

LÖSUNG 12

a.

- Die Datei FERIENANLG.SAV laden. "Analysieren", "Klassifizieren", "Diskriminanzanalyse" öffnet die Dialogbox "Diskriminanzanalyse". Die Variable BESUCH (der Ferienanlage mit den Variablenwerten 1 = ja und 2 = nein) wird in das Eingabefeld von "Gruppenvariable" übertragen. Klicken auf die Schaltfläche "Bereich definieren..." öffnet die Unterdialogbox "Diskriminanzanalyse: Bereich definieren...". In das Eingabefeld "Minimum" wird 1 und in "Maximum" 2 eingetragen. In das Eingabefeld "Unabhängige Variable(n)" werden die Variablen EINKOMM, REISEN, FAMURLB, HGROESSE und ALTER übertragen. Die Voreinstellung "Unabhängige Variable zusammen aufnehmen" wird belassen.

Klicken auf die Schaltfläche "Statistiken..." öffnet die Unterdialogbox „Diskriminanzanalyse: Statistik“. Im Feld "Deskriptive Statistik" fordern wir "Box´M" und im Feld "Funktionskoeffizienten" "Nicht standardisiert" an. Klicken auf die Schaltfläche "Klassifizieren..." öffnet die Dialogbox "Diskriminanzanalyse: Klassifizieren". Im Feld "Anzeige" fordern wir "Fallweise Ergebnisse" und "Zusammenfassungstabelle" an, im Feld "Diagramme" "Gruppenspezifisch" (und im Feld „Kovarianzmatrix verwenden“ „Innerhalb der Gruppen“).

Bei großen Fallzahlen sollte man auf "Fallweise Ergebnisse" verzichten (bzw. die Ergebnisausgabe nur auf wenige Fälle begrenzen), da zu jedem einzelnen Fall der Datendatei Tabellenergebnisse ausgegeben werden. Man kann sich alternativ die gleichen Ergebnisinformationen zu jedem einzelnen Fall in Form von neu erzeugten Variablen in den Dateneditor ausgeben lassen. Dazu klickt man auf die Schaltfläche "Speichern..." zur Öffnung der Dialogbox "Diskriminanzanalyse: Speichern". Angefordert werden kann "Vorhergesagte Gruppenzugehörigkeit", "Scores der Diskriminanzfunktion" und "Wahrscheinlichkeit der Gruppenzugehörigkeit". (Wir verzichten darauf.) Bei den vorgenommenen Einstellungen ergibt sich ein umfangreicher Output.

- Man sollte zuerst die Anwendungsvoraussetzungen prüfen. Eine Anwendungsvoraussetzung der Diskriminanzanalyse ist, dass die Merkmalsvariable in der Grundgesamtheit für alle Gruppen etwa gleiche Varianzen und Kovarianzen haben sollten (gleiche Kovarianz-Matrizen). Mit Hilfe des Box´M-Tests kann diese Voraussetzung geprüft werden. Der Test beruht auf einer F-verteilten Prüfvariablen. Bei einem angenommenen Signifikanzniveau von $\alpha = 0,05$ wird die Hypothese gleicher Kovarianz-Matrizen angenommen ("Signifikanz" = $0,141 > \alpha = 0,05$).

Textergebnisse

Box-M		25,964
F	Näherungswert	1,393
	df1	15
	df2	3156,632
	Signifikanz	,141

Testet die Null-Hypothese der Kovarianz-Matrizen gleicher Grundgesamtheit.

- Einschränkend muss man aber konzedieren, dass das Vertrauen in die Güte des Box´M-Tests nicht besonders hoch ist, da der Test sensitiv auf die Höhe des Stichprobenumfangs einerseits und die Abweichung von der Normalverteilung der Variablen andererseits reagiert. Daher wird empfohlen, sich die Kovarianz-Matrix ausgeben zu lassen (in der Dialogbox "Diskriminanzanalyse: Statistik" im Feld "Matrizen" die Option "Kovarianz der einzelnen Gruppen" anklicken)

und diese im Gruppenvergleich auf sehr starke Ungleichheiten in den Kovarianzen und auf Vorzeichenunterschiede der Kovarianzen zu prüfen.

Vergleicht man die Kovarianzen, so zeigt sich, dass in mehreren Fällen die Vorzeichen der Kovarianzen der Variablen für die beiden Gruppen verschieden sind und auch in einigen Fällen die Kovarianzen sich um mehr als den Faktor 10 unterscheiden. Dieses sind Indikatoren dafür, dass die Bedingungen für die Anwendung der Diskriminanzanalyse verletzt sind. Daher sollte man die Ergebnisse der Diskriminanzanalyse mit Vorsicht interpretieren. Man sollte mit weiteren Daten (mehr Fälle) die Voraussetzungen für die Anwendung der Diskriminanzanalyse prüfen.

Kovarianz-Matrizen

besuch		einkomm	reisen	famurlb	hgroesse	alter
1 ja	einkomm	96,642	7,170	9,397	3,207	28,249
	reisen	7,170	3,686	,157	-,214	-,243
	famurlb	9,397	,157	3,314	,357	-,057
	hgroesse	3,207	-,214	,357	1,524	-2,762
	alter	28,249	-,243	-,057	-2,762	76,924
2 nein	einkomm	57,020	-,469	-6,287	-1,497	-30,388
	reisen	-,469	3,810	,476	,143	-6,262
	famurlb	-6,287	,476	4,210	-,057	,633
	hgroesse	-1,497	,143	-,057	,886	1,957
	alter	-30,388	-6,262	,633	1,957	68,410

- Da es sich um eine Diskriminanzanalyse für den Fall von 2 Gruppen (Besuch der Ferienanlage ja oder nein) handelt, gibt es nur eine Diskriminanzfunktion. Dieses wird in der Ausgabetabelle "Eigenwerte" angezeigt. Auch kann man dieser Tabelle Maßzahlen entnehmen, die als Indikatoren für die Güte der Trennqualität der Diskriminanzfunktion dienen.

Die Höhe des Eigenwerts (= 1,786) entspricht dem Quotienten aus den summierten quadrierten Abweichungen der Diskriminanzwerte D_i zwischen den Gruppen (= erklärte Streuung) und den summierten quadrierten Abweichungen innerhalb der Gruppen (= nicht erklärte Streuung). Dieser Quotient entspricht dem Maximalwert der Maximierungsaufgabenstellung bei Bestimmung der Diskriminanzkoeffizienten (s. Kapitel 23). Je größer dieser Quotient ist, desto besser ist die Trenngüte.

Die kanonische Korrelation einer Diskriminanzfunktion errechnet sich als Quadratwurzel des Quotienten aus den summierten quadrierten Abweichungen der Diskriminanzwerte D_i zwischen den Gruppen (= erklärte Streuung) und den summierten quadrierten gesamten Abweichungen (= gesamte Streuung). Ihr Quadrat entspricht dem Anteil der Gesamtvariation, der durch Unterschiede zwischen den Gruppen erklärt wird (= eta einer Varianzanalyse). Ein hoher Wert spricht für eine gute Trennung der Gruppen durch die Diskriminanzfunktion. Etwa 64 % ($0,801^2 = 0,6416$) der Gesamtvariation werden durch die Unterschiede zwischen den Gruppen erklärt. Auch gilt, dass der kanonische Korrelationskoeffizient dem einfachen Korrelationskoeffizienten zwischen den Diskriminanzwerten D_i und der Gruppenvariable BESUCH entspricht.

Eigenwerte

Funktion	Eigenwert	% der Varianz	Kumulierte %	Kanonische Korrelation
1	1,786 ^a	100,0	100,0	,801

a. Die ersten 1 kanonischen Diskriminanzfunktionen werden in dieser Analyse verwendet.

Als gebräuchlichste Maß zur Beurteilung der Trennqualität der Diskriminanzfunktion dient Wilks Lambda. Diese Maßzahl wird aus dem Verhältnis der summierten quadrierten Abweichungen der Diskriminanzwerte D_i innerhalb der Gruppen (= nicht erklärte Streuung) und den summierten quadrierten gesamten Abweichungen (= gesamte Streuung) berechnet (Wilks Lambda ergänzt sich mit dem Quadrat der kanonischen Korrelation zu 1, s. Kapitel 23.2). Ein kleiner Wert spricht für die Güte der Trennqualität der Diskriminanzfunktion. Knapp 36 % (= 0,359) der Streuung der Diskriminanzwerte werden durch die Gruppenunterschiede nicht erklärt.

Mit einem Chi-Quadrattest (durch eine Transformation von Wilks Lambda entsteht eine annähernd chi-quadrat-verteilte Variable, s. Kapitel 23.2) kann man prüfen, ob die Mittelwerte der Diskriminanzwerte D_i der beiden Gruppen (auch Zentroide genannt) sich signifikant voneinander unterscheiden, d. h. ob die Diskriminanzfunktion eine Gruppenzugehörigkeit vorhersagen kann. Nimmt man ein Signifikanzniveau von $\alpha = 0,05$ an, ergibt sich aus einer Chi-Quadrattabelle (diese können Sie von den Internetseiten zum Buch herunterladen) bei 5 Freiheitsgraden ein kritischer Wert (Grenzwert zwischen Annahme- und Ablehnungsbereich von H_0) in Höhe von 11,07. Da der empirisch bestimmte kritische Wert größer ist ($26,13 > 11,07$), wird die H_0 -Hypothese (die Gruppenzentroide unterscheiden sich nicht) abgelehnt und die H_1 -Hypothese (die Gruppenzentroide unterscheiden sich) angenommen. Alternativ kann man zum Testen auch den Wert von "Signifikanz" mit dem Signifikanzniveau α vergleichen. Da "Signifikanz" = $0,00 < \alpha = 0,05$ ist, kommt man auch auf diese Weise zur Ablehnung von H_0 . Der Diskriminanzfunktion wird damit eine signifikante Trennung der Gruppen attestiert.

Wilks' Lambda

Test der Funktion(en)	Wilks-Lambda	Chi-Quadrat	df	Signifikanz
1	,359	26,130	5	,000

- In der folgenden Tabelle sind die Mittelwerte der Diskriminanzwerte D_i der beiden Gruppen (die Zentroide) aufgeführt. Der Mittelwert für Anlagenbesucher beträgt $\bar{D}_{ja} = 1,291$ und der für Nicht-Anlagenbesucher $\bar{D}_{nein} = -1,291$.

Funktionen bei den Gruppen-Zentroiden

	Funktion
besuch	1
ja	1,291
nein	-1,291

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

- Aus der Klassifikationsmatrix, überschrieben mit "Klassifizierungsergebnisse", kann man entnehmen, dass mit der Diskriminanzfunktion für 3 Familien keine richtige Vorhersage hinsichtlich eines Ferienanlagenbesuchs getroffen wird. Für diese 3 Familien, die tatsächlich eine Ferienanlage besucht haben, wird kein Besuch vorhergesagt. Immerhin ist in 90 % der Fälle (27 von 30) die Vorhersage richtig. Bei einer zufälligen Zuordnung könnte man wegen der gleichen Gruppengröße nur mit einer Trefferquote von 50 % rechnen. Für welche Familien die Vorhersage falsch ist, kann man dieser Tabelle nicht entnehmen. Dazu muss man sich entweder die mit "Fallweise Ergebnisse" gewonnene Ausgabetable oder die mit "Speichern" zu erzielende Informationen ansehen (s. Lösung Teilaufgabe c).

•

Klassifizierungsergebnisse^a

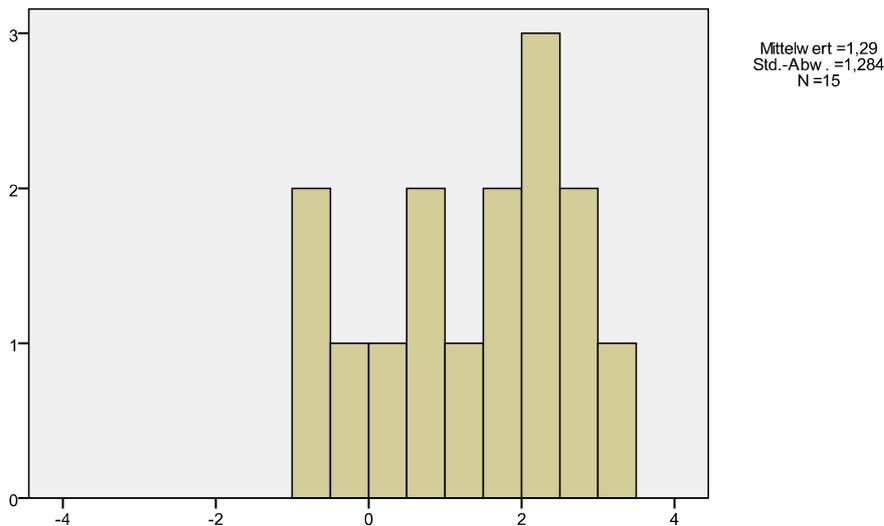
		besuch	Vorhergesagte Gruppenzugehörigkeit		Gesamt
			ja	nein	
Original	Anzahl	ja	12	3	15
		nein	0	15	15
	%	ja	80,0	20,0	100,0
		nein	,0	100,0	100,0

a. 90,0% der ursprünglich gruppierten Fälle wurden korrekt klassifiziert.

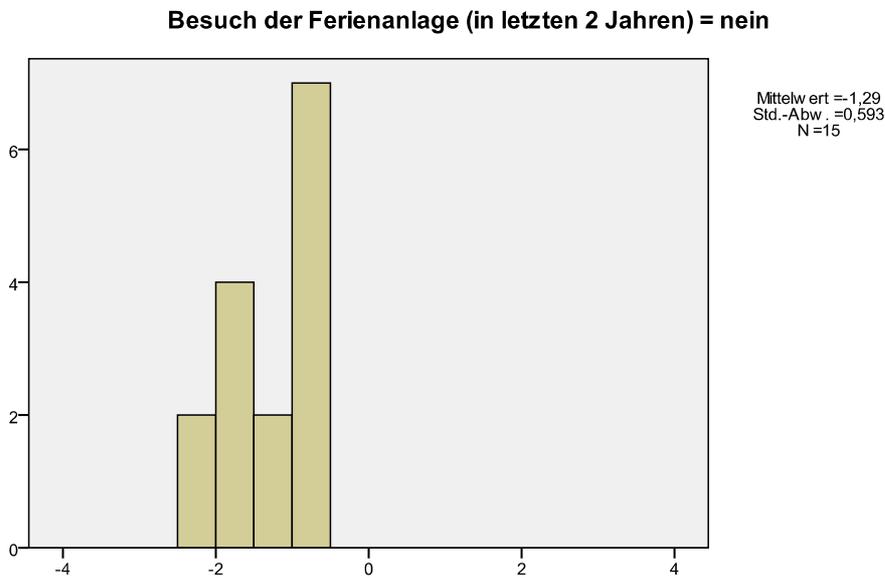
- Die beiden folgenden Abbildungen zeigen in Form von Histogrammen die Verteilung der Diskriminanzwerte in den beiden Gruppen (Nicht-Anlagenbesucher und Anlagenbesucher). Die oben angesprochenen Mittelwerte der Diskriminanzwerte D_1 werden in der Legende der Grafiken angegeben.

Kanonische Diskriminanzfunktion 1

Besuch der Ferienanlage (in letzten 2 Jahren) = ja



Kanonische Diskriminanzfunktion 1



b.

- In der Ausgabetable "Kanonische Diskriminanzfunktionskoeffizienten" werden die Koeffizienten der Diskriminanzfunktion angeführt:

Kanonische Diskriminanzfunktionskoeffizienten

	Funktion
	1
einkomm	,085
reisen	,050
famurlb	,120
hgroesse	,427
alter	,025
(Konstant)	-7,975

Nicht-standardisierte Koeffizienten

Standardisierte kanonische Diskriminanzfunktionskoeffizienten

	Funktion
	1
einkomm	,743
reisen	,096
famurlb	,233
hgroesse	,469
alter	,209

- Die Werte dieser Funktion sind Grundlage zur Berechnung der Diskriminanzwerte D_i und diese dienen zur Vorhersage der Gruppenzugehörigkeit. Der Diskriminanzwert eines Falles i wird wie folgt berechnet:

$$D_i = -7,975 + 0,025 * ALTER_i + 0,085 * EINKOMM_i + 0,12 * FAMURL_i + 0,427 * HGROESSE_i + 0,05 * REISEN_i$$

Setzt man für einen Fall i die Variablenwerte in die Gleichung ein, so ergibt sich D_i für diesen Fall. Für den ersten Fall ergibt sich $D_1 = 2,33$ (s. Tabelle „Fallweise Statistiken“ letzte Spalte).

Aber aus der Höhe dieser Koeffizienten können keine Rückschlüsse auf die relative Bedeutung einer Merkmalsvariablen für die Vorhersagekraft der Diskriminanzfunktion gewonnen werden. Dafür muss man sich auf die standardisierten Koeffizienten stützen.

- Die standardisierten Koeffizienten der Diskriminanzfunktion geben analog den Beta-Koeffizienten einer Regressionsgleichung Anhaltspunkte für die relative Bedeutung der Merkmalsvariablen für die Vorhersagekraft der Diskriminanzfunktion. Einschränkend muss aber hervorgehoben werden, dass dieses nur gilt, wenn die Merkmalsvariablen untereinander nicht korrelieren. Die Korrelation zwischen den Merkmalsvariablen ist nicht sehr stark, so dass eine vorsichtige Interpretation der Koeffizienten zulässig erscheint (der höchste Korrelationskoeffizient besteht zwischen EINKOMM und HGROESSE mit $r = 0,482$).

Die Variable EINKOMM hat den größten (absoluten) standardisierten Diskriminanzkoeffizienten (= 0,743). EINKOMM hat im Modell der Diskriminanzanalyse zur Vorhersage eines Ferienanlagenbesuchs die größte Bedeutung.

Die Variable REISEN (Einstellung zum Reisen) hat den kleinsten (absoluten) standardisierten Diskriminanzkoeffizienten (= 0,096) und trägt demgemäß zur Vorhersage des Ferienanlagebesuchs mittels der Diskriminanzfunktion am wenigsten bei. Wird die Diskriminanzanalyse ohne REISEN durchgeführt, dann wird die Trefferquote für die Vorhersage der Gruppenzugehörigkeit tatsächlich nicht schlechter. Auf die Variable kann wohl verzichtet werden.

c.

- Wegen der kleinen Fallzahl ist hier die Tabelle mit den Ergebnissen zu allen einzelnen Fällen angefordert worden (in der Dialogbox "Diskriminanzanalyse: Klassifizieren", im Feld "Anzeigen" die Option "Fallweise Ergebnisse" wählen). Durch den Vergleich der Informationen von "Tatsächlicher Gruppe" und "Vorhergesagte Gruppe" kann man entnehmen, für welche Fälle die Vorhersage mit der Diskriminanzfunktion nicht richtig ist. In den Fällen 13, 14 und 15 wird der tatsächliche Ferienanlagenbesuch nicht vorhergesagt ("falsch klassifizierter Fall").

Der Grund dafür sei am Fall 13 näher erläutert. Aus der Tabelle kann man die Wahrscheinlichkeiten für die Zuordnung der Fälle zu den beiden Gruppen entnehmen. Für den 13. Fall z.B. wird für die Zugehörigkeit zur 2. Gruppe (Besucher = nein) eine Wahrscheinlichkeit von 0,609 ausgewiesen, für die Zugehörigkeit zur 1. Gruppe beträgt diese 0,391. Der 13. Fall wird daher der 2. Gruppe zugeordnet. Er gehört tatsächlich aber zur 1. Gruppe.

Für den 13. Fall beträgt der Wert der Diskriminanzfunktion $D_{13} = -0,172$. Der Abstand zu $\bar{D}_{\text{nein}} = -1,291^1$ ist also kleiner als der zu $\bar{D}_{\text{ja}} = 1,291$. Auch daran kann man sehen, dass der 13. Fall der 2. Gruppe zugeordnet wird.

Der (für die metrischen Variablen) „Quadierte Mahalanobisabstand zum Zentroid“ für die beiden Gruppen entspricht der quadrierten Euklidischen Distanz von D_i zu den Zentroiden \bar{D}_{ja} bzw. \bar{D}_{nein} . Für den 13. Fall ist die quadrierte Euklidische Distanz zu \bar{D}_{nein} gleich $(-0,172 - -1,291)^2 = 1,252$ und zu \bar{D}_{ja} gleich $(-0,172 - 1,291)^2 = 2,14$.

¹ Siehe „Funktionen bei den Gruppen-Zentroiden“

Fallweise Statistiken

Fallnummer	Tatsächliche Gruppe	Höchste Gruppe					Zweithöchste Gruppe			Diskriminanzfunktion
		Vorhergesagte Gruppe	P(D>d G=g)		P(G=g D=d)	Quadrierter Mahalanobis-Abstand zum Zentroid	Gruppe	P(G=g D=d)	Quadrierter Mahalanobis-Abstand zum Zentroid	Funktion 1
			p	df						
Original 1	1	1	,299	1	,998	1,080	2	,002	13,115	2,330
2	1	1	,393	1	,996	,731	2	,004	11,814	2,146
3	1	1	,236	1	,567	1,407	2	,433	1,950	,105
4	1	1	,471	1	,814	,519	2	,186	3,467	,571
5	1	1	,033	1	1,000	4,562	2	,000	22,262	3,427
6	1	1	,969	1	,962	,001	2	,038	6,471	1,253
7	1	1	,822	1	,980	,050	2	,020	7,878	1,516
8	1	1	,414	1	,996	,667	2	,004	11,553	2,108
9	1	1	,190	1	,999	1,721	2	,001	15,166	2,603
10	1	1	,187	1	,999	1,740	2	,001	15,220	2,610
11	1	1	,690	1	,987	,159	2	,013	8,888	1,690
12	1	1	,446	1	,796	,582	2	,204	3,311	,528
13	1	2**	,263	1	,609	1,252	1	,391	2,141	-,172
14	1	2**	,519	1	,841	,416	1	,159	3,754	-,646
15	1	2**	,555	1	,859	,349	1	,141	3,968	-,701
16	2	2	,487	1	,824	,482	1	,176	3,564	-,597
17	2	2	,597	1	,991	,279	1	,009	9,676	-1,819
18	2	2	,394	1	,996	,725	1	,004	11,792	-2,143
19	2	2	,946	1	,971	,005	1	,029	7,024	-1,359
20	2	2	,699	1	,912	,149	1	,088	4,822	-,905
21	2	2	,646	1	,989	,211	1	,011	9,249	-1,750
22	2	2	,602	1	,879	,272	1	,121	4,246	-,769
23	2	2	,612	1	,883	,258	1	,117	4,304	-,783
24	2	2	,442	1	,794	,592	1	,206	3,287	-,522
25	2	2	,650	1	,897	,206	1	,103	4,531	-,837
26	2	2	,908	1	,974	,013	1	,026	7,281	-1,407
27	2	2	,520	1	,842	,414	1	,158	3,758	-,647
28	2	2	,609	1	,991	,261	1	,009	9,569	-1,802
29	2	2	,512	1	,993	,430	1	,007	10,485	-1,947
30	2	2	,431	1	,995	,621	1	,005	11,359	-2,079

** Falsch klassifizierter Fall

d.

- Man geht wie oben vor mit dem Unterschied, dass nun die Variable UAUSGABE (Ausgaben für den Urlaub mit den Variablenwerten 1 = niedrig, 2 = mittel und 3 = groß) in das Eingabefeld "Gruppenvariable" übertragen wird. Klicken auf die Schaltfläche "Bereich definieren..." öffnet die Unterdialogbox "Diskriminanzanalyse: Bereich". In das Eingabefeld "Minimum" wird 1 und in "Maximum" 3 eingetragen. In das Eingabefeld "Unabhängige Variable(n)" werden wie oben die Variablen EINKOMM, REISEN, FAMURLB, HGROESSE und ALTER übertragen. Die Voreinstellung "Unabhängige Variable zusammen aufnehmen" wird belassen. Klicken auf die Schaltfläche "Statistiken..." öffnet die entsprechende Unterdialogbox. Im Feld Deskriptive Statistiken" fordern wir "Box´ M" und im Feld "Funktionskoeffizienten" "Nicht standardisiert" an. Klicken auf die Schaltfläche "Klassifizieren" öffnet die Dialogbox "Diskriminanzanalyse: Klassifizieren". Im Feld "Anzeige" fordern wir "Zusammenfassungstabelle" an. Im Feld "Diagramme" markieren wir im Unterschied zu oben das Kontrollkästchen "Kombinierte Gruppen". Die fallweisen Ergebnisse wollen wir diesmal nicht in Tabellenform ausgeben, sondern als Variable der Datenmatrix anfügen. Dazu klicken wir "Speichern" zur Öffnung der entsprechenden Dialogbox und markieren die Optionen "Vorhergesagte Gruppenzugehörigkeit", "Scores der Diskriminanzfunktion" und "Wahrscheinlichkeiten der Gruppenzugehörigkeit". In der Datenmat-

rix werden die vorhergesagte Gruppenzugehörigkeit mit dem Namen DIS_1, die Diskriminanzwerte mit dem Namen DIS1_1 sowie die Wahrscheinlichkeit der Gruppenzugehörigkeit mit dem Namen DIS1_2 (für Gruppe 1) sowie DIS2_1 (für Gruppe 2) angefügt.

- Wie oben angesprochen, kann mit Hilfe des Box´ M-Tests die Anwendungsvoraussetzung gleicher Kovarianz-Matrizen für die Gruppen geprüft werden. Bei einem angenommenen Signifikanzniveau von $\alpha = 0,05$ wird die Hypothese gleicher Kovarianz-Matrizen angenommen ("Signifikanz" = $0,675 > 0,05$). Da aber das Vertrauen in die Güte des Tests nicht hoch ist, sollte man sich die Matrizen der Varianzen und Kovarianzen für die Gruppen ausgeben lassen und vergleichen (wie im Fall der zwei Gruppen, s. Teilaufgabe a).

Textergebnisse

Box-M		35,915
F	Näherungswert	,866
	df1	30
	df2	2309,985
	Signifikanz	,675

Testet die Null-Hypothese der Kovarianz-Matrizen gleicher Grundgesamtheit.

- Da es sich um eine Diskriminanzanalyse für den Fall von 3 Gruppen (Haushalte mit niedrigen, mittleren und großen Ausgaben für den Urlaub) handelt, gibt es 2 Diskriminanzfunktionen, die orthogonal zueinander sind (d. h. die Achsen der Diskriminanzvariablen D_i der beiden Funktionen stehen senkrecht zueinander). In der Ausgabetable "Eigenwerte" werden Informationen zu diesen zwei Diskriminanzfunktionen angeführt. Der Anteil des Eigenwerts einer Funktion an der Summe der Eigenwerte entspricht der relativen Bedeutung der Diskriminanzfunktionen. Diese Anteile werden in der Spalte "% der Varianz" aufgeführt. Für die erste Diskriminanzfunktion gilt, dass der Quotient aus der Summe quadrierter Abweichungen zwischen den Gruppen und der Summe quadrierter Abweichungen innerhalb der Gruppen = Eigenwert) am größten ist.

Eigenwerte

Funktion	Eigenwert	% der Varianz	Kumulierte %	Kanonische Korrelation
1	,833 ^a	70,4	70,4	,674
2	,351 ^a	29,6	100,0	,510

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

- Zu prüfen ist, ob die H_0 -Hypothese (gleiche Gruppenzentroide) anzunehmen oder abzulehnen ist. Für diesen Test wird das multivariate Wilks Lambda (= Produkt aus den einzelnen Lambdas) in einen Chi-Quadratwert transformiert (s. Kapitel23.2). Der Signifikanztest kann zweistufig erfolgen. In der ersten Stufe werden beide Diskriminanzfunktionen simultan einbezogen. Bei einem angenommenen Signifikanzniveau von $\alpha = 0,05$ und 10 Freiheitsgraden kann man einer tabellierten Chi-Quadratverteilung einen kritischen Wert (Grenzwert zwischen Annahme- und Ablehnungsbereich von H_0) von 18,307 entnehmen.

Da $22,658 > 18,307$ ist (alternativ: "Signifikanz" = $0,012 < \alpha = 0,05$), wird die Hypothese gleicher Gruppenzentroide abgelehnt. Die beiden Funktionen gemeinsam können also die 3 Gruppen signifikant unterscheiden. In der zweiten Stufe kann getestet werden, ob durch die zweite Diskriminanzfunktion alleine (also bei Weglassen der ersten Funktion) signifikante Unterschiede in den Zentroiden der 3 Gruppen bestehen. Der aus einer tabellierten Chi-Quadratverteilung entnommene kritische Wert beträgt für $\alpha = 0,05$ und 4 Freiheitsgraden 9,488. Da $7,516 < 9,488$ ist

(alternativ: "Signifikanz" = $0,238 > \alpha = 0,05$), wird die H_0 -Hypothese (gleiche Gruppenzentroide) angenommen.

Da die zweite Funktion zur Gruppenunterscheidung nicht signifikant ist, soll unten geprüft werden, ob bei Verwendung nur der ersten Diskriminanzfunktion eine ähnliche Trefferquote für die Vorhersage der Gruppenzugehörigkeit erreicht wird, wie bei Verwendung von zwei Funktionen.

Wilks' Lambda

Test der Funktion(en)	Wilks-Lambda	Chi-Quadrat	df	Signifikanz
1 bis 2	,404	22,658	10	,012
2	,740	7,516	4	,111

- In der folgenden Tabelle sind die Gruppenzentroide (d.h. die Mittelwerte der Diskriminanzwerte D_i für jede der beiden Funktionen) aufgeführt. Die Zentroide der ersten Funktion, insbesondere der Besucher mit niedrigen und der mit großen Urlaubsausgaben, unterscheiden sich deutlich voneinander. Die Unterschiede in den Zentroiden der zweiten Funktion sind hinsichtlich der Gruppen mit niedrigen und großen Urlaubsausgaben nur sehr klein. Dieses entspricht der obigen Schlussfolgerung, dass bei Einbeziehen nur der zweiten Funktion die 3 Zentroide sich nicht signifikant voneinander unterscheiden. Betrachtet man aber beide Funktionen gemeinsam, so wird eine deutlichere Unterscheidung der 3 Gruppen sichtbar: Familien mit niedrigen Urlaubsausgaben haben einen kleinen Gruppenzentroiden bei der ersten Funktion (-1,10) und die Familien mit hohen Urlaubsausgaben einen hohen (1,015). Die zweite Funktion hat für Familien mit mittleren Urlaubsausgaben (im Vergleich zu denen mit niedrigen bzw. hohen Ausgaben) einen kleinen Zentroiden. Diese Funktion dient also im Wesentlichen zur Unterscheidung der Familien mit mittleren von den Familien mit niedrigen und großen Urlaubsausgaben.

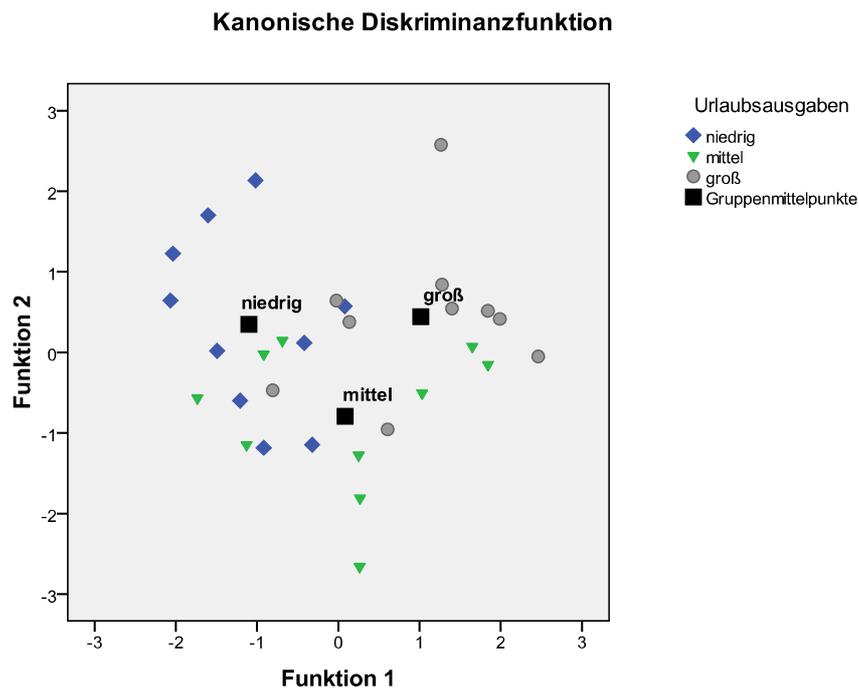
Funktionen bei den Gruppen-Zentroiden

Ausgabe	Funktion	
	1	2
1 niedrig	-1,100	,349
2 mittel	,084	-,793
3 groß	1,015	,444

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

- Die oben anhand der Zentroide der Gruppen dargelegten Gruppenunterscheidungen werden anschaulicher in der folgenden Abbildung sichtbar. Auf der horizontalen Achse des Koordinatensystems wird die erste und auf der senkrechten Achse die zweite Funktion dargestellt. Jede Familie wird im Koordinatensystem durch einen Punkt dargestellt (die Diskriminanzwerte D_i der beiden Funktionen für jede Familie sind die Koordinaten der Punkte). Dabei werden die Familien mit unterschiedlichen Urlaubsausgaben (niedrig, mittel, groß) durch unterschiedliche Symbole dargestellt. Es wird sichtbar, dass durch die zwei Funktionen die drei Gruppen zwar getrennt werden, diese Trennung aber nicht besonders gut ist. Die Punktwolken der 3 Gruppen überlappen sich.

Die Gruppenzentroide (Gruppenmittelwerte) sind ebenfalls zu sehen und damit kann insbesondere auch die Rolle der zweiten Funktion für die Trennung der 3 Gruppen verdeutlicht werden.



- Die folgenden Tabellen geben die unstandardisierten und die standardisierten Diskriminanzkoeffizienten der Diskriminanzfunktion an. Die unstandardisierten Koeffizienten bilden die Grundlage für die Berechnung der Diskriminanzwerte und damit für die auf diesen beruhende Vorhersage der Gruppenzugehörigkeit. Die standardisierten Koeffizienten sind Grundlage zur Beurteilung der relativen Bedeutung der Merkmalsvariablen für die Vorhersagekraft der Diskriminanzfunktion. Absolut hohe Koeffizienten weisen auf wichtige Merkmalsvariable für die Vorhersage und absolut kleine auf unwichtige hin. Man muss aber einschränkend bedenken, dass eine derartige Beurteilung nur zulässig ist, wenn die Merkmalsvariablen untereinander nicht korrelieren. Wie oben schon festgestellt worden ist, besteht zwischen den Variablen keine hohe Korrelation. Eine vorsichtige Interpretation erscheint daher zulässig.

Für die erste Funktion sind die Variable EINKOMM und HGROESSE, für die zweite Funktion EINKOMM, FAMURLB sowie HGROESSE die Variablen mit den höchsten absoluten standardisierten Koeffizienten und haben für die Vorhersage einer Gruppenzugehörigkeit demgemäß eine wichtige Bedeutung. REISEN hat für beide Funktionen relativ kleine absolute standardisierte Diskriminanzkoeffizienten. Demgemäß wäre die Variable REISEN für die Vorhersage der Gruppenzugehörigkeit eher unwichtig. Führt man eine Diskriminanzanalyse ohne REISEN durch, so ist die Trefferquote nur geringfügig schlechter.

Kanonische Diskriminanzfunktionskoeffizienten

	Funktion	
	1	2
einkomm	,095	-,060
reisen	-,086	-,071
famurlb	,009	,382
hgroesse	,194	,563
alter	-,019	,029
(Konstant)	-4,216	-1,991

Standardisierte kanonische Diskriminanzfunktionskoeffizienten

	Funktion	
	1	2
einkomm	,952	-,601
reisen	-,173	-,143
famurlb	,017	,754
hgroesse	,231	,671
alter	-,166	,258

Nicht-standardisierte Koeffizienten

- Der folgenden Tabelle ist das Ergebnis der Gruppenvorhersage durch die Diskriminanzfunktion zu entnehmen. Bei 11 Familien ist die Vorhersage der Höhe der Urlaubsausgaben mit den Kategorien niedrig, mittel und groß falsch. Damit wird für ca. 63 % ($19/30 = 0,6333$) eine richtige Vorhersage getroffen. Bedenkt man, dass bei zufälliger Zuordnung der Familien zu den 3 Gruppen (angesichts der gleichen Gruppengröße) die Wahrscheinlichkeit einer richtigen Zuordnung 33,33 % wäre, so kann die Vorhersagequalität des Diskriminanzanalysemodells als recht gut beurteilt werden.

Klassifizierungsergebnisse^a

		Vorhergesagte Gruppenzugehörigkeit			Gesamt	
		1 niedrig	2 mittel	3 groß		
Original	Anzahl	1 niedrig	7	2	1	10
		2 mittel	3	4	3	10
		3 groß	1	1	8	10
%		1 niedrig	70,0	20,0	10,0	100,0
		2 mittel	30,0	40,0	30,0	100,0
		3 groß	10,0	10,0	80,0	100,0

a. 63,3% der ursprünglich gruppierten Fälle wurden korrekt klassifiziert.

- Durch Gegenüberstellung der Werte in der Variablen UAUSGABE mit den Werten der im Dateneditor hinzugefügten Variablen DIS_1 kann man die tatsächliche mit der vorhergesagten Gruppenzugehörigkeit vergleichen (etwa durch „Analysieren“, „Berichte“, „Fallzusammenfassungen“, übertragen von UAUSGABE und DIS_1 in „Variablen:“ und Anklicken des Kontrollkästchens „Fälle anzeigen“ oder durch Erstellen einer Kreuztabelle mit diesen beiden Variablen). In der folgenden Tabelle werden für die fehlerhaft vorhergesagten Fälle die tatsächliche und die fehlerhaft vorhergesagte Gruppenzugehörigkeit vergleichend gegenübergestellt. Man sieht, dass es vorwiegend bei der Trennung der Fälle der Gruppe 2 (Familien mit mittleren Urlaubsausgaben) zur den der Gruppe 1 (Familien mit niedrigen Urlaubsausgaben) bzw. zur Gruppe 3 (Familien mit großen Urlaubsausgaben) zu fehlerhaften Zuordnungen kommt. In nur zwei Fällen kommt es zur falschen Trennung von Gruppe 1 und Gruppe 3.

Fall	Tatsächliche Gruppe	Vorhergesagte Gruppe
1	2	3
4	1	3
7	2	3
8	2	3
12	3	2
15	3	1
16	1	2
18	2	1
19	2	1
20	2	1
30	1	2

e.

- Oben wurde dargelegt, dass die zweite Diskriminanzfunktion alleine hinsichtlich der Unterscheidung der drei Gruppen nicht signifikant ist. Die darauf folgende Betrachtung der Gruppenzentroide (auch in der oben dargestellten Grafik) hat dieses Ergebnis insofern bestätigt, als dass sich die Zentroide der zweiten Funktion hinsichtlich der Gruppen mit niedrigen und mit großen Urlaubsausgaben kaum unterscheiden. Aber es wurde auch festgestellt, dass die zweite Funktion zur Unterscheidung der Familien mit mittleren Urlaubsausgaben von den Familien mit niedrigen bzw. hohen Urlaubsausgaben beiträgt.

Im Folgenden soll untersucht werden, ob man bei der Vorhersage der Gruppenzugehörigkeit sich eventuell nur auf die berechnete erste Diskriminanzfunktion stützen sollte.

Diese Analyse ist über das Menü nicht möglich. Daher muss man sich der Syntaxsprache von SPSS bedienen. Dabei geht man folgendermaßen vor. Zunächst wird die Dialogbox zur Diskriminanzanalyse erneut aufgerufen. Alle gewünschten Einstellungen sind noch aus der vorhergehenden Diskriminanzanalyse vorhanden. Nun klickt man die Schaltfläche "Einfügen" und übergibt damit die angeforderte Analyse in Form von Syntaxbefehlen in das Syntax-Fenster. Die Syntaxbefehle werden um den Befehl /FUNCTIONS = 1 ergänzt. Mit diesem Befehl wird SPSS angewiesen, nur eine (die erste) Diskriminanzfunktion zur Vorhersage der Gruppenzugehörigkeit zu verwenden. Die Reihenfolge von Syntaxbefehlen ist zu beachten. Deshalb wird der FUNCTIONS-Befehl vor dem Statistics-Befehl eingefügt (s. den Syntax Guide im Hilfesystem). Im Syntaxfenster werden die Befehle mit "Ausführen", "Alle" an SPSS übergeben.

DISCRIMINANT

```

/GROUPS=uausgabe(1 3)
/VARIABLES=einkomm reisen famurlb hgrosse alter
/ANALYSIS ALL
/SAVE=CLASS SCORES PROBS
/PRIORS EQUAL
/FUNCTIONS = 1
/STATISTICS=BOXM RAW TABLE
/PLOT=COMBINED
/CLASSIFY=NONMISSING POOLED.

```

- Die Ergebnisse dieser Diskriminanzanalyse unterscheiden sich nur hinsichtlich der Vorhersage der Gruppenzugehörigkeit (s. die Tabelle "Klassifikationsergebnisse"). Aus dem Vergleich der Klassifikationsergebnisse ergibt sich, dass für 2 weitere Familien (insgesamt 13) eine falsche

Vorhersage der Gruppenzugehörigkeit erfolgt. Damit reduziert sich die Trefferquote von ca. 63 % auf ca. 57 %. Daher sollte man für die Vorhersage der Gruppenzugehörigkeit wohl nicht auf die zweite Funktion verzichten.

Klassifizierungsergebnisse^a

		Ausgabe	Vorhergesagte Gruppenzugehörigkeit			Gesamt
			1 niedrig	2 mittel	3 groß	
Original	Anzahl	1 niedrig	7	3	0	10
		2 mittel	4	3	3	10
		3 groß	1	2	7	10
	%	1 niedrig	70,0	30,0	,0	100,0
		2 mittel	40,0	30,0	30,0	100,0
		3 groß	10,0	20,0	70,0	100,0

a. 56,7% der ursprünglich gruppierten Fälle wurden korrekt klassifiziert.

- Aus der im Dateneditor hinzugefügten Variablen DIS_2 kann man die vorhergesagte Gruppenzugehörigkeit entnehmen. In der folgenden Tabelle für die falsch vorhergesagten Fälle die tatsächliche und fehlerhaft vorhergesagte Gruppenzugehörigkeit vergleichend gegenübergestellt. Man sieht, dass auch hier (mit einer Ausnahme) die Trennung zwischen Fällen der Gruppen 2 zu den der Gruppen 1 bzw. 3 nicht gut gelingt.

Fall	Tatsächliche Gruppe	Vorhergesagte Gruppe
1	2	3
3	3	2
4	1	2
7	2	3
8	2	3
14	3	2
15	3	1
16	1	2
18	2	1
19	2	1
20	2	1
23	1	2
26	2	1