TALLINA PEDAGOOGIKAÜLIKOOL
MATEMAATIKA – LOODUSTEADUSKOND
INFORMAATIKA OSAKOND

PROSEMINARITÖÖ

DISKRIMINANTANALÜÜS PROGRAMMI SPSS ABIL ÕPPEMATERJALI LOOMINE.

KOOSTAS: KERLI KOPPEL

JUHENDAS: KATRIN NIGLAS

TALLINN 2002

SISUKORD

SISSEJUHATUS 3
DISKRIMINANTANALÜÜSI MÕISTE JA EELDUSED 4
DISKRIMINANTANALÜÜSI RAKENDUSVALDKONNAD 4
DISKRIMINANTANALÜÜSI LIIGID 5
DISKRIMINANTANALÜÜSI PÕHISAMMUD 5
DISKRIMINANTANALÜÜSI ÜLESANDE PÜSTITUS 5
DISKRIMINANTANALÜÜSI TELLIMINE SPSS'IS 6
DISKRIMINANTANALÜÜSI TULEMUSTE TÕLGENDAMINE 11
TULEMUSTE TÄPSUSTAMINE 19
DISKRIMINANTANALÜÜS SAMMPROTSEDUURIGA 21
LÕPPSÕNA 24
KASUTATUD KIRJANDUS 25

SISSEJUHATUS

Käesoleva proseminaritöö temaatika - diskriminantanalüüs, kuulub mitmemõõtmelise statistilise analüüsi st. klassikaliste meetodite hulka, mida kasutatakse paljudes eri valdkondades praktiliste ülesannete lahendamisel.

Proseminaritöö eesmärgiks oli koostada õppematerjal diskriminantanalüüsi läbiviimise kohta statistikaprogrammis SPSS.

Antud õppematerjal on mõeldud neile, kes ei ole oma õpingutes kokku puutunud kõrgema matemaatikaga ning seetõttu on materjal koostatud nii lihtsalt, et diskriminantanalüüsist kui statistika meetodist ülevaate saamine ning diskriminantanalüüsi läbiviimine statistikaprogrammis SPSS, ei sõltu oluliselt kasutaja varasematest matemaatilistest teadmistest.

Proseminaritöö ülesehituse eesmärgiks on ühelt poolt anda ülevaade diskriminantanalüüsi kui matemaatilise statistika meetodi olemusest ja selle raksendusvaldkondadest ning teiselt poolt vaadelda diskriminantanalüüsi kui andmeanalüüsi meetodit ning kirjeldada diskriminantanalüüsi läbiviimist statistika programmiga SPSS.

Antud temaatikaga proseminaritöö koostamist ajendas asjaolu, et diskriminantanalüüsile pühendatud materjale on eesti keeles välja antud ainult üks: S. Koskel, E. - M. Tiit, P. Arandi *"Diskriminantanalüüs"* ning, et mainitud raamat eeldab lugejalt suhteliselt sügavaid kõrgema matemaatika valdkonda puudutavaid teadmisi.

Materjali lihtsuse tagamiseks on diskriminantanalüüsi läbiviimist SPSS-is vaadeldud ainult lihtsamal erijuhul – kahe rühma eristamise korral. Kogu matejal on üles ehitatud nii, et diskriminantanalüüsi tellimist ning hiljem tulemuste tõlgendamist SPSS-is, on kirjeldatud panganduse valdkonda kuuluva näitega. Sellise metoodilise käsitluse valikul on silmas peetud just kasutajaks olevate inimeste rühma.

Proseminaritöö võib jagada kolmeks osaks, milles esimene osa hõlmab järgmisi teemasid: *diskriminantanalüüsi mõiste ja eeldused, diskriminantanalüüsi rakendusvaldkonnad, diskriminantanalüüsi liigid, diskriminantanalüüsi põhisammud, diskriminantanalüüsi ülesande püstitus,* teise osa moodustab *diskriminantanalüüsi tellimine SPSS-is* ning kolmas osa hõlmab *diskriminantanalüüsi tulemuste tõlgendamist* ning nende täpsustamist.

Et enne proseminaritöö kirjutamisele asumist, puudus mul igasugune eelteave diskriminantanalüüsi kohta, siis olen prosemianritöö koostamisel aluseks võtnud raamatu S. Koskel, E. - M. Tiit, P. Arandi *"Diskriminantanalüüs"* ning seda just eriti teoreetilise materjali osas. Veel olen kasutanud ühe eelpool nimetatud raamatu autori E. – M. Tiit internetis olevat loengumaterjali teemal diskriminantanalüüs. Samuti olen kasutanud proseminaritöö koostamisel SPSS-i abiotsingumootorit.

Proseminaritöö juhendamise eest tänan assistenti Katrin Niglast.

DISKRIMINANTANALÜÜSI MÕISTE JA EELDUSED

Diskrimineerimine on ladina keelest tuletatud sõna ning tähendab eristamist.
Diskriminantanalüüsi ülesandeks on leida **rühmitamise eeskiri** ehk **diskrimineerimiseeskiri.**
Seejuures on eeldatud, et üldkogum jaguneb lõplikuks hulgaks rühmadeks ning eesmärk on:

1. leida rühmitamise eeskiri, mis paigutaks olemasolevad objektid rühmadesse võimalikult õigesti;
2. määrata tundmatu kuuluvusega objekt ühte olemasolevasse rühma.

Diskriminantanalüüsi ülesande praktiliseks lahendamiseks kasutatakse statistilist andmestikku, millel on järgmised eeldused:

- Igast rühmast on olemas valim, kusjuures kõigi valimi objektide kohta on teada, millisesse rühma ta kuulub.
- Tundmatu objekti kohta eeldatakse, et ta kuulub ühte olemasolevasse rühma, kuid millisesse, see ei ole teada.
- Tunnuste vahel, mida kasutatakse dikrimineerimiseeskirja koostamisel, ei ole tugevat korrelatsiooni.
- Tunnuste, mida kasutatakse dikrimineerimiseeskirja koostamisel, väärtused alluvad normaaljaotusele.

Diskriminantanalüüsi meetodi käigus leitakse lineaarkombinatsioone tunnustest, mis kõige paremini võimaldavad rühmasid eristada. Neid kombinatsioone nimetatakse diskriminantfunktsioonideks ning on esitatud järgmise võrrandiga

$$d_k = b_{0k} + b_{1k} x_1 + b_{2k} x_2 + \ldots + b_{pk} x_p$$

kus d_k – on k-nda diskriminantfunktsiooni väärtus
p – tunnuste arv diskriminantfunktsioonis
x_i – i-nda tunnuse väärtus
b_{ik} – k-nda diskriminantfunktsiooni i-nda tunnuse kordaja ning b_{0k} on k-nda diskriminantfunktsiooni vabaliige.

Funktsioonide arv on võrdne MIN(rühmade arv – 1; tunnuste arv).

DISKRIMINANTANALÜÜSI RAKENDUSVALDKONNAD

Diskriminantanalüüsi kasutatakse paljudes valdkondades erinevate rakendusülesannete lahendamisel. Klassikaliseks diskriminantanalüüsi rakenduseks peetakse taime - või loomaliikide määramist, samuti kasutatakse antud analüüsi meetodit tehnika ning meditsiini erinevates valdkondades. Diskriminantanalüüsi saab kasutada ka paljude majandusülesannete lahendamisel ja prognoosimisel, samuti ühiskonnaelu nähtusi puudutavate ülesannete korral.

DISKRIMINANTANALÜÜSI LIIGID

Vastavalt sellel, millistele tingimustele vastavad tunnused, mida kasutatakse diskrimineerimiseeskirja koostamisel ning millised on eristatavate rühmade proportsioonid, on diskriminantanalüüsi metoodikad erinevad.

1. Kõige lihtsama variandi korral eeldatakse, et kõikides rühmades tunnuste väärtused alluvad normaaljaotusele. Rühmade kovariatsioonimaatriksid on sarnased, kuid rühmad erinevad keskväärtuste poolest. Sellisel juhul on tegemist **lineaarse diskriminantanalüüsiga**.

2. Mõnevõrra keerulisem ülesanne saadakse, kui eeldatakse jätkuvalt, et kõikides rühmades tunnuste väärtused alluvad normaaljaotusele, kuid lisaks sellele, et rühmad erinevad keskväärtuse poolest, on ka nende kovariatsioonimaatriksid erinevad. Sellisel juhul on tegemist **mittelineaarse diskriminantanalüüsiga**.

3. Eeldatakse, et erinevates rühmades on aprioorsed tõenäosused (hinnang, kui suure tõenäosusega objektid vastavasse rühma kuuluvad) erinevad ning diskrimineerimiseeskirja koostamisel kasutatakse neid tõenäosusi, sellisel juhul kõneletakse **diskriminantanalüüsi Bayesi käsitlusest**.

4. Mõnikord on põhjust eeldada, et eksimused, mis tekivad tundmatu objekti määramisel valesse rühma, ei ole samaväärsed, st. mõne rühma puhul on eksimuse hind kõrgem, teise puhul madalam. Kui need eksimuse hinnad on teada ning neid arvestatakse diskrimineerimiseeskirja koostamisel räägitakse **otsustusteoreetilisest diskriminantanalüüsist**.

DISKRIMINANTANALÜÜSI PÕHISAMMUD

Diskriminantanalüüsi läbiviimisel on võimalik välja tuua järgmised tegevuse etapid:

1. Valimi moodustamine.
2. Valimi põhjal diskrimineerimiseeskirja koostamine.
3. Rühmitamistulemuste hindamine.
4. Tundmatu objekti määramine ühte olemasolevasse rühma.

DISKRIMINANTANALÜÜSI ÜLESANDE PÜSTITUS

Edaspidi võtame vaatluse alla diskriminantanalüüsi kõige lihtsama alamjuhu, kus **eristatavaid rühmasid on** ainult **kaks**.

Käesolevas näites vaatame ühe panga andmestikku, mis sisaldab 850 laenutaotleja andmeid. Neist 700 on kliendid, kelle laenu taotlused on rahuldatud, ülejäänud 150 ootavad otsust laenu saamise kohta. Panga seisukohalt on oluline ära tunda kliendid, kellel tõenäoliselt võib tekkida maksejõuetus ehk teisisõnu identifitseerida madala ja

kõrge riskiga kliendid. Klientide andmestik sisaldab järgmisi andmeid: *kliendi vanust aastates, haridustaset, viimases töökohas töötatud aastate arvu, viimases elukohas elatud aastate arvu, leibkonna aastasissetulekut tuhandetes, võla suhe sissetulekusse, krediitkaardi võlga tuhandetes, teisi võlgu tuhandetes, eelneva maksekohuse mittetäitmist (0 – kui klient on maksnud makse õigeaegselt ning 1 – kui klient ei ole suutnud makse õigeaegselt pangale tasuda).*
Järgnevalt kasutatakse nende 700 kliendi andmeid, kelle laenu taotlus on rahuldatud. Vastavalt nende andmetele moodustatakse diskrimineerimiseeskiri, mille põhjal klassifitseeritakse ülejaanud 150 klienti, kes veel laenu saanud ei ole, kõrge ja madala riskiga laenutaotlejateks.

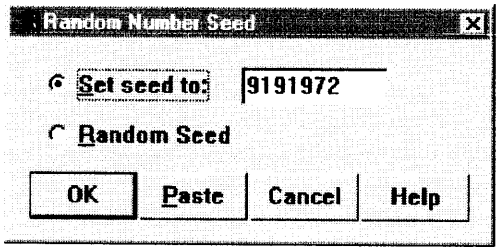
DISKRIMINANTANALÜÜSI TELLIMINE SPSS'IS

1.samm: Valimi moodustamine

Vaatleme antud näite 700 klienti, kellele on juba laenu antud, ning kelle maksekohuse õigeaegne täitmine/mittetäitmine on teada. Moodustame osadest eelnevalt mainitud klientidest valimi, kelle andmete põhjal koostame diskrimineerimiseeskirja. Ülejäänud klientide, kes valimisse ei pääsenud, andmeid kasutame hiljem selleks, et kontrollida, kui õigesti moodustatud diskrimineerimiseeskiri objekte rühmitab.
Et valimi moodustamisel peab olema igal objektil võrdne võimalus valimisse sattuda, kasutame täieliku juhuslikkuse saavutamiseks **juhuarvude generaatorit**.

Vali: ***Transform → Random Number Seed...***

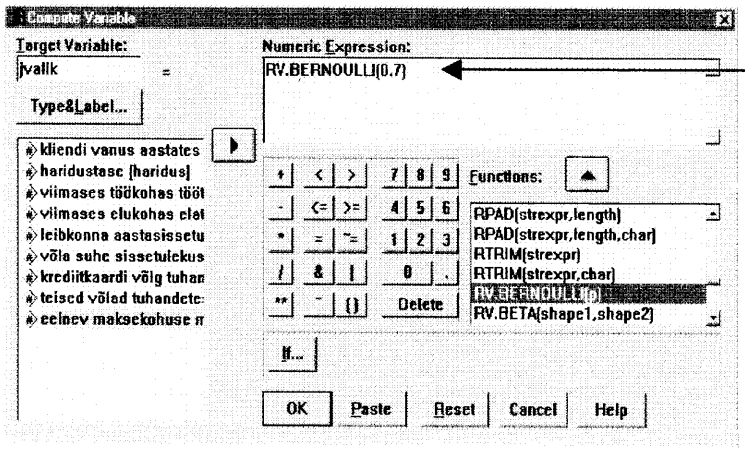
- Määra tingimus *Set seed to*. Ning sisesta täisarv vahemikus 1 kuni 2 000 000. Automaatselt on alati paika pandud arv 2 000 000.



Moodustame valimi selliselt, et ligikaudu 70% 700-st laenu saanud kliendist sattub valimisse ja ülejäänud 30% klientide andmete põhjal kontrollime diskrimineerimiseeskirja rühmitamise täpsust, selleks:

Vali: ***Transform → Compute...***

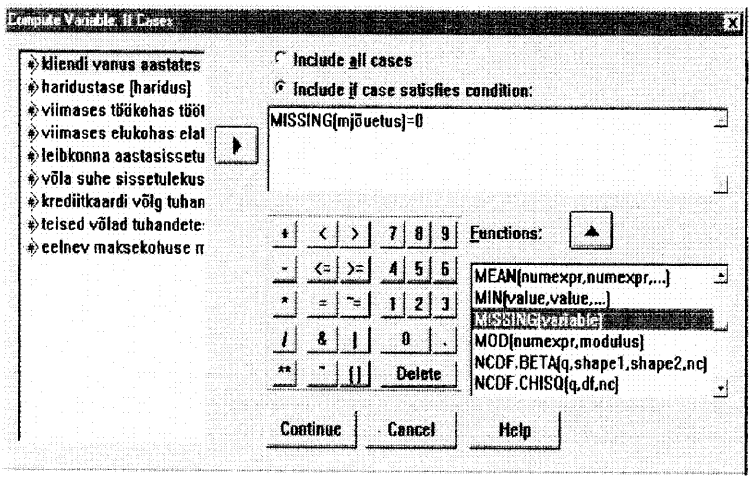
Väljale *Target Variable* kirjuta uue tunnuse nimetus.



Väljale *Numeric Expression* paiguta kasutada olevate funktsioonide valikust *RV.BERNOULLI(p)* ning määra tõenäosus parameeter. Tulemuseks väljastatakse objektidele väärtused 0 ja 1 vastavalt sisestatud tõenäosusele.

Selleks, et valimi moodustamisel ei osaleks need 150 objekti, mille on tunnuse *mjõuetus* (eelnev maksekohuse mittetäitmine) korral puuduv väärtus, tuleb need objektid juhuväljavõtul vaatluse alt välja jätta, selleks:

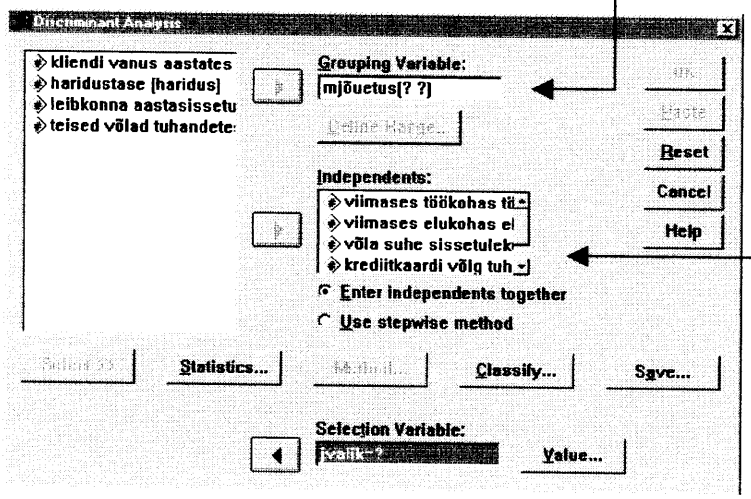
- Vajuta nuppu *If* ning määra tingimus *Include if case satisfies condition*, paiguta väljale kasutada olevate funktioonide seast MISSING (variable) ning määra, et juhuväljavõtul osalevad objektid, millel ei ole puuduvat väärtust rühmitamise aluseks oleva tunnuse osas.
  MISSING funktsioon toimib järgmiselt: 1-ga tähistatakse MISSING funktsiooni korral neid objekte, mille on puuduv väärtus antud tunnuse väärtusena ning 0-ga kõiki ülejäänuid. Järelikult, selleks, et meie näites toimuks juhuväljavõtt objektide korral, millel ei ole tunnuses *mjõuetus* (eelnev maksekohuse mittetäitmine) puuduvat väärtust, tuleb märkida MISSING (mjõuetus) = 0.



Kokkuvõttes on nüüd jäetud vaatluse alt välja 150 klienti kelle laenutaotlus on rahuldamata; 70%-le 700-st kliendist, kes laenu on saanud, antakse juhuväljavõtu väärtuseks 1 ning ülejäänutele 30%-le 0.

2.samm: Valimi põhjal diskrimineerimiseeskirja koostamine

Vali: ***Analyze → Classify → Discriminant...***

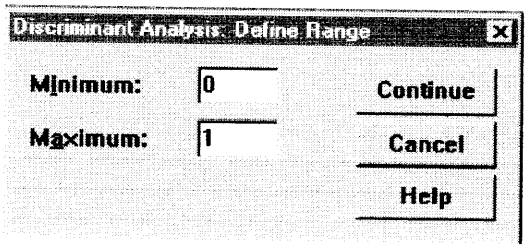


Paiguta tunnus, mille järgi rühmitamine teostatakse.

Paiguta tunnused, mille väärtusi kasutatakse diskrimineerimis-eeskirja koostamisel.

Antud näites on tunnuseks, mille põhjal klientide rühmitamine toimub *mjõuetus* (st. eelnev maksekohuse mittetäitmine). Tunnusteks, mille väärtusi kasutatakse diskrimineerimiseeskirja koostamisel, valime *viimases töökohas töötatud aastate arv*, *viimases elukohas elatud aastate arv*, *võla suhe sissetulekusse*, *krediitkaardi võlg tuhandetes* (Teine võimalus on lasta programmil otsustada, millised etteantud tunnustest sobivad kõige paremini diskrimineerimiseeskirja koostamiseks, vt. teema: diskriminantanalüüs sammprotseduuriga).

- Aktiveeri tunnus väljal *Grouping Variable* ning vajuta nuppu *Define Range...*, määra minimaalne ja maksimaalne täisarvuline väärtus vastava tunnuse väärtuste jaoks. Siin lubatakse kasutada ka tunnust, kus on enam väärtusi kui meid huvitavate gruppide arv. Objektid, mis antud tunnuse väärtusena omavad väärtust, mis sisestatud lõikku ei kuulu, jäetakse diskriminantanalüüsi käigus vaatluse alt välja. Väljal *Minimum* olev väärtus peab olema alati väiksem kui väljal *Maximum* olev väärtus.

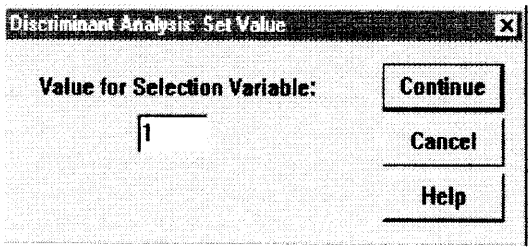


Et antud näites on väljal Grouping Variable tunnuseks *mjõuetus* (st. eelnev maksekohuse mittetäitmine) ning et andmestikku on sisestatud tunnuse väärtused selliselt, et maksekohuse täitmine on tähistatud 0-ga ja mittetäitmise 1-ga, on antud aknas vähimaks väärtuseks 0 ja suurimaks väärtuseks 1.

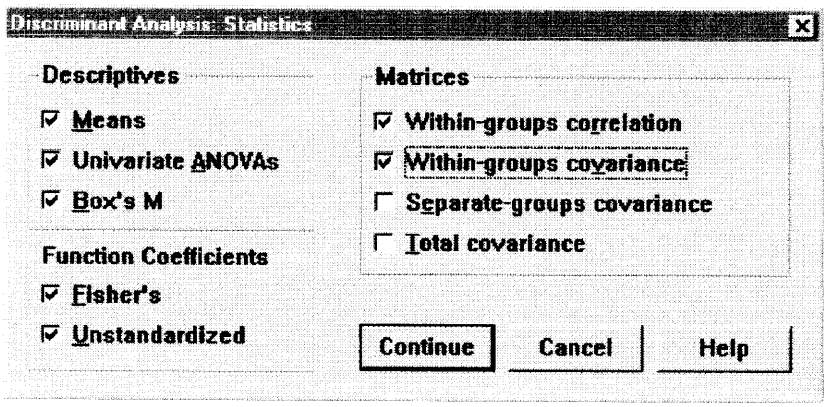
Diskrimineerimiseeskirja koostamisel otsustasime lähtuda objektidest, millel on tunnuse *jvalik* (juhuväljavõtt) korral antud väärtuseks 1. Seda tuleb siin ka märkida:

- Vajuta nuppu *Select>>* ning tekkinud väljale *Selection Variable:* paiguta selekteerimise aluseks olev tunnus. Meie näite korral on see tunnus *jvalik*. Seejärel ...

- Vajuta nuppu *Value*... ning sisesta selekteeriv väärtus. Diskrimineerimis-eeskirja moodustamisel arvestatakse ainult nende objektide andmeid, mis omavad selekteeriva tunnuse väärtusena vastavalt sisestatud väärtust, ülejäänud objektide andmeid kasutatakse automaatselt diskrimineerimis-eeskirja tõhususe kontrollimisel.



- Vajuta nuppu *Statistics*... ja vali soovitud kirjeldavad statistikud, funktsioonide kordajad ning maatriksid.



Kirjeldavatest statistikutest on kasutada:

Means – väljastab keskväärtuse koos standarthälbega

Univariate ANOVAs – olulisustest rühmade keskväärtuste võrdlemiseks

Box's M – võrdleb rühmade kovariatsioonimaatrikseid, vastavalt sellele osutuvad kovariatsioonimaatrikisd kas sarnasteks või erinevateks.

Funktsiooni kordajate leidmiseks on kasutada järgmised funktsioonid:

Fisher's – Fisher'i funktsiooni kordaja.

Unstandardized – mittestandardiseeritud diskriminantfunktsiooni kordajad.

Standardiseeritud diskriminantfunktsiooni kordajad väljastatakse automaatselt

Maatriksid:

Within - groups correlation – ühendatud rühmasisene korrelatsioonimaatriks

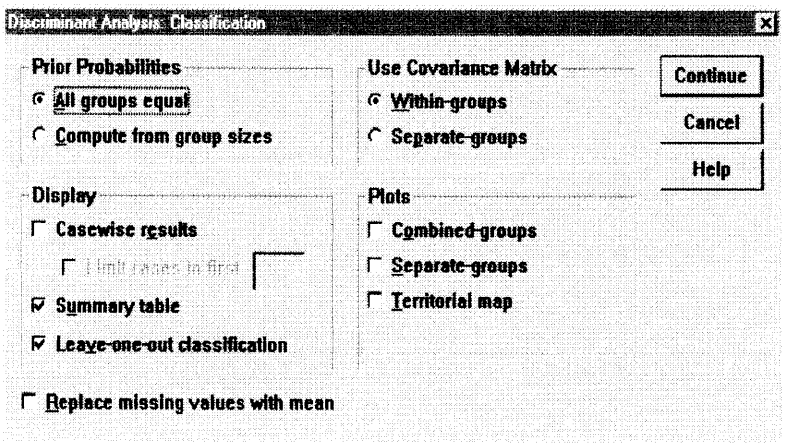
Within - groups covariance – ühendatud rühmasisene kovariatsioonimaatriks, mis on saadud erinevate rühmade kovariatsioonimaatriksite kaalutud keskmisena

Separate - groups covariance – erinevate rühmade kovariatsioonimaatriksid

Total covariance – summarne kovariatsioonimaatriks, mis saadakse rühmasisese ja rühmade vahelise hajuvuse liitmisel

- Vajuta nuppu *Classify*... ning määra tingimus aprioorsete tõenäosuste kohta, määra milliseid kovariatsioonimaatrikseid kasutatakse objektide rühmitamisel ning millised tabelid ning joonised väljastatakse.
  Automaatselt on paika pandud, et läbi viiakse lineaarne diskriminantanalüüs. Seetõttu on märgistatud *Prior Probabilities*, *All groups equal*, mis ütleb, et kõik aprioorsed tõenäosused on võrdsed ning samuti on märgistatud *Use Covariance Matrix*, *Within-Groups*, mille kohaselt kovariatsioonimaatriksid on sarnased. Pärast diskriminantanalüüsi läbiviimist võib osutuda, et kovariatsioonimaatriksid on siiski erinevad (vt. rühmade kovariatsioonimaatriksite võrdlemine, tabel 4 ja tabel 5), mistõttu võib osutuda otstarbekaks kasutada

  mittelineaarset diskriminantanalüüsi. Sellisel juhul peaks valima *All groups equal* ning *Separate – groups*.



Aprioorsete tõenäosuste määramiseks on võimalused:

All groups equal – kui puudub eelteave rühmade aprioorsete tõenäosuste kohta, siis loeb programm kõik aprioorsed tõenäosused võrdseks.

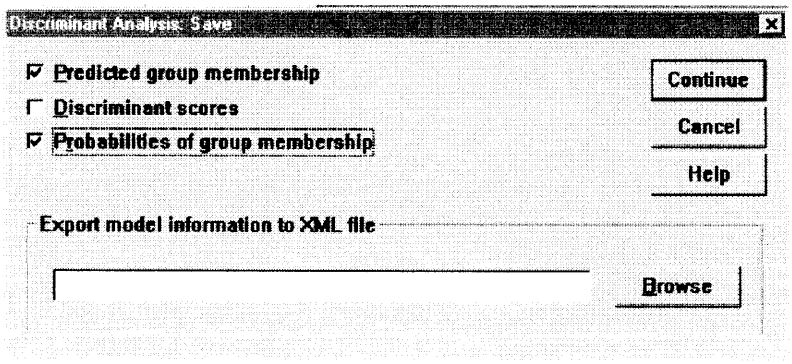
Compute from group size – kasutatakse diskrimineerimis-eeskirja koostamisel aprioorseid tõenäosusi, mis leitakse vastavalt valimi proportsioonidele. Sellisel juhul on tegemist diskriminantanalüüsi Bayesi käsitlusega.

Kovariatsioonimaatriksitest on kasutada:

Within - groups – ühendatud rühmasisest kovariatsioonimaatriksit kasutatakse *lineaarse diskriminantanalüüsi* korral

Separate – groups – iga rühma jaoks tellitakse eraldi kovariatsioonimaatriks, kasutatakse kui on tegemist *mitte-lineaarse diskriminantanalüüsiga.*

- Vajuta nuppu *Save...* ning võid lisada uusi tunnuseid oma andmefaili.



Kasutada olevad võimalused on järgmised:

Predicted group membership – objektide ennustatav rühmakuuluvus.

Discriminant scores - diskriminantfunktsioonide väärtused.

Probabilities of group membership - objektide rühmakuuluvuse tõenäosused.

DISKRIMINANTANALÜÜSI TULEMUSTE TÕLGENDAMINE

- **Klassifitseerimisfunktsiooni kordajate tabel:**

Classification Function Coefficients

	eelnev maksekohuse mittetäitmine	
	ei	ja
viimases töökohas töötatud aastate arv	,277	,109
võla suhe sissetulekusse(x 100)	,291	,386
krediitkaardi võlg tuhandetes	-,734	-,303
viimases elukohas elatud aastate arv	,145	8,489E-02
(Constant)	-3,485	-3,676

Fisher's linear discriminant functions

Tabel 1

Antud tabelis iga veerg esitab ühe rühma klassifitseerimisfunktsiooni kordajate väärtused.

Tähistame tunnuseid järgmiselt:

A – viimases töökohas töötatud aastate arv
B – võla suhe sissetulekusse (x 100)
C – krediitkaardi võlg tuhandetes
D – viimases elukohas elatud aastate arv

Tabelist järeldub, et rühma *"ei"* klassifitseerimisfunktsiooni näeb välja järgmine:

$$d_{\text{"ei"}} = 0{,}277 \times A + 0{,}291 \times B - 0{,}734 \times C + 0{,}145 \times D - 3{,}485$$

Objekt määratakse sellesse rühma, kus tema klassifitseerimisfunktsiooni väärtus on kõige suurem.

Tabelist on näha, et veerus *"ja"*, tunnuste *"viimases töökohas töötatud aastate arv"* ning *"viimases elukohas elatud aastate arv"* klassifitseerimisfunktsiooni kordajad on väikesemad võrreldes teiste kordajatega, mis viitab sellele, et kliendid, kes on töötanud ning elanud pikemat aega ühes kohas, suudavad laenu õigeaegselt tagasi maksta ning makseraskustesse sattumise tõenäosus on väiksem.

Lineaarse standardiseerimata diskriminantfunktsiooni kordajad saame, kui esimese rühma klassifitseerimisfunktsiooni kordajatest lahutame vastavad teise rühma klassifitseerimisfunktsiooni kordajad.
Antud näite korral arvutatakse tunnuste A, B, C ning D jaoks diskriminantfunktsioonikordaja b_a, b_b, b_c, b_d järgmiselt:

$$b_a = 0{,}277 - 0{,}109 = 0{,}168$$
$$b_b = 0{,}291 - 0{,}386 = -0{,}095$$
$$b_c = -0{,}734 - (-0{,}303) = -0{,}431$$
$$b_d = 0{,}145 - 8{,}489 \times 10^{-2} = 0{,}06011$$

Vastava diskriminantfunktsiooni vabaliige saadakse, kui teise rühma klassifitseerimisfunktsiooni vabaliikmest lahutada esimese rühma klassifitseerimisfunktsiooni vabaliige.
Antud näite korral saadakse diskriminantfunktsiooni vabaliige järgmiselt:

$$b = -3{,}676 - (-3{,}485) = -0{,}191$$

Lineaarse standardiseerimata diskriminantfunktsioon näeb antud näite korral välja järgmine:

$$d = 0{,}168 \times A - 0{,}095 \times B - 0{,}431 \times C + 0{,}06011 \times D - 0{,}191$$

- **Rühmade ruutmaatriksite tabel:**

Pooled Within-Groups Matrices[a]

		viimases töökohas töötatud aastate arv	võla suhe sissetulekusse(x 100)	krediitkaardi võlg tuhandetes	viimases elukohas elatud aastate arv
Covariance	viimases töökohas töötatud aastate arv	41,674	4,241	6,809	12,521
	võla suhe sissetulekusse(x 100)	4,241	40,152	6,690	6,013
	krediitkaardi võlg tuhandetes	6,809	6,690	4,319	4,094
	viimases elukohas elatud aastate arv	12,521	6,013	4,094	46,024
Correlation	viimases töökohas töötatud aastate arv	1,000	,104	,508	,286
	võla suhe sissetulekusse(x 100)	,104	1,000	,508	,140
	krediitkaardi võlg tuhandetes	,508	,508	1,000	,290
	viimases elukohas elatud aastate arv	,286	,140	,290	1,000

a. The covariance matrix has 497 degrees of freedom.

Tabel 2

Selles tabelis väljastatakse rühmadesisesed ruutmaatriksid: kovariatsioonimaatriks ja korrelatsioonimaatriks.
Kovariatsioonimaatriksis peadiagonaalil olevad väärtused näitavad dispersiooni ning ülejäänud väärtused kovariatsiooni. Tabelist on näha omadus, et kovariatsioonimaatriks on sümmeetriline.
Korrelatsioonimaatriksil on kõik peadiagonaali väärtused alati 1,0-id, sest iga tunnus on iseendaga maksimaalses seoses. Samuti nagu kovariatsioonimaatriks on ka korrelatsioonimaatriks sümmeetriline. Diskriminantanalüüsi korral **ei ole lubatud** erinevate **sõltumatute tunnuste vaheline tugev korrelatsioon**, sest selline olukord peegeldab asjaolu, et tunnuste seas võib leiduda alternatiivne tunnuste alamhulk, mis võimaldab erinevaid rühmi sama hästi eristada. Tabelist on näha, et kõige tugevam korrelatsioon on tunnuse *"krediitkaardi võlg tuhandetes"* ning teiste tunnuste vahel. Veel on raske öelda, kas see asjaolu võib diskrimineerimiseeskirja koostamist mõjutada (vt lk 16, tabel 7).

- **Rühmade statistiliste andmete tabel:**

Group Statistics

eelnev maksekohuse mittetäitmine		Mean	Std. Deviation	Valid N (listwise) Unweighted	Weighted
ei	viimases töökohas töötatud aastate arv	9,5840	6,67766	375	375,000
	võla suhe sissetulekusse(x 100)	8,8179	5,69545	375	375,000
	krediitkaardi võlg tuhandetes	1,2554	1,41769	375	375,000
	viimases elukohas elatud aastate arv	8,8800	6,94239	375	375,000
ja	viimases töökohas töötatud aastate arv	5,1855	5,72737	124	124,000
	võla suhe sissetulekusse(x 100)	14,4468	7,97554	124	124,000
	krediitkaardi võlg tuhandetes	2,3656	3,36732	124	124,000
	viimases elukohas elatud aastate arv	6,3548	6,27836	124	124,000
Total	viimases töökohas töötatud aastate arv	8,4910	6,72386	499	499,000
	võla suhe sissetulekusse(x 100)	10,2166	6,78238	499	499,000
	krediitkaardi võlg tuhandetes	1,5313	2,13087	499	499,000
	viimases elukohas elatud aastate arv	8,2525	6,86476	499	499,000

Tabel 3

Väljastatakse iga rühma jaoks kõikide tunnuste kirjeldavate statistikute väärtused. Diskriminantanalüüsi eelduse kohaselt ei tohi erinevate rühmade standarthälbed samade tunnuste korral oluliselt erineda.
Antud näite korral aga võib tekkida probleem, sest **rühmade standarthälbed erinevad teineteisest oluliselt tunnuste "võla suhe sissetulekusse" ja "krediitkaardi võlg tuhandetes" korral** (vt lk 16, tabel 7).
Veeru *Valid N Unweighted* väärtused näitavad selliste objektide arvu rühmas, mis ei oma puuduvaid väärtusi.

- **Rühmade kovariatsioonimaatriksite võrdlemine:**

Log Determinants

eelnev maksekohuse mittetäitmine	Rank	Log Determinant
ei	4	11,185
ja	4	12,253
Pooled within-groups	4	11,957

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Tabel 4

Antud tabelis veerus *Log Determinant* väljastatakse iga rühma jaoks rühmasisese kovariatsioonimaatriksi omaväärtus, mis võimaldab välja selgitada, millise rühma kovariatsioonimatriks erineb teistest kõige rohkem. Antud näite korral on eristatavaid

rühmasid ainult kaks ning veerus *Log Determinant* väljastatud väärtused näitavad, et **rühmade kovariatsioonimaatriksid erinevad teineteisest.** Selle väite paikapidavust kinnitab ka järgmine tabel.

Test Results

Box's M		252,117
F	Approx.	24,893
	df1	10
	df2	245917,2
	Sig.	,000

Tests null hypothesis of equal population covariance matrices.

Tabel 5

Antud tabelis võrreldakse kovariatsiooni maatrikseid. Nullhüpotees – kovariatsioonimaatriksid on sarnased. Sisukas hüpotees – kovariatsioonimaatriksid on erinevad. Kui olulisustõenäosus on väiksem kui 0,10 (tabelis Sig. väärtus), oleme lükanud ümber nullhüpoteesi ja tõestanud sisuka hüpoteesi olulisusnivool 10%, mis tähendab, et kovariatsioonimaatriksid on erinevad. Vastasel juhul, kui olulisustõenäosus on suurem kui 0,10, jääme nullhüpoteesi juurde, see aga ütleb meile, et kovariatsioonimaatriksid on sarnased.

Tabelist on näha, et antud näite korral olulisustõenäosus on 0%, mis näitab, et **kovariatsioonimaatriksid on erinevad**, see aga tähendab, et meil ei ole tegu lineaarse diskriminantanalüüsiga, vaid mittelineaarse diskriminantanalüüsiga. Edasise analüüsi käigus oleks mõttekas *Classification* aknas määrata erinevad kovariatsioonimaatriksid st. ***Separate-groups*** (vt. lk 19, tabel 12).

- **Tunnuste eristusvõime hindmine:**

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
viimases töökohas töötatud aastate arv	,920	43,262	1	497	,000
võla suhe sissetulekusse(x 100)	,871	73,534	1	497	,000
krediitkaardi võlg tuhandetes	,949	26,597	1	497	,000
viimases elukohas elatud aastate arv	,975	12,911	1	497	,000

Tabel 6

Veeru *Wilks' Lambda* väärtused on vahemikus 0 kuni 1,0. Mida väiksem on selles veerus toodud statistiku väärtus, seda paremini võimaldab antud tunnus rühmasid eristada. *Wilks' Lambda* väärtuste kohaselt on tunnuste paremusjärjestus hulkade eristamiseks järgmine:

1. "võla suhe sissetulekusse (x 100)"
2. "viimases töökohas töötatud aastate arv"
3. "krediitkaardi võlg tuhandetes"
4. "viimases elukohas elatud aastate arv"

Kui veerus *Sig.* olev olulisustõenäosus mingi tunnuse jaoks on suurem kui 0,10, siis eristusvõimel puudub statistiline usaldusväärsus. Meie näite puhul on kõikide tunnuste olulisustõenäosused 0%, järelikult **kõik tunnused on statistiliselt olulised rühmade eristamiseks.**

Standardized Canonical Discriminant Function Coefficients

	Function
	1
viimases töökohas töötatud aastate arv	-,784
võla suhe sissetulekusse(x 100)	,437
krediitkaardi võlg tuhandetes	,649
viimases elukohas elatud aastate arv	-,295

Tabel 7

Standardiseeritud diskriminantfunktsiooni kordajate tabel lubab võrrelda erinevates ühikutes mõõdetud tunnuseid. Absoluutväärtuselt suurimat väärtust omav tunnus eristab hulki kõige paremini.
Selle tabeli kohaselt tunnuste paremusjärjestus hulkade eristamiseks on järgmine:

1. "viimases töökohas töötatud aastate arv"
2. "krediitkaardi võlg tuhandetes"
3. "võla suhe sissetulekusse (x 100)"
4. "viimases elukohas elatud aastate arv"

Võrreldes eelmises tabelis saadud tulemustega on järjekord põhimõtteliselt sama, kuid tunnuse "võla suhe sissetulekusse (x 100)" tähtsust hindab antud tabel madalalt.
Põhjus on ilmselt selles, et korrealtsioonimaatriksi tabelis (Tabel 2) tunnuste *"krediitkaardi võlg tuhandetes"* ning teiste tunnuste vaheline **korrelatsioon on siiski piisavalt suur**, et mõjutada diskrimineerimiseeskirja koostamist.

Structure Matrix

	Function
	1
võla suhe sissetulekusse(x 100)	,644
viimases töökohas töötatud aastate arv	-,494
krediitkaardi võlg tuhandetes	,387
viimases elukohas elatud aastate arv	-,270

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

Tabel 8

Antud tabel peegeldab iga tunnuse ning diskriminantfunktsiooni vahelist korrelatsiooni. Selles tabelis on samuti nagu eelmise tabeli korralgi kõige paremini hulki eristavaks tunnuseks see, mis omab absoluutväärtuselt kõige suuremat väärtust. Selle

tabeli korral on tunnuste järjekord sama, mis hulkade keksväärtuste võrdsuse tabeli (Tabel 6) korral, kuid erineb standardiseeritud diskriminantfunktsiooni kordajate tabeli (Tabel 7) tunnuste järjekorrast.

Kokkuvõttes, et tunnuste vaheline tugev korrelatsioon on mõjutanud tabelis 7 tunnuste järjekorda ning, et tabelis 6 ja tabelis 8 on tunnuste järjestus sama, võime öelda, et kõige paremini rühmi eritavaks tunnuseks on *"võla suhe sissetulekusse (x 100)"*.

- **Diskriminantfunktsiooni klassifitseerimisvõime hindamine:**

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,357[a]	100,0	100,0	,513

a. First 1 canonical discriminant functions were used in the analysis.

Tabel 9

Omaväärtuste tabel annab informatsiooni diskriminantfunktsiooni tõhususest. Kahe rühma korral hindab diskriminantfunktsiooni eristamise võimet kõige paremini väärtus veerus *Canonical Correlation*, mis on ekvivalentne Pearsoni korrelatsioonikordajaga. See tähendab, et veerus Canonical Correlation olev väärtus paikneb –1 ja 1 vahel ning mida lähemal on antud väärtuse absoluutväärtus arvule 1, seda paremini diskriminantfunktsioon rühmi eristab.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	,737	151,007	4	,000

Tabel 10

Veerus *Wilks' Lambda* olevad väärtused näitavad, kui hästi diskriminantfunktsioonid rühmi eristavad. Väärtused asuvad vahemikus 0 kuni 1. Mida väiksem on funktsiooni väärtus antud veerus, seda paremini diskriminantfunktsioon rühmi eristab. Et antud tabelis *Wilks' Lambda* väärtus paikneb suhteliselt lähedal arvule 1, ei ole selle funktsiooni rühmade eristamisvõime kõige parem, kuid väike olulisustõenäosus veerus *Sig.* viitab sellele, et diskriminantfunktsioon suudab siiski rühmi eristada paremini kui juhuslikkuse alusel.

3.samm: Rühmitamistulemuste hindamine

Classification Results[b,c,d]

			eelnev maksekohuse mittetäitmine	Predicted Group Membership		Total
				ei	ja	
Cases Selected	Original	Count	ei	281	94	375
			ja	30	94	124
		%	ei	74,9	25,1	100,0
			ja	24,2	75,8	100,0
	Cross-validated[a]	Count	ei	278	97	375
			ja	31	93	124
		%	ei	74,1	25,9	100,0
			ja	25,0	75,0	100,0
Cases Not Selected	Original	Count	ei	106	36	142
			ja	10	49	59
			Ungrouped cases	95	55	150
		%	ei	74,6	25,4	100,0
			ja	16,9	83,1	100,0
			Ungrouped cases	63,3	36,7	100,0

← Tulemus 1

← Tulemus 2

← Tulemus 3

← Tulemus 4

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 75,2% of selected original grouped cases correctly classified.

c. 77,1% of unselected original grouped cases correctly classified.

d. 74,3% of selected cross-validated grouped cases correctly classified.

Tabel 11

Tabelist on näha, kui hästi saadud diskrimineerimisfunktsioon toimib ehk kui õigesti objekte vastavatesse rühmadesse paigutatakse.

- *Tulemus 1* - Tabel näitab, et juhusliku valikuga välja valitus 124 kliendist, kes ei suutnud laenu tagasi, paigutati õigesse rühma 94. Ning 375-st kliendist, kes suutsid laenu tagasi maksta õigeaegselt, paigutati õigesse rühma 281. See tähendab, et ligikaudu **75,2% diskriminantfunktsiooni moodustamisel kasutatud objektidest suudeti paigutada õigesse rühma**.

- *Tulemus 2* - Tihtipeale võib ridades *Original* olevad arvud olla liiga optimistlikud, seepärast on nende all esitatud *Cross-Validated* tulemused, mille korral iga objekt rühmitatakse vastavalt diskriminantfunktsioonile mingisse rühma, nii, et see objekt ise ei osale diskriminantfunktsiooni moodustamisel. Tabel näitab, et 124-st kliendist, kes sattusid makseraskustesse, suudeti paigutada õigesse rühma 93 ning 278 klienti 375st, kes ei sattunud makseraskustesse, paigutati õigesse rühma. Järelikult ligikaudu **74,3% diskriminantfunktsiooni moodustamisel kasutatud objektidest suudeti paigutada õigesse rühma**.

- *Tulemus 3* - Ridades *Cases Not Selected Original*, on esitatud nende objektide rühmitamise tulemused, mida juhuväljavõtu korral ei kaasatud diskrimineerimiseeskirja koostamisse. Tabel näitab, et **77,1% nendest objektidest paigutati õigetesse rühmadesse**. Mis tegelikult näitab, et 3 objekti 4-st paigutab saadud diskriminantfunktsioon õigetesse rühmadesse.

4.samm: tundmatu objekti määramine ühte olemasolevasse rühma

(vt. lk. 5, *diskriminantanalüüsi ülesande püstitus*)

Praeguseks oleme läbinud 3 sammu. Esimeses sammus moodustasime valimi 70%-st nendest objektidest, mis ei omanud tunnuse *mjõuetus* (eelnev maksekohuse mittetäitmine) väärtusena puuduvat väärtust. Teises sammus koostasime valimi põhjal diskrimineerimiseeskirja. Kolmandas sammus hindasime valimi põhjal koostatud diskrimineerimiseeskirja tõhusust. Käesolevas neljandas sammus vaatame, kuidas koostatud diskrimineerimiseeskiri paigutab objektid, mis omasid tunnuse *mjõuetus* väärtusena puuduvat äärtust, etteantud rühmadesse.

- *Tulemus 4* - Tabel 11. annab hinnagu 150-laenutaotleja kohta. Tabelist real *Ungrouped cases* on näha, et 55 klienti 150-st satub tõenäoliselt makseraskustesse ning 95 suudab oma laenu õigeaegselt tagasi maksta.

TULEMUSTE TÄPSUSTAMINE

Diskriminantanalüüsi tellimisel määrasime *Classification* aknas *Within-groups* st. et tegemist on lineaarse diskriminantanalüüsiga, kuid Tabel 4 kohaselt on antud andmete korral tegemist erinevate kovariatsioonimaatriksitega st. mittelineaarse diskriminant-analüüsiga, mistõttu oleks mõttekas veel kord programm käivitada ning *Classification* aknas määrata ***Separate-groups*** ning vaadata, kas see mõjutab tulemusi.

Classification Results[a,b]

			eelnev maksekohuse mittetäitmine	Predicted Group Membership		Total	
				ei	ja		
Cases Selected	Original	Count	ei	287	88	375	← Tulemus 1
			ja	31	93	124	
		%	ei	76,5	23,5	100,0	
			ja	25,0	75,0	100,0	
Cases Not Selected	Original	Count	ei	107	35	142	← Tulemus 2
			ja	10	49	59	
			Ungrouped cases	96	54	150	← Tulemus 3
		%	ei	75,4	24,6	100,0	
			ja	16,9	83,1	100,0	
			Ungrouped cases	64,0	36,0	100,0	

a. 76,2% of selected original grouped cases correctly classified.

b. 77,6% of unselected original grouped cases correctly classified.

Tabel 12

- *Tulemus 1* - Saame, et 76,2% objektidest, mis välja valiti, suudeti paigutada õigesse rühma (enne 75,2%).
- *Tulemus 2* - 77,6% objektidest, mis ei olnud kaasatud diskriminantfunktsiooni koostamisse (need 30% klientidest olid juba laenu saanud), paigutati õigesse rühma. (enne 77,1%).
- *Tulemus 3* - Tabelist realt *Cases Not Selected Ungrouped cases*, 150st tulevasest kliendist tõenäoliselt 96 ei sattu makseraskustesse ning 54 sattub.

Discriminant Anaysis: Classification aknas on alati automaatselt paikka pandud aprioorsete tõenäosuste kohta *All groups equal*. Aprioorne tõenäosus näitab hinnangut, kui suure tõenäosusega mingi objekt teatud rühma kuulub, kui mingit muud informatsiooni selle objekti kohta teada ei ole. *All groups equal* kohaselt mistahes objektil on võrdne võimalus igasse rühma kuuluda, see tähendab, et kõikide rühmade aprioorsed tõenäosused on võrdsed ning nende kogusumma on 1 (vt. tabel 13).

Prior Probabilities for Groups

eelnev maksekohuse mittetäitmine	Prior	Cases Used in Analysis	
		Unweighted	Weighted
ei	,500	375	375,000
ja	,500	124	124,000
Total	1,000	499	499,000

Tabel 13

Objektide rühmitamise tulemusi on võimalik täpsemaks muuta kohandades aprioorseid tõenäosusi rühmade suurustele. Sellisel juhul on tegemist diskriminantanalüüsi Bayes'i käsitlusega.
Selleks tuleb määrata *Classification* aknas *Compute from group sizes* ja *Within-groups*.

Prior Probabilities for Groups

eelnev maksekohuse mittetäitmine	Prior	Cases Used in Analysis	
		Unweighted	Weighted
ei	,752	375	375,000
ja	,248	124	124,000
Total	1,000	499	499,000

Tabel 14

Vastavalt meie täpsustusele on aprioorsed tõenäosused erinevate rühmade jaoks erinevad ning nende leidmisel on kasutatud valimi proportsioone. Tabeli kohaselt 75,2% klientidest ei sattu makseraskustesse laenu tagasi maksmisel, mille tulemusena diskriminantfunktsioon leitakse selliselt, et rohkem kliente rühmitatakse rühma "ei" .

Saame uue tabeli hindamaks objektide rühmitamist.

Classification Results[b,c,d]

			eelnev maksekohuse mittetäitmine	Predicted Group Membership		
				ei	ja	Total
Cases Selected	Original	Count	ei	356	19	375
			ja	75	49	124
		%	ei	94,9	5,1	100,0
			ja	60,5	39,5	100,0
	Cross-validated[a]	Count	ei	355	20	375
			ja	77	47	124
		%	ei	94,7	5,3	100,0
			ja	62,1	37,9	100,0
Cases Not Selected	Original	Count	ei	137	5	142
			ja	31	28	59
			Ungrouped cases	130	20	150
		%	ei	96,5	3,5	100,0
			ja	52,5	47,5	100,0
			Ungrouped cases	86,7	13,3	100,0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 81,2% of selected original grouped cases correctly classified.

c. 82,1% of unselected original grouped cases correctly classified.

d. 80,6% of selected cross-validated grouped cases correctly classified.

Tabel 15

Selle tabeli kohaselt on õigesti rühmitamise protsent suurem kui võrdsete aprioorsete tõenäosuste korral (Tabel 11). Kuid selliste tulemuste korral objektide hulk, mis rühmitatakse rühma "ei " on 24,2% asemel 60,5%, mis tähendab, et kliente, kellel tõenäoliselt maksejõuetust ei teki on 60,5%.

Olenevalt ülesande sisust ja eesmärkidest võib diskriminantanalüüsi läbiviimisel valida, kas objekte paigutatakse rühmadesse nii, et aprioorsed tõenäosused on võrdsed või on need leitud vastavalt valimi proportsioonidele. Antud näite korral, kus on tegemist pangaga sõltub meetodi valik panga eesmärkidest ja ülsisest laenupoliitikast. Konservatiivse laenupoliitika korral peaks pank kasutama võrdsete aprioorsete tõenäosuste meetodit, mis minimiseerib finantsriskid, kuid samas takistab klientuuri juurdekasvu. Juhul, kui panga laenupoliitika on agressiivne ning suunatud laenuklientide ringi suurendamisele, peaks ta kasutama meetodit, kus aprioorsed tõenäosused on leitud vastavalt valimi proportsioonidele. Sellise meetodiga kaasneb paraku finantsriskide oluline suurenemine.

DISKRIMINANTANALÜÜS SAMMPROTSEDUURIGA

Üks võimalus diskriminantanalüüsi ülesannet lahendada, on seda teha samm-protseduuriga. Sammprotseduuri korral valib programm etteantud tunnuste hulgast eristamisvõime alusel alamhulga.

Et SPSS-is sammprotseduuri kasutada, tuleb *Discriminant Analyses* aknas *Enter independents together* asemel märgistada *Use stepwise method* (vt lk. 8) ning edasi jätkada vastavalt eespool toodud õpetusele.

Sammprotseduuri korral on üks levinumaid tunnuste valimismeetodeid:

Lisamismeetod - selle meetodi korral **lisatakse** tunnuseid ükshaaval, alustades rühmi paremini eristavast tunnusest. Kui k tunnust on juba leitud, siis valitakse järgmiseks $k+1$ tunnuseks selline, mis parandab rühmade eristamist kõige rohkem. Valik lõpeb siis, kui ei suudeta tõestada, et $k+1$ tunnusest koosnev komplekt suudab eristada rühmasid paremini kui k tunnusest koosnev komplekt.

Järgmine tabel illustreerib lisamismeetodi toimimist diskriminantanalüüsi sammprotseduuri korral.

Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	võla suhe sissetulekusse(x 100)	1,000	73,534	
2	võla suhe sissetulekusse(x 100)	,989	79,529	,920
	viimases töökohas töötatud aastate arv	,989	48,992	,871
3	võla suhe sissetulekusse(x 100)	,710	17,903	,780
	viimases töökohas töötatud aastate arv	,710	77,226	,870
	krediitkaardi võlg tuhandetes	,533	26,575	,793
4	võla suhe sissetulekusse(x 100)	,709	18,229	,764
	viimases töökohas töötatud aastate arv	,690	62,065	,830
	krediitkaardi võlg tuhandetes	,523	30,398	,782
	viimases elukohas elatud aastate arv	,889	10,278	,752

Tabel 16

Tabelist on näha, et esimesel sammul valitakse tunnus *võla suhe sissetulekusse (×100)*, mis vastavalt lisamismeetodi toimimiseeskirjale ütleb, et valitud tunnus eristab rühmi kõigi teiste tunnustega võrreldes kõige paremini. Seejärel lisatakse teisel sammul tunnus *viimases töökohas töötatud aastate arv*, kolmandal sammul *krediitkaardi võlg tuhandetes* ning neljandal ning ühtlasi ka viimasel sammul valitakse järelejäänud tunnuste seast tunnus *viimases elukohas elatud aastate arv*. Asjaolu, et samme on ainult neli ning rohkem tunnuseid ei lisata, tähendab, et järelejäänud tunnuste lisamisega ei ole võimalik diskrimineerimiseekirja rühmade eristamisvõimet statistiliselt olulisel määral parandada.

Järgmine tabel näitab igal sammul tunnuseid, mis ei ole diskriminanteeskirja veel haaratud. Tabelist on näha, et viimaseks sammuks on neljas samm ning tunnused, mida diskriminanteeskirja ei kaasata on järgmised: *kliendi vanus aastates, leibkonna aastasissetulek tuhandetes, teised võlad tuhandetes, haridustase*.

Variables Not in the Analysis

Step		Tolerance	Min. Tolerance	F to Enter	Wilks' Lambda
0	kliendi vanus aastates	1,000	1,000	9,631	,981
	viimases töökohas töötatud aastate arv	1,000	1,000	43,262	,920
	leibkonna aastasissetulek tuhandetes	1,000	1,000	6,747	,987
	krediitkaardi võlg tuhandetes	1,000	1,000	26,597	,949
	teised võlad tuhandetes	1,000	1,000	5,599	,989
	võla suhe sissetulekusse(x 100)	1,000	1,000	73,534	,871
	viimases elukohas elatud aastate arv	1,000	1,000	12,911	,975
	haridustase	1,000	1,000	4,743	,991
1	kliendi vanus aastates	,985	,985	15,169	,845
	viimases töökohas töötatud aastate arv	,989	,989	48,992	,793
	leibkonna aastasissetulek tuhandetes	,998	,998	7,825	,858
	krediitkaardi võlg tuhandetes	,742	,742	,752	,870
	teised võlad tuhandetes	,663	,663	8,976	,856
	viimases elukohas elatud aastate arv	,980	,980	20,367	,837
	haridustase	1,000	1,000	4,829	,863
2	kliendi vanus aastates	,715	,715	,065	,793
	leibkonna aastasissetulek tuhandetes	,580	,575	4,683	,785
	krediitkaardi võlg tuhandetes	,533	,533	26,575	,752
	teised võlad tuhandetes	,483	,483	,528	,792
	viimases elukohas elatud aastate arv	,906	,906	6,563	,782
	haridustase	,980	,970	1,374	,791
3	kliendi vanus aastates	,703	,524	,825	,751
	leibkonna aastasissetulek tuhandetes	,475	,437	,001	,752
	teised võlad tuhandetes	,437	,437	,775	,751
	viimases elukohas elatud aastate arv	,889	,523	10,278	,737
	haridustase	,956	,520	,127	,752
4	kliendi vanus aastates	,491	,491	1,013	,736
	leibkonna aastasissetulek tuhandetes	,464	,435	,227	,737
	teised võlad tuhandetes	,433	,433	,341	,737
	haridustase	,947	,512	,427	,736

Tabel 17

Kokkuvõttes oleme saanud sama tulemuse, mis varem, kui valisime diskrimineerimis-eeskirja jaoks tunnused ise (vt tabel 6 ja tabel 8).

Sammprotseduuri korral valitakse tunnuseid nende väärtuste statistilistest seisukohtadest ning seetõttu tuleb silmas pidada, et antud meetod võib valida rühmade eristamiseks küll tugevalt seotud, kuid reaalselt mitte nii suurt mõju omavaid tunnuseid, mistõttu ei ole sammprotseduuri alati otstarbekas kasutada.

LÕPPSÕNA

Et keerukate ülesannete puhul on alati mõistlik ning kasulik alustada lihtsamatest meetoditest, siis olen käesoleva materjali korral tutvustanud diskriminantanalüüsi kahe rühma eristamise juhtu. Selle läbiviimiseks valisin esialgu diskriminantanalüüsi liikidest lineaarse diskriminantanalüüsi ning seejärel vaatlesin teisi diskriminant-analüüsi meetodeid tulemuste järkjärguliseks parandamiseks.

Et diskriminantanalüüsi kui statistika meetodit on võimalik kasutada erinevaid valdkondi puudutavate ülesannete lahendamisel ning tihtipeale on elust enesest välja kasvavad ülesanded matemaatiliselt küllaltki keerulise lahenduskäiguga, siis võib tekkida vajadus lahendada ülesannet mingisugusel üldisemal juhul (eristatavate rühmade arv on suurem kui 2). Et käesolev materjal ei kajasta diskriminantanalüüsi läbiviimist üldisematel juhtudel, on võimalik sellesisulise ülesande korral abi saada kasutatud kirjanduse loetelus toodud raamatust: Säde Koskel, Ene – Margit Tiit, Paul Arandi. *Diskriminantanalüüs*. Tartu Ülikool, 1998. Olgu seejuures märgitud, et antud raamat eeldab suhteliselt sügavaid matemaatilisi teadmisi. Samuti võib leida diskriminantanalüüsiülesande lahendamisel raamatus kasutatava programmi SAS ning käesolevas materjalis kasutatava programmi SPSS vahel olulisi erinevusi.

Et käesolev proseminaritöö on minu esimene kirjutis informaatika vallas, siis võib kindlasti leida momente, mida oleks võinud teisiti teha.

Antud proseminaritöö koostamisel sain oma juhendajalt hulga huvitavaid ja kasulikke nõuandeid, mis aitasid kaasa materjali koostamisele ning aitasid üle proseminaritöö kirjutamise ajal tekkinud kitsaskohtadest. Loodan, et antud materjal leiab ka reaalset kasutust.

KASUTATUD KIRJANDUS

1. Säde Koskel, Ene –Margit Tiit, Paul Arandi. *Diskriminantanalüüs*. Tartu Ülikool, 1998.

2. http://www.ms.ut.ee/ained/Diskrim/Diskrimkava.htm

3. A.A. Afifi, Virginia Clark. *Computer – Aided Multivariate Analysis. Second Edition.* Chapman& Hall.