Udo Bankhofer I Jürgen Vogel

Datenanalyse und Statistik

Eine Einfühftfri^'lur^ö'tenomen im Bachelor

GABLER

Teil 1: Beschreibende Statistik

1 Einführung ... 3

2 Grundbegriffe der Statistik ... 5

2.1 Grundgesamtheit und Merkmale ... 5

2.2 Skalenarten ... 8

2.3 Datenerhebung ... 10

2.4 Quellen wirtschaftsstatistischer Daten ... 11

3 **Häufigkeitsverteilungen ... 13**

3.1 Primäre Häufigkeitstabelle ... 13

3.2 Sekundäre Häufigkeitstabellen ... 16

ä.3 Grafische Darstellung von Häufigkeiten ... 21

4 Statistische Maßzahlen ... 27

4.1 Lageparameter ... 27

4.2 Streuungsparameter ... 32

4.3 Box-Whisker-Plots ... 37

4.4 Empirische Quantile ... 39

4.5 Empirische Momente ... 40

4.6 Konzentrationsmaße ... 44

5 Zusammenhänge zwischen Merkmalen ... 51

5.1 Der empirische Korrelationskoeffizient ... 52

5.2 Der Rangkorrelationskoeffizient von Spearman ... 55

5.3 Der Kontingenzkoeffizient ... 58

6 Lineare Regression ... 63

6.1 Lineare einfache Regression ... 64

6.2 Curvilineare Regression ... 70

7 **Indexzahlen** **75**
7.1 Einfache Indizes 75
7.2 Zusammengesetzte Indizes 76
7.3 Internationale Preisvergleiche 82
7.4 Einige wirtschaftlich bedeutsame Indizes 85

Teil 2: Schließende Statistik

8 Grundlagen der schließenden Statistik **91**
8.1 Grundbegriffe der schließenden Statistik 91
8.2 Wichtige Stichprobenfunktionen 94
8.3 Quantile von Prüfverteilungen 100
9 Punktschätzungen **105**
9.1 Wünschenswerte Eigenschaften von Punktschätzungen 106
9.2 Konstruktionsmethoden für Punktschätzungen 110
10 Bereichsschätzungen **117**
10.1 Definition des Begriffs Konfidenzintervall 117
10.2 Konfidenzintervalle für normalverteilte Merkmale 118
10.3 Konfidenzintervall für eine Wahrscheinlichkeit p 122
11 Signifikanztests **125**
11.1 Grundbegriffe der Testtheorie 125
11.2 Parametertests für normalverteilte Merkmale 127
11.3 Test auf Wahrscheinlichkeit 133
11.4 Anpassungstests 135
11.5 Tests auf Unabhängigkeit 140
11.6 Stichprobenvergleiche 144
11.7 Der Vorzeichentest 147
11.8 Signifikanztests in Statistiksoftware 151

Teil 3: Datenanalyse

12 Daten- und Distanzmatrizen ... 155
12.1 Objekte" Merkmale, Distanzen ... 155
12.2 Merkmalstypen und ihre Distanzen ... 158
12.3 Aggregation von Distanzen ... 164
13 Klassifikationsverfahren ... 173
13.1 Klassifikationstypen ... 173
13.2 Klassifikationsheuristiken ... 177
13.3 Bewertungskriterien ... 180
13.4 Partitionierende Verfahren ... 188
13.5 Hierarchische Verfahren ... 196
14 Repräsentationsverfahren ... 207
14.1 Mehrdimensionale Skalierung ... 208
14.2 Faktorenanalyse ... 219
15 Wentifikationsverfahren ... 225
15.1 Multiple Regression ... 227
15.2 Diskriminanzanalyse ... 234
15.3 Varianzanalyse ... 243

Teil 4: Data Mining

16 Gegenstand des Data Mining ... 253
16.1 Knowledge Discovery in Databases ... 253
16.2 Anwendungsbereiche und Methodenüberblick ... 255
16.3 Einsatzgebiete und Anwendungsbeispiele ... 258
17 Assoziationsanalyse ... 261
17.1 Grundlegende Begriffe ... 261
17.2 Generierung von Assoziationsregeln ... 263
17.3 Interessantheitsmaße ... 268
17.4 Sequenzanalyse ... 270

18 Entscheidungsbäume **273**

18.1 Klassifikationsbäume 273

18.2 Auswahlmaße 276

18.3. Entscheidungsbaumverfahren 282

18.4 Kritische Anmerkungen 283

Anhang

A Wahrscheinlichkeitstheorie **287**

A.1 Wahrscheinlichkeiten 287

A.2 Eindimensionale Verteilungen 292

A.3 Zweidimensionale Verteilungen 298

A.4 Grenzwertsätze 303

B Statistische Tafeln **307**

Literaturverzeichnis **315**

Stichwortverzeichnis **321**