00	.45	.58	.12	.22
y(9)	.08	.03	.18	.28	.12
y(10)	.14	.12	.18	.22	.17
y(11)	.04	.15	.04	.11	.02
y(12)	.01	.04	.26	.27	.04
y(13)	.06	.05	.55	.37	.03
y(14)	.43	.36	.30	.07	.20

((15)) Siotani, M. 1980 'Asymptotic Approximations to the Conditional Distributions of the Classification Statistic Z and Its Studentized Form Z*', Tamkang Journal of Mathematics, n:o 11

((16)) Hoel, P.G. & Peterson, R.P. 1979 'A Solution to the Problem of Optimum Classification', Annals Mathematical Statistics, n:o 20

((17)) Hand, D.J. 1981 'Discrimination and Classification' New York: John Wiley & Sons

((18)) Tatsuoka, M.M. 1970 'Multivariate Analysis' New York: John Wiley & Sons

((19)) Morrison, D.F. 1976 'Multivariate Statistical Methods', Second Edition, New York: McGraw-Hill

((20)) Mardia, K.V. & Kent, J.T. & Bibby, J.M. 1979 'Multivariate Analysis', London: Academic Press

((21)) SAS Institute Inc. SAS User's Guide: Statistic, Version 5 Edition Cary, NC: SAS Institute Inc., 1985

((22)) Klecka, W.R. 1980 'Discriminant Analysis' Sage University Paper series on Quantitative Applications, series no. 7-19, Beverly Hills: Sage Publications

((23)) Constanza, M.C. & Afifi, A.A. 1979 'Comparison of Stopping Rules in Forward Stepwise Discriminant Analysis' Journal of the American Statistical Association, 74

((24)) Lachenbruch, P.A. 1975 'Discriminant Analysis' Hafner Press, A Division of McMillan Publishing Co., Inc.

((25)) Rao, C.R. 1973 'Linear Statistical Inference and Its Applications Second Edition, New York: John Wiley & Sons

((26)) Cochran, W.G. 1964 'On the Performance of the Linear Discriminant Functions' Technometrics, n:o 6

((27)) Kettunen, P. & Mäkinen, V & Neilimo, K 1980 'Tilinpäätösanalyysi' Ekonomia-sarja 46. Tapiola: Weilin & Göös

((28)) Yritystutkimusneuvottelukunta 1987 'Tilinpäätösanalyysi' Helsinki

((29)) Laitinen, E.K. 1986 'Yrityksen tunnuslukuanalyysi' Yritysinnovaatio Oy, Jyväskylä

((30)) Laitinen, E.K. & Virtanen, V. 1986 'Vastaperustetun pienyrityksen eloonjääntі/konkurssiprosessin selittäminen' Liiketaloudellinen Aikakausikirja 3

((31)) Prihti, A. 1975 'Konkurssin ennustaminen taseinformaation avulla' Acta Academiae Oeconomicae Helsingiensis, Helsinki

((32)) Kaikkonen, R. 1988 'Konkurssit kaipaavat hälyytysjärjestelmää' Kauppalehti Optio 20

((33)) Jaakkola, J. & Ratia, U. 1984 'Yrityksestä ja sen tutkimisesta/ Yrityksen tervehdyttäminen' Yritystutkimusneuvottelukunta

((34)) Argenti, J. 1976 'Corporate Collapse' The Causes and Symptoms, London

((35)) Laitinen, E.K. 1987 'Konkurssin välttäminen vaatii uudelta yritykseltä hyvää kannattavuutta ja nopeaa kasvua' Yritystalous 8, Oy Rastor Ab

((36)) Hartigan, P.W.J. 1973 'Causes of Company Failure' Accountant 168

((37)) Liiketaloudellinen tutkimuslaitos, 1980 'Konkurssin taloudelliset seuraamukset yrityksen sidosryhmille' Helsinki

((38)) Neilimo, K. 1989 'Tilinpäätös antaa systemaattista tietoa luottopäätöksen pohjaksi' Suomen Luotonantajayhdistyksen Tiedotuslehti Majakka, no 3

((39)) Barthroy, A. 1987 'The Analysis of Credit' McGraw-Hill, London

((40)) Beaver, W. 1966 'Financial Ratios as Predictors of Failure' Journal of Accounting Research, vol 5

((41)) Altman, E. 1968 'Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy' Journal of Finance, September

((42)) Laitinen, E.K. 1989 'Yrityksen talouden mittarit' Espoo: Amer-Yhtymä Oy Weilin & Göös

((43)) Bartnett, W. & Lewis, T. 1978 'Outliers in Statistical Data' New York: John Wiley & Sons

((44)) Gnanadesikan, R. 1977 'Methods for Statistical Data Analysis of Multivariate Observations' New York: John Wiley & Sons

((45)) Kendall, M. & Stuart, A. 1961 'The Advanced Theory of Statistics' Vol.3 London:Charles Griffin and Company, Ltd

((46)) Afifi, A.A. and Clark, V.C. 1984 'Computer-Aided Multivariate Analysis' Belmont, CA: Lifetime Learning Publications

((47)) Desu, M.M. & Deisser, S. 1973 'Methods and Application of Equal Mean Discrimination', ss. 139-159