

TALLINNA ÜLICOOL

Diskriminantanalüüs

Gruppi kuuluvuse prognoosimine

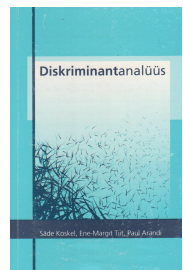
Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLICOOL

Kirjandus

Koskel, S., Tiit, E.-M., Arandi, P. **Diskriminantanalüüs** 1998 Tartu Ülikool

Koppel, K. **Diskriminantanalüüs programmi SPSS abil**. Õppematerjali loomine. 2002 Tallinna Ülikool
<http://www.tlu.ee/~kairio/7071mm/diskriminantanalyyys.pdf>




Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLICOOL

Diskriminantanalüüs

- "Diskrimineerimine"
 - Kellegi õiguste kitsendamine või kärpimine teistega võrreldes, kellessegi halvustavalt suhtumine (Eesti keele seletav sõnaraamat, EKSS)
 - Õiguste kärpimine, kellessegi halvasti suhtumine, kellegi kõrvale tõrjumine (Eesti õigekeelsussõnaraamat ÕS 2013)
 - Diskrimineerimine tööturul tähendab inimeste **eristamist** nende kvalifikatsiooni või võimete seisukohast ebaoluliste tunnuste alusel (vikipeedia)
- Diskriminantanalüüs tegeleb
 - olemasolevate objektide **gruppidesse paigutamise** (koostatakse mudel)
 - uutele objektidele (mudeli põhjal) sobiva **grupi prognoosimisega**
- Diskriminantanalüüs
 - on tehnika, mida kasutatakse objektide võimalikult täpselt eristamiseks ehk rühmitamiseks etteantud gruppidesse (eeldatakse, et üldkogum jaguneb lõplikuks hulgaks rühmadeks/klassideks ning eeskirja loomise aluseks oleva valimi puhul on rühma kuuluvus teada). (K.Niglasel loenguslaudid)

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLICOOL

Näide

Diskriminantanalüüs  õiguse ja ebasõbraliku suhtumise

Diskriminantanalüüsi teostamiseks analüüsitakse pärast loendust kahte inimeste rühma (nn õpperühmad):

1. „kindlad“ residendid – need, kes on püsielanikena loendatud ja kes kuuluvad RR-s Eesti elanike hulka
2. „kindlad“ mitteresidendid – loendatud lahkunutena, ei ela RR andmetel Eestis

ja võrreldakse nende esindatust registrites.

Kõigi registriandmete põhjal arvutatakse välja niisugune funktsioon, mis neid kahte rühma kõige paremini eristab. See on **diskriminantfunktsioon**.

Diskriminantfunktsiooni väärtus on keskmise mitteresidendi korral 0 ja keskmise residendi korral 1.

28.11.2012 Ene-Margit Tiit

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLICOOL

Näide 1

ÕS EESTI KEELEREGLID Otsustusreegel  õiguse ja ebasõbraliku suhtumise

Analüüsitakse neid inimesi, kes on **RR andmetel Eesti elanikud, aga keda ei ole loendatud**. Nende kohta tuleb otsus langetada – kas nad on residendid või mitte.

Iga niisuguse inimese jaoks arvutatakse diskriminantfunktsiooni väärtus – see on seda suurem, mida rohkemates olulistes registrites ta 2011. aastal esines.

Need inimesed, kelle puhul diskriminantfunktsiooni väärtus on 1 lähedal, loetakse residentide hulka. Need, kelle puhul funktsiooni väärtus on 0 lähedal, loetakse mitteresidentideks.

Residente mitteresidentidest eristav funktsiooni väärtus (lävend) määratakse vastavalt lubatavale vea tõenäosusele.

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLICOOL

Diskriminantanalüüs

- **Sõltuv tunnus DV**
 - Kirjeldab gruppikuuluvust
 - Grupeeriv tunnus
- **Sõltumatud tunnused IV**
 - 2 või enam sõltumatut tunnust
 - Nende tunnuste põhjal prognoositakse gruppi-kuuvust
 - IV väärtused on olemas ning nende teadasaamine on lihtsam kui DV mõõtmine, vastasel juhul ei oleks mõtet mudelit looma hakata
 - Mõõdetud intervallskaalal

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLIKOOL

Diskriminantanalüüs SPSS-s

psychology.sav

group	exper	stats	social	develop	person
1st Year	10.00	15.00	14.00	14.00	13.00
1st Year	7.00	8.00	10.00	9.00	10.00

- Sõltuv tunnus** – mitmenda kursuse üliõpilane
- Sõltumatud tunnused** – tulemused erinevates ainetes (15 palli skaalal)
- Kuivõrd aitavad erinevates ainetes saadud tulemused eristada gruppi kuuluvust** e olemasolevate andmete jaoks on meil teada gruppi kuuluvus, nüüd tahame teada, kui hästi on võimalik prognoosida ainet tulemuste põhjal, millisesse gruppi (milline kursus) vastav objekt kuulub?
- Esmalt** uuri andmeid kirjeldava statistika meetodite abil. Interpreteeri saadud tulemusi (kasulik edaspidiseks analüüsiks).

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLIKOOL

Diskriminantanalüüs SPSS-s

• Analyze/Classify/Discriminant...

Grouping Variable
sõltuv tunnus

Independents
Sõltumatud tunnused

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLIKOOL

Diskriminantanalüüs SPSS-s

Descriptives

- ☒ Means: sõltumatute tunnuste keskväärtused ja standardhälbed (Output: Group Statistics)
- ☒ Univariate ANOVA: gruppide keskväärtuste võrdlemine (võrreldakse(eraldi) iga sõltumatu tunnuse keskväärtuseid sõltuva tunnuse gruppide lõikes)
- (Output: Test of Equality of Group Means)
- ☒ Box's M: kovariatsioonide võrdsuse kontrolli (üks eeldustest)
- (Output: Box's Test of Equality of Covariance Matrixes)
- ☒ Unstandardized – mittestandardiseeritud diskriminantanalüüsi võrrand (Output: Canonical Discriminant Function Coefficients)

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLIKOOL

Diskriminantanalüüs SPSS-s

Prior Probabilities

- ☒ All groups equal: sõltuva tunnuse gruppide suurusel on võrdsed (või enam-vähem võrdsed)
- ☒ Compute from group sizes: mitte-võrdsed gruppide suurusel

Display

- ☒ Summary table: kokkuvõtte sellest, kui õigesti meie mudel väärtuseid prognoosib
- ☒ Leave-one-out classification: kui hästi prognoosib mudel väärtuseid, kui jätta objektid kordamööda välja e hinnang prognoosile

Use Covariance Matrix

- ☒ Within- groups: lineaarne DA (Box's M testi põhjal)
- ☒ Separate-groups: mittelineaarne DA

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLIKOOL

Diskriminantanalüüs SPSS-s

Predicted group membership:
prognoositav gruppikuuluvus

Probabilities of group membership: kui suure tõenäosusega kuulub objekt ühte või teise gruppi

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLIKOOL

Diskriminantanalüüs SPSS-s

		Group Statistics		Valid N (listwise)	
Group		Mean	Std. Deviation	Unweighted	Weighted
1st Year	Experimental Psychology	5.6364	2.15744	11	11.000
	Statistics	7.5455	3.55988	11	11.000
	Social Psychology	10.3636	2.73030	11	11.000
	Developmental	11.0000	2.64575	11	11.000
	Personality	10.6364	3.32484	11	11.000
2nd Year	Experimental Psychology	5.5000	1.59164	16	16.000
	Statistics	8.6875	2.38659	16	16.000
	Social Psychology	8.5625	2.80401	16	16.000
	Developmental	8.8750	1.70783	16	16.000
	Personality	8.4375	1.99896	16	16.000
Total	Experimental Psychology	5.5556	1.80455	27	27.000
	Statistics	8.2222	2.91328	27	27.000
	Social Psychology	9.2963	2.86645	27	27.000
	Developmental	9.7407	2.34673	27	27.000
	Personality	9.3333	2.78733	27	27.000

Kirjeldavad arvnäitajad (keskväärtus, standardhälve, objektide arv)

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLIKOO

Diskriminantanalüüs SPSS-s

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Experimental Psychology	.999	.036	1	25	.851
Statistics	.961	1.002	1	25	.326
Social Psychology	.901	2.747	1	25	.110
Developmental	.794	6.469	1	25	.018
Personality	.844	4.622	1	25	.041

Statistiliselt olulised erinevused 1. ja 2. kursuse üliõpilaste erinevate ainete tulemuste keskvaartuste vahel olid arengu- ja isiksusepsühholoogia ainetes ($p < 0,05$).

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLIKOO

Diskriminantanalüüs SPSS-s

Box's Test of Equality of Covariance Matrices

Log Determinants

Group	Rank	Log Determinant
1st Year	5	6.631
2nd Year	5	4.031
Pooled within-groups	5	5.960

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

Test Results

Box's M	22.211
F	1.135
Approx. df1	15
df2	1841.347
Sig.	.318

Tests null hypothesis of equal population covariance matrices.

Testib kovariatsioonide võrdust
 Ho: kovariatsioonide maatriksid on ühesugused (erinevus puudub)
 H1: kovariatsioonide maatriksid on erinevad
 $\alpha = 0,05$

Sig=0,318 => jääme Ho juurde
eeldus on täidetud

(kui ei oleks täidetud, siis peaksime diskriminantanalüüsi aknas valida:
Use Covariance Matrix
 ☉ Separate-groups: mittelineaarne DA

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLIKOO

Diskriminantanalüüs SPSS-s

Summary of Canonical Discriminant Functions

Kas kogu mudel on statistiliselt oluline e kas mudel eristab objekte?

Ho: sõtumatute tunnuste lisamine mudelisse ei aita paremini prognoosida
 H1: sõtumatute tunnuste lisamine mudelisse aitab paremini prognoosida
 $\alpha = 0,05$
 Sig=0,031 (Wilks Lambda)
 Mudel tervikuna on statistiliselt oluline e väärtuste prognoosimiseks ei vaja me alginfot.

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.728 ^a	100.0	100.0	.649

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

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.579	12.306	5	.031

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLIKOO

Diskriminantanalüüs SPSS-s

Classification Results^{b,c}

Original	Count	Predicted Group Membership		Total
		1st Year	2nd Year	
1st Year	9	2	11	
2nd Year	2	14	16	
Ungrouped cases	3	10	13	
%				
1st Year	81.8	18.2	100.0	
2nd Year	12.5	87.5	100.0	
Ungrouped cases	23.1	76.9	100.0	

Algandmete põhjal (Original) prognoositi (Predicted Group Member):
 1 kursuse üliõpilastest prognoositi õigesti 9 objekti väärtused e 81,8%
 valesti 2 objekti väärtused e 18,2%
 2 kursuse üliõpilastest prognoositi õigesti 14 objekti väärtused e 87,5%
 valesti 2 objekti väärtused e 12,5%

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLIKOO

Diskriminantanalüüs SPSS-s

Cross-validated ^a	Count	1st Year	2nd Year	Total
1st Year	5	6	11	
2nd Year	4	12	16	
%				
1st Year	45.5	54.5	100.0	
2nd Year	25.0	75.0	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. 85.2% of original grouped cases correctly classified.
 c. 63.0% of cross-validated grouped cases correctly classified.

Valimis prognoositi õigesti väärtuseid 85,2% juhtudest

☒ Leave-one-out classification: kui hästi prognoosib mudel väärtuseid, kui jätta objektid kordamööda välja (Tehakse mudel, millest jäetakse välja ühe objekti väärtused. Mudeli põhjal prognoositakse selle ühe objekti väärtust ja hinnatakse, kui õige on prognoos. Seda tegevust korratakse niikaua, kuni kõik objektid on mudelist ükskhaaval välja jäetud.)
 63% juhtudest prognoositi sel viisil õigesti

Mitmemõtteline statistika Kairi Osula 2013/kevad

TALLINNA ÜLIKOO

Diskriminantanalüüs SPSS-s

Canonical Discriminant Function Coefficients

	Function 1
Experimental Psychology	.209
Statistics	-.434
Social Psychology	.197
Developmental	.274
Personality	.096
(Constant)	-2.985

Unstandardized coefficients

$$D = -2,985 + 0,209X_1 - 0,434X_2 + 0,197X_3 + 0,274X_4 + 0,096X_5$$

X_1 – eksperimentaalne ps X_2 – statistika
 X_3 – sotsiaalps X_4 – arengups
 X_5 – isiksuseps

Mitmemõtteline statistika Kairi Osula 2013/kevad