

➤ Lineare Diskriminanzanalyse

Ein sehr kurzer Einblick

- 1 Beispiele
- 2 Grundgedanken der DA
- 3 Diskriminanzkriterium
- 4 Schätzung der Diskriminanzfunktion
- 5 Teststatistiken
- 6 Klassifizierung
- 7 Schätzung der Klassifikationsfehler

Literatur: Rudolf & Müller (2. Auflage 2012), Kapitel 4, S. 149-181

➤ Lineare Diskriminanzanalyse

Beispiel - Fahrleistungen

In einem fiktiven Beispiel soll eine Diskriminanzfunktion gebildet werden, die die Zuordnung von Autofahrern zu einer der drei Gruppen

- Fahranfänger (<200 km Fahrleistung)
- Unsicherer Fahrer (200-2000 km Fahrleistung)
- Sicherer Fahrer (>2000 km Fahrleistung auf Grund von Testergebnissen ermöglicht)

Folgende Parameter werden im Test erhoben:

- mittlere Reaktionszeit in unerwarteten Verkehrssituationen (zeit)
- Fehlerhäufigkeit in komplexen Verkehrssituationen (fehler)
- Sicherheit in Verkehrsregeln (sicher)
- Selbsteinschätzung der Fahrfähigkeiten (selbst)

Folgende Daten wurden in einer Stichprobe von 30 Autofahrern erfasst:

➤ Lineare Diskriminanzanalyse

Beispiel - Fahrleistungen

➤ Pb	Gruppe	Zeit	Fehler	sicher	selbst
➤ 1	Anfänger	45	13	17	34
➤ 2	unsicher	41	11	17	33
➤ 3	sicher	48	16	19	45
➤ 4	Anfänger	49	17	14	34
➤ 5	unsicher	42	13	15	33
➤ 6	sicher	38	11	15	35
➤ 7	Anfänger	46	18	15	39
➤ 8	unsicher	41	15	16	40
➤ 9	sicher	40	11	14	40
➤ 10	Anfänger	47	17	15	39
➤ 11	unsicher	42	14	16	38
➤ 12	sicher	37	10	14	35
➤ 13	Anfänger	48	18	16	36
➤ 14	unsicher	44	14	14	36
➤ 15	sicher	40	11	14	37

➤ Lineare Diskriminanzanalyse

Beispiel - Fahrleistungen

➤ Pb	Gruppe	zeit	fehler	sicher	selbst
➤ 16	Anfänger	46	13	17	34
➤ 17	unsicher	42	12	18	37
➤ 18	sicher	48	16	19	45
➤ 19	Anfänger	49	17	14	37
➤ 20	unsicher	42	13	15	37
➤ 21	sicher	38	11	15	35
➤ 22	Anfänger	46	18	15	39
➤ 23	unsicher	41	15	16	43
➤ 24	sicher	40	12	14	43
➤ 25	Anfänger	48	17	15	39
➤ 26	unsicher	42	14	16	38
➤ 27	sicher	37	10	14	36
➤ 28	Anfänger	48	18	16	36
➤ 29	unsicher	44	13	14	36
➤ 30	sicher	40	11	14	37

➤ Lineare Diskriminanzanalyse

Grundgedanken der DA

Hauptziel:

- Zuordnung „neuer“ Probanden (Versuchspersonen) zu einer der gegebenen Gruppen aufgrund der erfassten Merkmale

weitere Anliegen:

- Untersuchung der Bedeutung der untersuchten Variablen für die Unterscheidung der verglichenen Stichproben
- Ermittlung einer Linearkombination der erfassten Merkmale, die eine maximale Unterscheidbarkeit der gegebenen Gruppen gewährleistet (Diskriminanzfunktion)
- Beurteilung der Möglichkeiten zur Gruppeneinteilung „neuer“ Personen
- Auswahl des diagnostisch aussagekräftigsten Merkmalsatzes (optimale Merkmalsmenge)
- Bestimmung diagnostisch redundanter Merkmale
- statistische Beurteilung der gefundenen Klassifikation

➤ Lineare Diskriminanzanalyse

Diskriminanzkriterium

Lineare Diskriminanzfunktion:

$$d_i = c_0 + c_1 X_{1i} + c_2 X_{2i} + \dots + c_k X_{ki}$$

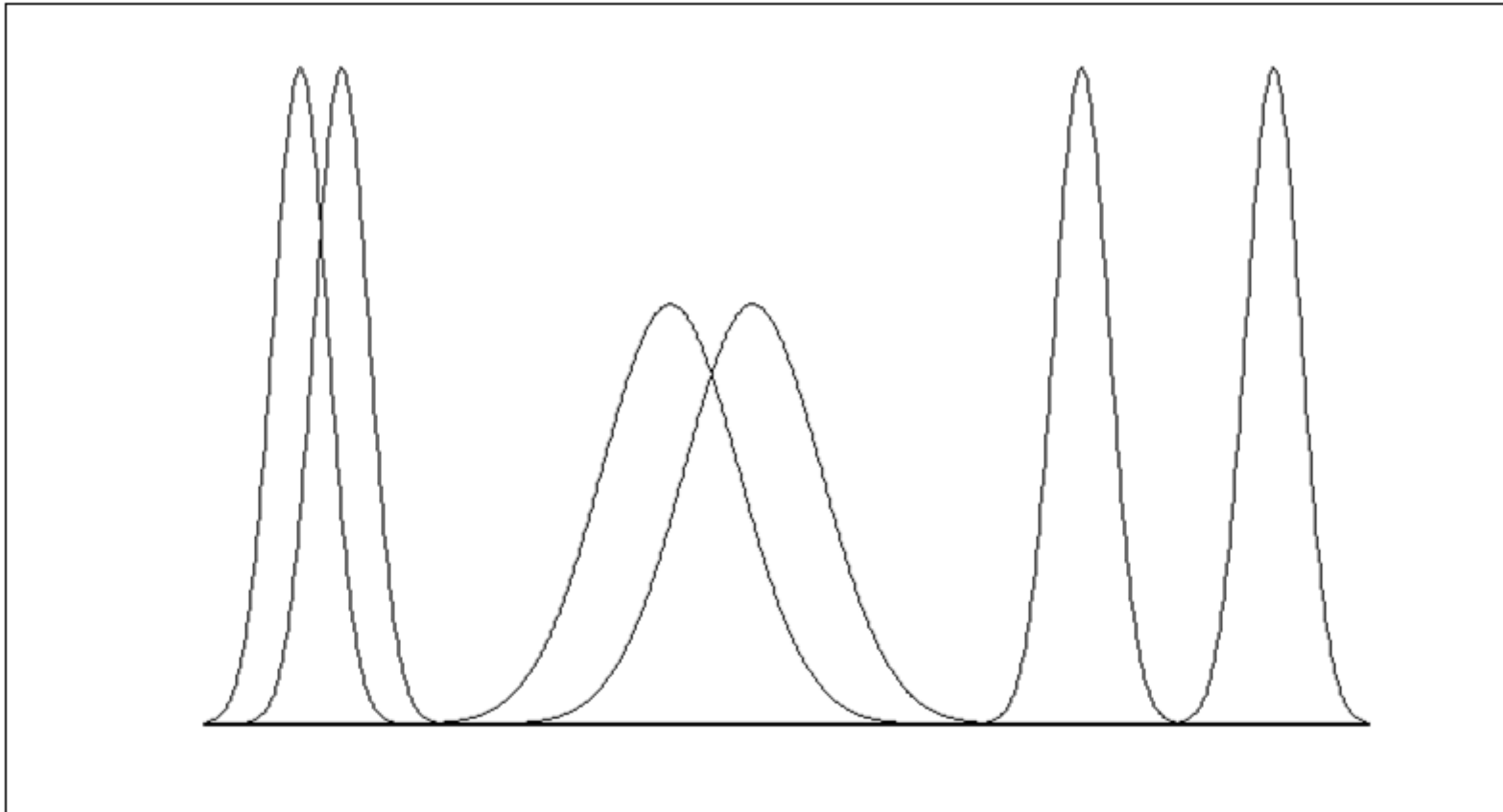
Kriterium für die Bestimmung der Diskriminationsfunktionskoeffizienten c_i :
maximale Trennung der gegebenen Gruppen

Kriterien für gute Trennbarkeit bei gegebenen Gruppen:

Differenz der Gruppenmittelwerte möglichst groß,
gleichzeitig Summe der Streuungen möglichst klein

➤ Lineare Diskriminanzanalyse

Diskriminanzkriterium



Verteilungen mit unterschiedlichen Abständen der Mittelwerte und unterschiedlichen Streuungen

➤ Lineare Diskriminanzanalyse

Diskriminanzkriterium

Mathematisch formuliert (am Beispiel von 2 Gruppen A und B):

$$|\bar{d}_A - \bar{d}_B| \rightarrow \text{MAX} \quad \text{gleichzeitig} \quad SQ_A + SQ_B \rightarrow \text{MIN}$$

$$\Gamma = \frac{(\bar{d}_A - \bar{d}_B)^2}{SQ_A + SQ_B} \rightarrow \text{MAX} \quad -$$

Kriterien für gute Trennbarkeit bei $g=3$ und $g>3$ gegebenen Gruppen:

$$\Gamma = \frac{\text{Streuung zwischen den Gruppen}}{\text{Streuung innerhalb der Gruppen}} \rightarrow \text{MAX},$$

$$\gamma = \Gamma_{\max} : \textit{Eigenwert}$$

➤ Lineare Diskriminanzanalyse

Diskriminanzkriterium

$$\Gamma = \frac{\text{Streuung zwischen den Gruppen}}{\text{Streuung innerhalb der Gruppen}} = \frac{SS_{\text{zwischen}}}{SS_{\text{innerhalb}}} = \frac{\sum_{l=1}^g n_g \cdot (\bar{d}_l - \bar{d})^2}{\sum_{l=1}^g \sum_{i=1}^{n_g} (d_{li} - \bar{d}_l)^2}$$
$$SS_{\text{gesamt}} = \sum_{l=1}^g \sum_{i=1}^{n_g} (d_{li} - \bar{d})^2$$

Quadratsummenzerlegung:

$$SS_{\text{gesamt}} = SS_{\text{zwischen}} + SS_{\text{innerhalb}}$$

Optimierungsproblem: $\max_{c_0, c_1, c_2, \dots, c_k} \Gamma$

➤ Lineare Diskriminanzanalyse

Teststatistiken

➤ Wilks' Lambda:

$$\Lambda = \frac{1}{1 + \gamma} = \frac{\text{Residualstreuung}}{\text{Gesamtstreuung}}$$

➤ Kanonischer Korrelationskoeffizient:

$$c = \sqrt{\frac{\gamma}{1 + \gamma}} = \sqrt{\frac{\text{erklärte Streuung}}{\text{Gesamtstreuung}}}$$

Im Fall von 2 Gruppen identisch mit der Korrelation zwischen den geschätzten Diskriminanzfunktionswerten und der Gruppenvariablen.

➤ Lineare Diskriminanzanalyse

Teststatistiken

- H_1 : Die Gruppen unterscheiden sich
- H_0 : Die Gruppen unterscheiden sich nicht

Test der Diskriminanzfunktion auf der Basis von Wilks' Lambda:

$$X^2 = - \left[n - \frac{k + g}{2} - 1 \right] \cdot \ln(\Lambda)$$

(k: Anzahl der Variablen, g: Anzahl der Gruppen)

- Bei Gültigkeit von H_0 Chi-Quadrat-verteilt mit $k \cdot (g-1)$ FG

➤ Lineare Diskriminanzanalyse

Ergebnisse Beispiel Fahrtstest

Gruppenstatistik

Fahrergruppe		Mittelwert	Standardabweichung	Gültige Werte (listenweise)	
				Ungewichtet	Gewichtet
Anfänger	Reaktionszeit	47.2000	1.39841	10	10.000
	Fehler	16.6000	1.95505	10	10.000
	Regelsicherheit	15.4000	1.07497	10	10.000
	Selbsteinschätzung	36.7000	2.21359	10	10.000
unsichere Fahrer	Reaktionszeit	42.1000	1.10050	10	10.000
	Fehler	13.4000	1.26491	10	10.000
	Regelsicherheit	15.7000	1.25167	10	10.000
	Selbsteinschätzung	37.1000	2.99815	10	10.000
sichere Fahrer	Reaktionszeit	40.6000	4.08792	10	10.000
	Fehler	11.9000	2.23358	10	10.000
	Regelsicherheit	15.2000	2.04396	10	10.000
	Selbsteinschätzung	38.8000	4.13118	10	10.000
Gesamt	Reaktionszeit	43.3000	3.79791	30	30.000
	Fehler	13.9667	2.68435	30	30.000
	Regelsicherheit	15.4333	1.47819	30	30.000
	Selbsteinschätzung	37.5333	3.23487	30	30.000

Gleichheitstest der Gruppenmittelwerte

	Wilks-Lambda	F	df1	df2	Signifikanz
Reaktionszeit	.428	18.065	2	27	.000
Fehler	.448	16.607	2	27	.000
Regelsicherheit	.980	.275	2	27	.761
Selbsteinschätzung	.918	1.205	2	27	.315

➤ Lineare Diskriminanzanalyse

Ergebnisse Beispiel Fahrtstest

Eigenwerte

Funktion	Eigenwert	% der Varianz	Kumulierte %	Kanonische Korrelation
1	4.548 ^a	96.6	96.6	.905
2	.162 ^a	3.4	100.0	.373

a. Die ersten 2 kanonischen Diskriminanzfunktionen werden in dieser Analyse verwendet.

Wilks' Lambda

Test der Funktion(en)	Wilks-Lambda	Chi-Quadrat	df	Signifikanz
1 bis 2	.155	47.520	8	.000
2	.861	3.829	3	.281

**Standardisierte kanonische
Diskriminanzfunktionskoeffizienten**

	Funktion	
	1	2
Reaktionszeit	.382	1.246
Fehler	1.232	-1.058
Regelsicherheit	-.053	-.945
Selbsteinschätzung	-1.265	.811

➤ Lineare Diskriminanzanalyse

Ergebnisse Beispiel Fahrtstest

Struktur-Matrix

	Funktion	
	1	2
Reaktionszeit	.535*	.480
Fehler	.519*	.189
Selbsteinschätzung	-.124	.343*
Regelsicherheit	.018	-.342*

Gemeinsame Korrelationen innerhalb der Gruppen zwischen Diskriminanzvariablen und standardisierten kanonischen Diskriminanzfunktionen. Variablen sind nach ihrer absoluten Korrelationsgröße innerhalb der Funktion geordnet.

*. Größte absolute Korrelation zwischen jeder Variablen und einer Diskriminanzfunktion

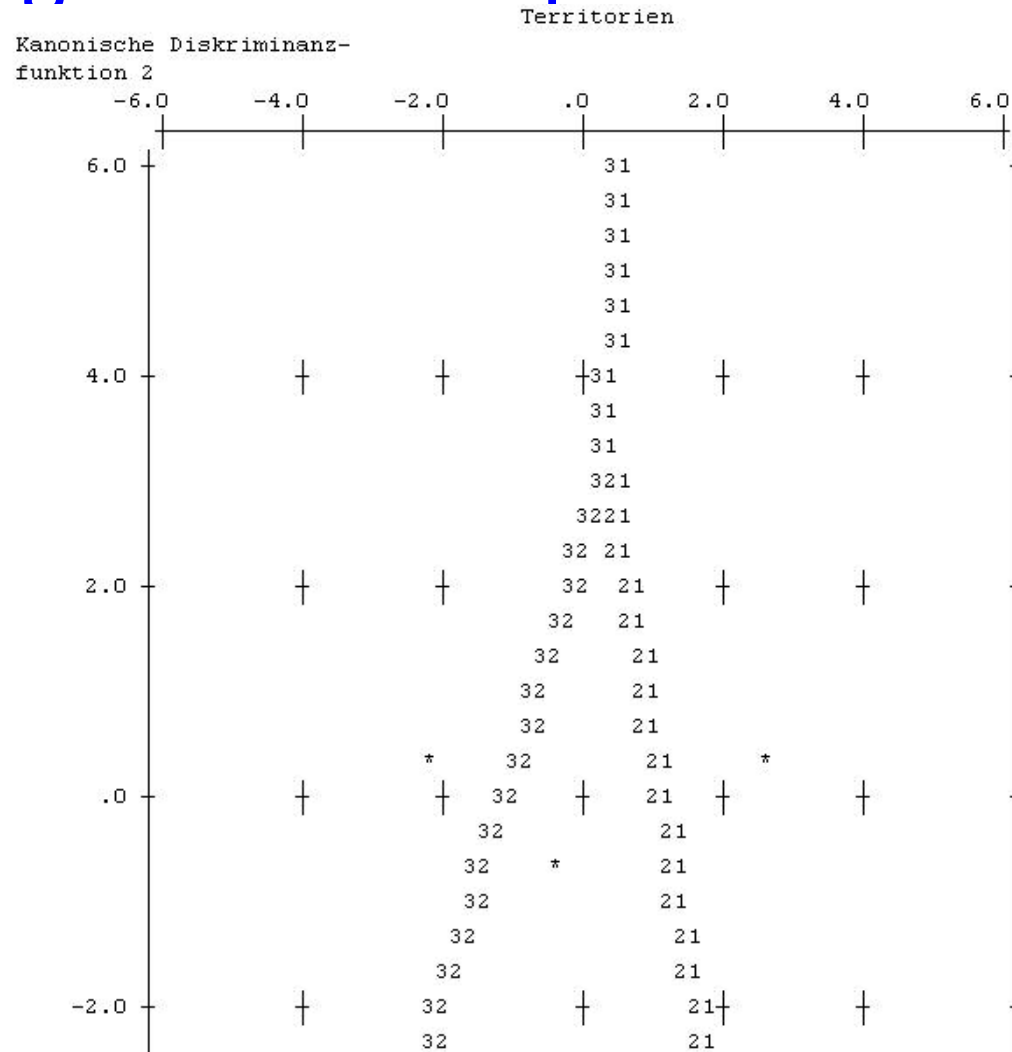
Kanonische Diskriminanzfunktionskoeffizienten

	Funktion	
	1	2
Reaktionszeit	.149	.484
Fehler	.661	-.568
Regelsicherheit	-.035	-.623
Selbsteinschätzung	-.394	.253
(Konstant)	-.346	-12.899

Nicht-standardisierte Koeffizienten

➤ Lineare Diskriminanzanalyse

Ergebnisse Beispiel Fahrtstest



Kanonische Diskriminanzfunktionskoeffizienten

	Funktion	
	1	2
Reaktionszeit	.149	.484
Fehler	.661	-.568
Regelsicherheit	-.035	-.623
Selbsteinschätzung	-.394	.253
(Konstant)	-.346	-12.899

Nicht-standardisierte Koeffizienten

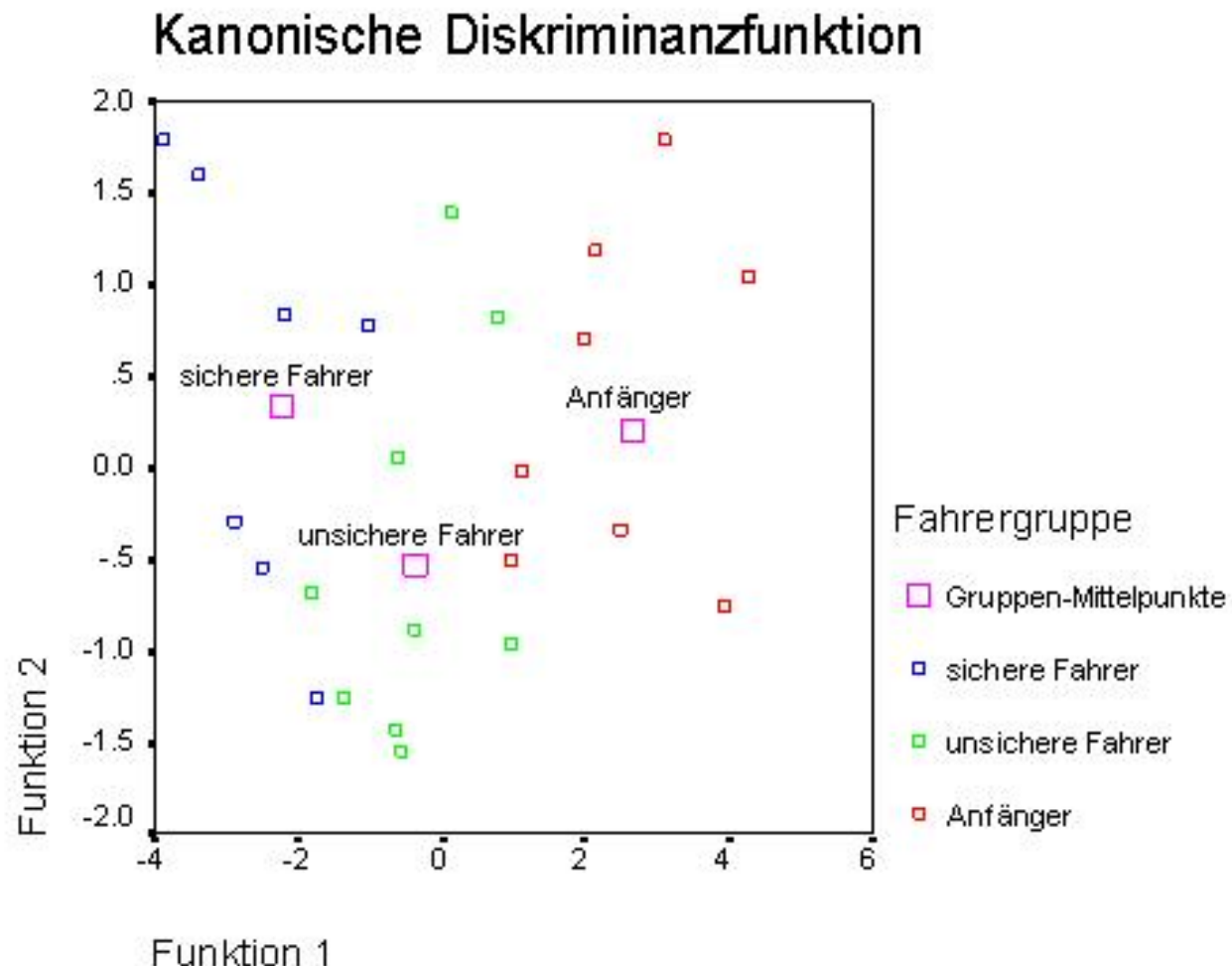
Funktionen bei den Gruppen-Zentroiden

Fahrergruppe	Funktion	
	1	2
Anfänger	2.650	.203
unsichere Fahrer	-.392	-.535
sichere Fahrer	-2.259	.331

Nicht-standardisierte kanonische Diskriminanzfunktionen, die bezüglich des Gruppen-Mittelwertes bewertet werden

➤ Lineare Diskriminanzanalyse

Ergebnisse Beispiel Fahrtstest



➤ Lineare Diskriminanzanalyse

Ergebnisse Beispiel Fahrtstest

Klassifizierungsergebnisse^{b,c}

			Vorhergesagte Gruppenzugehörigkeit			Gesamt
			Anfänger	unsichere Fahrer	sichere Fahrer	
Original	Anzahl	Anfänger	9	1	0	10
		unsichere Fahrer	0	9	1	10
		sichere Fahrer	0	2	8	10
	%	Anfänger	90.0	10.0	.0	100.0
		unsichere Fahrer	.0	90.0	10.0	100.0
		sichere Fahrer	.0	20.0	80.0	100.0
Kreuzvalidiert ^a	Anzahl	Anfänger	8	2	0	10
		unsichere Fahrer	1	8	1	10
		sichere Fahrer	0	4	6	10
	%	Anfänger	80.0	20.0	.0	100.0
		unsichere Fahrer	10.0	80.0	10.0	100.0
		sichere Fahrer	.0	40.0	60.0	100.0

a. Die Kreuzvalidierung wird nur für Fälle in dieser Analyse vorgenommen. In der Kreuzvalidierung ist jeder Fall durch die Funktionen klassifiziert, die von allen anderen Fällen außer diesem Fall abgeleitet werden.

b. 86.7% der ursprünglich gruppierten Fälle wurden korrekt klassifiziert.

c. 73.3% der kreuzvalidierten gruppierten Fälle wurden korrekt klassifiziert.