

Helge Toutenburg · Christian Heumann

Deskriptive Statistik

Eine Einführung in Methoden
und Anwendungen mit R und SPSS

6., aktualisierte und erweiterte Auflage

Mit Beiträgen von
Michael Schomaker und Malte Wißmann

Inhaltsverzeichnis

1. Grundlagen	1
1.1 Grundgesamtheit und Untersuchungseinheit	1
1.2 Merkmal oder statistische Variable	2
1.3 Datenerhebung	7
1.4 Datenaufbereitung	13
1.5 Aufgaben und Kontrollfragen	18
2. Häufigkeitsverteilungen	21
2.1 Absolute und relative Häufigkeiten	21
2.1.1 Qualitative Merkmale	21
2.1.2 Quantitative Merkmale	23
2.2 Empirische Verteilungsfunktion	28
2.2.1 Ordinale Merkmale und diskrete Merkmale	28
2.2.2 Stetige Merkmale	30
2.3 Grafische Darstellung	34
2.3.1 Stab- oder Balkendiagramme	34
2.3.2 Kreisdiagramme	36
2.3.3 Stamm-und-Blatt-Diagramme	37
2.3.4 Histogramme	40
2.3.5 Kerndichteschätzer	42
2.4 Aufgaben und Kontrollfragen	44
3. Maßzahlen und Grafiken für eindimensionale Merkmale	49
3.1 Lagemaße	49
3.1.1 Modus oder Modalwert	50
3.1.2 Median und Quantile	52
3.1.3 Quantil-Quantil-Diagramme (Q-Q-Plots)	57
3.1.4 Arithmetisches Mittel	59
3.1.5 Geometrisches Mittel	65
3.1.6 Harmonisches Mittel	69
3.2 Streuungsmaße	72
3.2.1 Spannweite und Quartilsabstand	73
3.2.2 Mittlere absolute Abweichung vom Median	74
3.2.3 Varianz und Standardabweichung	75

3.2.4	Variationskoeffizient	80
3.3	Schiefe und Wölbung	81
3.3.1	Schiefe	81
3.3.2	Wölbung	82
3.4	Box-Plots	83
3.5	Konzentrationsmaße	84
3.5.1	Lorenzkurven	86
3.5.2	Gini-Koeffizient	87
3.6	Aufgaben und Kontrollfragen	91
4.	Maßzahlen und Grafiken für den Zusammenhang zweier Merkmale	97
4.1	Darstellung der Verteilung zweidimensionaler Merkmale	97
4.1.1	Kontingenztafeln bei diskreten Merkmalen	97
4.1.2	Grafische Darstellung bei diskreten Merkmalen	101
4.1.3	Maßzahlen zur Beschreibung der Verteilung bei stetigen und gemischt stetig-diskreten Merkmalen	103
4.1.4	Grafische Darstellung der Verteilung stetiger bzw. gemischt stetig-diskreter Merkmale	105
4.2	Maßzahlen für den Zusammenhang zweier nominaler Merkmale	107
4.2.1	Pearsons χ^2 -Statistik	109
4.2.2	Phi-Koeffizient	112
4.2.3	Kontingenzmaß von Cramer	114
4.2.4	Kontingenzkoeffizient C	115
4.2.5	Lambda-Maße	116
4.2.6	Der Yule-Koeffizient	118
4.2.7	Der Odds-Ratio	120
4.3	Maßzahlen für den Zusammenhang ordinaler Merkmale	122
4.3.1	Gamma	123
4.3.2	Kendalls tau- b und Stuarts tau- c	125
4.3.3	Rangkorrelationskoeffizient von Spearman	126
4.4	Zusammenhang zwischen zwei stetigen Merkmalen	130
4.5	Explorative Grafiken für mehrere Variablen	137
4.5.1	Coplots	137
4.5.2	Chernoff Faces	142
4.6	Sachgemäße Gestaltung von Grafiken	144
4.6.1	Adäquate Skalierung	145
4.6.2	Einfluss von Extremwerten	147
4.6.3	Geschickte Wahl einer Grafik	150
4.6.4	Probleme bei der Berechnung einer linearen Regression	155
4.7	Maße zur Messung der Übereinstimmung von Beobachtern	156
4.7.1	Kappa-Koeffizient	158
4.7.2	Gewichtetes Kappa	162
4.8	Aufgaben und Kontrollfragen	165

5. Zweidimensionale Merkmale: Lineare Regression	169
5.1 Einleitung	169
5.2 Plots und Hypothesen	171
5.3 Prinzip der kleinsten Quadrate	173
5.3.1 Bestimmung der Schätzungen	175
5.3.2 Herleitung der Kleinste-Quadrate-Schätzungen	175
5.3.3 Eigenschaften der Regressionsgeraden	177
5.4 Güte der Anpassung	181
5.4.1 Varianzanalyse	181
5.4.2 Korrelation	184
5.5 Residualanalyse	187
5.6 Lineare Transformation der Originaldaten	189
5.7 Multiple lineare Regression und nichtlineare Regression	191
5.8 Polynomiale Regression	193
5.9 Lineare Regression mit kategorialen Regressoren	195
5.10 Spezielle nichtlineare Modelle – Wachstumskurven	199
5.11 Aufgaben und Kontrollfragen	200
6. Zeitreihen	205
6.1 Kurvendiagramme	205
6.2 Zerlegung von Zeitreihen	206
6.3 Fehlende Werte, äquidistante Zeitpunkte	207
6.4 Gleitende Durchschnitte	207
6.5 Saisonale Komponente, konstante Saisonfigur	209
6.6 Modell für den linearen Trend	213
6.7 Praktisches Beispiel mit SPSS	215
6.8 Aufgaben und Kontrollfragen	217
7. Verhältniszahlen und Indizes	219
7.1 Einleitung	219
7.2 Einfache Indexzahlen	221
7.2.1 Veränderung des Basisjahres	222
7.3 Preisindizes	224
7.3.1 Preisindex nach Laspeyres	225
7.3.2 Preisindex nach Paasche	226
7.3.3 Alternative Preisindizes	227
7.4 Mengenindizes	227
7.4.1 Laspeyres-Mengenindex	228
7.4.2 Paasche-Mengenindex	228
7.5 Umsatzindizes (Wertindizes)	228
7.6 Verknüpfung von Indizes	229
7.7 Spezielle Probleme der Indexrechnung	231
7.7.1 Erweiterung des Warenkorbs	231
7.7.2 Substitution einer Ware	232
7.7.3 Subindizes	233

7.8	Standardisierung von Raten und Quoten	235
7.8.1	Datengestaltung für die Standardisierung von Raten ..	238
7.8.2	Indirekte Methode der Standardisierung	238
7.8.3	Direkte Standardisierung	242
7.9	Ereignisanalyse	245
7.9.1	Problemstellung	245
7.9.2	Grundbegriffe der Lebensdaueranalyse	248
7.9.3	Empirische Hazardrate und Überlebensrate	250
7.10	Aufgaben und Kontrollfragen	254
8.	Fehlende Daten	259
8.1	Betrachtung eines einzelnen Merkmals	259
8.1.1	Behandlung fehlender Daten für ein binäres Merkmal	262
8.1.2	Behandlung fehlender Daten für ein nominales Merkmal	267
8.1.3	Behandlung fehlender Daten für ein ordinales Merkmal	268
8.1.4	Behandlung fehlender Daten für ein metrisches Merkmal	272
8.2	Betrachtung zweier Merkmale	277
8.2.1	Zwei binäre Merkmale	279
8.2.2	Zwei metrische Merkmale	283
9.	Einführung in SPSS	287
9.1	Grundaufbau des Programms	287
9.1.1	Das Datenfenster	288
9.1.2	Das Grafikfenster	289
9.1.3	Das Syntaxfenster	290
9.2	Ein praktisches Beispiel	290
9.2.1	Aufbau des Datensatzes	291
9.2.2	Deskriptive Analyse	291
9.2.3	Zusammenhangsanalyse	297
9.2.4	Lineare Regression	299
9.2.5	Weiterführende Analysen	300
10.	Einführung in R	301
10.1	Installation und Grundaufbau des Programmpakets R	301
10.1.1	R als überdimensionierter Taschenrechner	302
10.1.2	Programmiersprache R	303
10.2	Einige praktische Beispiele	305
10.2.1	Einlesen der Daten	305
10.2.2	Deskriptive Analyse	307
10.2.3	Zusammenhangsanalyse	315
10.2.4	Lineare Regression	320

Lösungen zu den Übungsaufgaben	325
Literatur	385
Sachverzeichnis	387