
Anhang

A.1	Verzeichnis der SPSS Datendateien	350
A.2	Datenzugriff via Internet	355
A.3	p-value-Konzept	356
A.4	Ausgewählte Grenz- und Elastizitätsfunktionen	357
A.5	Verzeichnis ausgewählter SPSS Funktionsgruppen	358
A.6	Lösungen zu den Übungs- und Klausuraufgaben	359
	Stichwortverzeichnis	369

A.1 Verzeichnis der SPSS Datendateien

ARIMA.sav Die SPSS Datendatei beinhaltet die Werte von simulierten stochastischen Prozessen zum Zwecke ihrer Nachbildung mit Hilfe von ARIMA-Modellen.

Aufgabe 2-1.sav Die SPSS Datendatei ist das Resultat einer Studierendenbefragung, die im Sommersemester 2007 am Fachbereich WiWi I der HTW Berlin mittels eines standardisierten Fragebogens durchgeführt wurde.

Aufgabe 2-2.sav Die SPSS Datendatei beinhaltet Daten von zufällig ausgewählten Mietwohnungen, die im Jahr 2007 auf dem Berliner Mietwohnungsmarkt angeboten wurden.

Aufgabe 2-3.sav Die SPSS Datendatei beinhaltet Daten von Lehrveranstaltungen, die im Wintersemester 2002/03 am Fachbereich WiWi I der HTW Berlin evaluiert wurden.

Aufgabe 3-1.sav Die SPSS Datendatei basiert auf einer Gästebefragung in Romantik-Hotels, die im vierten Quartal 2005 im deutschsprachigen Raum durchgeführt wurde.

Aufgabe 3-2.sav Die SPSS Datendatei beinhaltet die prozentualen Veränderungsraten des amtlichen Schlusskurses einer Aktie.

Aufgabe 3-3.sav Die SPSS Datendatei basiert auf einer Kundenbefragung, die im Sommersemester 2002 auf der Grundlage eines standardisierten Fragebogens durchgeführt wurde.

Aufgabe 3-4.sav Die SPSS Datendatei beinhaltet Angaben von Studierenden, die im Winter- und im Sommersemester 2005/06 am Fachbereich WiWi I der HTW Berlin in den Statistik-Kursen I, II und III erhoben wurden.

Aufgabe 3-5.sav Die SPSS Datendatei beinhaltet Daten über die Dauer von Mobilfunk-Gesprächen.

Aufgabe 4-1.sav Die SPSS Datendatei beinhaltet Daten von PKW, die im zweiten Quartal 2003 auf dem Berliner Gebrauchtwagenmarkt angeboten wurden.

Aufgabe 4-2.sav Die SPSS Datendatei basiert auf dem Automarkenindex AUTOMARXX, der vom ADAC für den Monat Dezember 2005 herausgegeben wurde.

Aufgabe 4-3.sav Die SPSS Datendatei beinhaltet Daten von Lehrveranstaltungen, die im Wintersemester 2002/03 am Fachbereich WiWi I der HTW Berlin evaluiert wurden.

Aufgabe 4-4.sav Die SPSS Datendatei beinhaltet die Bewertungen eines neuen Hochschullogos, die von zufällig ausgewählten Hochschulvertretern abgegeben wurden.

- Aufgabe 4-5.sav** Die SPSS Datendatei beinhaltet Angaben von Studierenden, die im Winter- und im Sommersemester 2005/06 am Fachbereich WiWi I der HTW Berlin in den Statistik-Kursen I, II und III erhoben wurden.
- Aufgabe 5-1.sav** Die SPSS Datendatei basiert auf einer Gästebefragung in Romantik-Hotels, die im vierten Quartal 2005 im deutschsprachigen Raum durchgeführt wurde.
- Aufgabe 5-2.sav** Die SPSS Datendatei beinhaltet Daten von Lehrveranstaltungen, die im Wintersemester 2002/03 am Fachbereich WiWi I der HTW Berlin evaluiert wurden.
- Aufgabe 5-3.sav** Die SPSS Datendatei beinhaltet Angaben von Studierenden, die im Winter- und im Sommersemester 2005/06 am Fachbereich WiWi I der HTW Berlin in den Statistik-Kursen I, II und III erhoben wurden.
- Aufgabe 5-4.sav** Die SPSS Datendatei beinhaltet Daten von Lehrveranstaltungen, die im Wintersemester 2002/03 am Fachbereich WiWi I der HTW Berlin evaluiert wurden.
- Aufgabe 5-5.sav** Die SPSS Datendatei beinhaltet Daten von PKW, die im zweiten Quartal 2003 auf dem Berliner Gebrauchtwagenmarkt angeboten wurden.
- Aufgabe 6-1.sav** Die SPSS Datendatei beinhaltet Daten von lebendgeborenen Babys, die 1998 in einem Berliner Geburtshaus „das Licht der Welt erblickten“.
- Aufgabe 6-2.sav** Die SPSS Datendatei beinhaltet Daten von zufällig ausgewählten VW Polo, die im Jahr 2003 auf dem Berliner Gebrauchtwagenmarkt angeboten wurden.
- Aufgabe 6-3.sav** Die SPSS Datendatei beinhaltet Informationen von PKW spezieller Marken, die im zweiten Quartal 2003 auf dem Berliner Gebrauchtwagenmarkt angeboten wurden.
- Aufgabe 6-4.sav** Die SPSS Datendatei basiert auf einer Kundenbefragung zweier Autobahn-Raststätten, die im Wintersemester 2002/03 realisiert wurde.
- Aufgabe 6-5.sav** Die SPSS Datendatei basiert auf Kennzahlen von zufällig ausgewählten und vergleichbaren Berliner Unternehmen, die aus dem jeweiligen Jahresabschluss für das Wirtschaftsjahr 1995 berechnet und banktechnisch zur Einstufung ihre Bonität benutzt wurden.
- Aufgabe 7-1.sav** Die SPSS Datendatei beinhaltet für die Zeitreihe der Tagesumsätze eines Reise-Shops auf einem Berliner Bahnhof.
- Aufgabe 7-2.sav** Die SPSS Datendatei beinhaltet die Zeitreihe der Anzahl der Neuzulassungen von PKW in Deutschland für den Zeitraum von Januar 2000 bis Mai 2006.

- Aufgabe 7-3.sav** Die SPSS Datendatei beinhaltet die Zeitreihe der täglichen Besuchszahlen der Ausstellung „Körperwelten“ in Berlin im Jahr 2001.
- Aufgabe 7-5.sav** Die SPSS Datendatei beinhaltet für den Zeitraum vom 2. August 2004 bis zum 29. Juli 2005 die amtlichen Schlusskurse für die Stammaktie der LUFTHANSA AG.
- Aufgabe 7-6.sav** Die SPSS Datendatei basiert auf Speichelmengendaten eines Patienten, die im Zuge eines ernährungsmedizinischen Experiments chronologisch erhoben wurden.
- Aufgabe 8-1.sav** Die SPSS Datendatei beinhaltet Informationen von zufällig ausgewählten Berliner Fahrschülern, die im Wirtschaftsjahr 1995 ihren Führerschein erwarben.
- Aufgabe 8-2.sav** Die SPSS Datendatei beinhaltet Informationen über Lehrveranstaltungen, die im Wintersemester 2002/03 am Fachbereich WiWi I der HTW Berlin evaluiert wurden.
- Aufgabe 9-1.sav** Die SPSS Datendatei beinhaltet sozio-ökonomische Daten für die Berliner Stadtbezirke aus dem Jahr 2002.
- Aufgabe 9-2.sav** Die Datei enthält arbeitsmarktbezogene Daten für die deutschen Bundesländer aus dem Wirtschaftsjahr 2002.
- Aufgabe 9-3.sav** Die SPSS Datendatei beinhaltet Evaluationsdaten von Lehrveranstaltungen, die an der HTW Berlin im Wintersemester 2002/03 erhoben wurden.
- Aufgabe 9-4.sav** Die SPSS Datendatei basiert auf einer Befragung von zufällig ausgewählten Nutzern von Berliner Parkhäusern aus dem Jahr 2006.
- Aufgabe 9-5.sav** Die SPSS Datendatei beinhaltet Angaben von Studierenden in den Bachelor-Studiengängen, die im Sommersemester 2007 am Fachbereich WiWi I der HTW Berlin fragebogenbasiert erhoben wurden.
- Aufgabe 10-1.sav** Die SPSS Datendatei enthält sozio-ökonomische Daten für die Berliner Stadtbezirke aus dem Jahr 2002.
- Aufgabe 10-2.sav** Die Datei enthält arbeitsmarktbezogene Daten für die deutschen Bundesländer aus dem Wirtschaftsjahr 2002.
- Aufgabe 10-3.sav** Die SPSS Datendatei basiert auf dem Automarkenindex, der vom ADAC für den Monat Dezember 2005 herausgegeben wurde.
- Aufgabe 10-4.sav** Die SPSS Datendatei basiert auf einem kommunalen Preisvergleich aus dem zweiten Quartal 2007, der für Städte und Gemeinden in den Bundesländern Berlin und Brandenburg durchgeführt wurde.
- Aufgabe 10-5.sav** Die SPSS Datendatei basiert auf einer Befragung von zufällig ausgewählten Nutzern von Berliner Parkhäusern aus dem Jahr 2006.

- Lesedauer.sav** Die Datei beinhaltet Daten von zufällig ausgewählten Personen, die im Sommersemester 2003 bezüglich favorisierter Berliner Tageszeitungen befragt wurden.
- Billet.sav** Die SPSS Datendatei basiert auf einem Marktforschungsprojekt aus dem Jahr 1995. Im Rahmen des Projektes wurden mittels einer geschichteten Zufallsauswahl Fahrgäste im Berliner Öffentlichen Personen-Nahverkehrs (ÖPNV) befragt.
- DAX.sav** Die SPSS Datendatei beinhaltet die an der Frankfurter Börse im zweiten Halbjahr 2009 börsentäglich erfassten Schlusskurse des Deutschen Aktienindex DAX.
- Diplom.sav** Die SPSS Datendatei beinhaltet Angaben von Studierenden in den Diplom-Studiengängen, die im Sommersemester 2007 am Fachbereich WiWi I der HTW Berlin mittels eines standardisierten Fragebogens erhoben wurden.
- Eier.sav** Die SPSS Datendatei beinhaltet Daten eines Loses von 1000 Hühnereiern.
- Evaluation.sav** Die SPSS Datendatei beinhaltet Informationen über Lehrveranstaltungen, die im Wintersemester 2002/03 am Fachbereich WiWi I der Hochschule für Technik und Wirtschaft Berlin evaluiert wurden.
- Fahrzeit.sav** Die SPSS Datendatei basiert auf einer Befragung von Mitgliedern der Professoren- und Mitarbeiterschaft des Fachbereichs WiWi I der HTW Berlin im Wintersemester 2003/04 bezüglich der Fahrzeit zur Hochschule.
- Faktorenanalyse.sav** Die SPSS Datendatei ist ein integraler Bestandteil der SPSS Datendatei *Fragebogen.sav*, die als Arbeitsdatei zudem die Ergebnisse einer Faktorenanalyse beinhaltet.
- FKK.sav** Die SPSS Datendatei basiert auf einer Umfrage unter zufällig ausgewählten Studierenden an Berliner Hochschulen, die im Sommersemester 1996 mit dem Ziel durchgeführt wurde, die Einstellung von Studierenden zur Freikörper-Kultur zu erforschen.
- Fluggaeste.sav** Die SPSS Datendatei beinhaltet für den Beobachtungszeitraum von Januar 2006 bis Dezember 2011 die strukturbruchbereinigte Zeitintervallreihe der monatlichen Fluggästezahlen auf den Berliner Flughäfen.
- Fragebogen.sav** Die SPSS Datendatei basiert auf semesterbezogenen Studienbefragungen, die seit dem Wintersemester 2007/08 am Fachbereich WiWi I der HTW Berlin in den wirtschaftswissenschaftlichen Bachelor-Studiengängen auf der Grundlage eines standardisierten Fragebogens durchgeführt wurden.
- Grundstudium.sav** Die SPSS Datendatei enthält Semesterabschlussbewertungen aus dem Wintersemester 2002/03 für die Grundstudienfächer in den be-

triebswirtschaftlichen Diplom-Studiengängen des Fachbereichs WiWi I der HTW Berlin.

Hochschullogo.sav Die SPSS Datendatei basiert auf einer statusgruppenbezogenen Blitzumfrage, die im Sommersemester 2009 an der Hochschule für Technik und Wirtschaft Berlin durchgeführt wurde und die Bewertung des neuen Hochschullogos zum Inhalt hatte.

Huehnereier.sav Die SPSS Datendatei beinhaltet Gewichts-, Breiten- und Höhendaten von 785 Hühnereiern, die von Hühnern der Rasse Loheimer Braun gelegt wurden.

Klausur.sav Die SPSS Datendatei beinhaltet die Ergebnisse der Abschlussklausur zur Deskriptiven Statistik aus dem Sommersemester 2000.

Klausurtraining.sav Die SPSS Datendatei beinhaltet Daten, die im Sommersemester 2003 im Kontext eines Klausurtrainings zur „Deskriptiven Statistik“ erhoben wurden.

Mietspiegel.sav Die SPSS Datendatei beinhaltet Daten von zufällig ausgewählten Mietwohnungen, die im Jahr 2011 auf dem Berliner Wohnungsmarkt angeboten wurden.

Nebenjob.sav Die SPSS Datendatei beinhaltet die Daten einer Befragung von zufällig ausgewählten Berliner Studierenden aus dem Sommersemester 2002.

Parkhausnutzer.sav Die SPSS Datendatei basiert auf einem Markforschungsprojekt. Im Rahmen des Projektes wurden im Herbst 2006 mittels einer geschichteten Zufallsauswahl Nutzer von Berliner Parkhäusern befragt.

Pruefungen.sav Die SPSS Datendatei enthält die Semesterabschlussbewertungen von Lehrveranstaltungen am Fachbereich WiWi I der HTW Berlin aus dem Wintersemester 2002/03.

Residual.sav Die SPSS Datendatei basiert auf der SPSS Datendatei *DAX.sav* und hat als Arbeitsdatei zudem noch die Analysebefunde für eine nichtlineare Trendfunktion zum Inhalt.

ZehnDutzend.sav Die SPSS Datendatei basiert auf der SPSS Datendatei *Huehnereier.sav* und beinhaltet 10 Zufallsstichproben jeweils mit einem Umfang von 12 Hühnereiern. ♣

A.2 Datenzugriff via Internet

Internet-Adresse. Alle in diesem Lehrbuch verwendeten und im alphabetisch geordneten Verzeichnis A.1 aufgelisteten SPSS Datendateien stehen im Internet unter der Adresse

<http://www.f3.htw-berlin.de/Professoren/Eckstein/Vorlesungsmaterial>

zur freien Verfügung. Der Downloadbereich ist in der Abbildung A.2-1 skizziert.

Abbildung A.2-1: Downloadbereich

HTW Berlin Fachbereich Wirtschaftswissenschaften I : Professoren/innen

Einleitung...

Fast 60 Professor/innen des Fachbereiches vermitteln mit über 100 Lehrbeauftragten Wissen und Erkenntnisse.

Die mehr als 2500 Studierenden, hiervon etwa 60 Prozent Frauen und rund acht Prozent ausländische Gäste, sind derzeit in elf Studiengängen eingeschrieben.

[weiterlesen](#)

Homepage : Prof. Dr. Peter Eckstein

Professor für Statistik, Ökonometrie und Empirische Wirtschaftsforschung

Vorlesungsmaterial – Download

Lehrbücher

Statistik für Wirtschaftswissenschaftler, 2. Auflage
 ↳ Downloadbereich Daten

Statistik für Wirtschaftswissenschaftler, 3. Auflage
 ↳ Downloadbereich Daten

Datenanalyse mit SPSS, 1. Auflage
 ↳ Downloadbereich Daten

Datenanalyse mit SPSS, 2. Auflage
 ↳ Downloadbereich Daten

Angewandte Statistik mit SPSS, 6. Auflage
 ↳ Downloadbereich Daten

Angewandte Statistik mit SPSS, 7. Auflage
 ↳ Downloadbereich Daten

Navigation

- ◆ Home
- ◆ Vita
- ◆ Lehre
- ◆ Publikationen
- ◆ Vorlesungsmaterial - Download
- ◆ Nachricht senden

Quicklinks

- ◆ LSF
- ◆ eCampus
- ◆ Bibliothek
- ◆ Fachbereich
- ◆ HTW Berlin

SPSS Datendateien, 7. Auflage

Download. Das „Herunterladen“ einer SPSS Datendatei aus dem Internet erfolgt in kommentierten Arbeitsschritten, die den Vorgang des Datentransfers begleiten und erleichtern. Eine „heruntergeladene“ SPSS Datendatei kann via Sequenz 2.2-1 geöffnet und in den SPSS Daten-Editor eingelesen werden.

Hotline. Für den Fall, dass beim Datentransfer Probleme auftreten, wende man sich an eine der folgenden eMail-Adressen:

Peter.Eckstein@HTW-Berlin.de oder Frank.Steinke@HTW.Berlin.de

Herr Diplom-Wirtschaftsinformatiker Frank STEINKE betreut den Downloadbereich und ist bestrebt, jederzeit sachdienliche Hinweise und Unterstützung bei auftretenden Problemen des Datentransfers zu gewähren. ♣

A.3 p-value-Konzept

In SPSS werden Testentscheidungen auf der Basis eines Vergleichs eines vorgegebenen Signifikanzniveaus α mit einem empirischen Signifikanzniveau α^* herbeigeführt. Während das Signifikanzniveau α stets im Vorfeld eines Tests zu vereinbaren ist, wird das empirische Signifikanzniveau α^* unter Einbeziehung der jeweiligen Prüfverteilung aus dem jeweiligen Stichprobenbefund ermittelt.

Empirisches Signifikanzniveau

Ist T eine Testvariable und t_n ein Testvariablenwert, der aufgrund einer Zufallsstichprobe vom Umfang n ermittelt wurde, dann heißt die Wahrscheinlichkeit dafür, dass unter einer Nullhypothese H_0 die Testvariable T Werte annimmt, die gleich oder kleiner bzw. gleich oder größer sind als der Testvariablenwert t_n , empirisches Signifikanzniveau α^* .



Hinweise. Für das Verständnis des Begriffs und Konstrukts eines empirischen Signifikanzniveaus erweisen sich die folgenden Hinweise als hilfreich: i) **Synonyme.** Ein empirisches Signifikanzniveau wird in SPSS synonym auch als Signifikanz, Alpha*, p-Wert, p(robability)-value bezeichnet. ii) **Semantik.** Da der Begriff Signifikanz in seiner synonymen Verwendung für ein empirisches Signifikanzniveau aus statistisch-methodischer Sicht missverständlich sein kann und semantisch streng genommen nicht exakt ist, wurde an Stelle des Begriffs *Signifikanz* der Begriff *empirisches Signifikanzniveau* vereinbart und mit dem Symbol α^* belegt. Das Attribut empirisch erklärt sich daraus, dass das Signifikanzniveau unter Annahme eines Verteilungsgesetzes bzw. einer Prüfverteilung aus dem (bzw. den) empirisch gewonnenen Stichprobenbefund(en) berechnet wird. ♦

Unabhängig davon, welchen statistischen Test man in SPSS praktiziert, es gelten im Kontext des sogenannten p-value-Konzepts stets die folgenden Testentscheidungen:

p-value-Konzept

Ist ein empirisches Signifikanzniveau α^* kleiner oder gleich einem vorgegebenen Signifikanzniveau α , dann wird die Nullhypothese H_0 verworfen. Ansonsten wird die Nullhypothese H_0 beibehalten.



Hinweise. Im Kontext einer Testentscheidung erweisen sich die folgenden Hinweise als hilfreich: i) **Interpretation.** Im Fall, dass $\alpha^* \leq \alpha$ gilt, sagt man auch: Das Testergebnis ist statistisch signifikant (zum Niveau α). Gilt $\alpha^* > \alpha$ dann sagt man: Das Testergebnis ist statistisch nicht signifikant (zum Niveau α). ii) **Semantik.** Aus der Interpretation wird deutlich, dass man mit Hilfe eines statistischen Tests lediglich erkennen kann, ob ein Testergebnis im statistischen Sinne bedeutungsvoll ist oder nicht. Statistische Signifikanz ist nicht ohne Weiteres gleichzusetzen damit, dass ein Testergebnis auch unter sachlogischen Gesichtspunkten bedeutend ist. Wohl kann ein Testergebnis statistisch signifikant, jedoch ökonomisch irrelevant sein. Die Umkehrung ist gleichsam möglich: Ein nicht signifikantes Ergebnis kann ökonomisch relevant sein. ♣

A.4 Ausgewählte Grenz- und Elastizitätsfunktionen

Motivation. Für eine sachbezogene und plausible Interpretation der Parameter einer Regressionsfunktion erweisen sich die zugehörige Grenz- und Elastizitätsfunktion als nützlich. Der Wert einer Grenzfunktion heißt marginale (lat.: *marginalis* → den Rand betreffend) oder Grenzneigung. Der Wert einer Elastizitätsfunktion heißt Punkt-Elastizität. Während eine marginale Neigung als eine Maßzahl für eine durchschnittliche absolute Veränderung einer abhängigen Variablen gedeutet werden kann, wird eine Elastizität als eine Maßzahl für eine durchschnittliche relative Veränderung einer abhängigen Variablen angesehen.

Bivariate Funktionen. In der Tabelle A.4-1 sind gemäß Tabelle 6.2-1 und in Anlehnung an das SPSS Dialogfeld *Kurvenanpassung* (vgl. Abbildung 6.2-2) für ausgewählte und in praxi häufig applizierte bivariate (stetige und differenzierbare) Funktionen vom Typ $Y = f(X)$ die zugehörige Grenzfunktion und die zugehörige Elastizitätsfunktion explizit dargestellt.

Tabelle A.4-1: Grenz- und Elastizitätsfunktion für bivariate Funktionen

Funktion (Modell)	Grenzfunktion	Elastizitätsfunktion
Linear	$Y' = \beta_1$	$\varepsilon = \beta_1 \cdot X \cdot (\beta_0 + \beta_1 \cdot X)^{-1}$
Logarithmisch	$Y' = \beta_1 \cdot X^{-1}$	$\varepsilon = \beta_1 \cdot (\beta_0 + \beta_1 \cdot (\ln X))^{-1}$
Invers	$Y' = -\beta_1 \cdot X^{-2}$	$\varepsilon = -\beta_1 \cdot (\beta_0 \cdot X + \beta_1)^{-1}$
Exponent (Power)	$Y' = \beta_0 \cdot \beta_1 \cdot X^{\beta_1 - 1}$	$\varepsilon = \beta_1$
Wachstumsfunktion	$Y' = \beta_1 \cdot \exp(\beta_0 + \beta_1 \cdot X)$	$\varepsilon = \beta_1 \cdot X$

Multiple Funktionen. Bei multiplen Funktionen $Y = f(X_j)$ mit $k \geq 2$ erklärenden Variablen X_j ($j = 1, 2, \dots, k$) beruhen die Betrachtungen zu den Grenz- und Elastizitätsfunktionen auf den partiellen Ableitungen erster Ordnung $\partial Y / \partial X_j$. Für eine multiple lineare Funktion $Y = f(X_1, X_2, \dots, X_k) = \beta_0 + \beta_1 \cdot X_1 + \dots + \beta_k \cdot X_k$ ist die j -te partielle Grenzfunktion $\partial Y / \partial X_j = \beta_j$ ($j = 1, 2, \dots, k$) eine Konstante, die mit dem Koeffizienten β_j für die erklärende Variable X_j identisch ist. Die zur multiplen linearen Funktion gehörende partielle Elastizitätsfunktion ist dann durch die Funktion $\beta_j \cdot X_j / (\beta_0 + \beta_1 \cdot X_1 + \dots + \beta_k \cdot X_k)$ gegeben. Für eine multiple Exponentialfunktion $Y = f(X_1, X_2, \dots, X_k) = \exp(\beta_0 + \beta_1 \cdot X_1 + \dots + \beta_k \cdot X_k)$ lautet die entsprechende partielle Grenzfunktion $\partial Y / \partial X_j = \beta_j \cdot \exp(\beta_0 + \beta_1 \cdot X_1 + \dots + \beta_k \cdot X_k)$ und die zugehörige partielle Elastizitätsfunktion $\beta_j \cdot X_j$ ($j = 1, 2, \dots, k$).



Hinweis. Eine elementare und paradigmatische Einführung in die Regressionsanalyse findet man u.a. bei ECKSTEIN, Peter P.: Repetitorium Statistik, Deskriptive Statistik – Stochastik – Induktive Statistik, Mit Klausuraufgaben und Lösungen, 6., aktualisierte Auflage, GABLER Verlag Wiesbaden 2006 sowie ECKSTEIN, Peter P.: Statistik für Wirtschaftswissenschaftler – Eine realdatenbasierte Einführung mit SPSS, 3., aktualisierte und erweiterte Auflage, Springer Gabler Wiesbaden 2012. ♣

A.5 Verzeichnis ausgewählter SPSS Funktionsgruppen

Die in der Tabelle A.5-1 angebotene Zusammenstellung soll das Auffinden von ausgewählten SPSS Funktionsgruppen erleichtern, die in der praktischen Arbeit mit SPSS häufig werden.

Tabelle A.5-1: Ausgewählte SPSS Funktionen

Funktionsgruppe	Funktion	Anmerkung
Arithmetisch	Exp	Exponent zur Basis e
	Ln	Logarithmus zur Basis e
	Rnd	Rundung eines Wertes
	Sqrt	Quadratwurzel
Verteilungsfunktionen (engl.: <i>cumulative distribution function</i>)	Cdf.Binom	Binomialverteilung
	Cdf.Exp	Exponentialverteilung
	Cdf.Normal	Normalverteilung
	Cdf.Poisson	Poisson-Verteilung
Fehlende Werte	Missing	
Quantilsfunktionen (engl.: <i>inverse distribution function</i>)	Idf.Chisq	Chi-Quadrat-Verteilung
	Idf. Exp	Exponentialverteilung
	Idf.F	F(ISHER)-Verteilung
	Idf. Normal	Normalverteilung
	Idf.T	STUDENT t-Verteilung
Dichtefunktionen (engl.: <i>probability density function</i>)	Pdf.Binom	Binomialverteilung
	Pdf.Poisson	Poisson-Verteilung
Signifikanz(niveau)	Sig.Chisq	Chi-Quadrat-Verteilung
	Sig.F	F(ISHER)-Verteilung
Statistisch	Mean	Arithmetisches Mittel
	Sd	Standardabweichung
	Sum	Summe
	Variance	Varianz
Suchen	Index	Stringsuche
	Rindex	Zahlensuche
Umwandlung	Number	String in numerische Variable
	String	Numerische Variable in String
Verschiedene	\$Casenum	Fälle nummerieren
Zufallszahlen (engl.: <i>random variable</i>)	Rv.Normal	normalverteilte Zufallszahlen
	Rv.Uniform	gleichverteilte Zufallszahlen

Erläuterung. Die Funktionen werden nach ihrem Aufruf im SPSS Dialogfeld *Variable berechnen* in der mittig platzierten Rubrik kurz erläutert. ♣

A.6 Lösungen zu den Übungs- und Klausuraufgaben

Die mit * gekennzeichneten Aufgaben sind Klausuraufgaben.

Aufgabe 2-1

a) analog zu Beispiel 2.3-3, b) analog zu Beispiel 2.3-4 mit $KMI = \text{Gewicht} / (\text{Größe}/100)^{**2}$, c) analog zu Beispiel 2.3-6 mit Umkodierungen $0 \rightarrow „m“$ und $1 \rightarrow „w“$ z.B. für Variable „Geschlecht“, d) analog zum Beispiel 2.3-7, in Bereiche eingeteilte Variable: Gewichtigkeit, e) etwa via *Analysieren, Tabellen, einfache Tabellen*, Rubrik *Auswerten*: Gewichtigkeit, Rubrik *Untergruppen*: Sex, 10 schwergewichtige Männer und 93 untergewichtige Frauen

Aufgabe 2-2

Die separaten SPSS Datendateien können jeweils via Sequenz 2.3-7 erstellt werden. a) WestOst ≤ 7 : 3583 Mietwohnungen, b) WestOst > 7 : 4005 Mietwohnungen, c) NordSüd > 4 : 5532 Mietwohnungen, d) NordSüd ≤ 4 : 2056 Mietwohnungen, e) WestOst > 5 & WestOst < 8 & NordSüd > 4 & NordSüd < 6 : 1832 Mietwohnungen. f) $\sim(\text{WestOst} > 5 \text{ \& WestOst} < 8 \text{ \& NordSüd} > 4 \text{ \& NordSüd} < 6)$: 5756 Mietwohnungen

Aufgabe 2-3*

a) Merkmalsträger: Lehrveranstaltung, Grundgesamtheit: alle Lehrveranstaltungen, die am Fachbereich Wirtschaftswissenschaften I der HTW Berlin im Wintersemester 2002/03 angeboten wurden, b) Studiengang: nominal, Stern: ordinal, Durchfaller: metrisch, c) Lehrveranstaltungsverteilung auf die fünf Studiengänge (Schichten): insgesamt: 120 LV, Wirtschaftsrecht: 25 LV, Betriebswirtschaftslehre: 30 LV, BWL-Immobilien: 25 LV, BWL-Banken: 20 LV, Public Management: 20 LV, d) Filter: *Studium = 0 & Teilnehmer ≥ 40 & Durchfaller > 20* , Anzahl: 9 Lehrveranstaltungen, e) analog zu Beispiel 2.3-10, Break-Variable Studium, Grund- bzw. Hauptstudium: 66 bzw. 54 Lehrveranstaltungen, 2816 bzw. 1297 Teilnehmer, durchschnittliche Durchfallerquote: 13,49 % bzw. 8,88 %, Standardabweichung der Durchfallerquoten: 9,83 % bzw. 11,18 %, kleinster mittlerer Punktwert: 2,28 bzw. 2,11 Punkte, größter mittlerer Punktwert: 4,47 bzw. 4,88 Punkte

Aufgabe 3-1*

a) SPSS Auswahlbedingung: *Sex = 1 & Grund = 0*, b) 302 Hotelgäste c) i) 12 Erwartungen, ii) dichotome Variablen mit der Ausprägung 1 bei einer Nennung und systemdefinierter fehlender Wert bei Nichtnennung, Variablenkonzept: multiple Dichotomien, Analysekonzept: Analyse von Mehrfachantworten, iii) 2 Hotelgäste, iv) 77,3 % der Hotelgäste, v) 12,6 % der Erwartungen, vi) 136 mal, vii) im Durchschnitt 6,38 Erwartungen

Aufgabe 3-2*

a) 160 Börsentage, b) KOLMOGOROV-SMIRNOV-Anpassungstest in der LILLIEFORS-Modifikation auf eine unvollständig spezifizierte Normalverteilung, Testentscheidung: wegen $\alpha^* \geq 0,2 > \alpha = 0,05$ besteht kein Anlass, die unvollständig spezifizierte Normalverteilungshypothese zu verwerfen, c) Mittelwert: $-0,43\%$, d.h. in den vergangenen 160 Börsentagen lag die durchschnittliche prozentuale börsentägliche Veränderungsrate des Schlusskurses bei $-0,43\%$, Standardabweichung: $3,39\%$, d.h. in den vergangenen 160 Börsentagen lag die durchschnittliche Streuung der prozentualen börsentäglichen Veränderung um den Mittelwert von $-0,43\%$ bei $3,39\%$, d) via Sequenz 2.3-4 berechnet man i) eine Wahrscheinlichkeit von $CDF.NORMAL(2, -0.43, 3.39) = 0,763$ und ii) ein 0,9-Quantil der börsentäglichen prozentualen Veränderungsrate von $IDF.NORMAL(0.9, -0.43, 3.39) = 3,91\%$

Aufgabe 3-3*

a) Filter: $Sex = 2 \ \& \ Einkommen = 2 \ \& \ Alter = 2$, Stichprobenumfänge: 22 Kundinnen für Region Nord, 19 Kundinnen für Region Ost, 26 Kundinnen für Region West, 28 Kundinnen für Region Süd, b) via Sequenz 3.2.1-1 kann für die „abhängige Variable“ Sortiment sowie für die „unabhängige Variable“ Region der unvollständig spezifizierte KOLMOGOROV-SMIRNOV-Anpassungstest auf eine Normalverteilung in der LILLIEFORS-Modifikation appliziert werden, Testergebnisse: für jede der vier Regionen besteht wegen $\alpha^* \geq 0,2 > \alpha = 0,05$ kein Anlass, an der Normalverteilungsannahme der gemessenen Kundenzufriedenheitswerte zu zweifeln, c) Filter: $Sex = 2 \ \& \ Einkommen = 2 \ \& \ Alter = 2 \ \& \ Region = 2$, Stichprobenmittel: 5,22 Punkte, Stichprobenstandardabweichung: 1,42 Punkte, realisiertes 99 %-Konfidenzintervall [4,28 Punkte; 6,16 Punkte] kann via Sequenz 3.2.1-1 angefordert werden; demnach ist es nahezu sicher, dass der „wahre, jedoch unbekannt“ durchschnittliche Preiszufriedenheitswert zwischen 4,28 und 6,16 Punkten liegt, d) via Sequenz 2.3-4 berechnet man i) eine Wahrscheinlichkeit von $1 - \text{CDF.NORMAL}(7,5,22,1.42) = 0,105$ und ii) ein 0,8-Quantil der Preiszufriedenheit von $\text{IDF.NORMAL}(0.8,5.22,1.42) = 6,42$ Punkten

Aufgabe 3-4*

a) Stichprobenmittel: im Durchschnitt 0,899 Prüfungswiederholungen je Semester und Student, Stichprobenvarianz: 1,051, Stichprobenstandardabweichung: 1,025, d.h. im Durchschnitt streuen die Anzahlen der Prüfungswiederholungen um 1,025 Prüfungswiederholungen um den Durchschnitt, b) wegen $\alpha^* = 0,629 > \alpha = 0,05$ besteht kein Anlass, für die diskrete Zufallsgröße A: Anzahl der Prüfungswiederholungen pro Student und Semester die vollständig spezifizierte Verteilungshypothese $H_0: A \sim \text{Po}(0,899)$ zu verwerfen, c) Ereigniswahrscheinlichkeiten: i) $\text{PDF.POISSON}(3,0.899) = 0,0493$, ii) $\text{CDF.POISSON}(3,0.899) = 0,9866$, iii) $1 - \text{CDF.POISSON}(3,0.899) + \text{PDF.POISSON}(3,0.899) = 1 - 0,9866 + 0,0493 = 0,0627$

Aufgabe 3-5*

a) analog zu Beispiel 3.2.2-2 vollständig spezifizierter KOLMOGOROV-SMIRNOV-Anpassungstest auf eine Exponentialverteilung, wegen $\alpha^* = 0,552 > \alpha = 0,05$ besteht kein Anlass, die Ausgangshypothese zu verwerfen, b) vollständig spezifizierte Verteilungshypothese, Ereigniswahrscheinlichkeiten via Sequenz 2.3-4 berechnen, wobei $2 \text{ min} = 120 \text{ sec}$, i) $\text{CDF.EXP}(120,0.021) = 0,9195$, ii) $1 - \text{CDF.EXP}(120,0.021) = 0,0805$

Aufgabe 4-1*

a) via Sequenz 2.3-4 z.B. die SPSS Variable $Jahresmittel = \text{Fahrleistung}/(\text{Alter}/12)$ in die Arbeitsdatei einfügen, Filter $Typ = 6$ (Ford Fiesta) setzen, Stichprobenmittel von 12,604 (1000 km/Jahr) und Stichprobenstandardabweichung von 5,634 (1000 km/Jahr) z.B. via Sequenz 3.1.1-1 ermitteln, b) realisiertes 90 %-Konfidenzintervall [11,669 (1000 km/Jahr); 13,540 (1000/Jahr)] via Sequenz 3.2.1-1 anfordern, c) KOLMOGOROV-SMIRNOV-Test auf eine Normalverteilung (LILLIEFORS-Modifikation), der via Sequenz 3.2.1-1 angefordert werden kann, Testergebnis: wegen $\alpha^* = 0,043 > \alpha = 0,01$ besteht kein Anlass, an der Normalverteilungshypothese zu zweifeln, d) wegen 15 (1000 km) gilt $1 - \text{CDF.NORMAL}(15,12.604,5.634) \approx 0,335$; e) t-Test für zwei unabhängige Stichproben bei ungleichen Varianzen (WELCH-Test): wegen $\alpha^* = 0,748 > \alpha = 0,05$ wird Mittelwerthomogenitätshypothese nicht verworfen

Aufgabe 4-2*

a) z.B. $\text{Gesamt} = \text{Image} * 0.25 + \text{Stärke} * 0.10 + \text{Zufrieden} * 0.15 + \text{Qualität} * 0.30 + \text{Technik} * 0.10 + \text{Trend} * 0.10$, b) Europa: 14 Automarken, 2,93 Punkte, 0,71 Punkte, Asien: 9 Automarken, 2,55

Punkte, 0,44 Punkte, Amerika: 10 Automarken, 2,03 Punkte, 0,61 Punkte, c) i) Varianzhomogenitätstest nach LEVENE: wegen $\alpha^* = 0,293 > \alpha = 0,05$ besteht für die drei kontinentalen Automarkengruppen kein Anlass, an der Varianzhomogenität der Gesamtbewertungen zu zweifeln, ii) einfaktorielle ANOVA: wegen $\alpha^* = 0,006 < \alpha = 0,05$ muss für die drei kontinentalen Automarkengruppen die Mittelwerthomogenitätshypothese bezüglich der Gesamtbewertungen verworfen werden, d) z.B. Post-Hoc-Test nach SCHEFFÉ: zwei homogene Untergruppen, Gruppe 1: Europa und Asien, Gruppe 2: Amerika

Aufgabe 4-3*

a) K-S-Test in der LILLIEFORS-Modifikation, abhängige Variable *Note* und unabhängige Variable *Studiengang*, Testentscheidungen: da für alle fünf studiengangsspezifischen Stichproben $\alpha^* > \alpha = 0,05$ gilt, können die lehrveranstaltungsbezogenen Durchschnittsnoten jeweils als Realisationen einer normalverteilten Zufallsgröße aufgefasst werden, b) Varianzhomogenitätstest nach LEVENE: wegen $\alpha^* = 0,865 > \alpha = 0,05$ besteht kein Anlass, die Varianzhomogenitätshypothese zu verwerfen, c) einfaktorielle ANOVA: wegen $\alpha^* = 0,000 < \alpha = 0,01$ muss die Mittelwerthomogenitätshypothese verworfen werden, d) z.B. SCHEFFÉ-Test: es gibt zwei homogene Untergruppen, Gruppe 1: BWL-Banken und BWL-Immobilien; Gruppe 2: BWL, Wirtschaftsrecht, Public Management

Aufgabe 4-4*

a) K-S-Test in der LILLIEFORS-Modifikation, abhängige Variable *Votum*, unabhängige Variable *Status*, Testentscheidungen: da für alle drei gruppenspezifischen Stichproben $\alpha^* > \alpha = 0,05$ gilt, können die abgegebenen Voten der Statusgruppenmitglieder jeweils als Realisationen einer normalverteilten Zufallsgröße aufgefasst werden, b) Varianzhomogenitätstest nach LEVENE: wegen $\alpha^* = 0,323 > \alpha = 0,05$ besteht kein Anlass, die Varianzhomogenitätshypothese zu verwerfen, c) einfaktorielle ANOVA: wegen $\alpha^* = 0,000 < \alpha = 0,02$ muss die Mittelwerthomogenitätshypothese verworfen werden, d) z.B. SCHEFFÉ-Test: es gibt zwei homogene Untergruppen, Gruppe 1: Professoren und Mitarbeiter; Gruppe 2: Studenten

Aufgabe 4-5*

a) 11 Knoten, 5 dichotome Knoten, 6 Endknoten (der Ordnung 3, 6, 7, 8, 9, und 10), b) Verbesserung: 2,158, c) Eta-Quadrat = 0,437, d.h. zu 43,7 % kann die Gesamtvarianz der erfassten Körper-Masse-Indizes aus den 6 Endknoten erklärt werden, d) Merkmalsträger 1001: Knoten 3, 19,630 kg/m², Merkmalsträger 1002: Knoten 9, 23,882 kg/m²

Aufgabe 5-1*

a) *Abschluss*: ordinal, *Einkommen*: originär metrisch, wegen Klassierung ordinal, b) wegen einer rechteckigen Kontingenztabelle vom Typ (4-3) bzw. (3-4) ist das Kontingenzmaß Tau-c nach KENDALL geeignet, wegen $\tau_c = 0,232$ ist eine mittelstark ausgeprägte positive statistische Kontingenz zwischen Bildungsabschluss und Einkommen für die privat logierenden Hotelgäste nachweisbar, c) da die vier bildungsspezifischen Konditionalverteilungen nicht identisch sind, ist eine Kontingenz angezeigt, d) wegen $\alpha^* = 0,000 < \alpha = 0,01$ wird die Unabhängigkeitshypothese verworfen, e) Quantil einer Chi-Quadrat-Verteilung der Ordnung $p = 1 - 0,01 = 0,99$ für $df = (4 - 1) \cdot (3 - 1) = 6$ Freiheitsgrade: $\chi^2_{0,99,6} = \text{IDF.CHISQ}(0,99,6) = 16,81$

Aufgabe 5-2*

a) rechteckige (2-3)- bzw. (3-2)-Kontingenztabelle, b) da die zwei studiums-spezifischen Konditionalverteilungen nicht „völlig“ identisch sind, ist dies ein Hinweis auf eine „schwache“ Kontingenz zwischen beiden Merkmalen, c) da das Merkmal *Studiumstufe* nominal und das Merkmal

Klassifikation ordinal ist, verwendet man das Kontingenzmaß V nach CRAMÉR, wegen $V = 0,163$ ist eine schwache statistische Kontingenz zwischen beiden Merkmalen empirisch nachzuweisen, d) χ^2 -Unabhängigkeitstest: wegen $\alpha^* = 0,203 > \alpha = 0,05$ besteht kein Anlass, die Unabhängigkeitshypothese zu verwerfen, demnach ist die angezeigte statistische Kontingenz nicht signifikant (verschieden von null), e) Quantilsberechnung mittels der SPSS Funktion $IDF.CHISQ(1 - 0,02, 2) = 7,824$, f) erwartete absolute Häufigkeit von 14,3 Lehrveranstaltungen kann im Unterdialogfeld *Kreuztabellen: Zellen anzeigen* angefordert werden

Aufgabe 5-3

a) 9 Knoten insgesamt, 6 Endknoten (Knoten 1, 3, 5, 6, 7 und 8), b) höchste Trefferquote von 95,7 % im Knoten 1 für die 177 weiblichen Personen (von insgesamt 185), die ein Körpergewicht von 69 kg oder weniger besitzen, niedrigste Trefferquote von 0 % im Knoten 7, in dem nur 54 männliche Personen vorkommen, die schwerer sind als 79 kg und einen Körper-Masse-Index unter 26 kg/m² besitzen, c) exakte Klassifikation: i) 86,7 % aller weiblichen Personen, ii) 88,7 % aller männlichen Personen, iii) 87,6 % aller Personen, d) Risiko für Fehlklassifikation: $(24 + 32)/425 = 0,124$ bzw. 12,4 %

Aufgabe 5-4

Filter *Studiengang = 16* setzen und eine Maßkorrelationsanalyse bewerkstelligen, a) wegen $-0,023$ ist nur ein sehr schwacher negativer linearer statistischer Zusammenhang zwischen den Evaluationspunkten und der Semesterabschlussnote empirisch nachweisbar, der Maßkorrelationskoeffizient von $-0,023$ ist wegen $\alpha^* = 0,914 > \alpha = 0,05$ nicht signifikant verschieden von null ist, b) wegen $0,861$ ist ein starker positiver linearer statistischer Zusammenhang zwischen der Durchfallerquote und der Semesterabschlussnote empirisch nachweisbar, der Maßkorrelationskoeffizient von $0,861$ ist wegen $\alpha^* = 0,000 < \alpha = 0,05$ signifikant verschieden von null

Aufgabe 5-5*

Filter *Typ = 9* setzen, a) Merkmalsträger: gebrauchter PKW von Typ Opel Vectra; Erhebungsmerkmale mit Skalierung: Typ, nominal; Alter, Fahrleistung, Hubraum, Zeitwert jeweils metrisch, b) SPSS Variable *Durch = Fahrleistung/(Alter/12)*, c) KOLMOGOROV-SMIRNOV-Anpassungstest auf eine vollständig spezifizierte Normalverteilung: während die Merkmale Alter, (bisherige) Fahrleistung und jahresdurchschnittliche Fahrleistung jeweils als eine normalverteilte Zufallsgröße gedeutet werden können, trifft dies auf das Merkmal Zeitwert nicht zu, d) Maßkorrelationsanalyse: wegen $0,806$ und $\alpha^* = 0,000 < \alpha = 0,05$ besteht zwischen dem Alter und der Fahrleistung ein signifikanter starker positiver linearer statistischer Zusammenhang, wegen $-0,859$ und $\alpha^* = 0,000 < \alpha = 0,05$ besteht zwischen dem Alter und dem Zeitwert ein signifikanter starker negativer linearer statistischer Zusammenhang, wegen $-0,118$ und $\alpha^* = 0,243 > \alpha = 0,05$ besteht zwischen dem Alter und der jahresdurchschnittlichen Fahrleistung kein signifikanter Zusammenhang, wegen $0,383$ und $\alpha^* = 0,000 < \alpha = 0,05$ besteht zwischen der (bisherigen) Fahrleistung und der jahresdurchschnittlichen Fahrleistung ein signifikanter, allerdings schwacher positiver linearer statistischer Zusammenhang, wegen $-0,138$ und $\alpha^* = 0,170 > \alpha = 0,05$ besteht zwischen dem Zeitwert und der jahresdurchschnittlichen Fahrleistung kein signifikanter statistischer Zusammenhang, e) partielle lineare Maßkorrelationsanalyse: wegen $-0,471$ und $\alpha^* = 0,000 < \alpha = 0,05$ besteht zwischen dem Zeitwert und der jahresdurchschnittlichen Fahrleistung bei Gebrauchtwagen gleichen Alters ein signifikanter negativer partieller linearer statistischer Zusammenhang, f) Streudiagramm-Matrix kann via Grafiken, Diagrammerstellung angefordert werden, obgleich zwischen Alter und Zeitwert bzw. zwischen Alter und logarithmiertem Zeitwert ein negativer statistischer Zusammenhang ersichtlich ist,

unterscheiden sich beide Korrelationskoeffizienten in Höhe von $-0,859$ bzw. in Höhe von $-0,879$ voneinander, dies erklärt sich daraus, dass der Maßkorrelationskoeffizient immer nur die Stärke und die Richtung eines linearen statistischen Zusammenhanges messen kann, der im konkreten Fall für die originären Altersangaben und den logarithmierten Zeitwerten stärker ausgeprägt ist als für die originären Altersangaben und Zeitwerte, dieser Zusammenhang wird gleichsam augenscheinlich für den Korrelationskoeffizienten von „nur“ $0,978$ zwischen den originären und den logarithmierten Zeitwerten, obgleich zwischen beiden Variablen ein (nicht-linearer) funktionaler Zusammenhang besteht

Aufgabe 6-1

a) Filter *Geschlecht* = 1 setzen, Streudiagramm mit linearer Regression analog zum Beispiel 6.1-1, b) Y: Körpergewicht, X: Körpergröße, lineare Kleinst-Quadrate-Regressionsfunktion $Y^*(X) = -3654 + 140 \cdot X$ des Körpergewichts Y über der Körpergröße X, Parameterinterpretation: die Regressionskonstante $b_0 = -3654$ g fungiert als bloße Ausgleichskonstante und ist im konkreten Fall nicht plausibel interpretierbar, die Regressionskoeffizient $b_1 = 140$ g/cm ist ein Maß für die durchschnittliche Veränderung des Körpergewichts je Körpergrößeneinheit, d.h. steigt (fällt) die Körpergröße um 1 cm, dann steigt (fällt) im Mittel das Gewicht um 140 g, c) mittels der Regression ist man wegen $R^2 = 0,574$ in der Lage, „nur“ zu 57,4 % die Varianz des Körpergewichts allein aus der Varianz der Körpergröße statistisch zu erklären, die vergleichsweise geringe Bestimmtheit der Regression koinzidiert mit der vergleichsweise stark streuenden Punktwolke in Streudiagramm, d) wegen $Y^*(50) = -3654 + 140 \cdot 50 = 3346$ g hätte ein lebendgeborenes und 50 cm großes Mädchen erwartungsgemäß ein Körpergewicht von 3346 g

Aufgabe 6-2*

a) Merkmalsträger: ein gebrauchter VW Polo, Grundgesamtheit: alle 2003 auf dem Berliner Gebrauchtwagenmarkt angebotenen VW Polo, b) jeweils metrisch, c) aufgrund des „höchsten“ Bestimmtheitsmaßes von $R^2 = 0,931$ wählt man das so genannte „Aufbaumodell“ $Z^*(A) = \exp(9,5156 - 0,0173 \cdot A)$ aus, Interpretation des Bestimmtheitsmaßes: mit Hilfe der bivariaten nichtlinearen Regression ist man bereits in der Lage, zu 91,3 % die Varianz des Zeitwertes Z allein aus der Varianz des Alters A statistisch zu erklären, d) Test auf Signifikanz des Zeitwertfaktors Alter ist identisch mit Test des Bestimmtheitsmaßes auf Signifikanz, wegen $\alpha^* = 0,000 < \alpha = 0,01$ wird das Bestimmtheitsmaß von $R^2 = 0,931$ als signifikant verschieden von null und damit auch das Alter A als ein signifikant von null verschiedener Zeitwertfaktor gedeutet, e) auf der Grundlage der Grenzfunktion berechnet man eine marginale Zeitwertneigung von $dZ^*/dA(120) = -0,0173 \cdot e^{9,5156 - 0,0173 \cdot 120} \approx -29,4$ €/Monat, demnach hat man bei 120 Monate bzw. $120/12 = 10$ Jahre alten Gebrauchtwagen vom Typ VW Polo im Verlaufe eines Monats im Mittel mit einem Zeitwertverlust von ca. 30 € zu rechnen, auf der Grundlage der Elastizitätsfunktion berechnet man eine Punkt-Elastizität von $\varepsilon(120) = -0,0173 \cdot 120 \approx -2,1$ und interpretiert sie wie folgt: wegen $|\varepsilon(120)| = 2,1 > 1$ hat man bei 120 Monate bzw. 10 Jahre alten VW Polo mit einem überproportionalen durchschnittlichen relativen Zeitwertverfall von 2 % bei einer 1 %-igen Alterszunahme zu rechnen, f) wegen $Z^*(12) = \exp(9,5156 - 0,0173 \cdot 12) \approx 11026$ besäße c.p. ein Einjahreswagen erwartungsgemäß einen Zeitwert von 11026 €

Aufgabe 6-3*

a) Symbole: Z(eitwert), Z^* geschätzter Zeitwert, A(Alter), F(ahrleistung), M(otorleistung); multiple quasilineare Regressionsfunktion $\ln Z^*(A, F, M) = 9,231 - 0,108 \cdot A - 0,008 \cdot F + 0,003 \cdot M$, multiple nichtlineare Zeitwertfunktion: $Z^*(A, F, M) = e^{9,231 - 0,108 \cdot A - 0,008 \cdot F + 0,003 \cdot M}$, b) da für alle drei Regressoren $\alpha^* < \alpha = 0,05$ gilt, können sie jeweils als signifikante Zeitwertfaktoren aufge-

deckt werden, c) mit Hilfe der multiplen quasilinearen Regression ist man in der Lage, zu 81,9 % die Varianz der logarithmierten Zeitwerte aus der Varianz des Alters, der Fahrleistung und der Motorleistung statistisch zu erklären, d) unter Verwendung der nichtlinearen Funktion berechnet man einen Zeitwert von $Z^*(5, 100, 90) = e^{9,231 - 0,108 \cdot 5 - 0,008 \cdot 100 + 0,003 \cdot 90} \approx 3500 \text{ €}$, e) partielle Grenzfunktionen: $\partial Z^*/\partial A = -0,108 \cdot \exp(9,231 - 0,108 \cdot A - 0,008 \cdot F + 0,003 \cdot M)$, $\partial Z^*/\partial F = -0,008 \cdot \exp(9,231 - 0,108 \cdot A - 0,008 \cdot F + 0,003 \cdot M)$, $\partial Z^*/\partial M = 0,003 \cdot \exp(9,231 - 0,108 \cdot A - 0,008 \cdot F + 0,003 \cdot M)$, f) wegen $\ln Z^*(5, 100, 90) = 9,231 - 0,108 \cdot 5 - 0,008 \cdot 100 + 0,003 \cdot 90 = 8,161$ ergeben sich die folgenden marginalen Zeitwertneigungen: $\partial Z^*/\partial A = -0,108 \cdot \exp(8,161) \approx -378 \text{ €/Jahr}$, d.h. ein mittlerer partieller Zeitwertverlust von 378 € pro Jahr, $\partial Z^*/\partial F = -0,008 \cdot \exp(8,161) \approx -28 \text{ €/1000 km}$, d.h. ein mittlerer partieller Zeitwertverlust von 28 € je weitere 1000 km, $\partial Z^*/\partial M = 0,003 \cdot \exp(8,161) \approx 11 \text{ €/PS}$, d.h. eine mittlere partielle Zeitwert-erhöhung von 11 € je weiteres PS, g) da alle drei VIF-Werte kleiner als 5 sind, kann die Kollinearität unter den drei Regressoren vernachlässigt werden

Aufgabe 6-4*

a) Merkmalsträger: Kunde, Erhebungsmerkmale: Raststätte, Zufriedenheit mit Preis-Leistung, Reisegrund jeweils nominal, Pro-Kopf-Ausgaben, Verweildauer und Getränkeanzahl jeweils metrisch, b) 440 Kunden, c) multiples Logit-Modell $P^*(A, D) = 1/(1 + \exp(-(-1,679 + 0,219 A - 0,093 D)))$ mit P^* für geschätzte Wahrscheinlichkeit, A für Pro-Kopf-Ausgaben und D für Verweildauer ermitteln, d) wegen $\alpha^* = 0,000 < \alpha = 0,05$ sind beide Regressionskoeffizienten signifikant verschieden von null und damit wesentliche Faktoren zur Unterscheidung von Privat- und Geschäftsreisenden, e) ja: höhere Ausgaben bei gleicher Verweildauer erhöhen partiell die Wahrscheinlichkeit, dass ein Kunde ein Geschäftsreisender (Grund = 1) ist, eine längere Verweildauer bei gleichen Ausgaben vermindert partiell die Wahrscheinlichkeit, dass ein Kunde ein Geschäftsreisender ist, f) mit dem multiplen binären Logit-Modell ist man in der Lage, „nur“ zu 45,7 % die Zuordnung von Kunden (Privat- oder Geschäftsreisender) allein aus der Varianz der Pro-Kopf-Ausgaben für Speisen und Getränke sowie der Verweildauer statistisch zu erklären, g) gemäß der Klassifikationstabelle sind es 78,4 % aller befragten Kunden, h) Kunde A: wegen $P^*(25, 30) = 1/(1 + \exp(-(-1,679 + 0,219 \cdot 25 - 0,093 \cdot 30))) \approx 0,732 > 0,5$ Einordnung als Geschäftsreisender, Kunde B: wegen $P^*(20, 45) = 1/(1 + \exp(-(-1,679 + 0,219 \cdot 20 - 0,093 \cdot 45))) \approx 0,185 < 0,5$ Einordnung als Privatreisender

Aufgabe 6-5*

a) binäre logistische Regressionsfunktion $P^*(K1, K2, K4, K5) = 1/(1 + \exp(-(-0,514 + 0,288 K1 + 0,279 \cdot K2 + 0,566 \cdot K4 - 0,282 K5)))$, für die vier signifikanten Bilanzkennzahlen gilt jeweils $\alpha^* < \alpha = 0,08$, b) 77,1 % aller erfassten Unternehmen, c) bei Festlegung eines Trennwertes von 0,5 für die Bonitätsentscheidung würde man wegen $P^*(8, 2, 1; 4) = 1/(1 + \exp(-(-0,514 + 0,288 \cdot 8 + 0,279 \cdot 2 + 0,566 \cdot 1 - 0,282 \cdot 4))) \approx 0,86$ das Unternehmen A „sehr wahrscheinlich“ als solvent (Bonität = 1) und analog wegen $P^*(-3, -2, -1; 5) = 1/(1 + \exp(-(-0,514 + 0,288 \cdot (-3) + 0,279 \cdot (-2) + 0,566 \cdot (-1) - 0,282 \cdot 5))) \approx 0,02$ das Unternehmen B als insolvent bzw. „sehr unwahrscheinlich“ als solvent einstufen

Aufgabe 7-1*

a) äquidistante Zeitintervallreihe, b) Beobachtungszeitraum $T_B = \{t \mid t = 1, 2, \dots, 77\}$, Prognosezeitraum $T_P = \{t \mid t = 78, 79, \dots, 84\}$ und Relevanzzeitraum $T_R = \{t \mid t = 1, 2, \dots, 84\}$, c) U bezeichnet den Umsatz und t die Zeit; lineare Trendfunktion $U^*(t) = 50,618 + 0,289 \cdot t$ für $t \in T_R$, Interpretation der Trendparameter: wegen $U^*(0) = 50,618 = b_0$ hätte man für Samstag, den 10. einen Umsatz von ca. 5062 € geschätzt, wegen $d U^*(t)/d t = 0,289 = b_1$ steigt von Tag zu Tag

der Umsatz im Durchschnitt um ca. 29 €, d) mittels der Berechnungsvorschrift $Modell = Trend + Saison$ können die Umsatzwerte auf der Basis des additiven Trend-Saison-Modells bestimmt werden, Umsatzprognose (Angaben in €) für die Wochentage der 12. Woche: 8762, 5878, 6554, 7155, 7091, 7423, 8981, e) über Berechnungsvorschrift $Residuen = Umsatz - Modell$ können die Modellresiduen berechnet werden, Residualstandardfehler: 4,465 (100 €), d.h. im Mittel weichen die Umsatzzahlen um 447 € von den Modellwerten nach oben und nach unten ab

Aufgabe 7-2

a) äquidistante Zeitintervallreihe, Beobachtungszeitraum $T_B = \{t \mid t = 1, 2, \dots, 77\} = \{t^* \mid t^* = \text{Januar 2000, Februar 2002, \dots, Mai 2006}\}$, b) einfaches Modell der exponentiellen Glättung, Prognose für Juno bis Dezember 2006, Angaben in 1000 PKW-Neuzulassungen: 313, 281, 245, 277, 281, 273, 248, c) da für volatile und schwach stationäre Zeitreihe die empirischen Autokorrelationskoeffizienten rasch aussterben und die empirischen partiellen Autokorrelationskoeffizienten in Folge zwei signifikant von null verschiedene Koeffizienten indizieren, diagnostiziert man für die schwach stationäre Zeitreihe ein AR(2)-Modell und für die originäre Zeitreihe letztlich ein ARIMA(2, 1, 0)(0, 1, 0)-Modell, d) autoregressives Modell zweiter Ordnung mit Trend- und Saisonkomponente erster Ordnung, wegen $\alpha^* < \alpha = 0,05$ können die beiden autoregressiven Parameter von -0,66 und -0,41 als signifikant verschieden von null gedeutet werden, demnach kann man davon ausgehen, dass die Anzahl der PKW-Neuzulassungen jeweils durch die beiden vorhergehenden Monate beeinflusst wird, e) ARIMA-Modell-Prognose für Juno bis Dezember 2006, Angaben in 1000 PKW-Neuzulassungen: 353, 275, 265, 298, 301, 291, 272, die unterschiedliche Prognosen liegen in den verschiedenen Modellen begründet

Aufgabe 7-3*

a) äquidistante Zeitpunktereihe, Beobachtungszeitraum $T_B = \{t \mid t = 1, 2, \dots, 70\} = \{t^* \mid t^* = \text{Samstag, 10. Februar, \dots, Freitag, 27. April}\}$, b) via Sequenz 7.2.1-1 die Zeitreihe für die zentrierten gleitenden Durchschnitt MA7 bzw. für die zurückgreifenden gleitenden Durchschnitte PMA7 erstellen und via Sequenz 7.1-2 gemeinsam mit der originären Zeitintervallreihe Besuche in einem Sequenzdiagramm darstellen, originäre Zeitreihe der Besuchszahlen ist volatil mit leichtem Anstieg und augenscheinlicher Saisonalität, beide Zeitreihen der gleitenden Durchschnitte lassen im Beobachtungszeitraum einen volatilen und steigenden Verlauf der Besuchszahlen erkennen, der von den saisonalen Schwankungen bereinigt ist, c) Prognosen der Besuchszahlen für die elfte Ausstellungswoche: i) additives Trend-Saison-Modell: 8289, 6638, 5307, 6024, 6674, 6523, 6998, ii) Holt-Winters-Modell der exponentiellen Glättung: 8993, 7357, 6059, 6859, 7564, 7432, 7946, iii) ARIMA-Modell: 7854, 7268, 6269, 6622, 7619, 7626, 7442, d) Filter Zeit > 70 setzen und für die jeweiligen Abweichungen der ex-post Besuchszahlen von den Modellwerten die Standardabweichung berechnen: i) Trend-Saison-Modell: 274 Besuche; ii) Holt-Winters-Modell: 317 Besuche; iii) ARIMA-Modell: 530 Besuche

Aufgabe 7-4

a) Variablengenerierung: Zeitvariable $t = \$casenum$ und Variable $v1 = RV.NORMAL(0,1)$, Trajektorie ist schwach stationär, volatil und simuliert „weißes Rauschen“, b) Zeitreihe $v2 = CSUM(v1)$ mittels der SPSS Funktion *Kumulierte Summe* erstellen, Trajektorie simuliert einen „Random Walk“, c) Zeitreihe mittels der „Zielvariable“ $v3$ und des „numerischen Ausdrucks“ $1 - 2*v2$ „berechnen“, Trajektorie simuliert einen „Random Walk with Drift“, d) Zeitreihe $v4 = CSUM(v3)$ analog zu b) erstellen, Trajektorie simuliert einen „Random Walk with Drift“ bzw. eine „Irrfahrt mit ausgeprägtem Trend“

Aufgabe 7-5*

Im Beobachtungszeitraum, der insgesamt 256 Börsentage im Zeitraum vom 2. August 2004 bis zum 29. Juli 2005 umspannt, gleicht die Trajektorie des amtlichen Schlusskurses der Lufthansa-Stammaktie einem „Random Walk with Drift“ bzw. einer „trendbehafteten Irrfahrt“, der bzw. die letztlich nur mittels eines ARIMA(0, 1, 0)-Modells beschrieben werden kann.

Aufgabe 7-6

a) stark trendbehafteter bzw. integrierter stochastischer Prozess, b) Trajektorie der ersten Differenzen ist volatil und indiziert einen trendbehafteten konvexen Verlauf, Trajektorie der zweiten Differenzen ist volatil und augenscheinlich schwach stationär, c) da die empirischen partiellen Autokorrelationskoeffizienten rasch aussterben und nur der empirische Autokorrelationskoeffizient erster Ordnung signifikant verschieden von null ist, diagnostiziert man für die „trendbereinigte“ Zeitreihe einen MA(1)-Prozess, so dass die originäre Zeitreihe letztlich mittels eines ARIMA(0, 2, 1)-Modells beschrieben werden kann, d) wegen $\alpha^* = 0,00 < \alpha = 0,05$ wird der geschätzte MA-Parameter von 0,834 als signifikant verschieden von null gedeutet

Aufgabe 8-1*

a) analog zum Beispiel 2.3-6 die originären Variablen Praxis bzw. Theorie in die 0-1-kodierten „Ausgabeveriablen“ Praxis1 bzw. Theorie1 „umkodieren“, b) für die Variablen Praxis1 und Theorie1 das arithmetische Mittel bestimmen, das seinem Wesen nach ein Anteil ist, der als ein Gradmesser für die Schwierigkeit einer Prüfung interpretiert werden kann, da „nur“ 48 % der Fahrschüler im ersten Anlauf die Praxisprüfung bestanden haben, indiziert der „niedrigere Wert“ einen „höheren Schwierigkeitsgrad“ für eine Praxis-Prüfung als für eine Theorie-Prüfung, für die ein „höherer Wert“ von 0,67 und somit ein „niedrigerer Schwierigkeitsgrad“ ausgewiesen wird, c) McNemar-Test: wegen $\alpha^* = 0,000 < \alpha = 0,05$ ist die Hypothese bezüglich der Gleichheit der Schwierigkeitsgrade zu werfen, d) obgleich der punktbiseriale Korrelationskoeffizient von 0,262 eine „geringe Trennschärfe“ indiziert, ist diese geringe Trennschärfe wegen $\alpha^* = 0,000 < \alpha = 0,05$ dennoch statistisch gesichert, demnach können die Merkmale Geschlecht und Fahrübungsbedarf als trennscharf gedeutet werden

Aufgabe 8-2*

a) Merkmalsträger: Lehrveranstaltung, Gesamtheit: alle im Wintersemester 2003/03 im Studiengang BWL-Banken an der HTW Berlin evaluierten Lehrveranstaltungen, Umgang: 20 Lehrveranstaltungen, b) Reliabilitätsanalyse, c) CRONBACHS Alpha: 0,949, demnach sind die Fragen 20 bis 32 (ohne Frage 31) zur Messung des Konstrukts „didaktische Fertigkeiten eines Dozenten“ geeignet, d) verbannt man die Frage 28 aus dem Fragekatalog, dann erhöht sich CRONBACHS Alpha schrittweise von 0,949 auf 0,958, die „Verbannung“ der Frage 28 aus dem Katalog ist auch sachlogisch nachvollziehbar: die Aufgeschlossenheit und Freundlichkeit eines Dozenten muss nicht in einem Zusammenhang mit seinen didaktischen Fertigkeiten stehen

Aufgabe 9-1*

a) Merkmalsträger: Berliner Stadtbezirk, metrische Erhebungsmerkmale: Arbeitslosenquote, Anteil der Sozialhilfeempfänger, Anteil der Personen mit weniger als 500 €/Monat, Rentneranteil, Anteil der Personen mit Hauptschulabschluss, Anteil der Personen mit Hochschulreife, Akademikeranteil, Anteil der Personen ohne beruflichen Abschluss, Anteil der Arbeiter an Erwerbstätigen, Analysekonzept: Faktorenanalyse, b) zwei extrahierte Faktoren mit 82,3 % Erklärungsanteil an der Gesamtvarianz, inhaltliche Bestimmung der Faktoren: Faktor 1 beschreibt die soziale Situation, der Faktor 2 den sozialen bzw. Bildungsstatus, c) Scree-Plot: da nur die ersten

beiden Eigenwerte größer als eins sind, extrahiert man nur zwei Faktoren aus den neun Erhebungsmerkmalen, Ladungs- oder Komponenten-Diagramm: auf dem Faktor bzw. auf der Komponente 1 „soziale Situation“ hoch geladen sind die Merkmale Rentneranteil, Arbeitslosenquote, Sozialhilfeempfängerquote und Anteil der Personen ohne beruflichen Abschluss, auf dem Faktor bzw. auf der Komponente 2 „sozialer bzw. Bildungsstatus“ sind die Merkmale Arbeiteranteil, Anteil der Personen mit Hauptschulabschluss; Anteil der Personen mit Hochschulreife und Akademikeranteil; d) die merkmalssträgerspezifischen Rangfolgen auf der Grundlage der Faktorwerte der beiden extrahierten Faktoren können via *Transformieren, Rangfolge bilden* erzeugt werden, der Grad der Übereinstimmung zwischen den faktorbasierten Rangfolgen und den beiden originären SPSS Variablen *Sozial* und *Status* kann z.B. mit dem SPEARMANSchen Rangkorrelationskoeffizienten gemessen werden, da beide Koeffizienten (in ihren absoluten Beträgen) nahe an eins liegen (0,91 bzw. 0,89), ist ein hoher Übereinstimmungsgrad zwischen den „vorgegebenen“ und den „faktoranalytisch ermittelten“ Stadtbezirksrängen zu konstatieren

Aufgabe 9-2*

a) Korrelationsmatrix für die 7 Erhebungsmerkmale analog zur Abbildung 9.2-3 anfordern, wegen 0,961 korrelieren die „Anzahl offener Stellen“ und „Anzahl von Kurzarbeitern“ am stärksten positiv und wegen $-0,72$ die „Insolvenzquote“ und „Erwerbstätigenquote“ am stärksten negativ miteinander, b) zwei Faktoren mit einem Varianzerklärungsanteil von insgesamt 75,4 %, Faktor 1 beschreibt den ersten und Faktor 2 den zweiten Arbeitsmarkt

Aufgabe 9-3*

a) Filter: (*Fachbereich = 3 | Fachbereich = 4*) & *Fragebögen* ≥ 20 ; Umfang: 36 Lehrveranstaltungen, b) neun (metrische) Evaluationsmerkmale: Aufbau, Praxisbezug, Verständlichkeit, Umgang mit Studierenden, Tempo, Einbeziehung der Studierenden, Weiterempfehlung, Teilnehmeranzahl, Anzahl auswertbarer Fragebögen, c) Faktorenanalyse, d) Varimax-Methode, Ladungsdiagramm: i) Kernaussage: zwei Gruppen von Merkmalen scharen sich jeweils um die orthogonalen und rotierten Komponentenachsen, ii) Faktorentaufe: der Faktor 1 beschreibt qualitative und der Faktor 2 quantitative Eigenschaften einer Lehrveranstaltung, iii) auf dem Faktor 1 ist das Evaluationsmerkmal „Dozent kann Kompliziertes verständlich machen“ mit einer Faktorladung von 0,967 am höchsten geladen, auf dem Faktor 2 ist das Evaluationsmerkmal „Anzahl der eingeschriebenen Teilnehmer“ mit einer Faktorladung von 0,973 am höchsten geladen, e) wegen der Faktorwerte 0,13 und 5,48 ist die Lehrveranstaltung mit Nr. 6200 überdurchschnittlich gut bewertet und sehr stark frequentiert worden, wegen der Faktorwerte $-2,10$ und $-0,50$ erfuhr die Lehrveranstaltung mit Nr. 8884 sehr geringe qualitative Bewertung, aus quantitativer Sicht wurde sie unterdurchschnittlich stark frequentiert bzw. besucht

Aufgabe 9-4*

a) Faktorenanalyse, b) drei Faktoren, c) 73,4 %, d) Varimax-Methode nach KAISER, e) Faktor 1: Fahrspurweite, Ladung: 0,895, Faktor 2: Belüftung, Ladung: 0,886, Faktor 3: Stellplatz wiederfinden, Ladung: 0,860, f) Faktor 1: Konstruktionsfaktor, Faktor 2: Zustandsfaktor, Faktor 3: Orientierungsfaktor, g) ANDERSON-RUBIN-Verfahren, h) weibliche Parkhausnutzer bewerten den Konstruktionsfaktor überdurchschnittlich (0,164), den Zustandsfaktor unterdurchschnittlich ($-0,120$) und den Orientierungsfaktor durchschnittlich (0,070)

Aufgabe 9-5*

a) Faktorenanalyse, b) zwei Faktoren, c) Varimax-Methode nach KAISER, d) Faktor 1: Selbststudium mit Faktorladung 0,789, Faktor 2: Vorlesungsbesuch mit Faktorladung 0,860, e) Faktor

1: curriculares Studium, Faktor 2: extracurriculares Studium, f) ANDERSON-RUBIN-Verfahren, g) weibliche Studierende bewerten das curriculare (0,358) und extracurriculare (0,031) Studium überdurchschnittlich, männliche Studierende bewerten wegen -0,436 und -0,038 beide Studienfaktoren unterdurchschnittlich

Aufgabe 10-1

a) Analysekonzept: hierarchische Clusteranalyse, Basis: standardisierte Daten, b) Klassifikationen: i) nächstgelegener Nachbar: Cluster 1: Charlottenburg-Wilmersdorf; Cluster 2: Mitte, Friedrichshain-Kreuzberg, Cluster 3: die restlichen neun Stadtbezirke, ii) entferntester Nachbar: Cluster 1: Mitte, Friedrichshain-Kreuzberg, Neukölln, Cluster 2: Pankow, Tempelhof-Schöneberg, Treptow-Köpenick, Marzahn-Hellersdorf, Lichtenberg; Cluster 3: die restlichen vier Stadtbezirke, iii) Ward-Verfahren: Cluster 1: Mitte, Friedrichshain-Kreuzberg, Neukölln, Cluster 2: Pankow, Tempelhof-Schöneberg, Treptow-Köpenick, Marzahn-Hellersdorf, Lichtenberg; Cluster 3: die restlichen vier (westlichen) Stadtbezirke

Aufgabe 10-2*

i) 3-Cluster-Klassifikation: Cluster 1: Baden-Württemberg, Bayern, Nordrhein-Westfalen; Cluster 2: Berlin, Brandenburg, Bremen, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt, Thüringen; Cluster 3: restlichen sechs (westlichen) Bundesländer; ii) 2-Cluster-Klassifikation: Cluster 1: Berlin, Bremen und die fünf neuen Bundesländer; Cluster 2: die restlichen neun alten Bundesländer

Aufgabe 10-3*

a) Auswahlbedingung: *Kontinent* = 1, b) Clusteranalyse, c) hierarchisch-agglomerative Klassifikation, d) 4 Automarkengruppen, e) 1 Automarke, Porsche, f) i) 38,82, d.h. Alfa Romeo und Porsche sind sich hinsichtlich der sechs Erhebungsmerkmale sehr unähnlich, ii) 2,36, d.h. Audi und Mercedes sind sich hinsichtlich der sechs Erhebungsmerkmale sehr ähnlich

Aufgabe 10-4*

a) 16 Kommunen, die zum Zwecke eines Preisvergleichs im zweiten Quartal 2007 erfasst wurden, für jede Kommune wurden sechs metrische Erhebungsmerkmale erfasst, b) Clusteranalyse, c) hierarchisch-agglomerative Klassifikation, d) sechs Kommunen-Cluster, e) sechs Kommunen, f) während sich wegen einer quadrierten euklidischen Distanz von 3,261 die Kommunen *Bernau* und *Calau* bezüglich der sechs metrischen Erhebungsmerkmale ähneln, sind sich wegen einer quadrierten euklidischen Distanz von 14,248 die Kommunen *Bernau* und *Luckau* bezüglich der sechs metrischen Erhebungsmerkmale unähnlich

Aufgabe 10-5*

a) drei Parkhaus-Cluster, b) drei Parkhäuser: Europa-Center, Rathaus-Passagen und Kultur-Brauerei, c) hinsichtlich der zehn Erhebungsmerkmale sind sich wegen 2,368 die Parkhäuser „DomAquaree“ und „Bebelplatz“ ähnlich und die wegen 57,987 die Parkhäuser „DomAquaree“ und Europa Center“ unähnlich, d) Parkhäuser „Bebelplatz“ und „Sony Center“ ♣

Stichwortverzeichnis

A

Abstand
 euklidischer 330
 quadrierter euklidischer 330
 ACF 256
 ANDERSON-RUBIN-Verfahren 321
 ANDREW-Plots 333
 AR(p)-Modelle 256
 ARIMA(p,d,q)-Modelle 270
 ARIMA-Modelle 255
 mit saisonalen Parametern 278
 ohne saisonale Parameter 283
 ARMA(p,q)-Modelle 267
 Ausreißerwerte 88
 Autokorrelationsfunktionen 257
 Autokorrelationsdiagramme 262

B

Balkendiagramm
 einfaches 68
 gestapeltes 66
 gruppiertes 138
 Bestimmtheitsmaß 191
 Binomialtest 74
 Box-and-Whisker-Plot 86
 BOX-JENKINS-Verfahren 276

C

CDF 71, 357
 CHAID-Entscheidungsbaum 160
 CHERNOFF-Faces 334
 Chi-Quadrat-
 Anpassungstest 67
 Unabhängigkeitstest 157
 Clusteranalyse 327
 Clusterstruktur 340
 Clusterzentrenanalyse 343
 Clusterzentren-Diagramm 346

COHENS Kappa 169
 CRONBACHS Alpha 299
 CRT-Entscheidungsbaum 143

D

Datei aufteilen 109
 Dateien zusammenfügen 32
 Daten
 aggregieren 49, 107
 bearbeiten 31
 eingeben 26
 einlesen 28
 speichern 27
 Datenansicht 19
 Datenbegriff 12
 Datenerhebung 12
 Datum definieren 231
 Dendrogramm 339
 Dichotome Methode 58
 Dichotomie, multiple 25
 Distanzmaß
 euklidisches 329
 quadriertes euklidisches 329
 Distanzmatrix 337
 Drei-Sigma-Regel 81
 Duplikate identifizieren 47

E

Eigenwerte 309
 Einheitsmatrix 318
 Ein-Stichproben-Verfahren 103
 Elastizitätsfunktion 198, 357
 Erhebungskonzept 14, 17
 Eta-Koeffizient 298
 Eta-Quadrat 147
 Explorative Datenanalyse 77
 Extremwerte 88

F

Faktorenanalyse 305
 Faktoren-Extraktion 307, 310
 Faktoren-Rotation 317
 Faktorentaufe 320
 Faktorladung 313
 Faktorwerte 321
 Fälle
 auswählen 44
 hinzufügen 32
 nummerieren 36
 Fehlende Werte 54
 Fehlerbalken 80
 Filter 46
 Fragebogen 16, 18
 Auswertung 52
 Konzept 16
 Funktionstypen, nichtlineare 193
 Fusionskoeffizient 338
 Fusionstabelle 338

G

Gleitende Durchschnitte 234
 zentrierte 236
 zurückgreifende 237
 Grenzfunktion 198, 357

H

Häufigkeitstabelle 53
 Histogramm
 normiertes 83
 Heterogenitätskoeffizient 338
 Homogenitätsindex Alpha 299

I

IDF 71, 357
 Interaktionsvarianz 142
 Inter-Item-Korrelation 292
 Interquartilsbereich 82

K

k-Stichproben-Verfahren 127
 Kernschätzer 223
 Klassierung
 äquidistante 43
 äquifrequente 43
 Drei-Sigma 43
 Klassifikation
 geometrische 331
 hierarchisch-agglomerative 335
 partitionierende 341
 Klassifizierungsbaum 143
 CHAID 160
 CRT 143
 Klassifikationsverfahren 330
 Kollinearität 201
 KOLMOGOROV-SMIRNOV-Test 93
 auf eine Exponentialverteilung 98
 auf eine Normalverteilung 94
 auf eine POISSON-Verteilung 96
 Kommunalität 315
 Konditionalverteilung 153
 Konfidenzintervall 79
 Kontingenzmaße 166
 nach COHEN 169
 nach CRAMÉR 167
 nach KENDALL 167
 Kontingenztafel 152
 Korrelation
 Maß- 174
 multiple 299
 partielle 180
 punktseriale 297
 Rang- 171
 Korrelationsmatrix 208, 306, 311
 Kreisdiagramm 64
 Kreuztafel 153
 KRUSKAL-WALLIS-Test 135
 Kurvenanpassung 195

L

Ladungsdiagramm 319
Lag-Zeitreihe 260
Lag-Operator 272
LEVENE-Test 119
LILLIEFORS-Modifikation 94
Linkage-Verfahren 335
LLR-Glättung 223
Logit 217

M

MA(q)-Modelle 264
MANN-WHITNEY-U-Test 120
Maßkorrelation 174
 partielle 180
MCNEMAR-Test 294
Median 81
Mehrfachantwortenanalyse 57, 60
 Sets definieren 59
Mittelwertvergleich 102
Modell
 Exponent 195
 Exponentiell 195
 Invers 195
 Linear 195
 Logarithmisch 195
 Power 195
 S 195
 Wachstum 195
 Zusammengesetzt 195
Monte-Carlo-Experiment 106
Multiple Dichotomien 58

O

Objektivität 292
Odds-Ratio 217
Operatoren, logische 44
Orthogonalität 318

P

PACF 256
PARETO-Diagramm 55, 333
Phi-Koeffizient 296
Post-Hoc-Test 131
Potenzfunktion 195
Proximitätsmaße 329
p-value-Konzept 356

Q

Q-Q-Plot 90, 91
Quantilsberechnung 71
Quartile 82

R

Random Walk 272
Rangkorrelation 171
Regression
 bivariate lineare 186
 bivariate nichtlineare 193
 logistische 210
 lokale lineare 223
 multiple 201
Reliabilität 291, 293
Residualanalyse 287
Residualstandardfehler 192
Rotationsverfahren 317

S

Saisonkomponente 243, 246, 250
SCHEFFÉ-Test 131
Schiefe 82
Scree-Plot 313
Sequenz 3
Sequenzdiagramm 230
Spannweite 82
SPSS Auswahlbedingung 45
SPSS beenden 3
SPSS Daten-Editor 4
SPSS Diagramm-Editor 7

SPSS Dialogfeld 8
 SPSS Expert Modeler 252
 SPSS Hilfesystem 9
 SPSS Optionen 10
 SPSS Pivot-Tabellen-Editor 6
 SPSS starten 3
 SPSS Variablendefinition 20, 24
 SPSS Viewer 4
 Standardabweichung 81
 Standardisierung 337
 Stationarität 254
 Stem-and-Leaf-Plot 84
 Stengel-Blatt-Diagramm 104
 Stochastischer Prozess 255, 279
 integrierter 279
 schwach stationärer 279
 Streudiagramm 175
 3D 332
 Matrix 207
 STUDENT t-Verteilung 107

T

Testentscheidung
 klassische 70
 unter SPSS 70, 256
 Trajektorie 231
 Trendfunktion
 lineare 238
 nichtlineare 240
 Trend-Saison-Modell 242
 additives 242
 multiplikatives 249
 t-Test
 doppelter 112
 einfacher 108
 gepaarte Stichproben 122
 unabhängige Stichproben 112

U

Unabhängigkeitstest 157, 173, 177

Urlistenkonzept 13

V

Variablen
 hinzufügen 34
 umkodieren 40
 Variablenansicht 21
 Variablenwerte berechnen 37, 40
 Varianz 81
 Varianzanalyse 127
 einfaktorielle 127
 zweifaktorielle 137
 Varianzhomogenitätstest 116
 Varimax-Verfahren 317
 Verteilungsparameter 79
 VIF-Werte 209
 Visuellen Klassieren 41

W

Wachstumsfunktion 193
 WARD-Verfahren 330
 Wahrscheinlichkeitsmodell
 lineares 211
 logistisches 212
 WELCH-Test 118
 WILCOXON-Test 125
 Wölbung 82

Z

Zeitreihe 231
 Zeitreihemodelle
 deskriptive 235
 stochastische 254
 Zufallsstichprobe ziehen 103
 Zufallszahlen 272, 357
 normalverteilte 272, 357
 Zuordnungsübersicht 337
 Zwei-Stichproben-Verfahren 112
 z-Transformation 341
 z-Werte 333, 341