

Inhaltsverzeichnis

1	Stichprobe, Zufallsvariable – Histogramm, Dichteverteilung	1
1.1	Stichprobenstatistiken	5
1.1.1	Zentralitätsmaße	5
1.1.2	Maße für Streuung	6
1.1.3	Stichprobenstatistiken am Beispiel	12
1.2	Häufigkeit, Dichte und Verteilung	14
2	Stichprobe, Zufallsvariable – Histogramm, Dichteverteilung in R	17
2.1	Datenerhebung	17
2.2	Dateneingabe	18
2.3	Daten importieren nach R	21
2.3.1	Kleine Datensätze manuell eingeben	21
2.3.2	Größere Datensätze aus einer Datei einlesen	22
2.4	Einfache, deskriptive Statistik und grafische Darstellung	25
2.4.1	Grafische Darstellung von Stichprobenstatistiken mit R	25
2.4.2	Deskriptive Statistik mit R	28
2.4.3	Dichtehistogramm und empirische Dichte	33
2.5	Übungen	35
3	Verteilungen, ihre Parameter und deren Schätzer	37
3.1	Verteilung	38
3.1.1	Zentraler Grenzwertsatz	40
3.2	Parameter einer Verteilung	41
3.3	Schätzer (für Parameter einer Verteilung)	44
3.3.1	Die <i>likelihood</i>	45
3.3.2	Maximierung der <i>likelihood</i>	46
3.3.3	<i>Maximum likelihood</i> – analytisch	48
3.3.4	<i>Maximum likelihood</i> – numerisch	51
3.3.5	<i>Maximum likelihood</i> – Eine Laudatio	51
3.4	Einige wichtige Verteilungen	52
3.4.1	Normalverteilung	52

3.4.2	Bernoulli-Verteilung	53
3.4.3	Binomialverteilung	54
3.4.4	Poisson-Verteilung	55
3.4.5	Negative Binomialverteilung	55
3.4.6	Lognormal-Verteilung	57
3.4.7	Uniforme Verteilung	58
3.4.8	β -Verteilung	59
3.4.9	γ -Verteilung	60
3.4.10	Beschränkte Verteilungen (<i>truncated distributions</i>)	62
3.5	Auswahl der Verteilung	62
3.5.1	Vergleich zweier Verteilungen: der Kolmogorov-Smirnov Test . .	63
3.5.2	Vergleich von Fits: <i>likelihood</i> und der informationstheoretische Ansatz	65
4	Verteilungen, Parameter und Schätzer in R	69
4.1	Darstellen von Verteilungen	71
4.1.1	Kontinuierliche Verteilungen	71
4.1.2	Diskrete Verteilungen	74
4.2	Berechnung der <i>likelihood</i> eines Datensatzes	76
4.3	Kolmogorov-Smirnov-Test	82
4.4	Test auf Normalverteilung	85
4.5	Übungen	86
5	Korrelation und Assoziation	89
5.1	Korrelation	90
5.1.1	Nicht-parametrische Korrelation	92
5.1.2	Korrelation mit und zwischen diskreten Variablen	93
5.1.3	Multiple Korrelationen	94
5.2	Test auf Assoziation – der χ^2 Test	95
6	Korrelation und Assoziation in R	97
6.1	Nicht-parametrische Korrelation	99
6.2	Multiple Korrelationen und die Korrelationsmatrix	101
6.3	Punktbiseriale und punktpolyseriale Korrelation	104
6.4	Der χ^2 -Test mit R	106
6.5	Übungen	108
7	Regression – Teil I	109
7.1	Regression	109
7.1.1	Regression: ein Verteilungsparameter variiert mit dem Prädiktor .	110
7.1.2	Regression und <i>maximum likelihood</i>	113
7.1.3	Die andere Skala und die <i>link</i> -Funktion	114
7.2	Kategoriale Prädiktoren	117

7.2.1	Ein kategorialer Prädiktor mit zwei Leveln	118
7.2.2	Ein kategorialer Prädiktor mit mehr als zwei Leveln	120
7.3	Ein paar Beispiele	122
7.3.1	Größe und Geschlecht – ein GLM ohne G	122
7.3.2	Raucher und Geschlecht – der χ^2 -Test als binomiales GLM	124
8	Regression in R – Teil I	127
8.1	Regression mittels GLM	127
8.2	Regression: <i>maximum likelihood</i> zu Fuß	136
8.2.1	Poisson-Modell zu Fuß	136
8.2.2	Nicht-lineare Regression zu Fuß	142
8.3	GLM mit VGAM	145
8.4	Modellierung beliebiger Verteilungsparameter	148
8.5	Übungen	152
9	Regression – Teil II	155
9.1	Modelldiagnostik	157
9.1.1	Analyse des Prädiktors	157
9.1.2	Analyse einflussreicher Punkte	158
9.1.3	Analyse der Dispersion	159
9.1.4	Analyse der Residuen	160
9.1.5	Analyse des funktionellen Zusammenhangs von y und x	163
10	Regression in R – Teil II	171
10.1	Modelldiagnostik	171
10.1.1	Analyse des Prädiktors	173
10.1.2	Analyse einflussreicher Punkte	174
10.1.3	Analyse der Dispersion	177
10.1.4	Analyse der Residuen	179
10.1.5	Analyse des funktionellen Zusammenhangs von y und x	181
10.2	Regressionsdiagnostik im linearen Modell (lm)	188
10.3	Übungen	193
11	Das Lineare Modell: t-Test und ANOVA	195
11.1	Der t -Test	196
11.1.1	Statistik mit nur einer Stichprobe (<i>one sample test</i>)	196
11.1.2	Vergleich zweier gepaarter Stichproben (<i>paired sample test</i>) . . .	198
11.1.3	Vergleich zweier ungepaarter Stichproben (<i>two sample test</i>) . . .	199
11.2	Varianzanalyse (ANOVA)	199
11.2.1	ANOVA mit kontinuierlichem Prädiktor: ein Beispiel	204
11.2.2	Annahmen der ANOVA	205
11.2.3	Nicht-normalverteilte Daten ANOVA-fähig machen	206
11.2.4	ANOVA für mehr als 2 Level	209

11.2.5	Post-hoc Vergleiche	210
11.3	Von der Regression zur ANOVA	213
11.3.1	ANOVA und Regression: Ein Ergebnisvergleich	213
11.3.2	Freiheitsgrade der ANOVA und ihre Erklärung durch die Regression	215
11.4	ANOVAs für GLMs	216
12	Das Lineare Modell: t-Test und ANOVA in R	219
12.1	t -Test und Varianten in R	219
12.2	ANOVA in R	223
12.2.1	Test auf Varianzhomogenität	224
12.2.2	Aus F -Werten die Signifikanz berechnen	227
12.2.3	Post-hoc Vergleiche mit R	229
12.3	ANOVA zur Regression und zurück	230
12.4	ANOVAs für GLM	231
12.5	Übungen	233
13	Hypothesen und Tests	235
13.1	Wissenschaftliche Methodik: Beobachtungen, Ideen, Modellvorstellungen, Hypothesen, Experimente, Tests und wieder von vorn	236
13.2	Das Testen von Hypothesen	238
13.2.1	Kochrezept für den Test von Hypothesen	238
13.2.2	Testfehler	239
13.3	Tests	242
13.3.1	Weitere Begriffe zum Thema Tests	243
13.3.2	Schlussbemerkungen zu Tests	244
13.4	Übungen	244
14	Experimentelles Design	247
14.1	Designprinzipien	248
14.1.1	Prinzip 1: Repräsentativität	248
14.1.2	Prinzip 2: Unabhängigkeit	249
14.1.3	Prinzip 3: Kontrollgruppe	251
14.2	Was alles schief gehen kann!	251
14.3	Wichtige Designs für manipulative Experimente	253
14.3.1	Vollständig randomisiertes Blockdesign	254
14.3.2	Split-plot Design	257
14.3.3	Nested design	264
14.4	survey Design	267
14.4.1	Einstufige Stichprobeverfahren: <i>simple random sampling</i>	268
14.4.2	Mehrstufige Stichprobeverfahren	270

15	Multiple Regression: mehrere Prädiktoren	279
15.1	Visualisierung von mehreren Prädiktoren	280
15.1.1	Visualisierung bei zwei kategorialen Prädiktoren	280
15.1.2	Visualisierung bei einem kategorialen und einem kontinuierlichen Prädiktor	282
15.2	Interaktionen zwischen Prädiktoren	284
15.3	Kollinearität	287
15.3.1	Hauptkomponentenanalyse	288
15.3.2	Cluster-Analyse	295
15.4	Modellselektion	297
15.4.1	Zwei-Wege ANOVA zu Fuß	301
15.5	Die Mathematik hinter dem linearen Modell	304
16	Multiple Regression in R	309
16.1	Interaktionen visualisieren und fitten	310
16.1.1	Zwei kategoriale Prädiktoren: Regression und ANOVA	310
16.1.2	Ein kontinuierlicher und ein kategorialer Prädiktor	314
16.1.3	Multiple Regression mit zwei kontinuierlichen Variablen	321
16.2	Kollinearität	328
16.2.1	Hauptkomponentenanalyse in R	329
16.2.2	Cluster-Analyse in R	333
16.3	Modellselektion	335
16.3.1	Modellselektion per Hand	336
16.3.2	Automatische Modellselektion	340
16.3.3	Alle Modelle fliegen hoch: <i>best-subset regression</i>	345
16.4	Übungen	346
17	Ausblick	347
	Literatur	349
	Sachverzeichnis	355