



Formelsammlung und Hilfsmittel Statistik



© Fabio Basler 2020. Alle Rechte vorbehalten, auch bzgl. jeder Verfügung, Verwertung, Reproduktion, Bearbeitung sowie Weitergabe.

Die veröffentlichten Informationen, insbesondere Daten und Kalkulationen wurden sorgfältig recherchiert und nach bestem Gewissen erstellt. Ich übernehme keine Gewähr, Garantie oder Zusicherung für die Richtigkeit der veröffentlichten Informationen und Berechnungen der Excel-Fallstudien.

Formelsammlung und Hilfsmittel Statistik

Deskriptiver Teil

1. Mittelwertkennzahlen	3
1.1 Modus.....	3
1.2 Median.....	3
1.3 Arithmetisches Mittel.....	3
1.4 Boxplot	4
2. Streuungskennzahlen	5
2.1 Spannweite und Lineare Streuung.....	5
2.2 Varianz und Standardabweichung	5
2.3 Variationskoeffizient.....	5
3. Zusammenhangsmaßkennzahlen	7
3.1 Kontingenzkoeffizient.....	7
3.2 Rangkorrelation	8
3.3 Korrelationskoeffizient	9
4. Regressionsanalyse.....	10
5. Zeitreihenanalyse	12

Zusammenstellung von Formeln und Hilfsmittel

Deskriptiver Teil

1. Mittelwertkennzahlen

Mittelwertkennzahlen gelten in der Statistik als statistische Maßzahlen. Die Kategorie „Mittelwertkennzahl“ kann unterteilt werden in die Berechnung der Kennzahlen Modus, Median und arithmetisches bzw. geometrisches Mittel. Entscheidend für die Berechnung einer Kennzahl für ein Merkmal ist die Bestimmung des jeweiligen Skalenniveaus.

Die nachfolgende Tabelle illustriert die Klassifizierung der Kennzahlen durch die Skalenniveaus:

Skalenniveau	Nominal	Ordinal	Metrisch
Mittelwertkennzahlen	Modus	Median	Arithmetisches Mittel
			Geometrisches Mittel

1.1 Modus

Der Modus M_o einer Beobachtungsreihe ist die Merkmalsausprägung, die am häufigsten vorkommt. Im Unterschied zu der Berechnung mit Rohdaten liegt der Modus bei klassierten Daten in der Klasse mit der größten Klassenhäufigkeit. Voraussetzung zur Berechnung ist mindestens ein ordinales Skalenniveau.

Berechnung in Excel über die Formel „*=Modus.einf(...)*“

1.2 Median

Der Median M_e (Zentralwert) ist die Merkmalsausprägung der Beobachtung, die in der nach der Größe geordneten Beobachtungsreihe in der Mitte steht.

Berechnung in Excel über die Formel „*=Median(...)*“

1.3 Arithmetisches und geometrisches Mittel

(1) Das arithmetische Mittel \bar{x} ist definiert als:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

Er kumuliert eine Reihe von Beobachtungswerten und dividiert die Summe durch die Anzahl der Beobachten. Voraussetzung zur Berechnung ist mindestens ein metrisches Skalenniveau.

Berechnung in Excel über die Formel „*=Mittelwert(...)*“

(2) Das geometrische Mittel ist derjenige Mittelwert, den man mithilfe der n -ten Wurzel aus dem Produkt der betrachteten positiven Zahlen erhält. Das geometrische Mittel ist stets kleiner oder maximal gleich dem arithmetischen Mittel.

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

1.4 Boxplots

Der Box-Plot (Box-Whisker-Plot, Kastengrafik) ist ein Diagramm, das zur grafischen Darstellung der Verteilung eines mindestens ordinalskalierten Merkmals verwendet wird.

Es fasst dabei verschiedene robuste Streuungs- und Lagemaße in einer Darstellung zusammen. Ein Box-Plot soll schnell einen Eindruck darüber vermitteln, in welchem Bereich die Daten liegen und wie sie sich über diesen Bereich verteilen.

Vor dem Hintergrund werden alle Werte der sogenannten Fünf-Punkte-Zusammenfassung, also der Median, die zwei Quartile und die beiden Ausreißer, dargestellt.

Kurze enge Blöcke eines Boxplots zeigen den Schwerpunkt der Daten an – hohe Dichte.

In jedem Block liegen ein Viertel (25%) der Beobachtungswerte.

Nachfolgend eine Übersicht zur Einordnung der Bestandteile eines Boxplots:

Kenngröße	Definition	Beschreibung
Minimum	Kleinsten Datenwert des Datensatzes	Ende eines Whiskers oder entferntester Ausreißer
Unteres Quartil	Die kleinsten 25% der Datenwerte sind kleiner als dieser oder gleich diesem Kennwert	Beginn der Box
Median	Die kleinsten 50% der Datenwerte sind kleiner als dieser oder gleich diesem Kennwert	Strich innerhalb der Box, gängiges Maß für die Lage der Daten. Die eine Hälfte der Beobachtungen ist kleiner oder gleich dem Wert, die andere Hälfte der Beobachtungen ist größer oder gleich dem Wert.
Oberes Quartil	Die kleinsten 75% der Datenwerte sind kleiner als dieser oder gleich diesem Kennwert	Ende der Box
Maximum	Größter Datenwert des Datensatzes	Ende eines Whiskers oder entferntester Ausreißer
Spannweite	Gesamter Wertebereich des Datensatzes	Länge des gesamten Box-Plots (inklusive Ausreißer)
Interquartilsabstand	Wertebereich, in dem sich die mittleren 50% der Daten befinden. (liegt zwischen dem 0,25- und dem 0,75-Quartil)	Ausdehnung der Box, gibt den Abstand zwischen dem ersten und dem dritten Quartil ($Q_3 - Q_1$) an.
Ausreißer	Datenwerte, die weit entfernt von den anderen Datenwerten liegen	können sich stark auf Ergebnisse auswirken

Ausreißer können im Rahmen einer Ausreißeranalyse (Outlier Detection) identifiziert werden durch die Berechnung eines Intervalls:

$$I = [\bar{x} - a \cdot s; \bar{x} + b \cdot s]$$

Die Parameter a und b können auf einen bestimmten Wert festgesetzt werden.

Alle Werte, die außerhalb von dem Intervall liegen, werden als Ausreißer deklariert.

2. Streuungskennzahlen

Streuungskennzahlen gelten in der Statistik als statistische Maßzahlen, welche die Streuung vom Mittelwert quantifizieren. Die Kategorie „Streuungskennzahlen“ kann unterteilt werden in die Berechnung der Kennzahlen Spannweite, Lineare Streuung, Varianz und Standardabweichung sowie Variationskoeffizient. Entscheidend für die Berechnung einer Kennzahl für ein Merkmal ist die Bestimmung des jeweiligen Skalenniveaus.

Die nachfolgende Tabelle illustriert die Klassifizierung der Kennzahlen durch die Skalenniveaus:

Skalenniveau	Nominal	Ordinal	Metrisch
Streuungskennzahlen		Spannweite	Lineare Streuung
			Varianz und Standardabweichung
			Variationskoeffizient

1.1 Spannweite und Lineare Streuung

- (1) Spannweite: Die Spannweite R ist das einfachste Streuungsmaß in der Statistik und berechnet sich als Distanz zwischen dem größten und dem kleinsten Messwert:

$$R = x_{\max} - x_{\min}$$

- (2) Lineare Streuung: Die durchschnittliche Abweichung d ist das arithmetische Mittel aus den absoluten Abweichungen der Beobachtungswerte einer Verteilung von einem beliebigen Wert:

$$d_b = \frac{1}{n} \cdot \sum_{i=1}^n |x_i - b|$$

Berechnung in Excel über die Formel „=Mittelabw(...)“

1.2 Varianz und Standardabweichung

- (1) Varianz: Die Varianz ist die Summe der Abweichungsquadrate der Beobachtungswerte einer Verteilung von ihrem arithmetischen Mittel, dividiert durch die Anzahl der Untersuchungen

$$s^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

Berechnung in Excel über die Formel „=Var.P(...)“ in der deskriptiven Statistik bzw. „=Var.S(...)“ in der induktiven Statistik.

- (2) Standardabweichung: Die Standardabweichung errechnet sich aus der Wurzel der Varianz.

Berechnung in Excel über die Formel „=Stabw.N(...)“ in der deskriptiven Statistik bzw. „=Stabw.S(...)“ in der induktiven Statistik.

2.3 Variationskoeffizient

Der Variationskoeffizient gilt als relatives Streumaß und beschreibt das Verhältnis der Standardabweichung s zum arithmetischen Mittelwert.

Der Variationskoeffizient wird durch die nachfolgende Formel berechnet:

$$v = \frac{s}{\bar{x}}$$

Das Ergebnis als Wert kann anhand der nachfolgenden Systematik interpretiert werden:

Wert	Interpretation	XYZ-Analyse
0 – 0,1	Geringe relative Schwankung	X konstanter Verbrauch, Schwankungen sind eher selten
0,1 – 0,25	Mittlere relative Schwankung	Y stärkere Schwankungen im Verbrauch, meist aus trendmäßigen oder saisonalen Gründen
>0,25	Hohe relative Schwankung	Z völlig unregelmäßiger Verbrauch

3. Zusammenhangsmaßkennzahlen

Zusammenhangsmaßkennzahlen gelten in der Statistik als statistische Kennzahlen, die eine Beziehung in Form eines Zusammenhangs zwischen zwei Merkmalen quantifizieren. Die Kategorie „Zusammenhangsmaßkennzahlen“ kann unterteilt werden in die Berechnung der Kennzahlen Kontingenzkoeffizient, Rangkorrelation und Korrelationskoeffizient. Entscheidend für die Berechnung einer Kennzahl für ein Merkmal ist die Bestimmung des jeweiligen Skalenniveaus.

Die nachfolgende Tabelle illustriert die Klassifizierung der Kennzahlen durch die Skalenniveaus:

Skalenniveau	Nominal	Ordinal	Metrisch
Zusammenhangsmaße	Chi-Quadrat/ Kontingenzkoeffizient	Rangkorrelation nach Spearman	Kovarianz -> Pearson'scher Korrelationskoeffizient

3.1 Kontingenzkoeffizient

Der Kontingenzkoeffizient K^* (nach Karl Pearson) ist ein statistisches Zusammenhangsmaß. Der Pearson'sche Kontingenzkoeffizient drückt die Stärke des Zusammenhangs zwischen zwei nominalen oder ordinalen Merkmalen aus. Er basiert auf dem Vergleich von tatsächlich ermittelten Häufigkeiten zweier Merkmale mit den Häufigkeiten, die man bei Unabhängigkeit dieser Merkmale erwartet hätte.

Die Berechnung in Excel sollte über eine Pivot-Tabelle erfolgen. Damit können die Häufigkeiten zweier Variablen in einer 2-dimensionalen Übersicht erstellt werden. Im ersten Schritt sollte die Berechnung des sog. Chi-Quadrat-Wertes erfolgen. Im zweiten Schritt kann der Kontingenzkoeffizient K^* final berechnet und interpretiert werden.

Berechnung des Kontingenzkoeffizienten K^* :

$$K^* = \sqrt{\frac{\chi^2}{n + \chi^2}} \cdot \sqrt{\frac{\min(k, l)}{\min(k, l) - 1}}$$

Mit vorheriger Berechnung des Assoziationsmaßes Chi-Quadrat:

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - n_{ij}^e)^2}{n_{ij}^e} \quad \text{mit } n_{ij}^e = \frac{n_{i.} \cdot n_{.j}}{n}$$

Interpretation für Kontingenzkoeffizient K^* zwischen [0;1]:

Wert (K^*)	Interpretation
0	Kein Zusammenhang
0 – 0,3	Schwacher Zusammenhang
0,3 – 0,7	Mittlerer Zusammenhang
0,7 – 1	Starker Zusammenhang
1	Vollständiger Zusammenhang

3.2 Rangkorrelation

Der Rangkorrelationskoeffizient (nach Spearman) ist ein Maß für Korrelationen, um den Zusammenhang zweier ordinal skalierten Variablen zu berechnen.

Die Berechnung erfolgt über die Klassifizierung der einzelnen Ränge der Variablen X und Y:

$$r_s = \frac{\sum_{i=1}^n (r(x_i) - \bar{r}(X)) \cdot (r(y_i) - \bar{r}(Y))}{\sqrt{\sum_{i=1}^n (r(x_i) - \bar{r}(X))^2 \cdot \sum_{i=1}^n (r(y_i) - \bar{r}(Y))^2}}$$

Berechnung in Excel erfolgt über die Formel „=RANG.MITTELW(...; 0 bzw. 1)“ für beide Variablen. Anschließend können die einzelnen Ausprägungen der Ränge von X und Y als Korrelation berechnet werden über die Excel-Formel „=KORREL“ bzw. „=PEARSON“. Alternativ kann auch eine Korrelationsmatrix erstellt werden über die Datenanalysefunktion in Excel.

Interpretation für den Rangkorrelationskoeffizienten: Er kann Werte zwischen -1 und +1 annehmen. Bei einem Wert von +1 bzw. -1 besteht ein vollständig positiver (bzw. negativer) linearer Zusammenhang zwischen den betrachteten Merkmalen. Wenn der Korrelationskoeffizient den Wert 0 aufweist, hängen die beiden Merkmale überhaupt nicht linear voneinander ab.

Wert (rs)	Interpretation
-1	Vollständiger negativer/entgegengesetztgerichteter Zusammenhang
[-0,7] – [-1]	Starker negativer/entgegengesetztgerichteter Zusammenhang
[-0,3] – [-0,7]	Mittlerer negativer/entgegengesetztgerichteter Zusammenhang
0 – [-0,3]	Schwacher negativer/entgegengesetztgerichteter Zusammenhang
0	Kein Zusammenhang
0 – 0,3	Schwacher positiver/gleichgerichteter Zusammenhang
0,3 – 0,7	Mittlerer positiver/gleichgerichteter Zusammenhang
0,7 – 1	Starker positiver/gleichgerichteter Zusammenhang
1	Vollständiger positiver/gleichgerichteter Zusammenhang

3.3 Korrelationskoeffizient

Der Korrelationskoeffizient, ist eine Zusammenhangsmaßkennzahl für den Grad des linearen Zusammenhangs zwischen zwei mindestens metrisch-skalierten Merkmalen.

$$r_{xy} = \frac{COV(X,Y)}{s_x \cdot s_y}$$

Die Korrelation errechnet sich aus der Kovarianz beider Variablen.

Berechnung in Excel erfolgt über die Excel-Formel „=KORREL“ bzw. „=PEARSON“. Alternativ kann auch eine Korrelationsmatrix erstellt werden über die Datenanalysefunktion in Excel. Hierzu alle Daten mit Überschrift selektieren.

Interpretation für den Pearson'schen Korrelationskoeffizienten:

Wert (r _{xy})	Interpretation
-1	Vollständiger negativer/entgegengesetztgerichteter Zusammenhang
[-0,7] – [-1]	Starker negativer/entgegengesetztgerichteter Zusammenhang
[-0,3] – [-0,7]	Mittlerer negativer/entgegengesetztgerichteter Zusammenhang
0 – [-0,3]	Schwacher negativer/entgegengesetztgerichteter Zusammenhang
0	Kein Zusammenhang
0 – 0,3	Schwacher positiver/gleichgerichteter Zusammenhang
0,3 – 0,7	Mittlerer positiver/gleichgerichteter Zusammenhang
0,7 – 1	Starker positiver/gleichgerichteter Zusammenhang
1	Vollständiger positiver/gleichgerichteter Zusammenhang

Er kann Werte zwischen -1 und +1 annehmen. Bei einem Wert von +1 bzw. -1 besteht ein vollständig positiver (bzw. negativer) linearer Zusammenhang zwischen den betrachteten Merkmalen. Wenn der Korrelationskoeffizient den Wert 0 aufweist, hängen die beiden Merkmale überhaupt nicht linear voneinander ab.

4. Regressionsanalyse

Die Regressionsanalyse (linearer Funktionszusammenhang) ist ein Instrumentarium statistischer Analyseverfahren, um Beziehungen zwischen einer abhängigen (oft auch erklärte Variable, oder Regressand genannt) und einer oder mehreren unabhängigen Variablen (oft auch erklärende Variablen, oder Regressoren bezeichnet) zu modellieren. Sie werden insbesondere zur quantitativen Beschreibung linearer Funktionszusammenhänge oder zur Prognose verwendet.

Nachfolgend wie wichtigsten Punkte der Regressionsanalyse:

- Bereits vor der Berechnung sollte die sachlogische Korrektheit der Hypothese überprüft werden, dass die unabhängige Variablen (X_1, X_2, \dots, X_n) die abhängige Variable (Y) in einem Funktionszusammenhang determinieren
- Evtl. Prüfung der Skalenniveaus der Variablen (metrisches Skalenniveau)
- Evtl. Korrelationsanalyse über Datenanalysefunktion mit grafischer Visualisierung mittels Punkt-XY-Diagramm, anschließende Interpretation der Zusammenhänge der Merkmale bzgl. Stärke, Existenz und Richtung
- Festlegung der abhängigen Variablen Y und den unabhängigen Variablen X_1, X_2, \dots, X_p

- Regressionsanalyse über Datenanalysefunktion ausführen:

- Inhaltliche Güte im ersten Ausgabeblock interpretieren durch Bewertung des Bestimmtheitsmaßes (analoge Interpretation zu Kontingenzkoeffizient im Wertebereich von 0 und 1) und Bewertung des adjustierten Bestimmtheitsmaßes - Je höher r^2 , umso besser determinieren die unabhängigen Variablen die abhängige Variable, umso besser ist der Regressionsansatz als Ganzes
- Stabilität berechnen: Standardfehler/Mittelwert(abhängige Variable) und Stabilität interpretieren–

$$\frac{s}{|\bar{y}|} \quad \text{mit} \quad s = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n - J - 1}}$$

- Interpretation (analog zu XYZ-Analyse):
 - (1) zwischen 0 und 0,1 geringe Schwankung und hohe Stabilität,
 - (2) zwischen 0,1 und 0,25 mittlere Stabilität und
 - (3) zwischen 0,25 und ∞ niedrige Stabilität und hohe Schwankung
- Statistische Signifikanz im zweiten Ausgabeblock interpretieren durch Bewertung von F-krit (Irrtumswahrscheinlichkeit) – Ist F-krit akzeptabel klein, kann H_0 abgelehnt werden.
- Im dritten Ausgabeblock P-Werte als Signifikanzniveau interpretieren

- Konfidenzintervalle (Obere und Untere 95%) analysieren (Kein Vorzeichenwechsel! Länge s_{bj} zu b_j als Variationskoeffizient interpretieren)
- Berechnung der BETA-Koeffizienten, Interpretation der relativen Wichtigkeit der unabhängigen Variablen bei der Erklärung der abhängigen Variablen

$$BETA_j = \left| \hat{b}_j \right| \cdot \frac{s_{x_j}}{s_y}$$

- Absolute Einflüsse der unabhängigen Variablen erklären
(z.b.: durch 1 Einheit mehr von X_1 , nimmt Y um 5 Einheiten ab)
- Regressionsfunktion \hat{Y} durch Schnittpunkt und unabhängige Variablen

5. Zeitreihenanalyse

Die Zeitreihenanalyse beschäftigt sich mit der inferenzstatistischen Analyse von Zeitreihen und der Prognose bzw. Vorhersage von Trends und ihrer zukünftigen Entwicklung.

Zur Berechnung von Trends gibt es eine Vielzahl an Methoden und Ansätzen. In der nachfolgenden Tabelle werden die zwei gängigsten Verfahren vorgestellt:

Verfahren zur Trendermittlung	Interpretation
Naive Trendverfahren	1) Fortschreiben von absoluten Änderungen: $*T_{t+1} = T_t + (T_t - T_{t-1})$ 2) Fortschreiben von relativen Änderungen: $T_{t+1}^* = T_t \cdot \frac{T_t}{T_{t-1}}$ 3) Trendprognose durch Regressionsfunktion
1-seitig gleitender Durchschnitt	Festlegung der Unterperioden und Berechnung über das arithmetische Mittel „=MITTELWERT(...)“
Exponentielles Glätten	Berechnung über die exponentielle Glättung 1. Ordnung: $T_t^e = \alpha \cdot y_t + (1 - \alpha) \cdot T_{t-1}^e$ Hierbei wird ein fixer Glättungsparameter α verwendet (oft von 0,15)
Schwankungskomponente	$sp_j = y_j - T_j$: Schwankungskomponente in der Unterperiode j
Saisonermittlung	$s_{p1}^e = \alpha \cdot s_{p1} + (1 - \alpha) \cdot s_{(p-1)1}^e$

Standardisierung von Daten

Die Standardisierung ermöglicht den Vergleich zweier Datenreihen hinsichtlich Ihrer Entwicklung über die Zeit. Insbesondere, wenn die Datenreihen unterschiedliche Niveaus aufweisen.

Sie kann mit der Formel „=STANDARDISIERUNG(x;Mittelwert;Standardabweichung“ in Excel berechnet werden.

Fehlerberechnung

Die Fehlerberechnung kommt insbesondere dann zum Einsatz, wenn neben den Prognosewerten (y_i^*) auch das tatsächliche Ereignis (y_i) eintreitet. Dadurch kann die Differenz (Fehlerwert e) der Werte quantifiziert werden und der Prognosefehler bewertet werden:

$$e_i = |y_i - y_i^*|$$

Eine geeignete Kennzahl zur Interpretation ist der Variationskoeffizient Root Mean Square Error (VKRMSE). Dieser kann analog zum Variationskoeffizient interpretiert werden.

$$VKRMSE(n) = \frac{\sqrt{\frac{1}{n} \cdot \sum_{t=1}^n e_t^2}}{|\bar{y}|}$$

Die Kennzahl MAPE (Mean absolute percentage error) kann ebenfalls verwendet werden:

$$MAPE(n) = \frac{100}{n} \cdot \sum_{i=1}^n \left| \frac{e_i}{y_i} \right|$$

Literaturverzeichnis

Backhaus, Klaus; Erichson, Bernd; Plinke, Wulff; Weiber, Rolf (Hg.) (2000): Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Neunte, überarbeitete und erweiterte Auflage. Berlin, Heidelberg: Springer (Springer-Lehrbuch).

Bättig, Daniel (2017): Angewandte Datenanalyse. Der Bayes'sche Weg. 2., überarbeitete und erweiterte Auflage. Berlin: Springer Spektrum (Statistik und ihre Anwendungen).

Berekoven, Ludwig; Eckert, Werner; Ellenrieder, Peter (2009): Marktforschung. Methodische Grundlagen und praktische Anwendung. 12., überarbeitete und erweiterte Auflage. Wiesbaden: Gabler Verlag / GWV Fachverlage GmbH Wiesbaden.

Bleymüller, Josef; Gehlert, Günther; Gülicher, Herbert (1996): Statistik für Wirtschaftswissenschaftler. 10., überarb. Aufl. München: Vahlen (WiSt-Studienkurs).

Bonart, Thomas; Bär, Jürgen (2018): Quantitative Betriebswirtschaftslehre, Band I. Grundlagen, Operations Research, Statistik. Wiesbaden: Springer Gabler.

Bourier, Günther (2018): Beschreibende Statistik. Praxisorientierte Einführung - mit Aufgaben und Lösungen. 13. Auflage. Wiesbaden: Springer Gabler (Lehrbuch).

Eckstein, Peter P. (2016): Angewandte Statistik mit SPSS. Praktische Einführung für Wirtschaftswissenschaftler. 8., überarbeitete und erweiterte Auflage. Wiesbaden: Springer Gabler.

Frost, Ira (2018): Einfache lineare Regression. Die Grundlage für komplexe Regressionsmodelle verstehen. Wiesbaden: Springer VS (essentials).

Hedderich, Jürgen; Sachs, Lothar (2018): Angewandte Statistik. Methodensammlung mit R. Springer Spektrum. 16., überarbeitete und erweiterte Auflage. Berlin, Germany: Springer Spektrum.

Jochen Schwarze: Beschreibende Verfahren. 11., vollst. überarb. Aufl. (Grundlagen der Statistik, / Jochen Schwarze ; 1).

Kohn, Wolfgang; Öztürk, Riza (2017): Statistik für Ökonomen. Datenanalyse mit R und SPSS. 3., überarbeitete Auflage. Berlin, Heidelberg: Springer Gabler (Springer-Lehrbuch).

Matthäus Heidrun; Matthäus Wolf-Gert (2015): Statistik und Excel: Elementarer Umgang mit Daten (Deutsch) Taschenbuch

Reiter, Joachim (2017): Statistik-Fallstudien mit Excel. Klausurenkurs für Studierende der Betriebswirtschaft im Bachelor. Wiesbaden: Springer Gabler (Lehrbuch).

Scharnbacher, Kurt (2004): Statistik im Betrieb. Lehrbuch mit praktischen Beispielen. 14., aktualisierte Auflage. Wiesbaden: Gabler Verlag.

Zwerenz Karlheinz (2007): Statistik verstehen mit Excel: Interaktiv lernen und anwenden Buch mit ExcelDownloads: 2. Auflage (Managementwissen für Studium und Praxis)