Lösungen zu den Kontrollfragen und Vernetzungsaufgaben

Kapitel 1: Ablauf einer Studie

Kontrollfragen

1. Bitte ergänzen Sie die Lückentexte.

   Das Gütekriterium der …..**Validität**…. prüft, inwiefern ein Konstrukt tatsächlich das misst, was es zu messen vorgibt. Die …..**Reliabilität**….. ist ein Maß für die Zuverlässigkeit eines Messinstrumentes ….**Objektivität**.…. liegt bei einer Messung dann vor, wenn mehrere Forscher unabhängig voneinander zum selben Ergebnis kommen.

2. Wie würden Sie "Kausalität" definieren? Welche Bedingungen müssen erfüllt sein, damit man Kausalität annehmen kann?

   Definition: Mit Kausalität bezeichnet man Ursache-Wirkungs-Zusammenhänge.
   Bedingungen, damit man Kausalität annehmen kann:

   - **Die vermutete Ursache (UV) und angenommene Wirkung (AV) sollten korrelieren.**
   - **Es darf keine Störgrößen geben, die die abhängige Variable beeinflussen, d.h. die Wirkung darf nur auf die AV zurückzuführen sein.**
   - **Die Ursache (UV) muss zeitlich vor der Wirkung (AV) liegen.**

3. Grenzen Sie den qualitativen und den quantitativen Forschungsansatz voneinander ab. Was ist ein Mixed-Methods-Ansatz? Warum kann es sinnvoll sein, einen Mixed-Methods-Ansatz zu verfolgen?
   - **Abgrenzung qualitativ vs. quantitativ:**
     Mit qualitativen Studien wird explorativ und anhand von kleinen Stichproben versucht, einen Sachverhalt besser zu verstehen und aufzuhellen. Die Repräsentativität der eher „weichen" Daten, die z. B. in Gruppendiskussionen gewonnen werden, ist gering. Quantitative Studien wollen Zusammenhänge entdecken und prüfen, sie haben deskriptiven und/oder explikativen Charakter. Es werden „harte", quantitative Daten gewonnen. Diese werden mithilfe von statistischen Analyseverfahren ausgewertet. Die Fallzahl ist meist groß und bei einer geeigneten Auswahl der Stichprobe kann diese auch repräsentativ für die Grundgesamtheit sein.
   - **Definition Mixed-Methods-Ansatz:**
     Forschungsansatz, bei dem qualitative und quantitative Studien kombiniert eingesetzt werden
   - **Zweck von Mixed-Methods-Ansätzen:**
     Beide Ansätze bringen Vorteile mit sich, die sich bei Mixed-Methods-Ansätzen verbinden lassen. Z. B. können mit einer qualitativen Vorstudie Problemfelder identifiziert werden, die dann in eine quantitative Studie einfließen und mit einer größeren Stichprobe näher analysiert werden. Manchmal bleiben auch bei quantitativen Studien Fragen offen, denen im Nachgang mit einer qualitativen Studie noch mal nachgegangen werden kann.

4. Welche der im Folgenden aufgeführten Antworten sind richtig?

- ■ Ob eine Variable normalverteilt ist, lässt sich per Augenschein in einem Histogramm erkennen.
- ☐ Der Kolmogorov-Smirnov-Anpassungstest prüft die Nullhypothese, dass die Beobachtungswerte einer Variablen in der Grundgesamtheit nicht der Normalverteilung folgen.
- ☐ Eine Verteilung ist rechtssteil und linksschief, wenn sich die Beobachtungswerte oberhalb des arithmetischen Mittels häufen.

Vernetzungsaufgaben

1. Berechnen Sie für die Variable mix 1 („Ich mag es, Mixgetränke zu trinken.“) folgende deskriptive Statistiken: Häufigkeiten, arithmetisches Mittel, Median, Varianz, Standardabweichung.

Statistiken

Ich mag es, Mix-Getränke zu trinken.

N	Gültig	284
	Fehlend	4
Mittelwert		1,03

Ich mag es, Mix-Getränke zu trinken.

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	lehne voll und ganz ab	13	4,5	4,6	4,6
	lehne ab	21	7,3	7,4	12,0
	lehne eher ab	23	8,0	8,1	20,1
	weder noch	20	6,9	7,0	27,1
	stimme eher zu	66	22,9	23,2	50,4
	stimme zu	93	32,3	32,7	83,1
	stimme voll und ganz zu	48	16,7	16,9	100,0
	Gesamt	284	98,6	100,0	
Fehlend	System	4	1,4		
Gesamt		288	100,0		

Deskriptive Statistik

			Statistik	Standardfehler
mix1 Ich mag es, Mix-Getränke zu trinken.	Mittelwert		1,03	,099
	95% Konfidenzintervall des Mittelwerts	Untergrenze	,83	
		Obergrenze	1,22	
	5% getrimmtes Mittel		1,14	
	Median		1,00	
	Varianz		2,805	
	Standardabweichung		1,675	
	Minimum		-3	
	Maximum		3	
	Spannweite		6	
	Interquartilbereich		2	
	Schiefe		-,904	,145
	Kurtosis		-,092	,288

2. Das „Alter“ ist im Datensatz als metrisch skalierte Variable gespeichert. Transformieren Sie die Variable „Alter“ in eine nominal skalierte Variable um. Nutzen Sie zur Bildung von zwei Gruppen den Median-Split, d.h. verwenden Sie den Median, um den Datensatz für die Variable „Alter“ in zwei etwa gleich große Gruppen aufzuteilen.

   Der Median beträgt 25. Es werden deshalb die folgenden zwei Gruppen gebildet:

   Gruppe 1 = bis 25 Jahre alt, Gruppe 2 = älter als 25

   Die Gruppe der bis 25-Jährigen umfasst 165 Befragte, 123 Probanden sind älter als 25 Jahre alt.

Alter 2 Gruppen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	bis 25 Jahre	165	57,3	57,3	57,3
	älter 25 Jahre	123	42,7	42,7	100,0
	Gesamt	288	100,0	100,0	

3. Berechnen Sie, ob ein statistisch signifikanter Unterschied zwischen dem Alter (bis 25 Jahre; > 25 Jahre) und der Einstellung zu Mixgetränken (Variable mix 1: „Ich mag es, Mixgetränke zu trinken.“) besteht. Nutzen Sie hierfür die Kreuztabellierung und den Chi-Quadrat-Test.

   Beachten Sie, dass für diese Analyse beide Variablen nominal skaliert sein müssen. Sie müssen deshalb zunächst die intervallskalierte Variable „mix1“ in eine nominal skalierte Variable umwandeln. Transformieren Sie dazu die Variable wie folgt: 1 = negative Einstellung zu Mixgetränken (Ausgangswerte -3, -2, -1, 0), 2 = positive Einstellung zu Mixgetränken (Ausgangswerte 1, 2, 3).

   77 Befragte (27,1%) haben eine eher negative Einstellung zu Mixgetränken, 207 Befragte (72,9%) haben hingegen auf dem positiven Teil der Skala geantwortet und damit eine eher positive Einstellung zu Mixgetränken.

Einstellung zu Mixgetränken 2 Gruppen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	negative Einstellung	77	26,7	27,1	27,1
	positive Einstellung	207	71,9	72,9	100,0
	Gesamt	284	98,6	100,0	
Fehlend	System	4	1,4		
Gesamt		288	100,0		

Die Kreuztabelle zeigt, dass beobachtete und erwartete Häufigkeiten nur wenig voneinander abweichen. Dies deutet darauf hin, dass sich die Altersgruppen nicht in ihrer Einstellung zu Mixgetränken unterscheiden. Der Chi-Quadrat-Test liefert ein nicht-signifikantes Ergebnis (p = ,349). Das heißt, die beiden Altersgruppen unterscheiden sich nicht in ihrer Einstellung zu Mixgetränken.

Einstellung zu Mixgetränken 2 Gruppen * Alter 2 Gruppen Kreuztabelle

			Alter 2 Gruppen		
			bis 25 Jahre	älter 25 Jahre	Gesamt
Einstellung zu Mixgetränken 2 Gruppen	negative Einstellung	Anzahl	41	36	77
		Erwartete Anzahl	44,5	32,5	77,0
	positive Einstellung	Anzahl	123	84	207
		Erwartete Anzahl	119,5	87,5	207,0
Gesamt		Anzahl	164	120	284
		Erwartete Anzahl	164,0	120,0	284,0

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	,877[a]	1	,349		
Kontinuitätskorrektur[b]	,642	1	,423		
Likelihood-Quotient	,872	1	,350		
Exakter Test nach Fisher				,418	,211
Zusammenhang linear-mit-linear	,874	1	,350		
Anzahl der gültigen Fälle	284				

a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 32,54.

b. Wird nur für eine 2x2-Tabelle berechnet

Kapitel 2: Varianzanalyse

Kontrollfragen

1. Ergänzen Sie die Lückentexte.

   Je nach Anzahl der abhängigen und unabhängigen Variablen lassen sich verschiedene Formen der Varianzanalyse einsetzen. Möchte man den Einfluss mehrerer unabhängiger Variablen auf eine abhängige Variable untersuchen, verwendet man die ..…**mehrfaktorielle ANOVA**…... Im Falle mehrerer abhängiger Variablen greift man auf die …..**MANOVA**…... zurück.

2. Zu welchem Zweck wird der F-Test im Rahmen der Varianzanalyse eingesetzt? Erläutern Sie die Logik des F-Tests.

   Zweck: Der F-Test lässt sich für jede unabhängige Variable berechnen. Er prüft, ob der Einfluss der unabhängigen Variablen auf die abhängige Variable statistisch signifikant ist, d.h. ob der Unterschied zwischen den Gruppen nicht nur zufälliger Natur ist.

   Logik: Es wird ein empirischer F-Wert berechnet, der mit einem theoretischen F-Wert verglichen wird. Wenn der empirische F-Wert größer ist als der theoretische F-Wert, kann die Nullhypothese verworfen werden. Den theoretischen F-Wert entnimmt man aus Tabellen der F-Verteilung, die in einschlägigen Statistik-Lehrbüchern enthalten sind. Um den theoretischen F-Wert ablesen zu können, müssen zunächst noch das Signifikanzniveau (= α-Fehler) festgelegt und die Zahl der Freiheitsgrade bestimmt werden.

3. Erklären Sie, unter welchen Voraussetzungen sich die einfaktorielle ANOVA als Testverfahren eignet.
   - **Metrisch skalierte AV, nominal skalierte UV**
   - **Normalverteilung der Werte der abhängigen Variablen in allen Grundgesamtheiten**
   - **Homogene Varianzen innerhalb der Gruppen**

4. Welche der im Folgenden aufgeführten Aussagen sind richtig?

   - ☐ Wenn der Levene-Test zu einem nicht-signifikanten Ergebnis führt, liegt keine Varianzhomogenität vor.
   - ☐ Aus einer Variablen werden Gruppen gebildet und deren Einfluss auf eine abhängige Variable untersucht. Der F-Test führt zu einem signifikanten Ergebnis. Dies lässt den Schluss zu, dass die Mittelwerte in den Gruppen gleich groß sind.
   - ■ Der t-Test (für unabhängige Stichproben) lässt sich nur für Designs mit einer abhängigen und einer unabhängigen Variablen mit 2 Fallgruppen anwenden.
   - ■ Ein Interaktionseffekt liegt dann vor, wenn die Wirkung einer unabhängigen Variablen auf die abhängige Variable von der Ausprägung einer anderen unabhängigen Variablen abhängt.

Vernetzungsaufgaben

Untersuchen Sie anhand der Varianzanalyse, inwiefern das Geschlecht die Einstellung zu Mixgetränken beeinflusst.

1. Prüfen Sie zunächst, ob die Variable „mix1“ („Ich mag es, Mixgetränke zu trinken.“) die Voraussetzung der Varianzhomogenität erfüllt.
   Mit dem Levene-Test wird geprüft, ob die Variable die Anforderung der Varianzhomogenität erfüllt. In unserem Fall (basiert auf Mittelwert) ist p < ,050, d.h. die Nullhypothese, wonach die Varianzen in den Gruppen homogen sind, kann verworfen werden. Die Voraussetzung der Varianzhomogenität ist damit nicht erfüllt.

Test der Homogenität der Varianzen

		Levene-Statistik	df1	df2	Signifikanz
Ich mag es, Mix-Getränke zu trinken.	Basiert auf dem Mittelwert	5,224	1	280	,023
	Basiert auf dem Median	3,242	1	280	,073
	Basierend auf dem Median und mit angepaßten df	3,242	1	271,926	,073
	Basiert auf dem getrimmten Mittel	3,716	1	280	,055

2. Berechnen Sie nun zunächst die Mittelwerte, die Männer und Frauen für die Variable „mix1“ aufweisen.
   Der Mittelwert der männlichen Probanden liegt bei ,89 und damit etwas unter dem Gesamtmittelwert von 1,02. Die weiblichen Probanden haben einen Mittelwert, der leicht über dem Mittelwert der Gesamtstichprobe liegt (1,14). Frauen scheinen etwas lieber Mixgetränke zu trinken als Männer.

Bericht

Ich mag es, Mix-Getränke zu trinken.

Geschlecht	Mittelwert	N	Std.-Abweichung
männlich	,89	139	1,805
weiblich	1,14	143	1,536
Insgesamt	1,02	282	1,676

3. Rechnen Sie nun eine einfache Varianzanalyse und prüfen Sie die Güte der Analyse mit dem F-Test.
   Der F-Test liefert für die Variable „sex“ einen Signifikanzwert von p = ,215. Der Unterschied zwischen Männern und Frauen ist nicht signifikant.

Einfaktorielle ANOVA

Ich mag es, Mix-Getränke zu trinken.

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	4,327	1	4,327	1,544	,215
Innerhalb der Gruppen	784,584	280	2,802		
Gesamt	788,911	281			

4. Zusätzlich zum Geschlecht soll nun noch das Alter als weitere unabhängige Variable in die Analyse einfließen. Da die unabhängigen Variablen nominal skaliert sein müssen, muss die metrisch skalierte Variable „Alter“ zunächst umkodiert werden.

   Transformieren Sie die Variable „Alter“ anhand des Medians (25 Jahre) in zwei Gruppen: 1 = bis 25 Jahre, 2 = älter als 25 Jahre. Berechnen Sie den Mittelwert für den Fall, dass das Geschlecht und das Alter gleichzeitig als unabhängige Variablen und die Variable „mix1“ als abhängige Variable betrachtet werden.

Zwischensubjektfaktoren

		Wertelabel	N
Geschlecht	0	männlich	139
	1	weiblich	143
Alter 2 Gruppen	1,00	bis 25 Jahre	162
	2,00	älter 25 Jahre	120

Deskriptive Statistiken

Abhängige Variable: Ich mag es, Mix-Getränke zu trinken.

Geschlecht	Alter 2 Gruppen	Mittelwert	Std.-Abweichung	N
männlich	bis 25 Jahre	1,20	1,628	79
	älter 25 Jahre	,48	1,953	60
	Gesamt	,89	1,805	139
weiblich	bis 25 Jahre	1,04	1,550	83
	älter 25 Jahre	1,28	1,519	60
	Gesamt	1,14	1,536	143
Gesamt	bis 25 Jahre	1,12	1,586	162
	älter 25 Jahre	,88	1,788	120
	Gesamt	1,02	1,676	282

Die deskriptiven Statistiken zeigen, dass Männer über 25 Jahre die negativste Einstellung gegenüber Mix-Getränken haben (,48). Die positivste Einstellung haben Frauen, die älter als 25 Jahre alt sind (1,28). Es ist kein deutlicher Effekt des Alters oder des Geschlechts zu erkennen. Die Mittelwerte lassen aber einen Interaktionseffekt von Alter und Geschlecht vermuten.

Die Ergebnisse der F-Tests bestätigen die Vermutung. Der Einfluss des Alters ist nicht signifikant (p = ,631). Auch die Variable „Geschlecht“ übt keinen signifikanten Einfluss aus (p = ,711). Der Interaktionseffekt von Geschlecht und Alter ist signifikant (p = ,016).

Tests der Zwischensubjekteffekte

Abhängige Variable: Ich mag es, Mix-Getränke zu trinken.

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Konstanter Term	Hypothese	276,410	1	.	.	.
	Fehler	.	.[a]	.		
sex	Hypothese	6,917	1	6,917	,430	,631
	Fehler	16,091	1	16,091[b]		
alter_2gruppen	Hypothese	3,839	1	3,839	,239	,711
	Fehler	16,091	1	16,091[b]		
sex * alter_2gruppen	Hypothese	16,091	1	16,091	5,849	,016
	Fehler	764,818	278	2,751[c]		

a. Fehler-Freiheitsgrade können mit der Satterthwaite- Methode nicht berechnet werden.

b. MS(sex * alter_2gruppen)

c. MS(Fehler)

Kapitel 3: Faktorenanalyse

Kontrollfragen

1. Bitte ergänzen Sie die Lückentexte.

   Die …..**Faktorladung**….. gibt an, wie hoch eine einzelne Variable auf einen Faktor lädt. Der …..**Eigenwert**….. wird pro Faktor berechnet und entspricht der Summe der quadrierten Faktorladungen eines Faktors über alle Variablen hinweg. Der …..**Faktorwert**…... ergibt sich als Produkt von Faktorladungs-Matrix und Ausgangsdatenmatrix. Er gibt an, wie ein Proband einen Faktor bewertet hätte.

2. Warum wird die explorative Faktorenanalyse als "explorativ" bezeichnet?

   Die Faktorenanalyse ist eine Interdependenzanalyse, d.h. es werden Zusammenhänge zwischen einer Vielzahl von Variablen aufgedeckt. Die Variablen werden dabei nicht in unabhängige und abhängige Variablen unterteilt. Es bestehen häufig auch keine theoretischen Vorannahmen dazu, wie die Variablen zusammenhängen. Der Forscher geht bei explorativen Faktorenanalysen empiriegeleitet vor, d.h. er trifft keine theoretischen Vorannahmen zur Faktorstruktur.

3. Anhand welcher Kriterien kann man entscheiden, wie viele Faktoren extrahiert werden sollten? Nennen Sie die Kriterien und erläutern Sie diese näher.
   - **Scree-Test: Der Scree-Test ist eine grafische Darstellungsform, bei der die Eigenwerte aller gebildeten Faktoren absteigend geordnet dargestellt werden. Es ergibt sich eine degressiv abfallende Kurve. Zu wählen ist die Anzahl an Faktoren, nach der die Eigenwertkurve noch einmal merklich abfällt (= 'Elbow').**
   - **Kaiser-Kriterium: Es sollen so viele Faktoren bestimmt werden, wie es Faktoren mit einem Eigenwert > 1 gibt.**

4. Welche der folgenden Aussagen sind richtig?

   ☐ Von der Einfachstruktur der Faktorladungsmatrix spricht man, wenn möglichst viele Variablen jeweils auf mehrere Faktoren hoch laden.

   ☐ Cronbachs Alpha ist ein Maß für die interne Konsistenz eines Faktors. Es handelt sich dabei um ein lokales Gütekriterium.

   ■ Der Bartlett-Test kann zum Prüfen der Voraussetzungen für eine Faktorenanalyse nur dann eingesetzt werden, wenn die Variablen normalverteilt sind.

   ☐ Faktoren mit einer Kommunalität von ≤ 0,5 sollten von der Analyse ausgeschlossen werden.

Vernetzungsaufgaben

Ihre Aufgabe ist es nun, eine Faktorenanalyse anhand der folgenden Variablen zur Messung der Einstellung gegenüber Marken zu rechnen: qb1, qb2, mt1, mt2, mb1, mb2, prb1, prb2.

1. Prüfen Sie zunächst, ob die Variablen für eine Faktorenanalyse geeignet sind. Betrachten Sie hierzu die Kommunalitäten der Ausgangsvariablen und berechnen Sie das KMO-Kriterium.

Das KMO-Kriterium beträgt ,588. Da es größer als ,500 ist, sind die Daten für eine Faktorenanalyse geeignet. Die Kommunalitäten sind für alle Variablen größer als der geforderte Wert von ,500. Die Analyse kann daher mit allen Variablen fortgeführt werden.

KMO- und Bartlett-Test

Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		,588
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	603,812
	df	28
	Signifikanz nach Bartlett	,000

Kommunalitäten

	Anfänglich	Extraktion
Qualitätsbewusstsein: Für besondere Qualität gebe ich gern etwas mehr aus.	1,000	,742
Qualitätsbewusstsein: Manchmal leiste ich mir ganz bewußt die beste Qualität.	1,000	,774
Markentreue: Bei vielen Produkten kaufe ich immer nur eine bestimmte Marke.	1,000	,676
Markentreue: Wenn ich mit einer Marke zufrieden bin, bleibe ich auch dabei.	1,000	,758
Markenbewusstsein: Markenprodukte sind qualitativ besser als markenlose Ware.	1,000	,912
Markenbewusstsein: Markenprodukte sind zwar teurer als andere Produkte, dafür aber auch besser.	1,000	,907
Prestigebewusstsein: Ich finde es wichtig, bei Freunden, Bekannten und Kollegen anerkannt zu sein.	1,000	,763
Prestigebewusstsein: Einige Produkte kaufe ich bewusst deshalb, weil mich meine Freunde darum beneiden werden.	1,000	,510

Extraktionsmethode: Hauptkomponentenanalyse.

2. Berechnen Sie nun die rotierte Faktorladungsmatrix. Verwenden Sie dabei die Hauptkomponenten-Analyse und das Varimax-Verfahren. Wie viele Faktoren sollten gebildet werden? Wie lassen sich die Faktoren inhaltlich interpretieren?

Faktor 1 = Markenbewusstsein
Faktor 2 = Qualitätsbewusstsein
Faktor 3 = Markentreue
Faktor 4 = Prestigebewusstsein

Rotierte Komponentenmatrix[a]

	Komponente			
	1	2	3	4
Qualitätsbewusstsein: Für besondere Qualität gebe ich gern etwas mehr aus.	,068	,838	,179	,058
Qualitätsbewusstsein: Manchmal leiste ich mir ganz bewußt die beste Qualität.	,064	,854	,201	-,009
Markentreue: Bei vielen Produkten kaufe ich immer nur eine bestimmte Marke.	,170	,195	,780	-,024
Markentreue: Wenn ich mit einer Marke zufrieden bin, bleibe ich auch dabei.	-,022	,143	,855	,073
Markenbewusstsein: Markenprodukte sind qualitativ besser als markenlose Ware.	,951	,058	,050	,044
Markenbewusstsein: Markenprodukte sind zwar teurer als andere Produkte, dafür aber auch besser.	,944	,081	,092	,035
Prestigebewusstsein: Ich finde es wichtig, bei Freunden, Bekannten und Kollegen anerkannt zu sein.	-,004	-,163	,154	,844
Prestigebewusstsein: Einige Produkte kaufe ich bewusst deshalb, weil mich meine Freunde darum beneiden werden.	,094	,352	-,131	,600

Extraktionsmethode: Hauptkomponentenanalyse.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 4 Iterationen konvergiert.

3. Prüfen Sie die globale Güte der Faktorenanalyse anhand der erklärten Varianz.

   Die erklärte Varianz beträgt 75,52%. Die vier Faktoren zusammen erklären damit mehr als 50% der Varianz der Variablen, was für die globale Güte der Faktorenanalyse spricht.

Erklärte Gesamtvarianz

Komponente	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	2,427	30,332	30,332	2,427	30,332	30,332	1,842	23,027	23,027
2	1,549	19,366	49,698	1,549	19,366	49,698	1,650	20,624	43,651
3	1,053	13,162	62,861	1,053	13,162	62,861	1,464	18,302	61,953
4	1,013	12,659	75,519	1,013	12,659	75,519	1,085	13,566	75,519
5	,888	11,106	86,625						
6	,529	6,616	93,242						
7	,371	4,632	97,873						
8	,170	2,127	100,000						

Extraktionsmethode: Hauptkomponentenanalyse.

Kapitel 4: Regressionsanalyse

Kontrollfragen

1. Ergänzen Sie die Lückentexte.

   Mit Hilfe von ..…**standardisierten Regressionskoeffizienten**…... lässt sich die Stärke des Einflusses zweier oder mehrerer unabhängiger Variablen überprüfen. Hohe absolute Beta-Werte sprechen für einen …..**starken**….. Einfluss eines Prädiktors auf die abhängige Variable. Ob eine weitere aufgenommene unabhängige Variable auch einen zusätzlichen Erklärungsbeitrag leistet, kann man weiterhin aus dem Anstieg des …..**korrigierten Bestimmtheitsmaßes**….. ablesen.

2. Nach welcher Maßgabe wird die Regressionsanalyse geschätzt? Welches Minimierungsproblem soll dabei gelöst werden?

   Die Regressionsanalyse wird nach Maßgabe der Methode der kleinsten Quadrate geschätzt. Demnach soll die Regressionsgerade so geschätzt werden, dass die Summe der quadrierten Abweichungen zwischen tatsächlichen und vorhergesagten Werten minimal ausfällt.

3. Was versteht man unter Autokorrelation? Bei welchen Daten ist es wichtig, auf Autokorrelation zu prüfen?

   Autokorrelation bedeutet, dass die Residuen miteinander korrelieren. Autokorrelation ist besonders bei Zeitreihenanalysen anzutreffen, da in diesem Fall die aufeinander folgenden Residuen miteinander korrelieren können.

4. Welche der folgenden Aussagen sind richtig?

   - ■ Der Regressionskoeffizient (b) entspricht bei der einfachen linearen Regressionsanalyse dem Wert der Steigung der linearen Regressionsgeraden.
   - ☐ Der t-Test prüft die lokale Güte einer Regressionsanalyse. Er testet für alle unabhängigen Variablen gemeinsam, ob sich die Regressionskoeffizienten von Null unterscheiden.
   - ■ Damit der F- und der t-Test interpretiert werden dürfen, müssen die Residuen bei einer multiplen Regressionsanalyse der Normalverteilung folgen.

Vernetzungsaufgaben

Sie sollen nun eine multiple Regressionsanalyse rechnen. Die abhängige Variable ist „mix1" („Ich mag es, Mixgetränke zu trinken."). Als unabhängige Variablen fließen die folgenden Variablen ein: fak1_reg_hedo: Spaß am Einkauf/Hedonismus, fak2_reg_mb: Markenbewusstsein, fak3_reg_qb: Qualitätsbewusstsein, fak4_reg_inno: Innovationsakzeptanz, fak5_reg_mt: Markentreue, fak6_reg_kf: Kontaktfreudigkeit.

1. Berechnen Sie die Regressionskoeffizienten und stellen Sie die Regressionsgleichung auf. Verwenden Sie dabei die „Schrittweise Methode". Wie beurteilen Sie Richtung und Stärke des Einflusses, den die einzelnen UV auf die AV ausüben?
   Einstellung zu Mixgetränken = 1,026 + ,342 * Innovationsakzeptanz +,268 * Kontaktfreudigkeit

Es werden zwei Prädiktoren in das Modell aufgenommen. Beide üben einen positiven Einfluss auf die abhängige Variable aus. Die Innovationsakzeptanz übt einen etwas stärkeren Einfluss aus (,203) als die Kontaktfreudigkeit (,152).

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressionskoeffizient B	Std.-Fehler	Beta	T	Sig.
1	(Konstante)	1,028	,098		10,533	,000
	Faktor Innovationsakzeptanz (inno1, inno2)	,333	,099	,197	3,382	,001
2	(Konstante)	1,026	,097		10,630	,000
	Faktor Innovationsakzeptanz (inno1, inno2)	,342	,097	,203	3,506	,001
	Faktor Kontaktfreudigkeit (kf1, kf2)	,268	,098	,158	2,729	,007

a. Abhängige Variable: Ich mag es, Mix-Getränke zu trinken.

2. Betrachten Sie nun das Bestimmtheitsmaß und das korrigierte Bestimmtheitsmaß. Welche Aussagen lassen sich aus beiden Werten ableiten?
Das Bestimmtheitsmaß für das Modell 2 beträgt ,064. Das korrigierte Bestimmtheitsmaß liegt bei ,057. Insgesamt erklären die Faktoren Innovationsakzeptanz und Kontaktfreudigkeit damit 6,4% der abhängigen Variablen „mix 1“: „Ich mag es, Mixgetränke zu trinken.“ Anhand des korrigierten Bestimmtheitsmaßes lässt sich erkennen, ob durch die Aufnahme einer weiteren unabhängigen Variablen in das Modell der Anteil der erklärten Varianz der abhängigen Variablen steigt. Dies ist hier der Fall (von 3,6% auf 5,7%).

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,197[a]	,039	,036	1,645
2	,253[b]	,064	,057	1,626

a. Einflußvariablen : (Konstante), Faktor Innovationsakzeptanz (inno1, inno2)

b. Einflußvariablen : (Konstante), Faktor Innovationsakzeptanz (inno1, inno2), Faktor Kontaktfreudigkeit (kf1, kf2)

3. Prüfen Sie die globale und die lokale Güte des berechneten Regressionsmodells.
Die globale Güte wird mit dem F-Test beurteilt, der die Nullhypothese testet, dass alle Regressionskoeffizienten in der Grundgesamtheit Null sind. In Modell 2 ist der F-Test hoch signifikant ($p \leq ,001$). Die Nullhypothese kann abgelehnt werden. D.h. der Einfluss von mindestens einer der unabhängigen Variablen Innovationsakzeptanz und Kontaktfreudigkeit auf die Variable Einstellung zu Mixgetränken ist nicht zufällig.

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	30,940	1	30,940	11,438	,001[b]
	Nicht standardisierte Residuen	762,834	282	2,705		
	Gesamt	793,775	283			
2	Regression	50,636	2	25,318	9,573	,000[c]
	Nicht standardisierte Residuen	743,139	281	2,645		
	Gesamt	793,775	283			

a. Abhängige Variable: Ich mag es, Mix-Getränke zu trinken.

b. Einflußvariablen : (Konstante), Faktor Innovationsakzeptanz (inno1, inno2)

c. Einflußvariablen : (Konstante), Faktor Innovationsakzeptanz (inno1, inno2), Faktor Kontaktfreudigkeit (kf1, kf2)

Die lokale Güte des Modells wird mit dem t-Test geprüft. Der t-Test prüft für jeden Regressionskoeffizienten einzeln, ob der Einfluss der jeweiligen UV auf die AV signifikant ist. In unserem Beispiel erreichen alle zwei unabhängigen Variablen ein p ≤ ,010. Es besteht also jeweils ein statistisch hoch signifikanter Einfluss der UV auf die AV.

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressionskoeffizient B	Std.-Fehler	Beta	T	Sig.
1	(Konstante)	1,028	,098		10,533	,000
	Faktor Innovationsakzeptanz (inno1, inno2)	,333	,099	,197	3,382	,001
2	(Konstante)	1,026	,097		10,630	,000
	Faktor Innovationsakzeptanz (inno1, inno2)	,342	,097	,203	3,506	,001
	Faktor Kontaktfreudigkeit (kf1, kf2)	,268	,098	,158	2,729	,007

a. Abhängige Variable: Ich mag es, Mix-Getränke zu trinken.

Kapitel 5: Clusteranalyse

Kontrollfragen

1. Ergänzen Sie die folgenden Lückentexte.

   Bei …..**agglomerativen hierarchischen**….. Fusionierungsalgorithmen werden pro Analyseschritt diejenigen Objekte zusammengefasst, welche die geringste Distanz zueinander aufweisen. Die …..**Single Linkage**….. – Methode eignet sich besonders dazu, um Ausreißer zu identifizieren. Der Vorteil der …..**Ward**….. – Methode besteht darin, dass in etwa gleich große Gruppen gebildet werden.

2. Wie kann die Proximität (Ähnlichkeit) bzw. Distanz zwischen Objekten bestimmt werden? Was versteht man unter der Minkowski-Metrik?

   Es können je nach Fragestellung und Skalenniveau der Variablen unterschiedliche Distanz- bzw. Ähnlichkeitsmaße berechnet werden. Die Minkowski-Metrik bildet die Grundlage für die Berechnung von Distanzen. Sie enthält als entscheidende Größen die Minkowski-Konstanten c und r, mit denen die Art der Minkowski-Metrik festgelegt wird.

3. Wie bestimmt man die Anzahl der Cluster und wie können die Cluster inhaltlich interpretiert werden?

   Zur Bestimmung der Anzahl der Cluster können u.a. das Dendrogramm und das Elbow-Kriterium herangezogen werden. Es sollten zudem nur so viele Cluster bestimmt werden, wie inhaltlich sinnvoll interpretiert werden können. Um die Cluster zu interpretieren, werden die Mittelwerte der beschreibenden Variablen pro Cluster berechnet (Clusterlösung = UV, beschreibende Variablen = AV). Indem man die Eigenschaften der Cluster vergleicht, lassen sich die Gruppen interpretieren und passende Bezeichnungen finden.

4. Welche der im Folgenden aufgeführten Aussagen sind richtig?

   - ■ Die Clusteranalyse verfolgt das Ziel, Variablen so zu Gruppen zusammenzufassen, dass die Gruppen in sich möglichst homogen und untereinander möglichst heterogen sind.
   - ☐ Bei der Clusteranalyse handelt es sich im Gegensatz zur Faktorenanalyse um eine Interdependenzanalyse. Inferenzstatistische Verfahren können deshalb nicht zur Güteprüfung herangezogen werden.
   - ☐ Nach dem inversen Scree-Test liegt die optimale Zahl der zu bildenden Cluster da, wo in der Grafik ein Knick zu erkennen ist, d.h. die Fehlerquadratsumme fällt sprunghaft ab.
   - ■ Die Streuung innerhalb eines Clusters wird analog zur Bestimmung der Streuung bei der Varianzanalyse ermittelt.

Vernetzungsaufgaben

Rechnen Sie nun eine Clusteranalyse, um Kundentypen zu identifizieren. Verwenden Sie die folgenden Variablen: fak1_mb: Markenbewusstsein, fak2_qb: Qualitätsbewusstsein, fak3_mt: Markentreue, fak4_pb: Preisbewusstsein.

1. Verwenden Sie als Fusionierungsalgorithmus die Ward-Methode und als Distanzmaß die quadrierte euklidische Distanz. Speichern Sie die gebildete Clusterlösung als Variable. Lassen Sie sich eine 5-Cluster-Lösung ausgeben.

Bericht

Ward Method		Qualitätsbewusstsein: Für besondere Qualität gebe ich gern etwas mehr aus.	Qualitätsbewusstsein: Manchmal leiste ich mir ganz bewußt die beste Qualität.	Preisbewusstsein:Ob ich etwas kaufe oder nicht, entscheide ich in erster Linie nach dem Preis.	Preisbewusstsein: Ich bin immer auf der Suche nach Billigangeboten.	Preisbewusstsein: Ich achte darauf, das beste Preis/Leistungs- Verhältnis zu bekommen.	Markentreue: Bei vielen Produkten kaufe ich immer nur eine bestimmte Marke.	Markentreue: Wenn ich mit einer Marke zufrieden bin, bleibe ich auch dabei.	Markenbewusstsein: Markenprodukte sind qualitativ besser als markenlose Ware.	Markenbewusstsein: Markenprodukte sind zwar teurer als andere Produkte, dafür aber auch besser.	Markenbewusstsein: Ich wähle meistens die teuren Marken aus.	Markenbewusstsein: Die bekannten Markenprodukte sind die besten.
1	Mittelwert	1,56	1,50	,92	,02	2,44	-,53	,69	-,66	-,57	-1,30	-1,65
	N	80	80	80	80	80	80	80	80	80	80	80
	Std.-Abweichung	,926	1,091	1,329	1,449	,613	1,242	1,218	1,440	1,448	1,141	1,032
2	Mittelwert	2,08	2,25	,02	-,77	2,46	1,29	1,94	,81	,81	,15	-,33
	N	48	48	48	48	48	48	48	48	48	48	48
	Std.-Abweichung	,679	,838	1,139	1,519	,617	,824	,665	1,045	,867	,945	,996
3	Mittelwert	-,13	-1,07	1,37	,46	1,74	-,63	,67	,22	,13	-1,72	-,91
	N	46	46	46	46	46	46	46	46	46	46	46
	Std.-Abweichung	1,408	1,237	,997	1,471	,828	1,306	1,034	1,519	1,408	,911	1,297
4	Mittelwert	1,22	,59	1,88	1,24	2,29	1,33	2,08	,00	,04	-1,61	-1,16
	N	51	51	51	51	51	51	51	51	51	51	51
	Std.-Abweichung	,856	1,403	,791	1,408	,576	,931	,595	1,549	1,371	1,021	1,206
5	Mittelwert	1,71	1,51	-,56	-1,37	,70	1,02	1,84	,16	,16	-,56	-,92
	N	63	63	63	63	63	63	63	63	63	63	63
	Std.-Abweichung	,869	1,203	1,377	1,274	1,026	1,238	,827	1,298	1,382	1,089	1,324
Insgesamt	Mittelwert	1,35	1,06	,69	-,13	1,92	,43	1,39	,02	,04	-1,02	-1,07
	N	288	288	288	288	288	288	288	288	288	288	288
	Std.-Abweichung	1,183	1,561	1,460	1,670	1,020	1,437	1,121	1,463	1,400	1,228	1,243

2. Wie lassen sich die Cluster inhaltlich interpretieren?
   - **Cluster 1: „Qualitätsbewusste ohne Markenpräferenz“**
     Dieses Cluster achtet gleichzeitig auf Qualität und Preis und ist nicht markenbewusst.
   - **Cluster 2: “Qualitätsbewusste Markenkäufer“**
     Dieses Cluster ist sowohl markenbewusst wie auch markentreu.
   - **Cluster 3: „Preisbewusste“**
     Das Preisbewusstsein ist stark ausgeprägt. Das Cluster ist nicht markenbewusst und nicht markentreu.
   - **Cluster 4: „Smart Shopper“**
     Das Cluster ist markentreu und hat gleichzeitig ein hohes Preisbewusstsein (d.h. sie warten darauf, dass „ihre Marken“ günstig angeboten werden.
   - **Cluster 5: „Experimentierfreudige“**
     Die Preissensibilität ist gering. Die Käufer achten auf Qualität, sind aber nicht sehr markenbewusst.

Bericht

Ward Method		Qualitätsbewusstsein: Für besondere Qualität gebe ich gern etwas mehr aus.	Qualitätsbewusstsein: Manchmal leiste ich mir ganz bewußt die beste Qualität.	Preisbewusstsein:Ob ich etwas kaufe oder nicht, entscheide ich in erster Linie nach dem Preis.	Preisbewusstsein: Ich bin immer auf der Suche nach Billigangeboten.	Preisbewusstsein: Ich achte darauf, das beste Preis/Leistungs- Verhältnis zu bekommen.	Markentreue: Bei vielen Produkten kaufe ich immer nur eine bestimmte Marke.	Markentreue: Wenn ich mit einer Marke zufrieden bin, bleibe ich auch dabei.	Markenbewusstsein: Markenprodukte sind qualitativ besser als markenlose Ware.	Markenbewusstsein: Markenprodukte sind zwar teurer als andere Produkte, dafür aber auch besser.	Markenbewusstsein: Ich wähle meistens die teuren Marken aus.	Markenbewusstsein: Die bekannten Markenprodukte sind die besten.
1	Mittelwert	1,56	1,50	,92	,02	2,44	-,53	,69	-,66	-,57	-1,30	-1,65
	N	80	80	80	80	80	80	80	80	80	80	80
	Std.-Abweichung	,926	1,091	1,329	1,449	,613	1,242	1,218	1,440	1,448	1,141	1,032
2	Mittelwert	2,08	2,25	,02	-,77	2,46	1,29	1,94	,81	,81	,15	-,33
	N	48	48	48	48	48	48	48	48	48	48	48
	Std.-Abweichung	,679	,838	1,139	1,519	,617	,824	,665	1,045	,867	,945	,996
3	Mittelwert	-,13	-1,07	1,37	,46	1,74	-,63	,67	,22	,13	-1,72	-,91
	N	46	46	46	46	46	46	46	46	46	46	46
	Std.-Abweichung	1,408	1,237	,997	1,471	,828	1,306	1,034	1,519	1,408	,911	1,297
4	Mittelwert	1,22	,59	1,88	1,24	2,29	1,33	2,08	,00	,04	-1,61	-1,16
	N	51	51	51	51	51	51	51	51	51	51	51
	Std.-Abweichung	,856	1,403	,791	1,408	,576	,931	,595	1,549	1,371	1,021	1,206
5	Mittelwert	1,71	1,51	-,56	-1,37	,70	1,02	1,84	,16	,16	-,56	-,92
	N	63	63	63	63	63	63	63	63	63	63	63
	Std.-Abweichung	,869	1,203	1,377	1,274	1,026	1,238	,827	1,298	1,382	1,089	1,324
Insgesamt	Mittelwert	1,35	1,06	,69	-,13	1,92	,43	1,39	,02	,04	-1,02	-1,07
	N	288	288	288	288	288	288	288	288	288	288	288
	Std.-Abweichung	1,183	1,561	1,460	1,670	1,020	1,437	1,121	1,463	1,400	1,228	1,243