

h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Kumuliertes Vorlesungsskript zu den Vorlesungen

Statistische Methoden / KI
Qualitätsmanagement
Qualitätsmanagement II

Stand: 27. Januar 2021

Prof. Dr. Andreas Thümmel

Contents

1 Vorwort **13**

2 Der Qualitätsbegriff **14**
2.1 Was bedeutet Qualität ? 14
2.2 Wie Qualität gemessen wird 15
2.3 Die Informationsgewinnung 18
2.4 Die Dimensionen der Qualität 21
2.5 Kosten, Zeit und Qualität 22

3 Die Systemanalyse **22**

4 Der Hypothesentest **25**
4.1 Konfidenzintervalle 29
4.2 Statistischer Test 33
4.2.1 Hypothesentest und Stichprobengröß 37
4.3 Der kritische Wert 37
4.4 Der Fehler erster und zweiter Art 38
4.4.1 Exkurs: Berechnung der Irrtumswahrschenlichkeit 40
4.5 Die t-Verteilung im Hypothesentest 41
4.6 Weitere Hypothesentests: Voraussetzungen prüfen 42
4.7 Nicht-parametrische Tests 44

5 Die Vorab-Analyse **44**
5.1 Zeitreihendiagramm 45
5.1.1 Test auf Ausreißer 48
5.1.2 Test auf Trend 50
5.1.3 Test auf Runs 52
5.1.4 Test auf Strukturbruch 53
5.2 Prozessdaten als stationärer stochastischer Prozess bzw. Zeitreihe 54
5.3 Grafiken für nominale Daten 54
5.4 Histogramm und Verteilungsmodell 55
5.4.1 Test auf Verteilung 57
5.4.2 Wahrscheinlichkeitsnetz 60
5.4.3 Anmerkungen zur Problematik von Histogrammen 62
5.4.4 Quantil-Aussagen anhand der Verteilung 63

6 Qualität von Prozessen **64**
6.1 Wesentliche Kriterien der Qualität der Produkte eines Prozesses . 65
6.2 Prozessfähigkeitsanalyse 66
6.3 Beispiel zur Prozessfähigkeitsanalyse (PFA) 67

7 Qualitätsregelkarten **71**
7.1 Definition einer Qualitätsregelkarte für ein stetiges Merkmal . . . 72
7.2 Definition einer Qualitätsregelkarte für ein attributives Merkmal . 85

8 Annahmeprüfung **88**
8.1 Einstufiger Prüfplan 90
8.2 Durchschlupf 94
8.3 Aufwand 96
8.4 Der Algorithmus von Guenther 97
8.5 Zweistufiger Prüfplan 98
8.6 Mittlerer Aufwand eines zweistufigen Prüfplanes 101
8.7 Äquivalente Prüfpläne 102
8.8 Sequentieller Prüfplan 103

9 Regressionen für stetige Antwortvariablen **108**
9.1 Varianzanalyse, ANOVA 110
9.1.1 Die Grundidee der Varianzanalyse 110
9.1.2 Anwendungsbeispiel ANOVA 112
9.1.3 Trennschärfe der ANOVA 120
9.2 Mehrfaktorielle ANOVA 123
9.2.1 Prinzip der Sparsamkeit, Modell-Robustheit und Prognose 128
9.2.2 Wechselwirkungen 129
9.2.3 Beispiel: 130
9.2.4 Residuum-Analyse, VIF-Werte, ungewöhnliche Beobachtungen, AIC und BIC 137

10 Lineare Regression **141**

11 Anwendung ANOVA: Gage R& R **148**
11.0.1 Beispiel zur MSA 151
11.0.2 Kriterien der AIAG 153
11.0.3 Linearität und Bias 157
11.0.4 Exkurs: Messsystemanalyse für attributive Daten 160

12 GLM **163**

13 Anwendung GLM: Design of Experiments **171**
13.1 Full Factorial Designs . . . 171
13.2 Reduziertes Design, Teilfaktorieller Versuchsplan . . . 175
13.3 Saturierung, Blockvariablen und Blöcke, Zentralpunkte, Auflösung, Aliasstruktur, Vermengung, Faltung . . . 176
13.4 Split-Plot-Design . . . 180
13.5 Praktisches Erstellen eines Full-Factorial-DoE . . . 180
13.6 Optimierung . . . 196

14 Lebensdauer- und Zuverlässigkeitsanalyse **200**
14.1 Lebensdauerdiagramme . . . 201
14.2 Zensur . . . 202
14.2.1 Allgemeine und Intervall-Zensur . . . 203
14.2.2 Rechtszensur . . . 204
14.3 Ausfallverhalten . . . 204
14.3.1 Verteilungsfreie Schätzungen . . . 204
14.3.2 Verteilungsbasierte Schätzungen . . . 207
14.3.3 WeiBayes-Schätzung . . . 216
14.3.4 Maximum-Likelihood-Schätzung für Daten mit Zensur . . . 218
14.3.5 Parametertests . . . 219
14.4 Felddaten- und Garantieanalyse . . . 221
14.5 Testplanung . . . 224
14.5.1 Schätzungstests . . . 224
14.5.2 Nachweistests . . . 227
14.5.3 Ausfallschätzung unter Vorwissen: Beyer/Lauster . . . 229
14.6 Ausfallursachenanalyse . . . 229
14.7 Regression mit Lebensdauerdaten . . . 232
14.8 Beschleunigte Lebensdauertest . . . 235
14.8.1 Testplanung für beschleunigte Lebensdauertests . . . 245
14.9 Probit-Analyse . . . 248
14.9.1 Probit-Analyse als logistische Regression . . . 250

15 Transformationen **252**
15.0.1 Johnson-Transformation . . . 252
15.0.2 Anwendung auf Lebensdauerdaten . . . 253

16 Anwendung von Multivariate Analysen im QM 258
16.1 Hauptkomponentenanalyse . . . 258
16.1.1 Einleitung . . . 258
16.1.2 Das Prinzip der Methode . . . 259
16.1.3 Beispiel . . . 261
16.2 Clusteranalyse . . . 265
16.2.1 Clusterbeobachtungen . . . 265
16.2.2 Clusterzentrenanalyse (k-means) . . . 270
16.3 Diskriminanzanalyse . . . 270
16.3.1 Ein Diskriminanzmodell für Clustervariablen . . . 270

17 Data Mining-Methoden 272
17.1 Entscheidungsbäume: CART . . . 272
17.2 Bayes'scher Schätzer . . . 275
17.3 Treenets . . . 275
17.4 MARS . . . 275
17.5 GPS/Generalized Lasso . . . 275
17.6 Random Forrests . . . 275

18 Techniken von Industrie 4.0 275
18.1 Big Data-Datenbanken und -abfragen . . . 275
18.1.1 Hadoop . . . 275
18.1.2 Spark . . . 275
18.2 Cloudservices und -analytics . . . 275

19 Anhang 275
19.1 Hypothesentest . . . 275
19.2 Die Operationscharakteristik . . . 279
19.3 Qualität der Produktion von Einzelstücken oder im Prozess . . . 280
19.4 Schätzung der Gesamt-Ausschussquote bei Kenntnis der Teil-Ausschussquoten . . . 281
19.5 χ^2-Test . . . 282

Zuordnung der Kapitel zu den aktuellen Vorlesungen an der hda:

Kapitel	QM I	QM II	SKI
1	x		x
2	x		x
3	x	x	x
4	x	(x)	x
5	x	(x)	x
6	x		x
7	x		x
8	x		
9	x		x
10	x		x
11	x		x
12	x		x
13	x	(x)	x
14		x	
15		x	
16	(x)	x	x
17	(x)	x	x
18	(x)	x	x
19	x	x	y

x: ist vollständig relevant, (x): ist nur teilweise relevant (ggf. als Wiederholung im Programm)

List of Figures

1 Merkmalsbaum . . . 17
2 Beispiel für Merkmalausprägungstypen in MTB . . . 18
3 Systemanalyse lt. Wikipedia . . . 23
4 Black Box-Analyse für eine Systemanalyse . . . 24
5 Stichprobengröß 10 für die Schätzung des Mittelwertes . . . 27
6 100 Stichproben a 10 für die Schätzung des Mittelwertes . . . 27
7 Mittelwert von 100 Stichproben a 10 . . . 28
8 Histogramm der Stichproben-Mittelwerte . . . 28
9 Standard-Normalverteilung mit 95%-Üerdeckung . . . 30
10 Konfindenz-Asymptotik: Abhängigkeit von der Stichprobengröße . 31
11 t-Verteilung mit 9 Freiheitsgraden und 95%-Üerdeckung . . . 32
12 95%-Konfidenzintervall der t-Verteilung . . . 32
13 Konfidenzintervalle zu 100 Stichproben . . . 33
14 Vergleich Wert im Konfidenzintervall . . . 34
15 Geschätze Grundgesamtheit . . . 36
16 Fehler erster und zweiter Art im Hypothesentest . . . 39
17 t-Verteilungen unterschiedlicher Freiheitsgrade . . . 42
18 Beispieldatensatz - erste Zeilen . . . 46
19 XY-Plot . . . 47
20 Pareto-Diagramm . . . 55
21 Histogramm zum Zeitreihendiagramm . . . 56
22 Histogramm . . . 56
23 Empirische Verteilungsfunktion . . . 58
24 Best Fit Normal und Weibull . . . 59
25 Tabelle AD-Test . . . 59
26 estFit emp. Vert. und Histogramm . . . 60
27 Wahrscheinlichkeitsnetz für die Beispieldaten . . . 61
28 Perzentil-Schätzung anhand emp. Verteilungsfunktion und WS-Netz . . . 62
29 Histogramm mit 11 Klassen . . . 63
30 Quantil-Schätzung . . . 64
31 Quantile der Normalverteilung . . . 65
32 Prozessfähigkeitsanalyse . . . 68
33 Prozessfähigkeitsanalyse normalverteilt, s instabil . . . 69
34 Prozessfähigkeitsanalyse: Six Pack . . . 70
35 Prozessfähigkeitsanalyse: Six Pack für Mittelwertversatz . . . 71

36 xs-Shewhart-Qualitätsregelkarte . . . 72
37 Standard-Normalverteilungsquantile für 3 σ . . . 74
38 $\bar{x} - s$ - Karte mit unterschiedlichen Stichprobengrößen . . . 75
39 $\bar{x} - s$ - Karte mit Fehlern . . . 76
40 Fehlerinformation zur X-S-Karten . . . 76
41 Tests zu den Stabilitätskriterien einer $\bar{x} - s$ - Karte . . . 77
42 QRK mit Ausreißer provoziert Folgefehler . . . 78
43 QRK mit Ausreißer und Abschnitten - ohne Folgefehler . . . 80
44 I-MR-Karte . . . 81
45 I/MR - R/S - Karte . . . 83
46 Gleitender Durchschnitt - Karte . . . 84
47 EWMA - Karte . . . 84
48 CUSUM - Karte . . . 85
49 p - Karte für eine Quote p . . . 86
50 Tests zu den Stabilitätskriteren einer p - Karte . . . 86
51 Np - Karte . . . 87
52 u - Karte . . . 88
53 Prozessbild für den einstufigen Prüfplan . . . 90
54 Eingabe Minitab Annahmeprüfung - Prüfplan . . . 92
55 Minitab Annahmeprüfung - Prüfplan . . . 93
56 Zweistufiger Prüfplan . . . 100
57 Intervall I_n zum sequentiellen Prüfplan . . . 107
58 Regressionen - 4+2-Schema . . . 109
59 Daten für die ANOVA . . . 112
60 Test auf NV für Faktoren in ANOVA . . . 113
61 F-Test ANOVA Output . . . 114
62 Einfache ANOVA . . . 115
63 Einfache ANOVA - Fehlerbalken . . . 116
64 Residuum und Anpassungen ANOVA . . . 117
65 Verteilungsanalyse ANOVA Residuum Gesamt . . . 117
66 4 in 1 Diagramm Residuum ANOVA . . . 118
67 Boxplot ANOVA . . . 119
68 Intervallplot ANOVA . . . 120
69 Trennschärfe ANOVA . . . 121
70 Trennschärfe ANOVA - theoretisch . . . 122
71 Beispieldatenerfassung für das GLM-Beispiel . . . 125
72 Haupteffekte der 4 Faktoren des Beispiels . . . 126
73 GLM mit 4 Faktoren . . . 130

74 2fach-Wechselwirkungen der 4 Faktoren des Beispiels 131
75 Aktivieren des Prinzips der Sparsamkeit 132
76 PdS-Methoden 133
77 GLM-4F-PdS1 134
78 GLM-4F-PdS2.jpg 135
79 GLM-4F-PdS3.jpg 136
80 GLM-4F-PdS4.jpg 136
81 GLM-Speichern des Residuums 138
82 GLM-Res-NV 139
83 GLM-Prog 140
84 GLM-Prog 141
85 Streudiagramm für das Beispiel Alter vs. Gehalt 142
86 Streudiagramm für das Beispiel hohe Korrelation 143
87 Regressionsgeraden für schwach korrelierte Größen 144
88 Regressionsgeraden für stark korrelierte Größen 145
89 Einflüsse auf Messwerte - Gründe für eine Streuung 148
90 Streuungsaufteilung bei der Messsystemanalyse mittels ANOVA . 149
91 Daten MSA Typ 2 151
92 Grafik MSA Typ 2 156
93 MSA: Linearität und Bias 157
94 MSA: Linearität und Bias - Einzelwertdiagramm 158
95 MSA: Linearität und Bias - Grafik 159
96 MSA: attributive Daten - analytisch 161
97 MSA: attributive Daten - analytisch, Ergebnis 162
98 MSA: attributive Daten - Prüferübereinstimmung, Daten 163
99 MSA: attributive Daten - Prüferübereinstimmung, Ergebnis . . . 163
100 GLM: Daten 165
101 GLM: Intervallplot der Daten 165
102 GLM: Residuum-Grafiken 170
103 GLM: Test Residuum auf Normalverteilung 170
104 Min. Anzahl Messungen 1 Faktor, lineares Modell 171
105 2-Faktoren-Schema 173
106 3-Faktoren-Schema 174
107 Reduziertes 3-Faktoren-Schema 176
108 DOE Faktor-Tabelle 180
109 DOE Datenblatt teil-faktorieller Versuchsplan 181
110 DOE Datenblatt teil-faktorieller Versuchsplan 182
111 DOE Datenblatt voll-faktorieller Versuchsplan mit Daten 183

112 DOE Pareto-Diagramm der Effekte für Werte . . . 191
113 DOE Pareto-Diagramm der Effekte für Kosten . . . 191
114 DOE Residuendiagramm für Werte . . . 192
115 DOE Residuendiagramm für Kosten . . . 192
116 DOE Residuum: Test auf Normalverteilung . . . 193
117 DOE Haupteffektediagramm Werte . . . 193
118 DOE Haupteffektediagramm Kosten . . . 194
119 DOE Wechselwirkungsdiagramm Werte . . . 194
120 DOE Wechselwirkungsdiagramm Kosten . . . 195
121 DOE Würfeldiagramm Werte . . . 195
122 DOE Würfeldiagramm Kosten . . . 196
123 DOE Optimierung Dialog - Einstellungen . . . 197
124 DOE Optimierung Ergebnis . . . 198
125 DOE Optimierung Ergebnis mit geänderter Bedeutung . . . 198
126 DOE Kontur-Diagramm . . . 199
127 DOE Wirkungsflächen-Diagramm . . . 199
128 DOE überlagertes Konturdiagramm . . . 200
129 Ausfalldiagramm ohne Zensur . . . 201
130 Ausfalldiagramm mit Zensur . . . 202
131 Histogramm von Ausfalldaten . . . 202
132 Beispiel: Intervallzensur . . . 203
133 Beispiel: Rechtszensur . . . 204
134 Ausfallschätzungen nach Kaplan-Meier . . . 206
135 Histogramm der Ausfalldaten . . . 208
136 AD-Werte für die Bestimmung der passenden Ausfallverteilung . . . 209
137 Bestimmung der bestpassenden Ausfallverteilung . . . 210
138 Übersicht über die ausgewählte Ausfallverteilung . . . 211
139 Schätzung Teil 1 . . . 213
140 Schätzung Teil 2 . . . 214
141 Schätzung Teil 3 . . . 215
142 Daten für die Ausfallschätzung mit WeiBayes . . . 216
143 WeiBayes in MTB . . . 217
144 Ausfallschätzung mit WeiBayes . . . 218
145 Zwei Weibull-Vert.: Form = 1,5, Skalen = 70 und 90 . . . 220
146 Testergebnis Parametertest Weibull . . . 220
147 Garantiedatenstruktur . . . 221
148 Garantiedaten in Intervallzensur . . . 222
149 Garantieprognose - Zahlen . . . 223

150 Garantieprognose - Grafik . . . 223
151 Verteilung mit Konfidenzintervallen . . . 225
152 MTB-Dialog zur Testplanung - Schätzung . . . 226
153 Testplanung - Schätzung über Prozente . . . 226
154 Testplanung - Schätzung über Werte . . . 227
155 Testplanung: Gleichung für erlaubte Ausfälle . . . 228
156 MTB Dialog für Nachweistestplanung . . . 228
157 MTB-Ergebnis für Nachweistestplanung . . . 229
158 MTB-Daten für Ausfallursachenanalyse . . . 230
159 MTB-Dialoge für Ausfallursachenanalyse . . . 230
160 Ausfallursachenanalyse - Gesamtverteilung . . . 231
161 Ausfallursachenanalyse - einzelne Verteilungen . . . 231
162 Ausfallursachenanalyse - totale Wahrscheinlichkeit . . . 232
163 Ausfallanalyse Regression - Daten . . . 232
164 Ausfallanalyse Regression - MTB Dialog . . . 233
165 Ausfallanalyse Regression - erste Analyse . . . 233
166 Ausfallanalyse Regression - finale Analyse . . . 234
167 Wahrscheinlichkeitsnetze bei Beschleunigung . . . 236
168 MTB-Dialog für Test gleiche Formparameter . . . 237
169 Testergebnis gleiche Formparameter . . . 237
170 Dialog in Minitab für die Beschleunigungsrechnung . . . 238
171 Ergebnis Beschleunigungsrechnung . . . 238
172 Beschleunigung - Wahrscheinlichkeitsnetz . . . 239
173 Relationsdiagramm der Beschleunigung . . . 240
174 Dialog in MTB für die Extrapolation auf Designwertgröße . . . 241
175 Dialog in MTB für die Angabe des Designwertes . . . 242
176 Ergebnis der Beschleunigungsrechnung für den sog. Designwert . 243
177 Ergebnis der Beschleunigungsrechnung für den sog. Designwert mit Konfidenz . . . 244
178 Testplanung unter Beschleunigung . . . 246
179 Testplanung unter Beschleunigung - Teil 2 . . . 246
180 Testplanung unter Beschleunigung - Ergebnis . . . 247
181 Probit - Daten für Fallbeispiel . . . 248
182 MTB Probitanalyse Dialog 1 . . . 249
183 MTB Probitanalyse Dialog 2 . . . 249
184 Probitanalyse Ergebnis Regressionstabelle . . . 250
185 Probitanalyse Ergebnis Regressionstabelle . . . 250
186 Probitanalyse als logistische Regression im Dialog . . . 251

187 Ergebnis der logistischen Regression . . . 251
188 Funktionsfamilie der Johnson-Transformation . . . 252
189 Ausfalldaten für Johnson-Transformation . . . 253
190 Ergebnis für die Johnson-Transformation . . . 254
191 Rechner explizit für die Johnson-Transformation . . . 255
192 Vgl. JT Minitab - JT Rechner . . . 255
193 10%-Perzentile der JT-transformierten Daten . . . 256
194 Überprüfung der Rücktransformationsformel . . . 256
195 10%-Perzentile der JT-transformierten Daten rücktransformiert . 257
196 10%-Perzentile der Weibull-Verteilung . . . 257
197 Ergebnistabellen der Hauptkomponentenanalyse . . . 261
198 Screeplot zur Hauptkomponentenanalyse . . . 262
199 Ausreißer-Diagramm zur Hauptkomponentenanalyse . . . 262
200 Ladungs-Diagramm zur Hauptkomponentenanalyse . . . 263
201 Werte der 3 Hauptkomponenten . . . 264
202 Koeffizienten zu den 3 Hauptkomponenten . . . 264
203 Unabhängigkeit der Hauptkomponenten . . . 265
204 Dendrogramm für einen Cluster . . . 267
205 Dendrogramm für die finale Clusterzuordnung . . . 268
206 Spalte mit den finalen Clusterzuordnungen . . . 269
207 Ergebnis der Diskriminanzanalyse - Güte . . . 271
208 Ergebnis der Diskriminanzanalyse - Güte nach Kreuzvalidierung . 272
209 CART - Entscheidungsbaum Beispiel . . . 273
210 CART - Split-Punktfindung . . . 273
211 CART - Entscheidungsbaum Ergebnis . . . 274
212 Hypothesentest: Richtigkeit und Fehler . . . 277
213 Hypothesentest: Wahl der Fehlerwahrscheinlichkeiten . . . 279
214 Bundesliga-Abschlusstabelle Saison 2009/2010 . . . 283

1 Vorwort

Dieses Skript ist aufgrund meiner ersten QM-Vorlesungen beginnend im Jahre 2004 für den Diplomstudiengang mit Schwerpunkt 'Management Support'entstanden und immer wieder überarbeitet worden. So fand es auch Anwendung in den Vorlesungen 'QM I' (Bachelor Mathematik), 'QM II' (Master Mathematik), 'Technische Analyse und Optimierung (TAO)' (div. Masterstudiengänge für Ingenieure) und 'Statistische Methoden / KI' (Fortsetzung von 'TAO'). Diese Weiterentwicklung liegt daran, dass sich die Software, mit der hier Ergebnisse usw. dargestellt werden, immer wieder geändert hat. Für statistische Analysen wird das Programm 'Minitab' verwendet, für komplexere Analysen des Data Minings 'Knime'. Die letzte Überarbeitung ist zudem eine Erweiterung der Texte. Aufgrund der Coronaviruskrise war es erforderlich, nicht nur ein Skript als Begleitung zur Vorlesung, sondern jetzt noch mehr zum Selbststudium zu präsentieren. Das leistet auch nach der Pandemie nun gute Dienste zum Erlernen des Stoffes.

Das Skript orientiert sich also an einem Semester-Programm und soll die Studierenden beim Erlernern der angewandten Methoden unterstützen.

In diesem Skript werden die Grundlagen der Statistik knapp dargestellt, sondern lediglich auf praxisrelevante Problemstellungen (konkrete Fallstudien und Datenanalysen) angewendet. Es wird also erwartet, dass der Leser des Skriptes statistisches (Vor-) Wissen mitbringt oder selbst erarbeitet. Dieses (Vor-) Wissen wird an der h_da in den Mathematikstudiengängen durch die Vorlesungen Wahrscheinlichkeitstheorie, Statistik I bzw. Stochastik I und Statistik II bzw. Stochastik II gebildet. Hier nun steht der praxisnahe Umgang zu realistischen Fragen im Vordergrund. Insofern bitte ich einen gewissen laxen Umgang mit Begrifflichkeiten manchmal zu entschuldigen. Bei Unklarheiten müssen diese z.B. mit Hilfe klassischer Lehrbücher oder dem entsprechenden Vorlesungsskript vom Leser gelöst werden.

Die Themenvielfalt, die in diesem Programm hier besprochen wird, ist für eine Vorlesung allein relativ hoch. Somit können an der einen oder anderen Stelle Vertiefungen nicht stattfinden; ggf. wird lediglich ein Verfahren erwähnt, kurz skizziert oder ad hoc gerechnet, so dass die Idee zur Lösung und die Ergebnisse der Berechnung aufgezeigt werden. Wer Details benötigt oder sich dafür weitergehend interessiert, der ist aufgefordert, dieses Verfahren selbst zu recherchieren und zu erarbeiten (wobei natürlich gern geholfen wird, bitte also in

solchen Fällen zwanglos Kontakt mit dem Autor aufnehmen).

2 Der Qualitätsbegriff

Für beide Gruppen, die Angewandten Mathematiker und vor allem auch die Ingenieure, ist der Qualitätsbegriff ein zentraler Zugang zu den in der Praxis relevanten Verfahren und Methoden. Aus diesem Grund im Sinne einer Einführung wird erläutert, was man unter 'Qualität' allgemein in Industrie und Wirtschaft versteht.

2.1 Was bedeutet Qualität ?

Mit dem Begriff Qualität wird zunächst lediglich eine Eigenschaft eines Produktes oder einer Dienstleistung (im Folgenden wird der Einfachheit halber nur noch von Produkten gesprochen. Dienstleistungen sind abstrakt ebenfalls Produkte ...) benannt. Eine solche Bezeichnung ist also zunächst nicht wertend und dementsprechend rein formal. Das widerspricht etwas dem allgemeinen Sprachgebrauch. Dort wird Qualität typischerweise gleichgesetzt mit 'hoher' Qualität. Das ist wie gesagt hier so nicht gemeint.

Qualität muss erfasst werden können. Die Eigenschaften, die also als Qualität angesehen werden, müssen wahrgenommen werden. Rein technisch z.B. durch Messungen, zu denen Messmethoden zu definieren sind. Zudem gibt es aber auch eine empfundene Qualität. Gemessene Qualität erscheint also objektiv, Empfindungen sind natürlich subjektiv - und somit sehr viel schwerer zu greifen und zu bewerten.

Ob gemessen oder empfunden, Qualität wird durch Merkmale definiert und wahrgenommen. Es sind natürlich nur solche Merkmale von Interesse, die aus Sicht des Herstellers oder der Kunden auch relevant sind. Auf diese beschränkt man sich also bei der Betrachtung weitestgehend. Diese dann zu bestimmen, und zwar vollständig (um das Produkt entsprechen vollumfänglich in allen Qualitäten zu beschreiben) ist also somit erste Aufgabe im Qualitätsmanagement.

Beispiel: Stahlkugeln

Die Eigenschaften von Stahlkugeln, die Hinweise auf die Brauchbarkeit und deren Einsatzmöglichkeiten geben, also den Produktsinn beschreiben, sind z.B.:

- Durchmesser
- Anzahl Kratzer oder Oberflächenfehler
- Oberflächenfarbe (gibt Hinweise auf die innere Struktur und den Härtungsprozess)
- Rundheit
- Härte
- Eigenfrequenz
- und einige mehr ...

Selbst bei einem so gewöhnlichen erscheinenden Industrieprodukt hat man viele Merkmalen zu definieren, bei unserem Kugelbeispiel liegt die Merkmalsanzahl so in der Größenordnung von ca. 40. Diese Flut von Eigenschaften ist also zu erkennen, zu messen und zu verarbeiten, um somit Aussagen über die aktuelle Qualitätslage zu bekommen. Insofern muss der Analytiker, der sich mit der Produktbeschreibung und Analyse beschäftigt, diese Vielzahl von Merkmalen auch datentechnisch organisieren und mit Analysemethoden effizient arbeiten.

Die Brücke der Merkmale eines Produktes nach reiner Bezeichnung hinein die Mathematik und insb. in die Statistik ist nun schon auch rein verbal durch die Bezeichnung 'Merkmal' gegeben. Merkmale sind in der Statistik ein Fachbegriff, der hier unter den Aspekten des Qualitätsmanagements wiederholt und vollumfänglich dargestellt wird (letzteres wird in den Ausbildungsprogrammen oft so nicht durchgeführt, man diskutiert Verfahren usw. recht speziell, so das das große Ganze insb. in Bezug zu den Messgrößen, also die Merkmale, nicht immer zu sehen ist).

2.2 Wie Qualität gemessen wird

Die Qualität von Produkten muss also erkannt werden; hierzu benötigt man einen möglichst eindeutigen und konstanten Messprozess, der insb. wiederholbar und gegenüber weiteren Einflüssen (Umwelt: Temperatur, Luftdruck, ...,

Organisatorisch: Personalwechsel durch Schichtarbeit, Wechsel des Lieferanten, usw.) robust, also unempfindlich bzw. nicht massgeblich sensitiv sein sollte. Die oben angesprochenen Merkmale als Manifestierung der Eigenschaften, also der Qualität der Produkte, werden somit durch Messungen erfasst. Aus Sicht der Statistik kann man hier schon mal vorab von einer Stichprobe sprechen. Die Bedeutung der Stichprobe als Informationsträger in Bezug auf Merkmale und somit als Abbild der Qualitäten von Produkten ist somit erläutert.

Anm.: der Messprozess zum Erkennen von Merkmalswerten muss selbst eine möglichst hohe Qualität besitzen in Form von Präzision und Genauigkeit. Ob dem bei einem gegebenen Verfahren so ist wird in einem Kapitel der Vorlesung im Detail besprochen (Messsystemanalysen hier insb. die sogenannte Gage R& R). Grundlage hierfür sind natürlich ebenfalls Stichproben.

Mathematisch abstrakt gesehen können die Informationen, die die Merkmale nun besitzen, unterschiedlich ausgeprägt sein. Man unterscheidet in der Statistik dazu folgende Merkmalsausprägungstypen:

- nominal: rein namensgebend
- ordinal: bewertend, jedoch subjektiv
- diskret: zählend
- stetig: messend

Am Beispiel der Stahlkugel wären dann z.B.:

- Durchmesser: stetig
- Anzahl Kratzer auf der Oberfläche: diskret
- Oberflächenfarbe: nominal
- Rundheit: stetig
- Härte: ordinal (Messergebnis wäre dann z.B. eine Härteklasse)
- Eigenfrequenz: stetig

Diese Zuordnung ist nicht eindeutig, also selbstverständlich. Die Oberflächenfarbe könnte man ebenfalls stetig messen (Farbspektrum), die Härte ggf. stetig (Kerbschlagzähigkeit). Das Verfahren bestimmt schlussendlich, in welcher Merkmalsausprägungskategorie das konkret gemessene Merkmal zu finden ist.

Die folgende Grafik teilt diese Ausprägungsarten zudem noch in die übergeordneten Ebenen qualitativ und quantitativ ein. Zudem werden später verwendete Fachbegriffe (z.B. Minitabnotation) hier den Ausprägungstypen vorab zugeordnet:

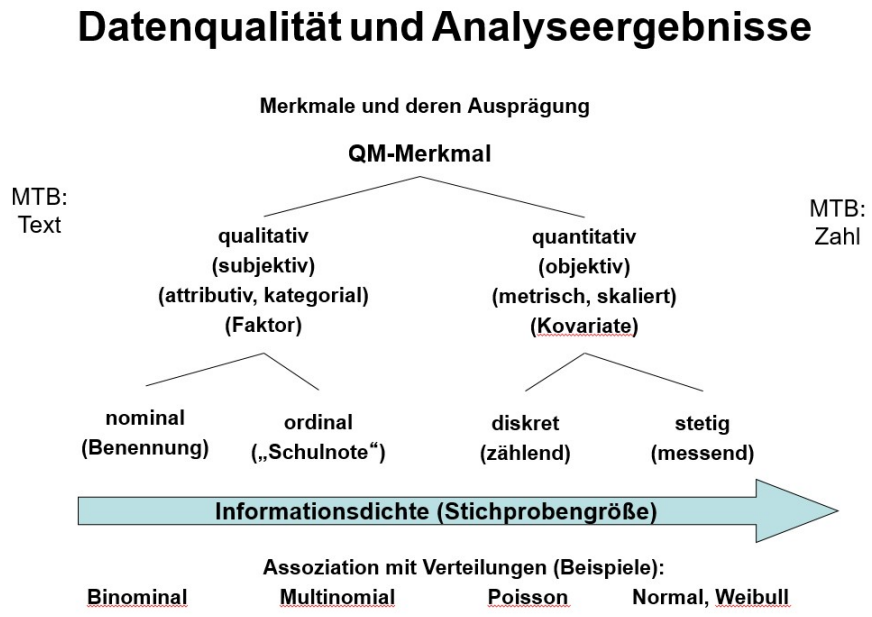


Figure 1: Merkmalsbaum

Jedes Merkmal, egal wie es definiert ist, kann und muss eine der angegebenen Ausprägungstypen zugeordnet werden. Es gibt also keinen unbekannten fünften Ausprägungstyp, die Erfassung von Informationen ist somit vollständig diesen Klassen zugeordnet.

Nun sind statistische Verfahren lediglich für z.T. sehr spezielle Kombinationen von Ausprägungstypen definiert bzw. konstruiert. Insofern hier eine Warnung: der Analyst muss den grundsätzlichen Umgang mit diesen Ausprägungstypen sicher kennen, also gut trainieren, um auch korrekte und sinnvolle Auswertungen

durchzuführen. Andernfalls wären Auswertungen zumindest fachlich falsch - was dann sicher nicht nur 'unangenehm' wäre. Es ist halt nicht jedes Verfahren sinnvoll, auch wenn die Daten es rein formal zulassen würden. Minitab und auch andere statistische Werkzeuge helfen einem bei dieser Zuordnung nur sehr wenig, das liegt in der Verantwortung des Analysten.

Dass bestimmte Begriffe, die hier zu sehen sind, direkt in Minitab Verwendung finden, das sei hier beispielhaft dargestellt; der Leser sollte sich hier selbst kontrollieren, ob er diese Begrifflichkeiten sicher beherrscht:

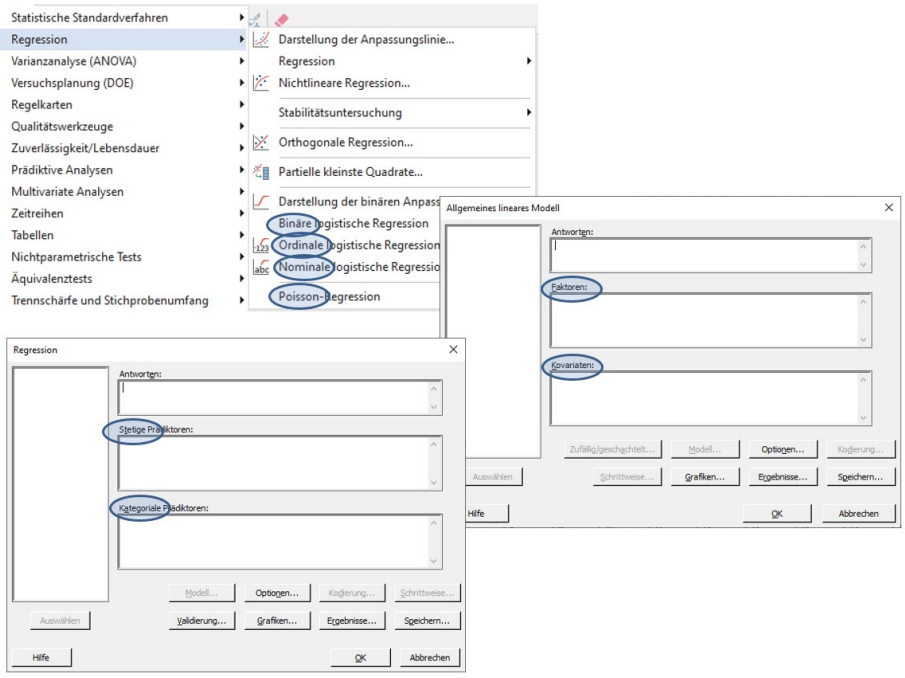


Figure 2:

2.3 Die Informationsgewinnung

In der Praxis werden die Merkmale anhand von Stichproben gemessen. Die Stichprobenproblematik, insb. im Zusammenhang mit der statistischen Aussagekraft eines Hypothesentests im Sinne eines statistischen Beweises. basierend auf Signifikanz, Trennschärfe, usw., wird im weiteren Verlauf des Skriptes noch erläutert

werden.

Der Informationsgehalt steigt in der Grafik oben, siehe Grafik 1, von links nach rechts an. Das bedeutet in Bezug auf die Statistik konkret, dass man bei stetigen Merkmalen z.T. erheblich weniger Stichproben mit zudem deutlich geringeren Umfängen benötigt, um fundierte Aussagen über die Eigenschaften und deren Qualitäten treffen (bewerten) zu können. Insofern zahlen sich technisch aufwendige, ggf. sehr teure Messmethoden u.U. bei einer permanenten Qualitätsanalyse (Prozessanalyse) aus.

Besonders schwierig ist die subjektive Messung von Qualität. Das ist z.B. der Fall bei Umfragen mittels Fragebogen oder bei Messungen, die auf Einschätzungen von Prüfern (z.B. Weinprobe oder auch Prüfungen an einer Hochschule ...) beruhen. Bei derartigen Messungen ist z.B. bei wiederholtem Messen mit Abweichungen der Messergebnisse zueinander zu rechnen. Die Ergebnisse sind also nur bedingt wiederholbar. Im Extremfall ist der Messprozess zur Gewinnung der Daten ggf. nicht mehr brauchbar. Die Problematik der Subjektivität von Messungen wird hier im Skript nicht weiter behandelt. D.h. hier werden im Folgenden praktisch nur technische Messungen, die vordergründig objektive Messergebnisse liefern, betrachtet. Die dennoch vorhandenen subjektive Einflüsse auf diese technischen Messverfahren (z.B. von Prüfern) werden in der sog. Messsystemanalyse herausgefunden.

Nominale oder ordinale Merkmale liegen in der Praxis typischerweise kodiert vor. Das insb. aus dem Grund, damit man mit diesen dann 'rechnen' kann. So werden binominale Größen oft in den Zahlen '0' und '1' kodiert. Der Informationsverlust ist dann offensichtlich: was bedeutet '0' oder '1' genau ? Beispiel Prüfung: bestanden - oder durchgefallen ?

Das fast fatale an dieser Kodierung ist, dass man mit dessen Hilfe, also der Zahlenwerte, nun tatsächlich 'rechnen' kann (was mit den üblichen reinen Namensbezeichnungen wie z.B. 'OK' oder 'Nicht OK' sicher nicht funktioniert). So zum Beispiel bei der Annahmeprüfung, in der mit '0' und '1' OK- und Nicht OK-Bauteile gekennzeichnet werden. Der Ungeübte oder auch Nichtwissende ist ggf. aufgrund der Kodierung dazu verleitet, z.B. so etwas wie einen Mittelwert auszurechnen. Rein technisch funktioniert das, jedoch ist dieses Vorgehen natürlich zu hinterfragen, sobald z.B. binominale Daten nicht mehr in einem einfachen Sinn interpretiert werden dürfen. Ich empfehle, z.B. binominale Merkmale

auch als 'Text' zu definieren, damit man halt nicht solche eigentlich unzulässigen Berechnungen durchführt.

Die gleiche Problematik gilt noch viel stärker für ordinale Merkmale. Diese nur subjektiv zu messenden Eigenschaften besitzen u.U. zwei Effekte, die ein 'Rechnen' mit ihnen eigentlich unmöglich machen:

- Skalenverschiebung
- Nullpunktverschiebung

Als anschauliches Beispiel für eine typische ordinale Merkmalsausprägung, die diese beiden Effekte nachweislich aufzeigt, und die jeder Studierende auch kennt, sei ein klassisches Schulnotensystem genannt. Eine 'Note' wird halt eine Zahl angesehen ... ist es aber nicht.

Nun hat die Pisa-Studie ergeben (d.h. es wurde statistisch bewiesen), dass Schüler in unterschiedlichen Bundesländern in Deutschland unterschiedliches Leistungsniveau besitzen (Nullpunktverschiebung). Es ist also fraglich, ob denn eine gleiche Note für Schüler aus unterschiedlichen Bundesländern auch den gleichen Lernerfolg aufzeigen.

Zudem ist es z.B. bei Noten an der Hochschule üblicherweise so, dass ein Leistungsnachweis für eine Lehrveranstaltung ab einer Prozentzahl von 50 % als 'bestanden' gilt. 'Nicht bestanden' ist die Note '5', während für 'bestanden' eine der Noten '1' bis '4' (mit Zwischennoten: 1,3, 1,7, usw., insg. 10 Stück !) gilt. Das bedeutet also, dass die Note '5' 50 % des gesamten Notenspektrums abdeckt, und die anderen 50 % sich 10 (Teil-) Noten teilen; hier liegt eine sog. Skalenverschiebung vor.

Problematisch wird es, wenn dann mit Hilfe solcher Noten 'gerechnet' wird; also z.B. eine Durchschnittsnote angegeben wird (und aufgrund dieses Durchschnittes dann weitere Entscheidungen getroffen werden ...). Dieses Vorgehen ist ja in Deutschland gebräuchlich. Alternativen gibt es, so besteht z.B. das amerikanische Notensystem lediglich aus den Buchstaben 'A' bis 'F'. Mit diesen Bezeichnungen ist die Gefahr eines 'Berechnens' nicht mehr gegeben, die Bewertungen sind anders vorzunehmen.

Man erkennt, dass die Problematik sicher eine eigene Statistk-Veranstaltung braucht. Und es gibt auch Spezialisten, die sich mit diesen Fragestellungen ganz spezialisiert auseinandersetzen müssen, z.B. Marktforscher oder Psychologen. Jedenfalls wird dieses Thema der ordinalen Merkmale und der zugehörigen Statistik hier nicht weiter diskutiert.

2.4 Die Dimensionen der Qualität

Die Qualität von Produkten wird zunächst nicht konkret an tatsächlich gemessenen Merkmalen definiert. Zunächst sind, z.B. bei der Definition von Qualitäten bei neuen Produkten 'höhere' Eigenschaften im Sinn der Entwickler und Ingenieure. Da ist die Rede von 'Haltbarkeit' oder 'Zuverlässigkeit' im Allgemeinen. So bleibt die Frage, wie sind diese Eigenschaften dann eigentlich konkret zu messen - und so zu bewerten sind.

Es handelt sich z.B. bei 'Haltbarkeit' also zunächst um ein Meta-Merkmal. Diese Meta-Merkmale sind als die sog. Dimensionen von Qualität bekannt. Hier sollen folgende acht wichtigste Dimensionen erwähnt werden, die typischerweise Qualität von Produkten zugrunde liegen:

- Haltbarkeit (Durability)
- Gebrauchsfähigkeit (Serviceability)
- Ästhetik (Aesthetics)
- Zuverlässigkeit (Reliability)
- Standardisierung (Conformance)
- Eigenschaften (Features)
- Erfüllungsgrad (Performance)
- Reputation (Perceived Quality)

Diese Liste dient somit zunächst nur als Ideengeber, was grundsätzlich einem Produkt als relevante Eigenschaften angesehen werden können. Zu diesen Meta-Merkmalen findet man dann in einem weiteren kreativen Prozess auch konkret

messbare Merkmale, die dann auch durch Statistik bewertet werden können. Hinter 'Zuverlässigkeit' kann dann z.B. die Härte der Stahlkugel gesehen werden, oder auch die Zeit, die die Kugel in einem (Langzeit-) Belastungstest den Kräften darin stand hält. Es sind also einige Merkmale den jeweiligen Meta-Merkmalen zuzuordnen.

2.5 Kosten, Zeit und Qualität

Im Rahmen der Vorlesung wird an einigen Stellen deutlich, dass die Verfahren des Qualitätsmanagements insb. auch Rücksicht auf die Ökonomie nehmen müssen. Kosten spielen in unserer hoch effizienten und industrialisierten Welt, in der Fertigungsprozesse mit z.T. sehr wertvollen Materialien und Ressourcen arbeiten, eine große Rolle. Indirekt aufwandsbezogen ist zudem der Faktor Zeit, der ebenfalls kritisch ist in einer innovationsgetriebenen Welt. An verschiedenen Stellen wird im Skript auf diese Problematik dann im Einzelnen hingewiesen bzw. es werden Verfahren auch unter diesen ökonomischen Gesichtspunkten entsprechend diskutiert. Hierzu zählt insb. die statistische Versuchsplanung, die durch ein sehr gezieltes Vorgehen Informationen eines komplexen Systems sammelt und einer statistischen Analyse zuführt - bis hin zu einer Optimierung.

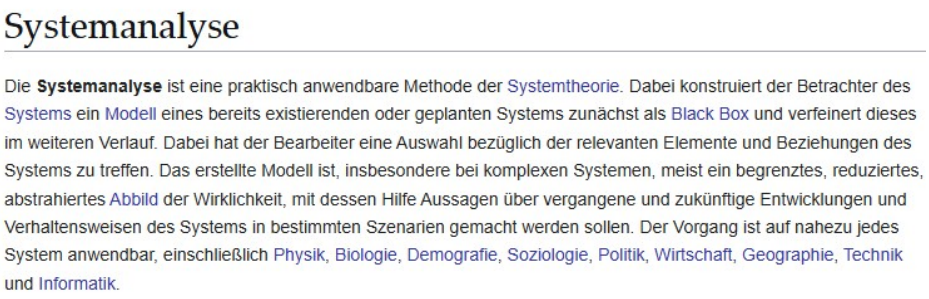
3 Die Systemanalyse

Die Merkmalsausprägungen sind wichtig, um den Informationsgehalt von etwas, was bemessen wird, zu kategorisieren. Wenn nun viele solcher Merkmale, also Variablen, statsistisch auch Variate (Variablen, die variieren ...) zur Beschreibung eines Systems notwendig sind, dann benötigt man diese Kategorien, um das Systemverhalten oder den Systemzustand zu beschreiben. Die in der Praxis zu messenden oder zu erlebenden Größen müssen dann z.B. in Software korrekt abgebildet werden.

Systemisch sind Vorgänge, wenn einge Merkmale von anderen abhängen. Generell ist hierbei diese Aufteilung nach Ursache und Wirkung nicht strikt. In den Ansätzen der Technischen Analyse jedoch sind die Ursachen und die Wirkungen klar getrennt. Es sei hier nur angemerkt, dass das so natürlich nicht sein muss, diese Vereinfachung jedoch in vielen Fällen ausreicht, um Vorgänge zu erfassen

(Analyse) und zu optimieren (basierend auf einem aus der Analyse gewonnen Modell).

Der Begriff 'Systemanalyse' klingt zunächst recht kompliziert oder theoretisch. Diesen Eindruck widerlege ich einfach, indem ich aus Wikipedia zitiere (wikipedia.de, Suchbegriff: Systemanalyse, Abfragezeitpunkt: 16.12.2024, 14:31h):



Systemanalyse

Die **Systemanalyse** ist eine praktisch anwendbare Methode der Systemtheorie. Dabei konstruiert der Betrachter des Systems ein Modell eines bereits existierenden oder geplanten Systems zunächst als Black Box und verfeinert dieses im weiteren Verlauf. Dabei hat der Bearbeiter eine Auswahl bezüglich der relevanten Elemente und Beziehungen des Systems zu treffen. Das erstellte Modell ist, insbesondere bei komplexen Systemen, meist ein begrenztes, reduziertes, abstrahiertes Abbild der Wirklichkeit, mit dessen Hilfe Aussagen über vergangene und zukünftige Entwicklungen und Verhaltensweisen des Systems in bestimmten Szenarien gemacht werden sollen. Der Vorgang ist auf nahezu jedes System anwendbar, einschließlich Physik, Biologie, Demografie, Soziologie, Politik, Wirtschaft, Geographie, Technik und Informatik.

Figure 3: Systemanalyse lt. Wikipedia

Die wichtigsten Aspekte in diesem Ausschnitt sind

- die Systemanalyse ist eine 'praktisch anwendbare Methode'
- die 'Black Box'-Analyse ist der Start der Betrachtungen
- es ist eine Auswahl der Beziehungen der Elemente zueinander zu wählen (!)
- es wird ein Modell erstellt, dass als Abbild der Wirklichkeit dient - und somit lediglich begrenzt, reduziert, vereinfacht, abstrahiert ist

Ich möchte die Black Box-Analyse nun sehr zielgerichtet für die finale Zielsetzung der Vorlesung aufstellen und zwar so, dass somit der Zusammenhang aus den einzelnen Schritten gut ersichtlich wird. Hierzu betrachte man das folgende Bild, das einen abstrakten Prozess mit den (vereinfachten) Einflüssen und Zielen darstellt:

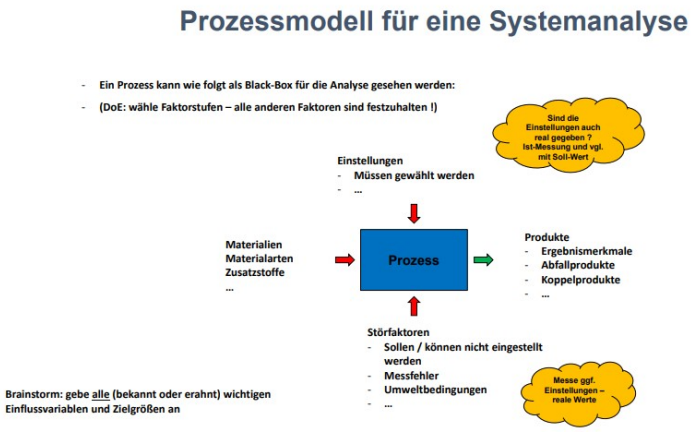


Figure 4: Black Box-Analyse für eine Systemanalyse

Das Zusammentragen der Merkmale, die auf einen Prozess einwirken, geschieht in der Praxis z.B. mit Hilfe einer Brainstorm-Sitzung. In dieser Sitzung sollten alle Prozess- und Produktverantwortlichen anwesend sein. Eine Anzahl von 40 oder 50 Variablen ist dann nicht ungewöhnlich, was aufzeigt, dass man es in den meisten Fällen solcher Untersuchungen mit komplexen Prozessen, d.h. komplexen Zusammenhängen zu tun hat.

Sobald das Erfassen möglicher Einflüsse beendet ist, werden die Einflussvariablen priorisiert. Bei diesem Schritt ist Erfahrung, aber auch pyhsikalisch-technisches Verständnis erforderlich. Erste Versuche sollten dann anhand der Variationen der am meisten Priorisierten Variablen durchgeführt werden.

Ein wichtiger Aspekt ist dabei ebenfalls sehr wichtig und darf nicht ausser Acht gelassen werden: die Variablen, die nicht variiert werden, sind (möglichst) konstant zu halten. Sofern das nicht möglich ist (z.B. bei Rauschgrößen) müssen diese dennoch gemessen und der Analyse zugeführt werden. Die Logik dahinter ist recht einfach und anhand der Systemanalyse oben abzulesen: ist eine Variable in der Balck Box-Analyse ein Kandidat für einen wichtigen Einfluss, dann ist diese ggf. auch wichtig (signifikant) - und darf somit nicht ignoriert werden. Einstellungen, also Messwerte dieser Variablen, können also Zusammenhänge ggf. gut erklären - auch wenn sie in den ersten Schritten keine hohe Priorität zugewiesen bekamen. Die Volständigkeit der Daten in Bezug auf die in der Systemanalyse erfassten Variablen ist also ein sehr wichtiger Punkt !

Ferner sei noch angemerkt, dass die Produkte des Prozesses typischerweise gewisse Anforderungen erfüllen sollen, z.B. durch vom Kunden vorgegebene Spezifikationen. Ein Produkt wird also zunächst in der Analyse durch seine formalen Ziele (z.B. Kerbschlagzähigkeit, Elektrische Leitfähigkeit) beschrieben. Diese müssen dann gemessen werden (hier z.B. durch einen Kerbschlagbiegeversuch nach ISO 148-1, in $[J/cm^2]$, oder die Leitfähigkeit in $[\mu S/cm]$).

Mathematisch werden dann die Input-Variablen der Systemanalyse aufgeteilt in Faktoren und Kovariaten. Die Faktoren stellen die priorisierten Größen dar, für die Versuche systematisch gezielt durchgeführt werden sollen. Die Kovariaten werden zusätzlich notiert, um ggf. bei nur relativ schwachem Erkennen von Zusammenhängen diese in zusätzliche Analysen einfliessen lassen zu können.

$$y = f(\underbrace{x_1, \ldots, x_m}_{Faktoren}, \underbrace{x_{m+1}, \ldots, x_n}_{Kovariate}) \tag{1}$$

Hier im abstrakten Beispiel liegen also m Faktoren mit hoher Priorität vor, zu denen die Versuche systematisch geplant werden, sowie $n - m$ Kovariate, die während die Versuche zu den m Faktoren durchgeführt werden, gemessen oder konstant gehalten werden.

4 Der Hypothesentest

Das Ergebnis von Messungen muss dabei natürlich nicht immer aus Zusammenhangsanalysen gewonnen werden. Die Messungen einzelner Zielgrößen zu Produkten ist zunächst recht einfach. Da es sich lediglich um einzeln zu betrachtende Größen handelt, spricht man von 'univariater Analyse'. In diesen Fällen sind einfache Vergleiche, z.B. mit Vorgabewerten oder von Werten zueinander, sinnvoll. Für diese relativ einfachen Fragestellungen sind die einfachen Hypothesentests gut geeignet.

Fragen an Daten, also die Herleitung von Informationen, soll möglichst präzise sein. Das ist erreichbar, wenn man z.B. bei der Frage, wie alt die Studierenden an der Hochschule Darmstadt sind, und ob ggf. im Sinne eines Benchmark-Vergleiches diese Studierenden 'im Mittel' jünger sind als 22 Jahre, alle Studieren-

den nach ihrem Alter befragt. Es ist klar, dass es dann kein 'Hinterfragen' der so gewonnen Zahl gibt. Es gibt also in dem Fall, wenn man eine Grundgesamtheit vollständig befragen kann, keine Unsicherheit über die Messung und die daraus gewonnene Information.

Ist man nicht in der Lage, eine Grundgesamtheit wie hier die 'Studierenden' der h_da vollständig zu befragen, dann wird man also lediglich einen Teil dieser Gruppe erfassen (Stichprobe). Das Ergebnis wird dann auf die Grundgesamtheit übertragen.

Jedoch ist klar, dass man mit der entsprechenden Behauptung, z.B. der Mittelwert (gemeint ist hier das arithmetische Mittel ...) der Studierenden 'ist' 22 Jahre, falsch liegen kann.

Bei der Arbeit mit Stichproben muss man dieser Unsicherheit also Ausdruck verleihen. Mathematisch erwartet man also eine Unsicherheitsaussage bzgl. der z.B. wie hier im Beispiel zu messenden Eigenschaft (Mittelwert des Alters).

Hierzu nun ein Gedankenexperiment:

Situation: Man befragt unterschiedliche Anzahlen n von Studierenden nach deren Alter (gemessen bis auf den Tag genau!) und berechnet den Mittelwert aus den Altersangaben. Für z.B. n $= 10$ ergibt sich dann folgendes Bild (das Alter der einzelnen befragten Studierenden ist die y-Achse):

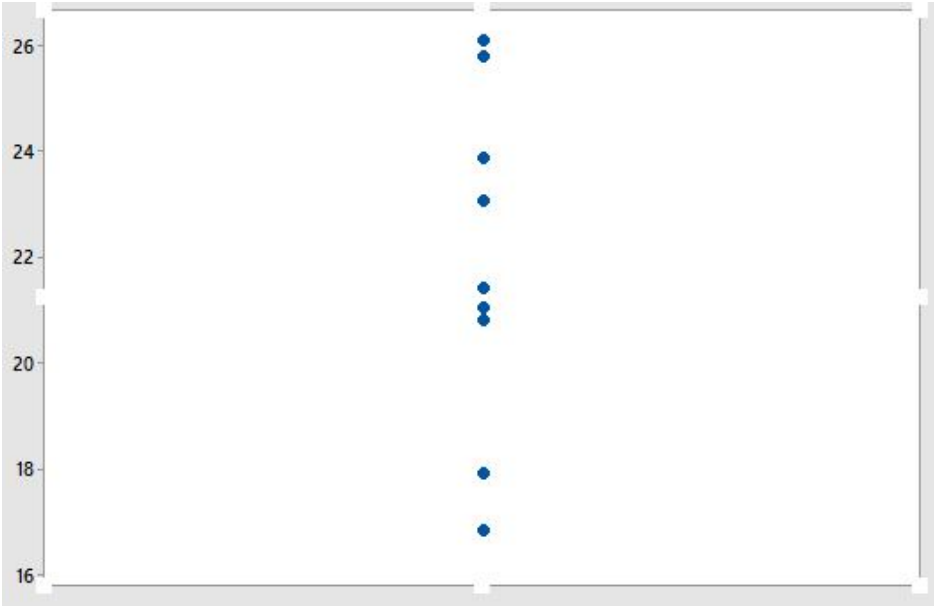


Figure 5: Stichprobengröß 10 für die Schätzung des Mittelwertes

Wiederholt man diese Befragung nun (fiktiv ...) 100 Mal, so sind die Messungen vielleicht wie folgt:

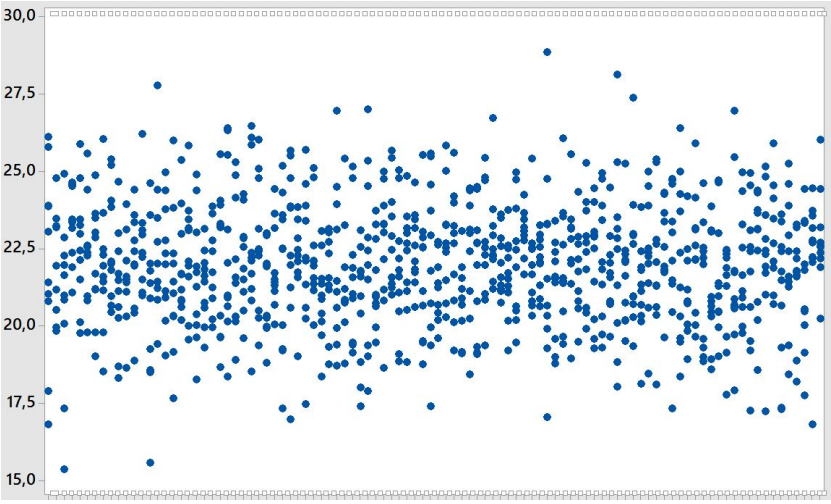


Figure 6: 100 Stichproben a 10 für die Schätzung des Mittelwertes

Berechnet man nun die Mittelwerte der Stichproben, dann sind diese wie folgt:

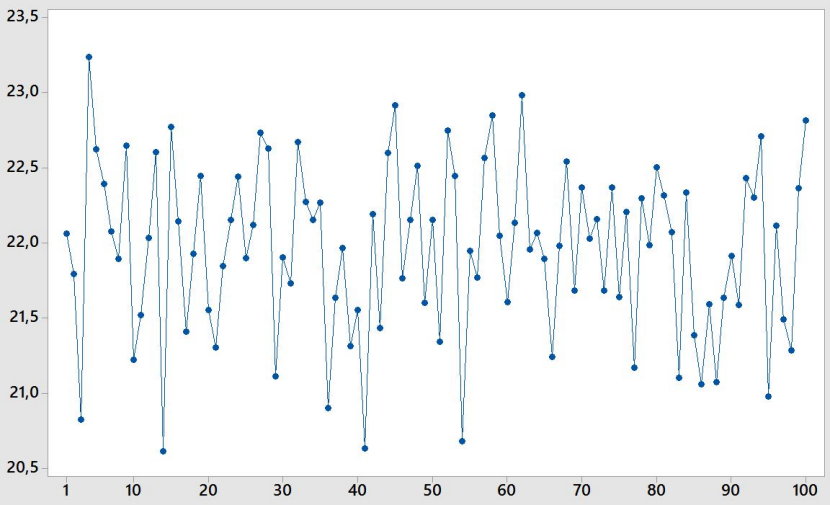


Figure 7: Mittelwert von 100 Stichproben a 10

Oder als Histogramm dargestellt:

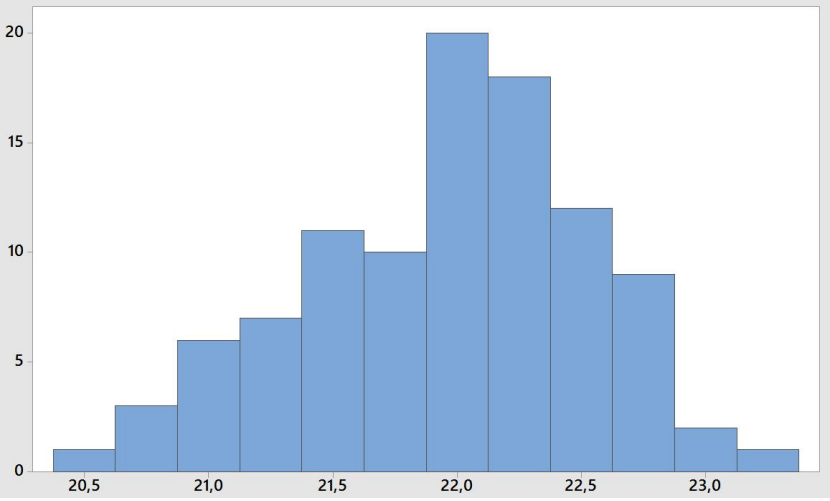


Figure 8: Histogramm der Stichproben-Mittelwerte

Die Mittelwerte schwanken - wie erwartet. Es lässt sich in etwa ein Bereich der

Schwankungen erahnen. Bei einer Aussage über den aus einer Stichprobe heraus berechneten Mittelwert ist demnach die Kenntnis über die Schwankungsbreite (die Unsicherheit) sinnvollerweise anzugeben.

Umgekehrt ist so ein aus einer Stichprobe heraus berechneter Mittelwert nur eben bis auf ein gewisses Intervall genau.

Dieses Intervall heißt Konfidenzintervall .

4.1 Konfidenzintervalle

Ist nun (fiktiv ...) das Alter der Studierenden einer Normalverteilung folgend, dann kann man das Konfidenzintervall angeben. Jedoch bleibt ein 'Restrisiko', dass man gerade eine so ungünstige Stichprobe von Studierenden erwischt hat, dass sogar das Unsicherheitsintervall nicht stimmt.

Die Unsicherheit wird nun mit Hilfe einer Wahrscheinlichkeit angegeben. Als Standard-Maßwird typischerweise 95% genommen. D.h. die Wahrscheinlichkeit, dass selbst in den Konfidenzintervall der Mittelwert der Grundgesamtheit nicht enthalten ist, ist 5%.

Es ist zudem festzustellen, dass die Mittelwerte für Stichproben aus einer normalverteilten Grundgesamtheit wiederum normalverteilt sind (bei 'bekannter' Standardabweichung).

Für die Normalverteilung ist nun folgende Eigenschaft bekannt: 95% der Ereignisse sind im Intervall

$$(\bar{x} - 1,96 * s, \bar{x} + 1,96 * s) \tag{2}$$

zu finden.

Für die sog. Standard-Normalverteilung ($\mu = 0$, $\sigma = 1$) bedeutet das grafisch:

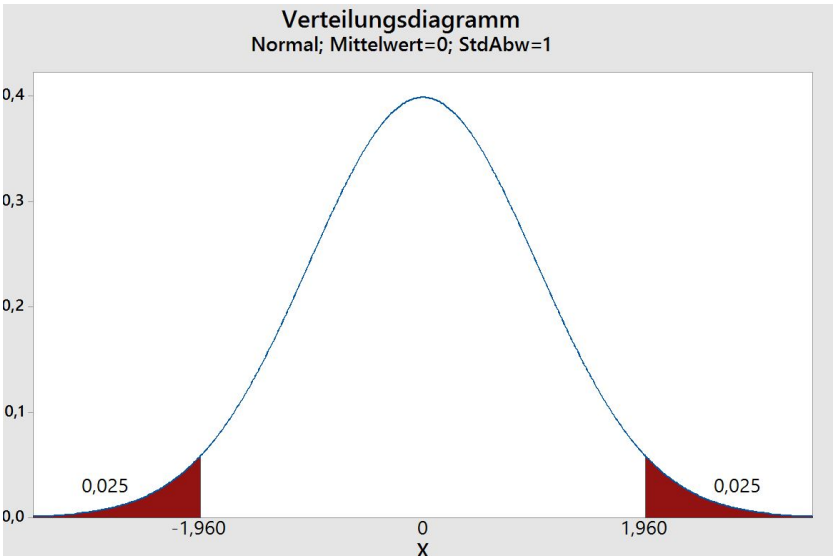


Figure 9: Standard-Normalverteilung mit 95%-Überdeckung

Die Unsicherheit nimmt mit der Größe der Stichprobe ab. Es kann festgestellt werden, dass die Unsicherheit, ausgedrückt durch die Konfidenz für den Mittelwert, mit $1/\sqrt{n}$ asymptotisch abnimmt.

Daraus ergibt sich für die Formel für z.B. die Schätzung des Mittelwertes des Alters der Studierenden (bei 'bekannter', also nicht statistisch unsicherer und geschätzter Standardabweichung) mit Konfidenz 95%:

$$\mu \in (\bar{x} - 1,96 * s/\sqrt{n}, \bar{x} + 1,96 * s/\sqrt{n}) \tag{3}$$

wobei μ hier nun der Mittelwert der Grundgesamtheit, s die Standardabweichung einer Stichprobe und $\bar{x}$ der Mittelwert des Alters der Studierenden der Stichprobe (also die Schätzung des mittleren Alters) ist.

Grafisch lässt sich das wie folgt darstellen:

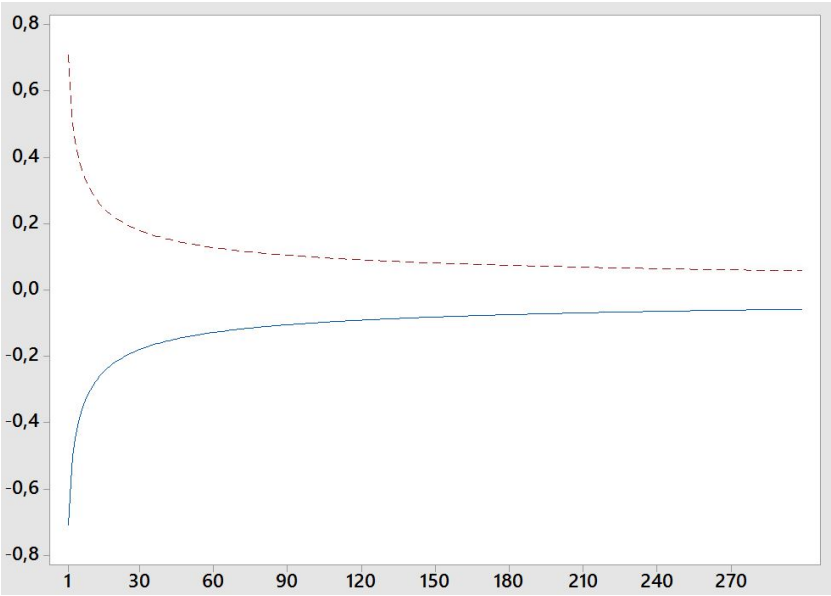


Figure 10: Konfindenz-Asymptotik: Abhängigkeit von der Stichprobengröße

Wird die Standardabweichung zudem aus den Stichprobendaten heraus geschätzt (Normalfall), dann bekommt man durch diese Schätzung eine weitere Unsicherheit in der Aussage bzgl. des Mittelwertes. In diesem Fall muss nicht die Normalverteilung, sondern die sog. t-Verteilung herangezogen werden, um die Intervallbreite der Konfidenz zu berechnen. Ganz konkret wird hierzu ein (von der Stichprobengröße abhängiger !), etwas größerer Vorfaktor als 1,96 eben aus der t-Verteilung heraus berechnet.

Für die Stichprobengröße 10 ergibt sich die 95%-Überdeckung der korrespondierenden t_9-Verteilung (die Stichprobengröße -1 heiß hier der Freiheitsgrad der t-Verteilung):

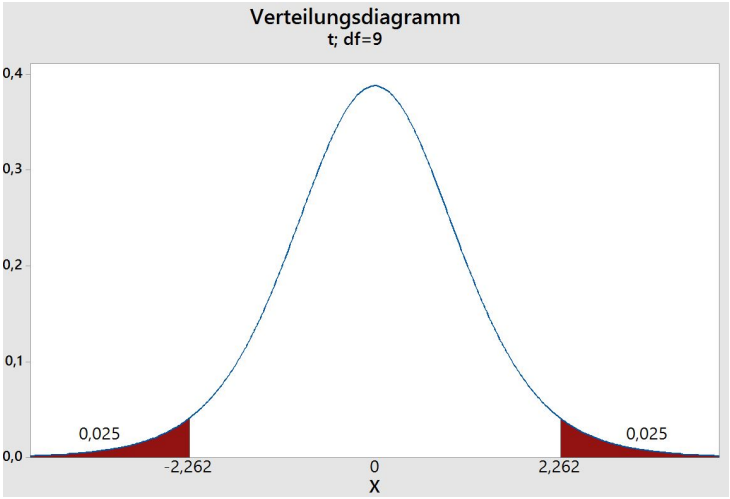


Figure 11: t-Verteilung mit 9 Freiheitsgraden und 95%-Üerdeckung

D.h. statt des Faktors 1,96 ergibt sich das Konfidenzintervall nun zu:

$$(\bar{x} - 2,262 * s, \bar{x} + 2,262 * s) \tag{4}$$

Die Asymptotik lässt sich wie folgt darstellen:

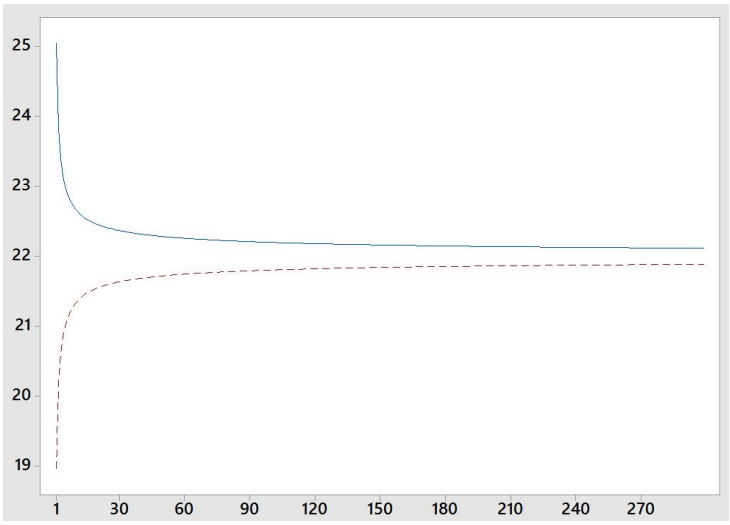


Figure 12: 95%-Konfidenzintervall der t-Verteilung

Die Aussage dieser Grafik ist: mit 95% Sicherheit liegt ein gemessener Mittelwert aus einer Stichprobe für die Altersschätzung der Studierenden zwischen den Linien.

Betrachtet man nun noch die Konfidenzintervalle der Stichproben aus dem (fiktiven) Experiment, dann würde in etwa ein solches Bild entstehen:

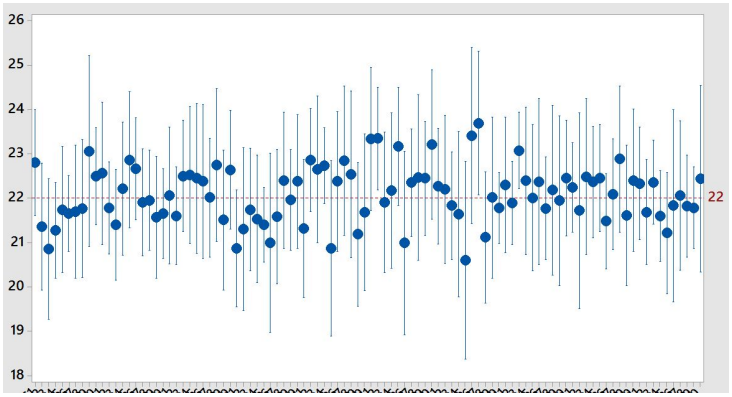


Figure 13: Konfidenzintervalle zu 100 Stichproben

Wie zu sehen ist, sind alle Konfidenzintervalle zu den Schätzungen so, dass der Wert 22 darin enthalten ist.

In der Praxis würde man eine Stichprobe ziehen, den daraus ermittelten Mittelwert dann als Mittelwert der Grundgesamtheit annehmen, und die Unsicherheit abhängig vom Mittelwert-Wert bestimmen.

4.2 Statistischer Test

Ein Test soll nun bezogen auf das Alter entwickelt werden. Ausgangspunkt kann die folgende Frage sein:

Sind die Studierenden an der h_da 'im Schnitt' 22 Jahre alt ?

Diese Frage ist so, wie sie gestellt ist, nicht sinnvoll. Zum einen wird sich sicher ergeben, dass das durchschnittliche Alter nicht genau 22 ist. Zum Anderen soll

das Alter der Studierenden mit einer Stichprobe ermittelt werden. Insofern ist aufgrund dessen schon klar, dass man die 22 Jahre in Bezug auf die Konfidenz relativieren muss.

Aus Sicht der Statistik unter Berücksichtigung des oben bereits beschriebenen würde man als Zwischenschritt zunächst fragen:

Ist das Alter, das sich aus einer Stichprobe ergibt, deutlich zu 22 unterschiedlich ?
Oder ist der aus der Stichprobe ermittelte Mittelwert nur wenig von 22 entfernt ?

Diese beiden Fragestellungen können nun anhand des Konfidenzintervalles konkretisiert werden:

Ist 22 im Konfidenzintervall des Mittelwertes der Stichprobe enthalten - oder nicht ?

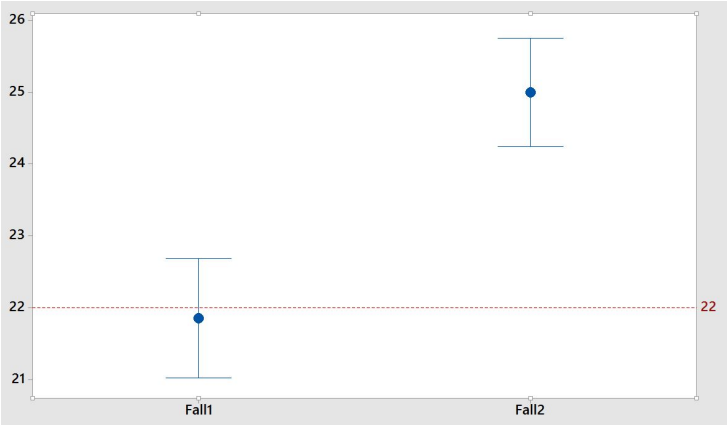


Figure 14: Vergleich Wert im Konfidenzintervall

Im ersten Fall ist zwar der gemessene Mittelwert unterschiedlich zu 22, die Differenz ist aber sehr klein. Anders im zweiten Fall.

Hier spricht man dann auch von einem signifikanten Unterschied, basierend auf der 95%-Konfidenz.

Diese Unterschiedsfeststellung, basierend auf dem Konfidenzintervall, ist nun die

grundlegende Idee für einen Hypothesentest .

Die anfängliche Frage nach dem signifikanten Unterschied des Mittelwertes wird nun in eine vordergründig wieder exakt wirkende Frage umgewandelt, wobei die beiden Antwortmöglichkeiten als Hypothese und Gegenhypothese definiert sind.

D.h. konkret, die Frage 'Ist der Mittelwert der Studierenden an der h_da gleich 22 ?' lautet nun in Form dieser beiden verbalen Aussagen:

$$\begin{array}{rcl} Nullhypothese\ H_0 : \mu & = & 22 \\ Gegenhypothese\ H_1 : \mu & \neq & 22 \end{array} \tag{5}$$

Durch die Berechnung von Mittelwert und zugehörigem Konfidenzintervall lässt sich dann entscheiden, welche Aussage zutrifft (konkret: zutreffend bis auf die Irrtumswahrscheinlichkeit ... diese stellt sozusagen das 'Restrisiko' für eine Fehleinschätzung dar).

Beispiel:

Es sei folgende Situation für eine Stichprobe bzgl. des Alters der Studierenden gegeben:

$$n = 9, \quad \bar{x} = 23, \quad s = 2 \tag{6}$$

Die Grundgesamtheit sieht dann (anhand der ermittelten Parameter) geschätzt wie folgt aus:

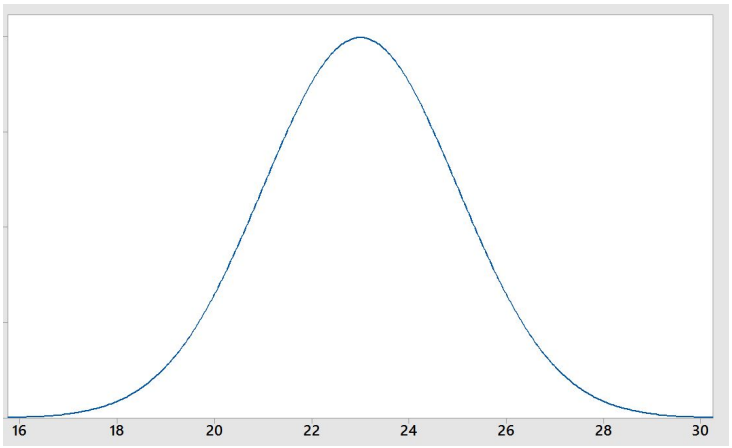


Figure 15: Geschätze Grundgesamtheit

Das Konfidenzintervall des Mittelwertes ist nun unter Annahme der exakten Gleichverteilung (Standardabweichung ist bekannt):

$$(23 - 1,96 * 2/\sqrt{9}, 23 + 1,96 * 2/\sqrt{9}) \approx [21,693...; 24,307...] \quad (7)$$

Die Zahl 22 liegt im Intervall. Also ist H_0 anzunehmen (dieses wird kritischer formuliert in der Form: es liegen nicht genügend Argumente vor, um die Null-Hypothese abzulehnen).

Kritisch anzumerken ist, dass die Entscheidung im Hypothesentest von der Stichprobengröße abhängt.

Hätte man fiktiv statt der 9 25 Studierende gefragt, so wäre bei exakt gleichem Ergebnis für $\bar{x}$ und s folgendes Konfidenzintervall berechent worden:

$$(23 - 1,96 * 2/\sqrt{25}, 23 + 1,96 * 2/\sqrt{25}) \approx [22,216...; 24,431...] \quad (8)$$

Also würde man sich nun für die H_1-Hypothese entscheiden !

4.2.1 Hypothesentest und Stichprobengröß

Das Ergebnis der Berechnungen im Abschnitt oben ist also von der Stichprobengröße abhängig. Das ist ein prinzipielles Problem, denn es stellt sich die Frage, wie groß nun eine sinnvoll gewählte Sztichprobengröße ist.

Im Umkehrschluss muss also die Stichprobengröße vorher (!) ermittelt werden, damit der statistische Test sinnvoll interpretiert werden kann. Die Formel für das Konfidenzintervall wird nun so aufgelöst dass als Bestimmungsgröße die Stichprobengröße abhängig dann vom Konfidenzintervall (im Sinne der Differenz zu den Intervallgrenzen, diese sei mit $Delta$ bezeichnet) zu lesen ist:

$$n = \left(\frac{1,96 * s}{\Delta} \right)^2 \tag{9}$$

Die Größe Δ ist also nun zu wählen. Dadurch wird die gewünschte Präzision des Tests festgelegt in dem Sinn, dass nun ein Unterschied von Δ zum Vorgabewert (hier im Beispiel 22) mit 95% Sicherheit auch festgestellt wird.

Legt man für das Beispiel hier nun die Größe Δ auf ein halbes Jahr fest (was Sinn macht bzgl. der Alterangaben in Jahren ...), dann folgt nach der Formel (9):

$$n = \left(\frac{1,96 * 2}{0,5} \right)^2 \approx 64 \tag{10}$$

Um der geforderten Genauigkeit im statistischen Sinne ausreichend und plausibel zu genügen, müssen also 64 Studierende befragt werden.

Das Konfidenzintervall lässt sich dann, sofern die Werte der Stichprobe weiterhin exakt wie angegeben sind, berechnen:

$$(21,510; 22,490) \tag{11}$$

4.3 Der kritische Wert

Ein wichtiger Bestandteil der Konfidenzintervall-Formel ist der Faktor $1,96$. Mit dessen Hilfe wird das Konfidenzintervall zu einem berechneten Mittelwert bes-

timmt. Ist nun die Distanz des zu testenden Wertes zum Mittelwert kleiner als die Distanz der Konfidenzintervallgrenzen zum Mittelwert, dann fällt die Entscheidung zugunsten von H_0 aus - ansonsten für H_1.

Das bedeutet im Umkehrschluss, man kann auch zunächst die Distanz zwischen dem berechneten und dem zu testenden Mittelwert ausrechnen; anschließend bestimmt man, welcher Faktor sich bzgl. dieser Distanz ergibt; das dann anhand der Normal- oder der t-Verteilung.

Ist dann für die Normalverteilung der Faktor größer als $1,96$, dann ist der Testmittelwert also nicht im Konfidenzintervall von H_0. Man entscheidet sich für H_1.

Anhand des so berechneten Faktors kann nun auch mit Hilfe der Normalverteilung die Wahrscheinlichkeit ausgerechnet werden, mit welcher Wahrscheinlichkeit der Testmittelwert im Konfidenzintervall liegt.

Würde sich z.B. für die hier erläuterte Differenz unter Annahme der Normalverteilung genau ein Wert von $1,96$ ergeben. dann ist die Wahrscheinlichkeit genau 5%.

Bei $1,96$ ist demnach die Lage unentschieden, es ist nicht sicher, ob H_0 oder H_1 ausgewählt werden soll. Insofern wird dieser Wert $1,96$ als $kritischer\ Wert$ bezeichnet, da eine Entscheidung für eine der beiden Hypothesen nicht eindeutig getroffen werden kann.

Diese Wahrscheinlichkeit wird im Folgenden also das Kriterium sein, mit der ein Wert innerhalb oder ausserhalb des Konfidenzintervalles liegt. Dieser Wert wird nun in einem etwas anderen Sinn (um-) interpretiert.

4.4 Der Fehler erster und zweiter Art

Ein Hypothesentest testet auf zwei mögliche Alternativen, die sich normalerweise gegenseitig ausschließen. Diese werden typischerweise als 'Nullhypothese' und 'Alternativhypothese' , Abk.: H_0 und H_1 oder auch H_A bezeichnet.

Dieser Test ist 'schließend', d.h. man wählt eine der Aussagen des Tests aufgrund des Stichprobenergebnisses aus. Da man die Grundgesamtheit (also die 'Realität') nicht kennt, unterstellt man, dass das Ergebnis der Stichprobe auch

in der Grundgesamtheit so gilt.

Es ist klar, dass man auch einem Trugschluss unterliegen kann, man also mit der Entscheidung aufgrund der Stichprobe eine Fehlentscheidung trifft. Man betrachte also, welche Fehler im Bezug auf die Nullhypothese eintreten können:

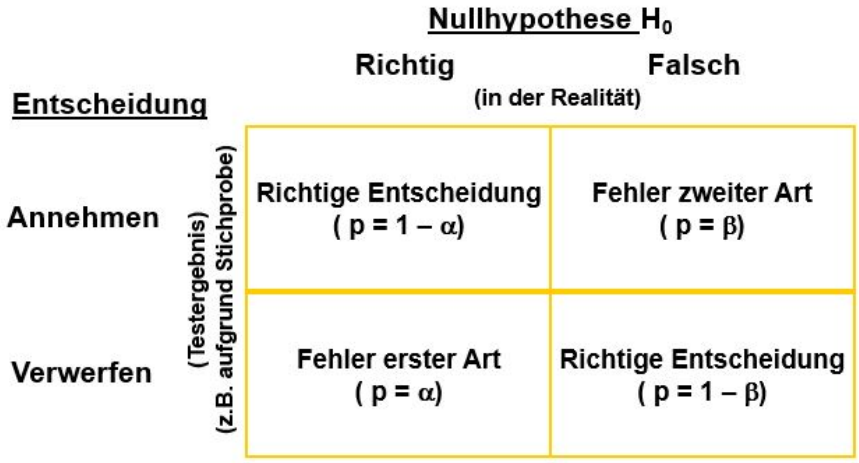


Figure 16: Fehler erster und zweiter Art im Hypothesentest

Der Fehler erster Art ist die Fehlerwahrscheinlichkeit dafür, dass wenn ich die Alternativ-Hypothese auswähle, diese Auswahl falsch ist.

Der Fehler zweiter Art ist die Umkehrung für die Auswahl der Nullhypothese.

Fehler sollen so klein wie möglich sein; daher definiert man nun für den Fehler erster Art eine Fehlentscheidungstoleranzlevel, das *Signifikanzniveau* (Bez.: α).

Dieses Signifikanzniveau muss VOR dem Test festgelegt werden; typischerweise setzt man dieses Signifikanzniveau auf 5%, es sind aber durchaus auch 1% oder 10% möglich bzw. sinnvoll.

Der mathematische Hypothesentest ist so aufgebaut, dass die Irrtumswahrscheinlichkeit für den Fehler erster Art berechnet wird.

Die Entscheidung bei einem Hypothesentest, also die Entscheidung, welche der Hypothesen H_0 oder H_1 nun (mit genügend statistischer Sicherheit) angenommen werden soll, kann nun also durch Vergleich der berechneten Irrtumswahrscheinlichkeit mit der max. zulässigen Fehlertoleranz in Gestalt des Signifikanzneiveaus getroffen werden: ist $p < \alpha$, dann entscheidet man sich für H_1 - und umgekehrt: ist $p \geq \alpha$, dann entscheidet man sich für H_1.

4.4.1 Exkurs: Berechnung der Irrtumswahrschenlichkeit

Einzusehen ist, dass die Wahrscheinlichkeit p mit Hilfe des Konfidenzintervalls berechnet werden muss. Das Konfidenzintervall entscheidet, ob ein Wert signifikant unterschiedlich von einem anderen ist.

Da p, die Irrtumswahrscheinlichkeit, eben eine Wahrscheinlichkeit ist, benötigt man eine Wahrscheinlichkeits-Funktion, die z.B. aufgrund der Differenz der Werte, wie es beim Mittelwertvergleich oben erläutert wurde, diese aufgrund der Differenz unter Bezugnahme zum Konfidenzintervall berechnet werden muss.

Beim Mittelwertvergleich ist das Konfidenzintervall:

$$\mu \in (\bar{x} - 1,96 * s/\sqrt{n}, \bar{x} + 1,96 * s/\sqrt{n}) \tag{12}$$

und man testet, ob ein Wert μ sich in dem Intervall befindet.

Umgekehrt kann man nun die Differenz vom Wert μ zum Wert $\bar{x}$ ausrechnen.

Beim Vorgehen zuvor war der Wert 1,96 entscheidend für die Signifikanz. Dieser Wert repräsentiert gerade die 95% - Überdeckung bzgl. der Normalverteilung. Umgekehrt soll nun dieser Faktor vor $s/\sqrt{n}$ über die Differenz ausgerechnet werden.

Ist das Ergebnis dann größer als 1,96, dann ist die Differenz zwischen $\bar{x}$ und μ also so groß, dass ein signifikanter Unterschied vorliegt. Und umgekehrt: Ist der Wert kleiner also 1,96, dann liegt kein signifikanter Unterschied vor.

Umstellen der Formel für das Konfidenzintervall für 1,96 eben auf den zu berechnenden Vorfaktor liefert:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \tag{13}$$

Mit diesem Wert wird nun die Überdeckungswahrscheinlichkeit bzgl. der Normalverteilung ausgerechnet.

4.5 Die t-Verteilung im Hypothesentest

Das Problem der im vorigen Abschnitt geschilderten Vorgehensweise ist, dass aus der Stichprobe zwei Parameter, die beide relevant für die Normalverteilung sind (nämlich $\bar{x}$ und s) geschätzt werden.

Die Standardabweichung kann sich also von Stichprobe zu Stichprobe ebenfalls verändern.

Das führt zu einer weiteren Unsicherheit in der Bestimmung der Konfidenz.

Führt man dieses mathematisch aus, so kommt man zu dem Ergebnis, dass statt der Normalverteilung die sog. t-Verteilung für die Konfidenzintervalle und somit auch für den Hypothesentest (p Wert-Berechnung) verwendet werden muss.

Die Schwankung der Standardabweichung ist ebenfalls von der Stichprobengröße abhängig (analog der Darstellung bzgl. des Mittelwertes und der Herleitung des Konfidenzintervalles). Insofern ist die t-Verteilung ebenfalls von der Stichprobengröße abhängig.

Ist die Stichprobengröße n, so definiert man mit $n-1$ den sog. Freiheitsgrad. Die t-Verteilung wird dann bzgl. dieses Freiheitsgrades zur Stichprobe zugehörend definiert.

Ist also $n = 10$, so sind die Schätzungen für die Konfidenz und auch zur Berechnung der Irrtumswahrscheinlichkeit p über die t_9-Verteilung zu berechnen.

Hier eine Darstellung zu verschiedenen t-Verteilungen:

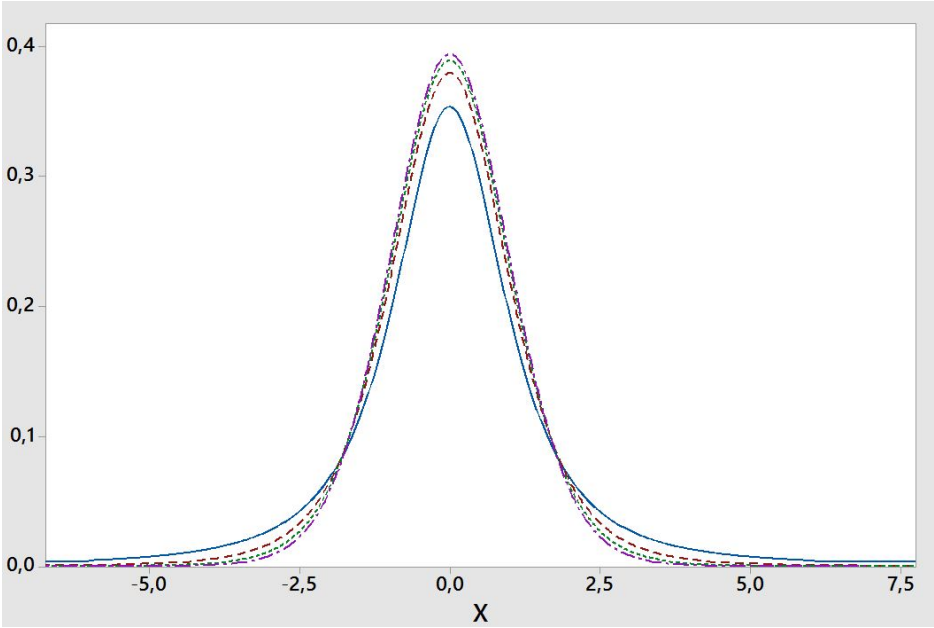


Figure 17: t-Verteilungen unterschiedlicher Freiheitsgrade

Die Breite der Verteilung ist für $n = 2$ am größten; mit zunehmendem n nähert sich die t-Verteilung der Normalverteilung an.

4.6 Weitere Hypothesentests: Voraussetzungen prüfen

Die so am Beispiel des Mittelwertvergleiches aufgezeigte Vorgehensweise kann nun auf unterschiedliche weitere Fälle übertragen werden. Der Vorteil ist, dass bei Kenntnis der Hypothesenaussagen man anhand des berechneten p-Wertes sofort eine Entscheidung treffen kann.

Vorab sind jedoch das Signifikanzniveau festzulegen, sowie ggf. Voraussetzungen zu prüfen.

Die Voraussetzungen sind Eigenschaften, die aufgrund von Stichprobendaten nachzuweisen sind. Diese Nachweise werden wiederum ebenfalls mit Hypothe-

sentests geführt.

Als Beispiel diene wieder der t-Test: alle Betrachtungen und Berechnungen sind anhand der Normalverteilung durchgeführt worden. Insofern ist die Normalverteilung eine Voraussetzung für die Anwendbarkeit des t-Tests.

Mit Hilfe eines Tests auf Normalverteilung muss also vorab (!) überprüft werden, ob die Voraussetzungen für den Test erfüllt sind.

Aus den Formeln ist zudem ersichtlich, dass die Standardabweichung eine wichtige Rolle spielt; dieses entscheidet über die Breite der Konfidenzintervalle und ist somit entscheidend für das Testergebnis.

Bei einem 2-Stichproben t-Test sollten somit die Standardabweichungen für die beiden Stichproben in etwa gleich groß sein.

Diese Eigenschaft ist wiederum vor (!) dem t-Test zu überprüfen, es gibt hierfür Hypothesentests, die die Varianzgleichheit testen.

Insofern ist ersichtlich, dass man Eigenschaften jeder Art, die im statistischen Sinn relevant sind, mit Hypothesentests testen möchte.

Beispiele:

- Test auf Anteil, z.B. bei nominalen Merkmalen (Quoten)
- Test auf Trend (d.h. bei metrischen Merkmalen Test auf rein auf- oder absteigende Werte)
- Test auf Runs
- Test auf Zufälligkeit
- Test auf Strukturbruch
- Test auf Verteilung
- Test auf Additivität

- Test auf Linearität
- Regressionen
- Zeitreihenanalysen

Diese Eigenschaftentests sind typischerweise als zusätzliche Tests in den Statistik-Werkzeugen wie Minitab optional auswählbar.

4.7 Nicht-parametrische Tests

Ein Test in Bezug auf eine Voraussetzung kann zu einem negativen Ergebnis führen. So kann z.B. bei einem t-Test die Voraussetzung der Normalverteilung der Daten durchaus nicht erfüllt sein. Beim 2-Stichproben t-Test reicht es, dass eine der Stichproben nicht der Normalverteilung genügt.

Für diese Situation muss man Alternativen wissen; diese Alternativen sind die sog. nicht-parametrischen Tests.

Typischerweise werden Hypothesentests nach deren 'Erfinder' benannt. Insofern hier die Liste der nicht-parametrischen Tests als Tabelle:

t-Test-Alternative für 1 Stichprobe	Wilcoxon-Test
t-Test-Alternative für 2 Stichproben	Mann-Whitney-Test

Der Mittelwert ist im Übrigen lediglich bei symmetrischen Verteilungen eine sinnvoll zu betrachtende Größe. Sobald die Verteilung beliebig angenommen wird, ist als Mittelwert im statistischen Sinn der Median zu wählen. Insofern ergibt sich als Hypothesen beim Wilcoxon-Test dementsprechend:

H_0: der Median $\tilde{x}$ ist gleich η_0 H_1: der Median $\tilde{x}$ ist ungleich η_0

5 Die Vorab-Analyse

Nun wird es Zeit, sich konkreter mit dem Umgang mit Messwerten (zu Merkmalen) auseinanderzusetzen. Die Situation ist also so, dass Daten aus Stich-

proben von Merkmalen von Prozessen vorliegen.

Der folgende Analyseprozess bezieht sich auf eine erste Vorabanalyse dieser Daten, die einem zunächst im Detail nicht weiter bekannt sind. Also wenn z.B. eine Stichprobe zur Darstellung der aktuellen Qualitätslage erhoben wird, dann werden diese Messwerte vom Analysten i.d.R. nicht selbst erhoben, sondern z.B. per Email vom (z.B.) Messingenieur zugesendet.

Zu diesem Zeitpunkt, wenn Daten zugeliefert werden, geht man davon aus, dass alles funktioniert hat und alles gewissenhaft durchgeführt wurde. Aber ist das auch der Fall ? Eine durchaus kritische Überprüfung ist nun erforderlich.

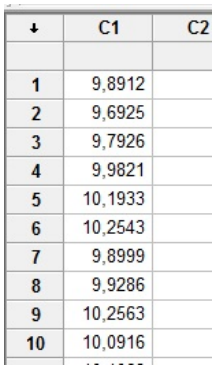
Aus der Praxis ist halt bekannt, dass Daten Messfehler, Eintragungsfehler (Zahlendreher) oder auch nur tatsächlich gemessene Auffälligkeiten enthalten können. Diese Fehler oder Auffälligkeiten müssen vorab - von Analyst - gefunden und hinterfragt werden können, bevor man weitere tiefgreifendere und komplexere Analysen durchführt. Die zugrundeliegende Datenqualität ist für Aussagen z.B. ggü. einem Management, halt entscheidend für alles, was daraus abgelesen werden soll. Schließlich übernimmt der Analyst ab dem Zeitpunkt, in dem er die Daten akzeptiert, die Verantwortung für die dann folgenden Analysen und die daraus abgeleiteten Aussagen. Diese Akzeptanz der Daten sollte also unbedingt sehr kritisch sein.

Als erstes empfiehlt es sich tatsächlich einen Einblick in die zeitliche Abfolge der Messungen, die in Minitab in einer Datenspalte eines Arbeitsblattes vorhanden sind, zu verschaffen.

Anm.: Hier im Folgenden wird das für ein stetiges Merkmal getan, das ist aber genauso hilfreich bzw. notwendig (aus oben genannten Gründen) für anders ausgeprägte Merkmale.

5.1 Zeitreihendiagramm

Ein einfaches Zeitreihendiagramm gibt für einen bislang unbekannten Datensatz einen ersten Einblick. Hier sind beispielhaft die ersten Zeilen eines Beispieldatensatzes in Minitab zu sehen:



	C1	C2
1	9,8912	
2	9,6925	
3	9,7926	
4	9,9821	
5	10,1933	
6	10,2543	
7	9,8999	
8	9,9286	
9	10,2563	
10	10,0916	

Figure 18: Beispieldatensatz - erste Zeilen

Anm.: der hier als Beispiel verwendete Datensatz wird auch für die nächsten Beispiele beibehalten. Der Datensatz ist im Downloadbereich zur Vorlesung vorhanden.

Die Datenpunkte werden mit einer Linie miteinander verbunden, man spricht dann von einer Spur.

Bei kontinuierlichen Prozessen sollte man zudem möglichst auf der X-Achse den Zeitpunkt der Messung (bzw. präziser: der Probenahme, die Messung kann durchaus (viel) später erfolgen) notieren und die Daten dann als sog. Zeitreihe darstellen. Evtl. sind Zeiteffekte in den Daten vorhanden, z.B. bei einer Betriebsaufnahme nach einem Wochenende oder mit Trend zum Ende eines Tages. Diese Auffälligkeiten sind natürlich immer interessant zu wissen.

Für diese Daten wird nun das Zeitreihendiagramm erzeugt, dieses ist wie folgt:

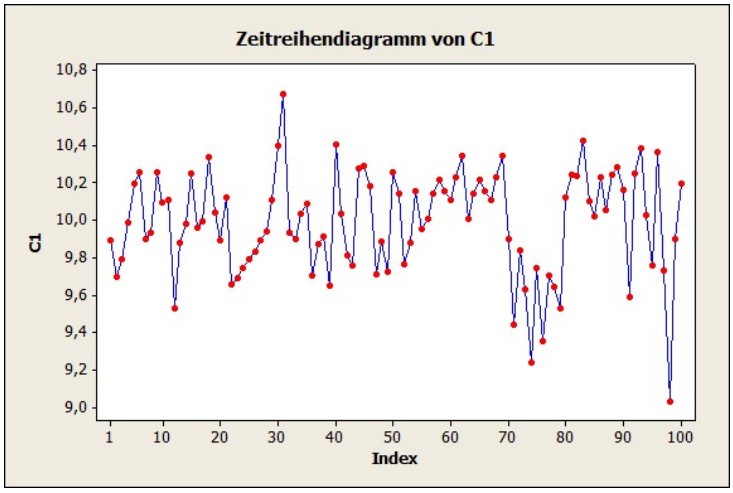


Figure 19: XY-Plot

Folgende Auffälligkeiten sind in den Daten zu diesem Beispiel sofort zu erkennen:

- die Punkte 30 und 98 sind relativ deutlich von den anderen Punkten entfernt; es handelt sich wohl um sog. Ausreißer nach oben bzw. unten
- die Punkte 23 bis 30 sind nacheinander aufsteigend, es handelt sich evtl. um einen sog. Trend
- die Punkte 70 bis 79 bilden eine verschobene Gruppe von Messwerten, eine solche Gruppe wird als Run bezeichnet

Bislang sind diese Auffälligkeiten Vermutungen, also nicht bewiesen oder - wie in der Statistik üblich - mit hoher Wahrscheinlichkeit anzunehmen.

Aussagen werden in der Statistik mit Hilfe von Hypothesentests 'bewiesen' (in dem Sinn, dass man mit Hilfe eines Tests eine Aussage - unter Beachtung der zugrunde liegenden Voraussetzungen - mit genügend kleiner Irrtumswahrscheinlichkeit annehmen kann. Für eine kurze Erläuterung siehe den Exkurs zum Hypothesentest im Anhang.

Insofern benötigen wir also nun Hypothesentests für die im Beispieldatensatz zu sehenden Auffälligkeiten. Folgende Tests wären dafür beispielhaft (es gibt zu den hier vorgestellten Tests durchaus Varianten oder Alternativen) hilfreich:

5.1.1 Test auf Ausreißer

Als ein Test auf Ausreißer sei hier der Test von Grubbs angegeben. Dieser Test setzt voraus, dass die Daten annähernd normalverteilt sind.

Anm.: 'annähernd normalverteilt' wären Daten, die monomodal, halbwegs symmetrisch und ohne zu viele Ausreißer sind. Diese vage Beschreibung hier drückt Voraussetzungen aus, die in der Praxis rein qualitativ (subjektiv) 'gesehen' werden können.

Die Prüfgröße des Tests ist:

$$P = \frac{|x_i - \bar{x}|}{s}$$

Die Null-Hypothese ist: x_i ist kein Ausreißer. Die Alternativ-Hypothese ist entsprechend das Gegenteil, also x_i ist ein Ausreißer.

Die Testverteilung ist:

$$z_\alpha = \frac{n-1}{\sqrt{n}} \sqrt{\frac{t^2_{\frac{\alpha}{2n},n-2}}{n-2+t^2_{\frac{\alpha}{2n},n-2}}}$$

Mit diesem Test sollten alle x_i untersucht werden und die Ausreißer sollten ggf. entfernt werden (wobei nach jeder Entfernung eines Wertes ein neuer Grubbs-Testlauf erforderlich ist). Bei Ausreißern spricht man immer von EINEM Ausreißer. Der Test findet also lediglich einen Wert, der mit höchster Wahrscheinlichkeit ein Ausreißer ist. Andere, evtl. ähnlich zu bewertende Daten werden nicht angezeigt.

Tipp: die Daten nicht einfach durch löschen des 'unangenehmen' Wertes manipulieren. In Minitab kann man im Menü 'Daten' auch einen Dialog zum kopieren von Arbeitsblättern oder von Spalten aus Arbeitsblättern finden. Die Originaldaten sollten auf jeden Fall so wie zugeliefert erhalten bleiben. Eine schöne Möglichkeit einen Ausreißer zu eliminieren ist, in einem Zeitreihendiagramm mit dem 'Pinsel' (Markierung von Punkten) die auffälligen Werte zu markieren und dann in den hier angesprochenen Dialogen für Kopieren über eine Bedingung die

markierten Zeilen auszuschließen (weiteres Untermenü).

Als zweiter Test sei der David-Hartley-Pearson - Ausreißertest vorgestellt.

Dieser Test setzt ebenfalls annähernde Normalverteilung voraus.

Die Null-Hypothese ist: der kleinste Wert x_U bzw. der größte Wert x_O von Punkten $x_1, \ldots, x_n$ sind keine Ausreißer. Die Alternativ-Hypothese ist dann entsprechend: entweder x_U, oder x_O, oder beide Werte sind Ausreißer.

Die Prüfgröße ist:

$$T = \frac{R}{s}$$

(Anm.: R ist die Spannweite der $x_1, \ldots, x_n$.)

H_0 wird abgelehnt, wenn:

$$T < Q_{n,\frac{\alpha}{2}} \text{ oder } \quad T > Q_{n,1-\frac{\alpha}{2}}$$

wobei $Q_{n,\alpha}$ für die kritischen Werte tabelliert vorliegt.

Weitere Tests auf Ausreißer:

- Dean Dixon Test
- Fuchs-Kenett Test
- Hampel Test
- Nalimov Test

In Minitab ist im Bereich 'Statistik → Statistische Standardverfahren' als vorletzter Eintrag der Grubbs- und der Dixon-Test zu finden.

5.1.2 Test auf Trend

Der Jonckheere-Terpstra-Test testet für Stichprobenwerte $x_1, \ldots, x_n$ auf das Vorliegen eines Trends.

Die Formulierung des Tests wird zumeist für Mediane von k Datengruppen definiert.

Gegeben sei (errechnet durch Regression):

$$x_{ij} = c + t_i + \epsilon_{ij}$$

wobei die x_{ij} alle aus einer Verteilung (nicht notwendig Normalverteilung) stammen.

Die Hypothesen des Tests sind nun:

H_0: $t_1 = t_2 = \cdots = t_k$

H_1: $t_1 \leq t_2 \leq \cdots \leq t_k$

wobei mindestens ein Mal die strikte Ungleichung gelten muss.

Anm.: die Formulierung des Tests hier ist etwas unpräzise; die eigentliche Formulierung der Alternativhypothese müsste stochastisch formuliert werden:

$$P(x_j \leq x) \geq P(x_{j+1} \leq x)$$

Anm.: für lediglich 2 Gruppen ($k = 2$ also) entspricht der Test dem Mann-Whitney-Test; dieser Test wird im Kapitel 13.6 beschrieben.

Der Test sei nun im Folgenden konkret beschrieben:

Gegeben seien k Gruppen von jeweils n Messungen. Dann sei weiter:

$$J := \sum_{i<j}^{k} U_{ij} = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} U_{ij}, \quad U_{ij} := \sum_{s=1}^{n_i} \sum_{t=1}^{n_j} \Psi(x_{jt} - x_{is}), \quad \Psi(x) := \begin{cases} 1 & \text{falls} \quad u > 0 \\ 0 & \text{falls} \quad u \leq 0 \end{cases}$$

Die Prüfgröße ist regulär normalverteilt mit Erwartungswert μ_J und Standardabweichung σ_J wie folgt:

$$\mu_J = \frac{N^2 - \sum_{i=1}^{k} n_i^2}{4}, \quad \sigma_J = \sqrt{\frac{N^2(2N+3) - \sum_{i=1}^{k} n_i^2(2n_i+3)}{72}}$$

Die Größe

$$Z := \frac{J - \mu_J}{\sigma_J}$$

wird als standard-normalverteilt angenommen, wenn die Anzahl der Stichproben größer ist als 12.

Der Umbrella-Test ist eine Verallgemeinerung des Jonckheere-Terpstra-Tests und testet auf Gipfel (auf- und absteigender Trend) in den Daten.

Weitere Tests auf Trend:

- Meyer-Bahlburg Trendtest
- Moore Trendtest
- Neumann Trendtest
- Page Test
- Reverse Arrangement Test
- Mil-HDBK Test
- Laplace Test

5.1.3 Test auf Runs

Der Test von Swed-Eisenhart analysiert den Verlauf von Werten im Vergleich zum Median. Dazu wird die Anzahl der sog. Runs (ein Run ist eine Abfolge von Daten ober- oder unterhalb des Medians) bestimmt.

H_0: Die Anordnung der Werte in der Stichprobe ist zufällig

H_1: Die Anordnung der Werte in der Stichprobe ist nicht zufällig

Die Prüfgröße ist:

$$T_{SE} := \frac{|r - \mu_r|}{\sigma_r}, \quad \mu_r := \frac{2n_1n_2)}{n_1 + n_2} + 1, \quad \sigma_r := \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

Die Prüfgröße wird dann wieder auf Standard-Normalverteilung getestet.

Der Test auf Runs fällt in die Klasse der Verkaufstests. Hier eine Liste weiterer solcher Tests:

- Cochran Q Test
- Iterationshäufigkeitstest
- Iterationslängen Test.
- Krauth Test (T1)
- Krauth Test (T2)
- Log-Rank Test
- Mantel-Cox Test
- Mantel-Haenszel Test
- Sprungstellendetektionstest (Cochran)
- Stevens' Run Test
- Wallis Moore Folgevorzeichen Iterationstest

5.1.4 Test auf Strukturbruch

Als Ergänzung soll noch ein Test auf Strukturbruch vorgestellt werden. Ein Strukturbruch liegt vor, wenn die Datenlinie z.B, einen Knick oder Sprung aufweist.

Für Daten $x_{i1}, \ldots, x_{ik}$ mit $i = 1, \ldots, N$ sei eine Regression gerechnet:

$$y_i = \sum_{j=1}^{k} c_j x_{ij} + c_0$$

Man testet nun, ob sich die Daten x_{ik} in zwei Gruppen so aufteilen lassen, dass jede Gruppe wesentlich besser durch eine (lineare) Regression approximiert werden kann.

Dann wäre:

$$\begin{aligned} y_i &= \sum_{j=1}^{k} a_j x_{ij} + a_0 \text{ fuer } i = 1, \ldots, N_a \\ y_l &= \sum_{m=1}^{k} b_m x_{lm} + b_0 \text{ fuer } l = N_a + 1, \ldots, N \end{aligned}$$

Die Hypothesen lauten nun:

H_0: $(a_0, a_1, \ldots, a_k) = (b_0, b_1, \ldots, b_k)$

H_1: $(a_0, a_1, \ldots, a_k) =\neq (b_0, b_1, \ldots, b_k)$

Es sei S die quadrierte Summe der Residuen für die Gesamt-Regression, S_a und S_b die quadrierten Summen der Residuen der Teil-Regressionen, dann ist die Prüfgröße:

$$T_C := \frac{S - \frac{S_a + S_b}{k+1}}{\frac{S_a + S_b}{N_a + N_b - 2(k+1)}}$$

Die so definierte Testgröße folgt einer F-Verteilung mit den Freiheitsgraden $k+1$ und $N_a + N_b - 2(k+1)$.

Ein Test auf Sprünge ist:

- Sprungstellendetektionstest (Cochran)

Mit dieser Reihe von Tests ist es nun müglich, die verdächtigen Muster oder Auffälligkeiten entsprechend zu testen. Ggf. können so gezielt Rückfragen an die Prüfer, die die Messdaten übermittelt haben, gestellt werden, und evtl. müssen Datensätze eliminiert werden, bevor man mit den Analysen fortschreitet.

5.2 Prozessdaten als stationärer stochastischer Prozess bzw. Zeitreihe

Die Streudiagramm-Analyse betrachtet insb. den Ablauf der Messdaten, also deren Dynamik (d.h. in der Zeit). Insofern sind solche Messreihen stochastische Prozesse und somit auch als Zeitreihe auffassbar.

Jedoch verfolgt man mit einem Produktionsprozess ein gewisses Ziel: die Merkmalsdaten sollten - bei unveränderter Einstellung - in einem gewissen Sinn 'gleich' bleiben; also sich in einem 'Werteband' bewegen. Das bedeutet also statistisch gesehen, es sollten dann bei einem stetigen Merkmal der Mittelwert und die Streuung in der Zeit nahezu konstant bleiben. Man spricht dann von einem stationären Prozess; für Einzelheiten diesbzgl. muss hier aber auf die Vorlesung 'Zeitreihenanalyse' verwiesen werden.

5.3 Grafiken für nominale Daten

Für rein nominale Daten sind folgende Analysen hilfreich:

- Auszählung oder auch eine Kreuztabelle

- Pareto-Diagramm

Ein Paretodiagramm ist von der folgenden Gestalt (hier im Beispiel sind 4 Pareto-Teildiagramme zu sehen):

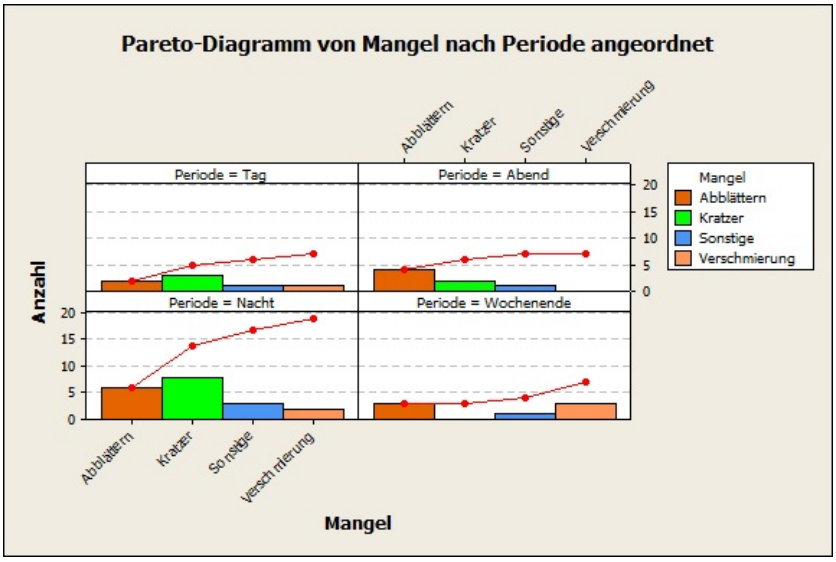


Figure 20: Pareto-Diagramm

Mit Hilfe eines Pareto-Diagramms werden die Attribute von links nach rechts nach Häufigkeit aufgelistet. So erkennt man grosse Häufigkeiten von wenigen attributiven Merkmalen schnell; es dient auch zur Darstellung von sog. '80/20-Diagrammen'.

5.4 Histogramm und Verteilungsmodell

Ist der Prozess stationär, bewegen sich die Daten also in einem gewissen Werteband, dann ist die Frage sinnvoll, mit welcher Häufigkeit die Werte in Intervallen liegen; es stellt sich also die Frage nach der Verteilung, die die Daten gut beschreiben kann. Es wird also ein Verteilungsmodell für die Messwertdaten gesucht.

Um eine Idee für eine Verteilung zu bekommen, werden die Daten klassiert und die Häufigkeit (Anzahl) der Klassenzählungen als sog. Histogramm dargestellt.

Um den Übergang von der dynamischen zur statischen Darstellung aufzuzeigen, ist folgende Darstellung der Daten des Beispieldatensatzes hilfreich:

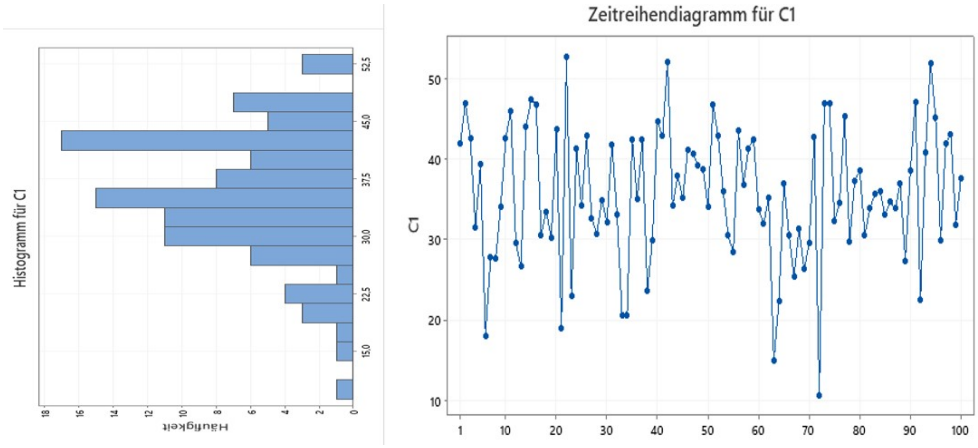


Figure 21: Histogramm zum Zeitreihendiagramm

Zum hier verwendeten Datensatz ergibt sich zunächst folgendes Histogramm:

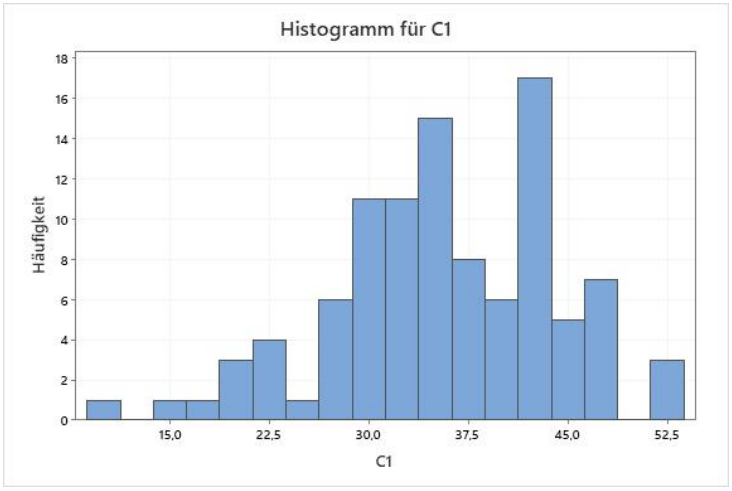


Figure 22: Histogramm

Es sind (weniger bedeutende) Ausreißer durch die isolierten kleineren Balken an den Rändern des Histogramms zu erkennen. Dieser Effekt ist im Übrigen für

kleinere Stichprobenumfänge typisch.

Hier ist aber auch der Effekt zu sehen, dass in einer statischen Analyse Zeiteffekte nicht mehr wahrgenommen werden können. Der Run zu Beginn in etwa des zweiten Drittels der Werte kann höchstens dadurch erahnt werden, dass die Verteilung mehrgipfelig ist. Ein Trend ist nicht ohne weiteres zu erkennen in diesem statischen Bild.

Die Daten sehen aber schon so aus, dass ein Verteilungsmodell zu finden ist. Die Daten sind halbwegs zentriert um einen mittleren Wert herum (d.h. die häufigsten Werte sind in einem relativ kleinen Bereich zu finden), von daher kann nun der Frage nach einem Verteilungsmodell nachgegangen werden. Als mögliche Kandidaten wären z.B. die Normal- oder die Weibull-Verteilung geeignet.

Tipp: in Minitab kann man sich theoretische Verteilungen einzeln, mit Parametervariationen oder auch vergleichend von zwei Verteilungen ansehen. Zudem sind Perzentile / Quantile berechenbar. Man findet den entsprechenden Dialog in 'Grafik $\rightarrow$ Darstellung der Wahrscheinlichkeitsverteilung'.

5.4.1 Test auf Verteilung

Es sind einige Tests auf Verteilung bekannt, darunter solche, die verteilungsunabhängig funktionieren (χ^2-Test), solche, bei denen eine (beliebige) Verteilung angenommen werden muss (Kolmogoroff-Smirnov, Anderson-Darling), oder Tests, die lediglich auf eine spezielle Verteilungen testen (Normalverteilung: Shapiro-Wilk).

In der Vorlesung wird ein Test verwendet, der verteilungsunabhängig ist und gleichzeitig mit einer Best Fit-Approximation arbeitet, so dass die Parameter der Verteilung in diesem approximativen Sinn bestmöglich geschätzt werden. Das Vorgehen ist also zweistufig: erst werden die Parameter für eine Liste vorgeschlagener (oder möglicher) Verteilungen geschätzt, dann wird anhand der so ermittelten Parameter ein Hypothesentest auf Verteilung duchgeführt. Das ist die Arbeitsweise z.B. des Anderson-Darling-Tests (kurz: AD-Test).

Die H_0-Hypothese des AD-Tests lautet: die Daten folgen der Verteilung; H_1 ist entsprechend das (logische) Gegenteil dieser Aussage.

Der ausgegebene α-Wert ist die Irrtumswahrscheinlichkeit für die Annahme von H_1, sofern H_0 in der Realität tatsächlich gilt. Insofern muss α möglichst klein sein.

Der AD-Wert ist die Prüfgröße des AD-Tests; dieser Wert stellt eine Art Abstandsmessung zwischen der empirischen Verteilungsfunktion und der Modell-Verteilungsfunktion dar. Dieser Wert muss dem entsprechend möglichst klein werden. Diese Vorgehensweise ist typisch für die Tests nach von Mises.

Zu den Daten oben bekommt man folgende empirische Verteilungsfunktion:

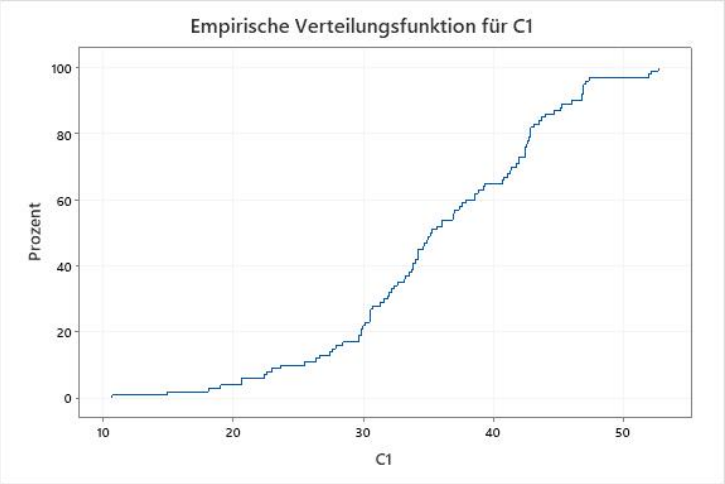


Figure 23: Empirische Verteilungsfunktion

Es gilt bekanntlich: jede Treppenstufe ist $1/n$, wenn n die Stichprobenanzahl ist.

Für die Normal- und die Weibull-Verteilung sieht eine Best Fit-Berechnung dann wie folgt aus:

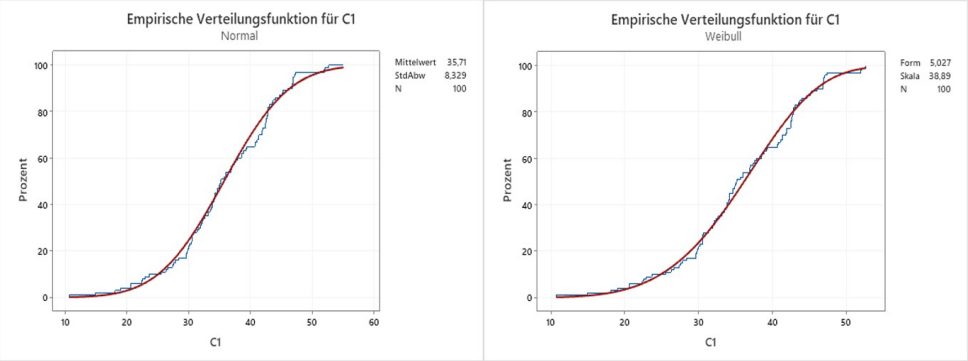


Figure 24: Best Fit Normal und Weibull

Ein Unterschied ist hier nicht wirklich zu erkennen. Jedoch der AD-Test liefert:

Test auf Güte der Anpassung

Verteilung	AD	p	LVT p
Normal	0,522	0,180	
Box-Cox-Transformation	0,553	0,150	
Lognormal	1,971	<0,005	
Lognormal mit 3 Parametern	0,525	*	0,000
Exponential	27,159	<0,003	
Exponential mit 2 Parametern	20,437	<0,010	0,000
Weibull	0,402	>0,250	
Weibull mit 3 Parametern	0,397	0,279	0,478
Kleinster Extremwert	0,730	0,055	
Größter Extremwert	2,546	<0,010	
Gamma	1,284	<0,005	
Gamma mit 3 Parametern	0,668	*	0,001
Logistisch	0,544	0,120	
Loglogistisch	1,048	<0,005	
Loglogistisch mit 3 Parametern	0,545	*	0,001

Figure 25: Tabelle AD-Test

Der sogenannte AD-Wert ist eine Art Unterschiedsmessung. Je kleiner der Unterschied zur Modellvorgabe (also der jeweils betrachteten theoretischen Verteilung) ist, desto besser passt diese Verteilung. In diesem Fall hier ist die 3-parametrige

Weibullverteilung die best passendste Verteilungsfunktion. Die empirische Verteilungsfunktion und der Best Fit im Histogramm sehen wie folgt aus:

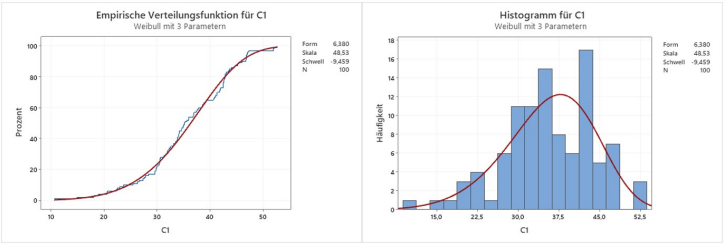


Figure 26: estFit emp. Vert. und Histogramm

Die Anpassung sieht im Histogramm stets schlechter aus, als bei der Anpassung der empirischen Verteilungsfunktion.

Anmerkungen:

1. der Anderson-Darling-Test ist aus der Klasse der von Mises-Test und so konstruiert, dass er besonders in den sog. Tails der Verteilung bestmöglich (also überproportional gewichtet) approximiert.

2. für den Anderson-Darling-Test benötigt man eine gewisse Mindestanzahl an Daten (ca. 50) - aber auch eine sehr grosse Zahl von Datensätzen ist problematisch. Sind mehr als ca. 200 Daten gegeben, dann wird u.U. keine Verteilung als Best Fit-Approximation gefunden, sofern die Daten nicht sehr gut zu mind. einer Verteilung passen.

5.4.2 Wahrscheinlichkeitsnetz

Das Wahrscheinlichkeitsnetz für diese Verteilung versus der Messwertdaten ist:

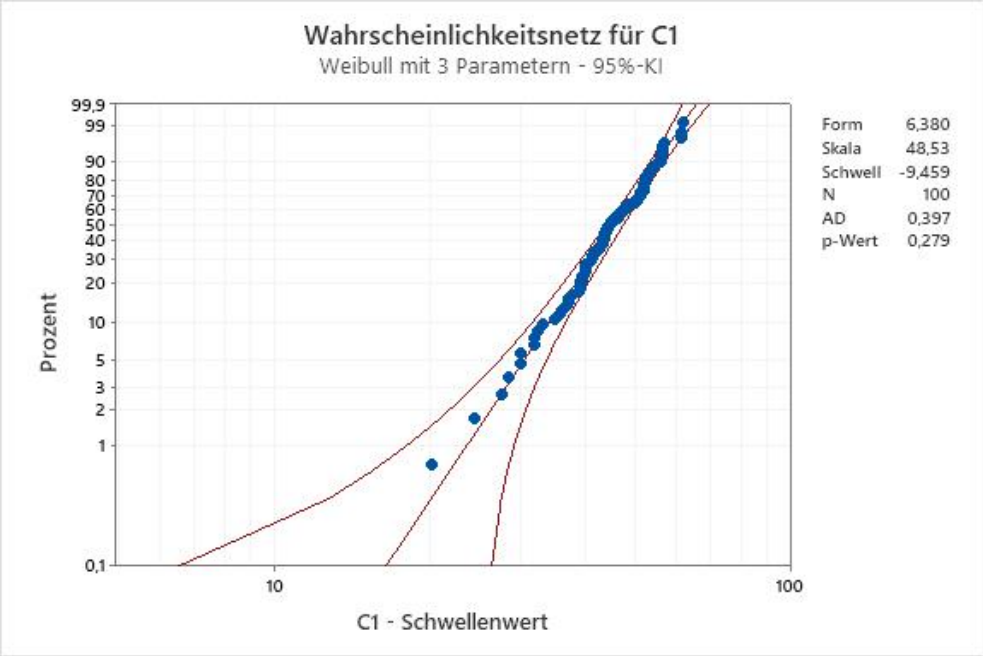


Figure 27: Wahrscheinlichkeitsnetz für die Beispieldaten

Die S-kurvenförmige Linie der Verteilungsfunktion (am besten zu sehen in der Grafik der empirischen Verteilungsfunktion) wird für ein Wahrscheinlichkeitsnetz so transformiert, dass diese eine Gerade wird. Man beachte jedoch die unterschiedlichen Skalen der y-Achsen. Die x-Achsenwerte jedoch bleiben gleich.

Anhand der empirischen Verteilungsfunktion oder des Wahrscheinlichkeitsnetzes kann nun eine Wahrscheinlichkeit für das Eintreten einer Messung geschätzt werden. Die Verteilungsfunktion liefert ja Prozent-Wahrscheinlichkeiten auf der y-Achse für vorgegeben x-Werte - oder umgekehrt. Hier ein Beispiel für die Berechnung:

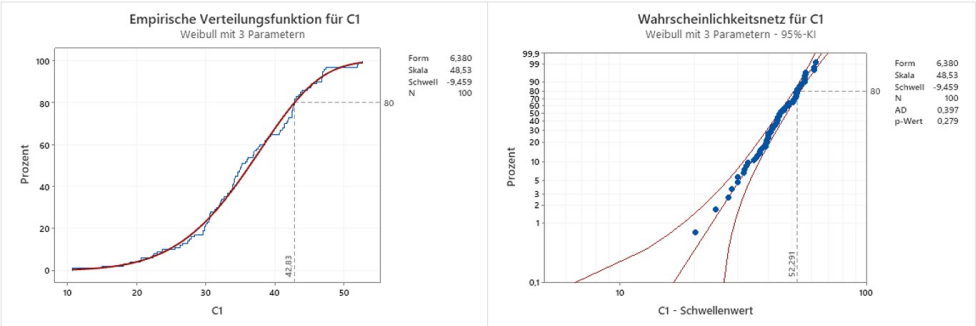


Figure 28: Perzentil-Schätzung anhand emp. Verteilungsfunktion und WS-Netz

Achtung: hier im Wahrscheinlichkeitsnetz ist lediglich eine 2-parametrige Weibull-funktion zu sehen. Der dritte Parameter (Schwellwert) muss zu den x-Werten addiert werden, um die korrekten Werte für die eigentliche Verteilungsfunktion zu bekommen.

Tipp: in Minitab auf die Grafik (z.B. das WS-Netz), dann in die Grafik einen Rechte-Maus-Klick durchführen. Dann in diesem Menü den Punkt 'Perzentillinie' auswählen.

5.4.3 Anmerkungen zur Problematik von Histogrammen

Histogramme dienen lediglich einer oberflächlichen (Sicht-) Kontrolle der Daten. Es sind kaum tiefere Erkenntnisse aus Histogrammen zu ziehen. Das liegt u.a. daran, dass bei einer anderen Klasseneinteilung das Histogramm eine völlig andere Gestalt annehmen kann.

Wählt man beispielsweise für die Beispieldaten hier eine Klassenanzahl von 11 (statt 18), dann ist das 11-klassige Histogramm etwas glatter.

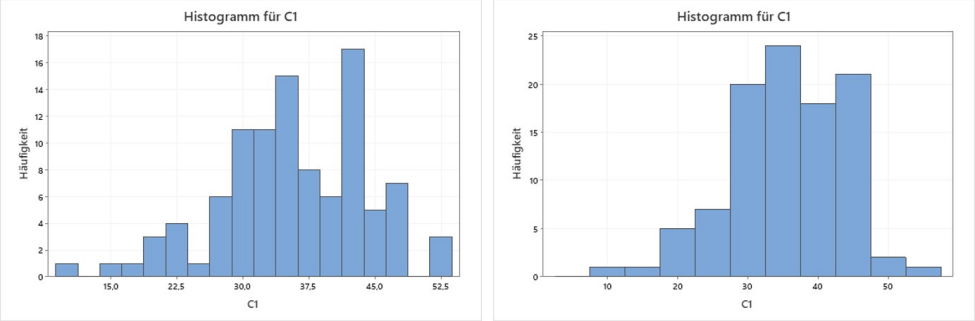


Figure 29: Histogramm mit 11 Klassen

Die Ausreißer-Klassen sind verschwunden und die weitere Gestalt stellt die Datenlücken nicht mehr so drastisch dar. Für die Erstellung eines Histogramms gibt es eine (internationale) Norm. Die Darstellungen sollen so für gleiche Daten auch immer die gleiche Gestalt haben. Die Anzahl der Klassen also bitte nicht ändern (d.h. manuell anpassen).

5.4.4 Quantil-Aussagen anhand der Verteilung

Mit Hilfe einer identifizierten Verteilung können nun Wahrscheinlichkeiten für das Auftreten der Merkmalswerte bis zu einem Wert oder in Intervallen geschätzt werden. Diese Quantilschätzungen werden anhand der Verteilungsfunktion abgelesen.

Für die Beispieldaten folgt für das 5, 50 und 95% - Quantil:

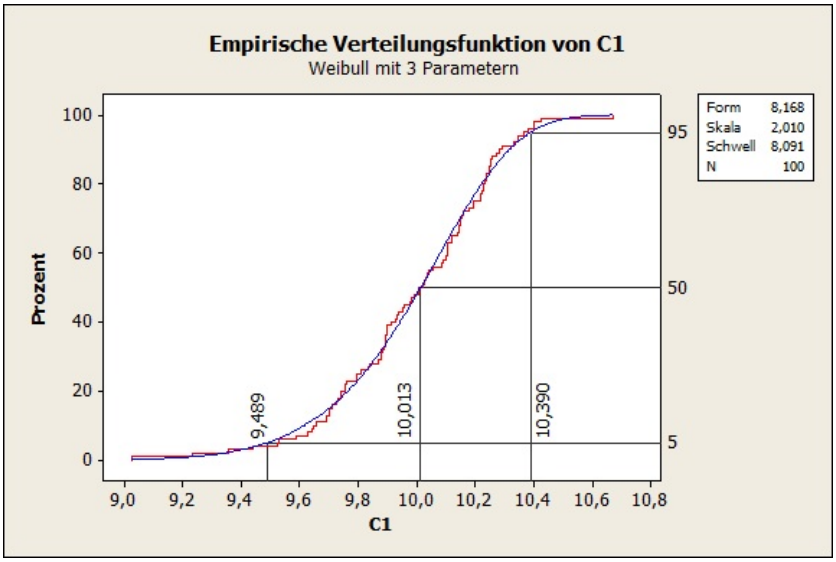


Figure 30: Quantil-Schätzung

6 Qualität von Prozessen

Wie in 5.4.4 bereits dargestellt, ist das Ziel bei einem kontinuierlichen Produktionsprozess, die Qualität auf einem gleichen Niveau zu halten. Das führt zu der Idee, dass Mittelwert und Varianz nahezu gleich bleiben sollten. Dadurch ist aber noch nicht geklärt, ob die Ergebnisse des Prozesses auch den Anforderungen an sie genügen. Es muss also noch untersucht werden, ob Produkte des Prozesses mit Vorgaben, z.B. kundenseitigen Spezifikationen, vereinbar sind.

Das 'ob' aus dem vorigen Absatz ist nun aus statistischer Sicht noch durch die Frage 'wie viele', z.B. gemessen in Prozent, oder noch gebräuchlicher in 'Parts per Million' (PPM), entsprechen den Vorgaben.

Umgekehrt wird so ein Ausschuss, also die Menge oder Quote der nicht brauchbaren Teile bestimmt. Dieser Anteil kann nun mit Hilfe der zugrundeliegenden Verteilungsfunktion geschätzt werden. Genau dieses leistet jetzt die Prozessfähigkeitsanalyse (PFA).

6.1 Wesentliche Kriterien der Qualität der Produkte eines Prozesses

Die Idee nun ist, dass für einen Prozess, dessen Messwerte der Normalverteilung folgen, Mittelwert und Varianz die beiden charakteristischen Werte des Prozesses sind, mit denen man dann die Qualität des Prozesses in Bezug auf die eigentliche Ziel-Qualität, nämlich die Ausschussquote, ggf. sofort einschätzen kann.

Ist z.B. für das Merkmal die Normalverteilung ein gültiges Modell, dann kennt man in Bezug auf Lage (Mittelwert) und Streuung (Standardabweichung) die Wahrscheinlichkeit des Auftretens eines Messwertes. Diese wird für Distanzen mit Faktor der Standardabweichung vom Mittelwert angegeben wie folgt:

Genauigkeit in %	Sigma-Wert	Fehler je 1 Mio. Möglichkeiten (DPMO)
69,15	2	308.537
93,32	3	66.807
99,38	4	6.210
99,98	5	233
99,9997	6	3,4

Figure 31: Quantile der Normalverteilung

Anm.: die hier angegebenen σ-Werte gelten für die gesamte Breite des Wahrscheinlichkeitsintervalls.

Es ist auch möglich, dass die Merkmale nicht der Normalverteilung folgen. In einem solchen Fall sind die Maßzahlen Mittelwert und Varianz bzw. Standardabweichung nicht mehr aussagekräftig. Das gilt insb. für schiefe Verteilungen. In diesem Fall muss die Quantil- bzw. Perzentilschätzung unmittelbar anhand der Verteilungsfunktion durchgeführt werden.

6.2 Prozessfähigkeitsanalyse

Die oben dargestellten Quantilschätzungen sind besonders dann interessant, wenn z.B. eine Produktspezifikation des Kunden als Intervall angegeben werden. Bei nahezu jeder Konstruktion ist klar, dass der ideale Zielwert i.d.R. nicht genau erreicht werden kann. Mit Abweichungen ist also zu rechnen. Die Konstrukteure müssen somit neben dem Zielwert auch ein Intervall definieren, in dem die Werte für das Merkmal toleriert werden können.

Dieses Intervall definiert somit Toleranz- oder Spezifikationsgrenzen. Über- oder unterschreitet ein Merkmalsmeßwert diese obere oder untere Grenze, dann kann das Produkt nicht verwendet werden, ist also evtl. Ausschuss und muss ggf. aufwendig nachbearbeitet werden.

In der Praxis ist i.d.R. die Lage, also der Mittelwert als Zielgröße eines Prozesses relativ gut justierbar. Die Streuung ist typischerweise ein wesentlich größeres Problem.

Anm.: die Prozessfähigkeitsanalyse wird also für jedes Merkmal einzeln durchgeführt. Zu bedenken ist, dass ein Produkt in Gänze sehr viele Merkmale besitzt. Die Verletzung eines einzigen Merkmales kann dazu führen, dass das Komplett-Produkt nicht mehr brauchbar ist. Eine Hochrechnung über die Ausschussrate der (vollständig bemessenen) Produkte sei hier als 'Aufgabe' gestellt (wobei insb. noch eine Abhängigkeit der Merkmale die Berechnung verkompliziert ...).

Um die Qualität eines Prozesse mit Hilfe der beiden Kennzahlen nun einzuordnen wird zunächst die Streuung in Relation zur vorgegebenen Toleranz betrachtet. Man definiert:

$$C_p := \frac{Tol}{6s}$$

wobei Tol die Breite des (vom Kunden vorgegebenen) Toleranzintervalls und s die gemessene Standardabweichung ist.

Klassisch gilt: Ist $p_p > 1,33$, dann heißt der Prozess 'beherrscht'.

Die Bemessung mit $1,33$ ist hier willkürlich. Manche Hersteller (z.B. OEM)

gehen zu einer Verschärfung auf $1,5$ oder sogar $1,66$ über.

In diesem Fall ist die Streuung also klein genug, um prinzipiell eine sehr kleine Ausschussquote zu produzieren.

Anm.: die Umstellung der Ungleichung oben ergibt die Bedingung $Tol > 8s$. In Bezug auf die Normalverteilung wäre dann $\sigma = 4$ in der Tabelle 31. Man liest dann für die (von der Lage unabhängige !) Ausschussquote von höchstens 0,62% ab.

Bei der Prozessbeherrschung geht es zunächst lediglich um die Streubreite der (Messungen der) Merkmale. Ist ein Prozess beherrscht, dann ist die nächste natürliche Frage, ob denn ggü. einer gewünschten Lage (eines Zielwertes) die Qualität auch eingehalten werden kann.

Zu diesem Zweck definiert man:

$$C_{p_k} := \min\left(C_{p_U}, C_{p_O}\right), \quad \text{wobei } C_{p_U} := \frac{\bar{x} - Tol_U}{3s}, \quad C_{p_O} := \frac{Tol_O - \bar{x}}{3s}$$

Tol_O und Tol_U sind hierbei die obere und untere Grenze des Spezifikationsintervalles für die Toleranz.

Der Prozess heißt (klassisch) 'fähig', wenn $C_{p_k} > 1,33$.

6.3 Beispiel zur Prozessfähigkeitsanalyse (PFA)

Für unsere Weibull 3P-verteilten Beispieldaten folgt als Ergebnisoutput für die PFA folgendes Bild (die Spezifikationsgrenzen sind auf 15 und 52 gesetzt):

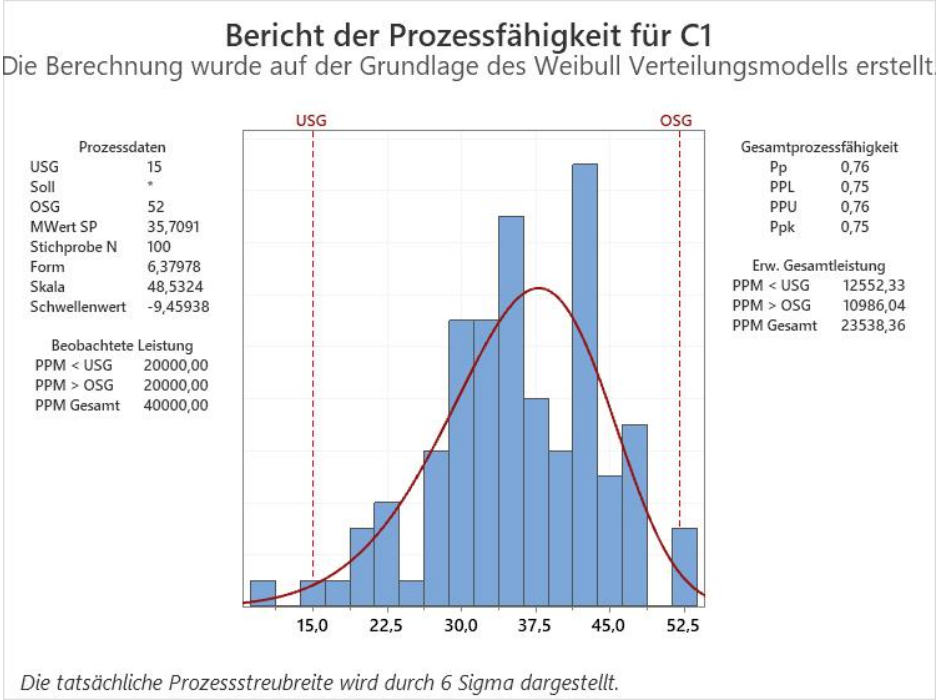


Figure 32: Prozessfähigkeitsanalyse

Der Prozess ist deutlich nicht beherrscht ($p_p = 0,76 < 1,33$).

Anm.: in Minitab wird statt C_p und C_{p_k} nun P_p und P_{p_k} geschrieben. Dieses sind die im amerikanischen gebräuchlichen Schreibweisen.

Die Gesamt-Streuung für Daten, die einer Normalverteilung folgen und die aus $i = 1, \ldots, k$ Stichproben vom Umfang $n_i > 1$ stammen, kann prinzipiell auf zwei Wegen berechnet (geschätzt) werden:

Mit

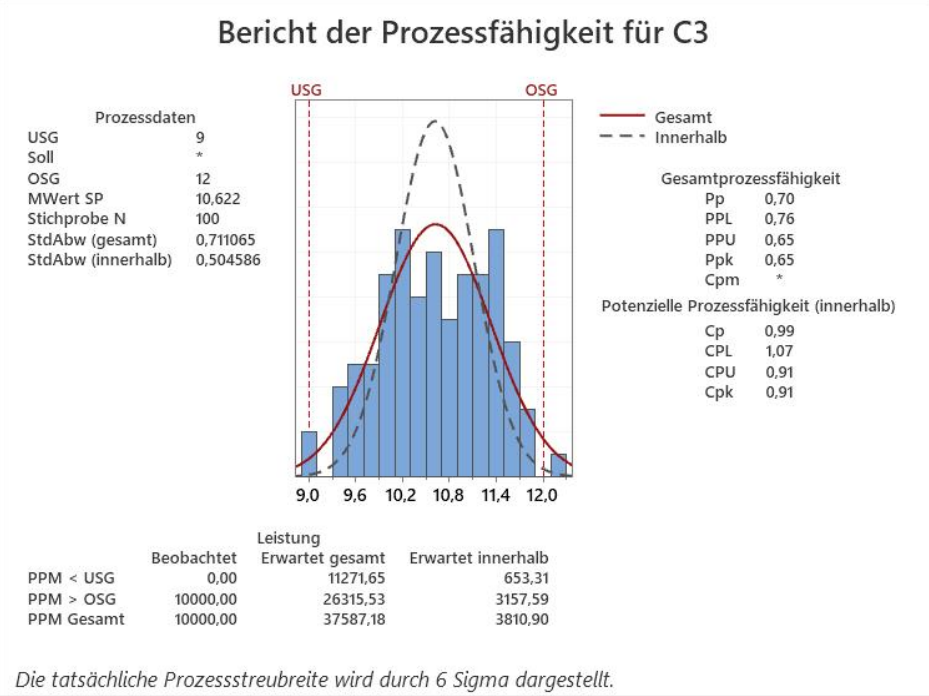
$$s_o^2 = \frac{1}{\sum_{i=1}^{k} n_i - 1} \sum_{j=1}^{k} \sum_{l=1}^{n_j} (x_{jl} - \bar{\bar{x}})^2$$

wird die äussere Streuung bezeichnet, während

$$s_i^2 = \frac{1}{\sum_{i=1}^{k}(n_i - 1)} \sum_{j=1}^{k} \sum_{l=1}^{i} (x_{ik} - \bar{\bar{x}})^2$$

die innere Streuung darstellt.

Das sind zwei mögliche Schätzer für die Varianz bzw. für die Standardabweichung. In den Grafiken werden die jeweiligen Verteilungsfunktionen mit beiden so berechneten Standardabweichungen berechnet und ausgewiesen. Die Bezeichnungen sind dann entsprechend. Hier ein Beispiel, um die Darstellung dann aufzuzeigen:



Prozessdaten	
USG	9
Soll	*
OSG	12
MWert SP	10,622
Stichprobe N	100
StdAbw (gesamt)	0,711065
StdAbw (innerhalb)	0,504586

Gesamtprozessfähigkeit	
Pp	0,70
PPL	0,76
PPU	0,65
Ppk	0,65
Cpm	*

Potenzielle Prozessfähigkeit (innerhalb)	
Cp	0,99
CPL	1,07
CPU	0,91
Cpk	0,91

Leistung	Beobachtet	Erwartet gesamt	Erwartet innerhalb
PPM < USG	0,00	11271,65	653,31
PPM > OSG	10000,00	26315,53	3157,59
PPM Gesamt	10000,00	37587,18	3810,90

Figure 33: Prozessfähigkeitsanalyse normalverteilt, s instabil

Hilfreich für eine weitere Betrachtung ist in Minitab zudem das sog. 'Capability Six Pack':

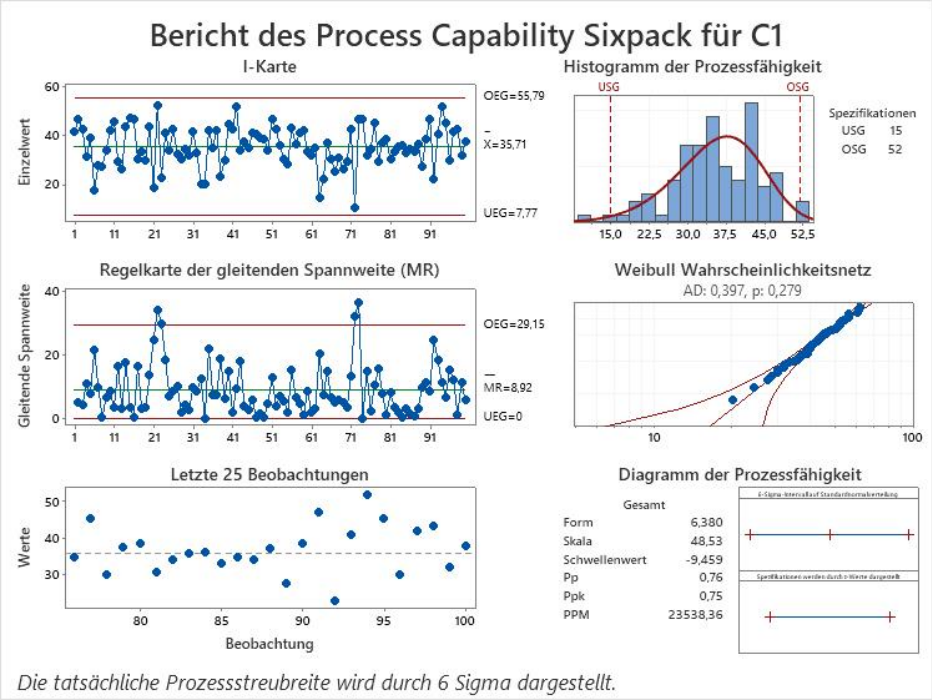


Figure 34: Prozessfähigkeitsanalyse: Six Pack

Man erkennt als erste Teilgrafik eine Zeitreihengrafik (vgl. Abb.19); darunter eine Streuungsgrafik für die sog. gleitende Spannweite (MR = Moving Range). Das Wahrscheinlichkeitsnetz (vgl. Abb.27), die Verteilungsgrafik aus der PFA (vgl. Abb.32) und die da berechneten Prozessparameter C_p und C_{p_k} sind ebenfalls noch in der Grafik dargestellt. Insg. bekommt man hier weitere Informationen darüber, wie die Prozessdaten in der Zeit zustande gekommen sind.

Anhand des 6Packs für die PFA ist nun der Grund für die unterschiedlichen Streuungswerte erkennbar: der Prozess erleidet zur Hälfte der betrachteten Zeit einen Mittelwert-Versatz (deutlich zu erkennen an der Einzelwert-Grafik):

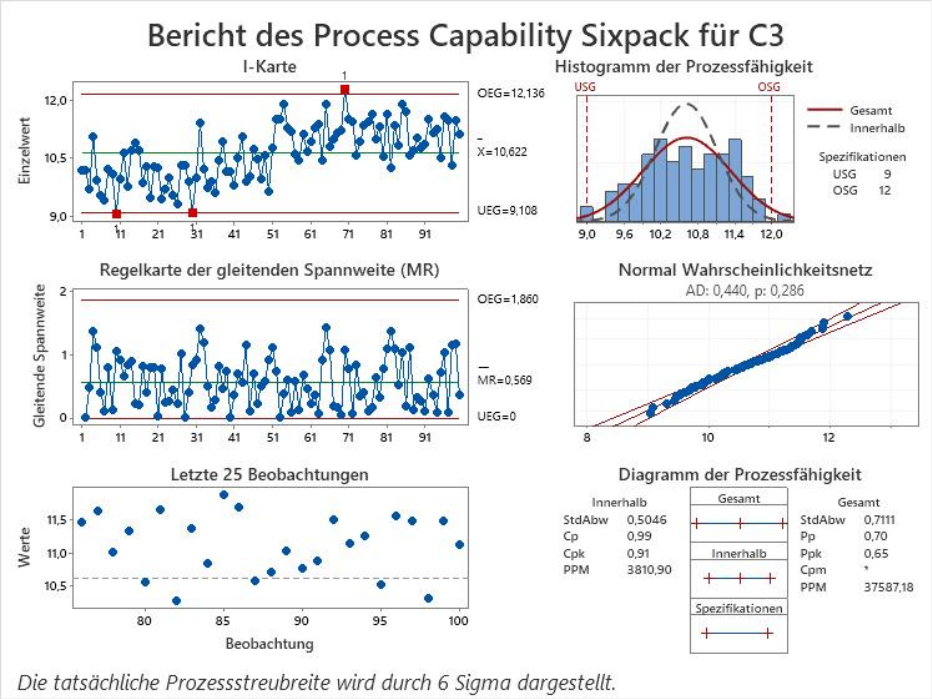


Figure 35: Prozessfähigkeitsanalyse: Six Pack für Mittelwertversatz

7 Qualitätsregelkarten

Die Prozessfähigkeitsanalyse analysiert die Daten eines Merkmales für ein Produkt eines Fertigungsprozesses zu einem Zeitpunkt (bzw. 'ignoriert' die Zeit, da die keine Rolle spielt). Die PFA ist also nicht in der Lage, Veränderungen in der Zeit festzustellen - außer man wiederholt die komplette PFA zu unterschiedlichen Zeitpunkten.

Das wiederum ist insofern problematisch, da man für die Verteilungsfunktionsidentifikation relativ viele Daten benötigt (siehe Abschnitt 2).

Es besteht also der Wunsch, den Prozess in der Zeit, also ggf. in regelmäßigen Abschnitten, mit Stichproben geringen Umfanges zu beobachten, und das zudem grafisch. Um das zu leisten wird eine Qualitätsregelkarte definiert.

7.1 Definition einer Qualitätsregelkarte für ein stetiges Merkmal

Wesentliche Elemente zur Prozessbeobachtung sind wiederum Mittelwert und Streuung.

Die klassische $\bar{x} - s$ - Shewhart-Karte (benannt nach ihrem Erfinder Walter A. Shewhart) ist für unseren Beispieldatensatz der Weibull-Daten wie folgt (als Stichprobengröße wird hierbei 5 angenommen; das bedeutet: zu einem Zeitpunkt werden 5 Proben aus dem Prozess genommen und die Merkmale vermessen):

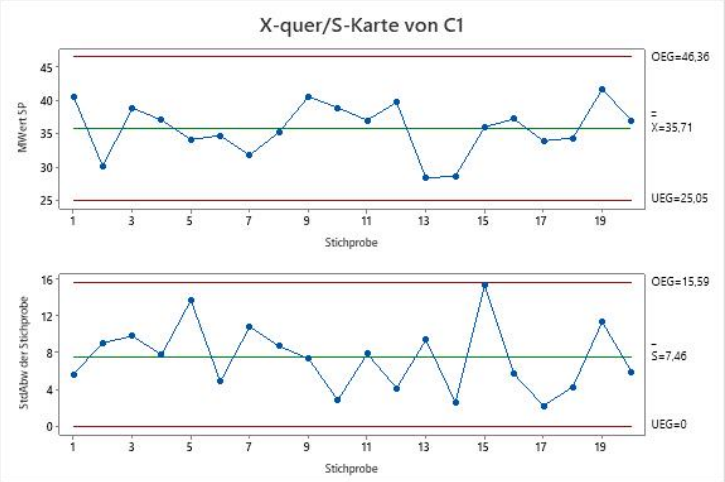


Figure 36: xs-Shewhart-Qualitätsregelkarte

Die in den beiden Grafiken zu sehenden Elemente sind wie folgt definiert:

- jeder Punkt ist ein Mittelwert der Werte bzw. die Standardabweichung zu einer Stichprobe.
- die Linie, welche die Messmittelwerte bzw. die Streuungswerte der Stichproben verbindet, heißt 'Spur'.
- die Mittellinien ist die Linie des Mittelwertes der Mittelwerte (der Stichproben), also $\bar{\bar{x}}$, bzw. der Mittelwert der Standardabweichungen (der Stichproben).

- OEG und UEG bedeuten Obere bzw. Untere Eingriffsgrenze; deren Konstruktion und Bedeutung wird hier jetzt noch im Detail erläutert.

Zur $\bar{x}$-Karte:

Die Abfolge der Mittelwerte ist nun als Zeitreihendiagramm dargestellt.

Die Frage ist nun: wie sind die Mittelwerte verteilt ? Kann man zu diesen so verteilten Mittelwerten eine Quantilanalyse durchführen ? Sind dann 'Toleranzgrenzen' definierbar ?

In diesem Sinn sind die OEG und UEG als Quantile zu verstehen, jedoch (stets) unter der Vereinfachung der Annahme einer Normalverteilung.

Man nimmt also für die Schätzungen der Quantile an, die Daten der Messungen sind normalverteilt (besser: die Stichprobenwerte entstammen einer normalverteilten Grundgesamtheit).

Als Toleranz akzeptiert man einen Bereich so, dass erwartet wird, dass max. in etwa 1 auf 1000 Messungen oberhalb oder unterhalb des Toleranzbereiches liegen. Das führt bei der Standard-Normalverteilung auf 3σ Abweichung nach unten und oben vom Mittelwert (vgl. 31, dort ist $\sigma = 6$ einzusetzen). Die folgende Grafik zeigt dieses auf.

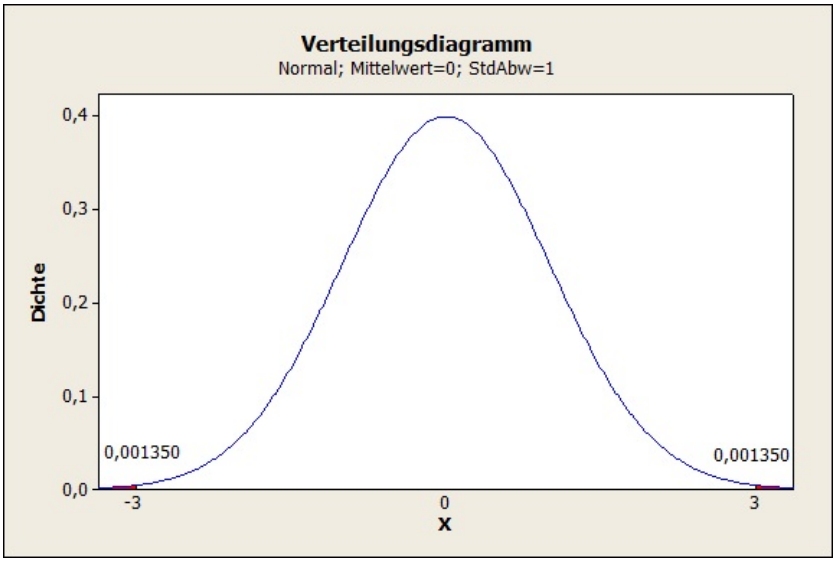


Figure 37: Standard-Normalverteilungsquantile für 3 σ

Die Frage ist nun jedoch: wie sind die Mittelwerte von Messwerten aus Stichproben aus einer normalverteilten Grundgesamtheit eigentlich verteilt ?

Die Verteilung der Mittelwerte ist normalverteilt, **wenn die Standardabweichung bekannt ist.**

Das ist hier nicht der Fall, die Standardabweichung wird aus den laufenden Messungen heraus geschätzt. Dann gilt folgendes:

Der standardisierte Mittelwert normalverteilter Daten ist t-verteilt, wenn die zur Standardisierung des Mittelwerts benötigte Varianz des Merkmals unbekannt ist (also wie der Mittelwert ebenfalls aus der Stichprobenvarianz geschätzt werden muss).

Die Berechnung der OEG und der UEG erfolgt also mit Hilfe der t-Verteilung:

$$t_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

Die OEG und UEG sind für die erste Grafik also Konfidenzintervalle für die

Mittelwerte. Diese hängen von der Stichprobengröße ab und werden dynamisch berechnet. Ändert sich die Stichprobengröße, dann verändert sich auch das zugehörige Konfidenzintervall. Diesen Effekt kann man hier in der nächsten Grafik sehen. Bezogen auf die Beispieldaten sind hier zwei Stichproben 'zusammengelegt' worden, diese zusammengelegte Stichprobe ist also von der Größe 10. Das ergibt folgendes Bild:

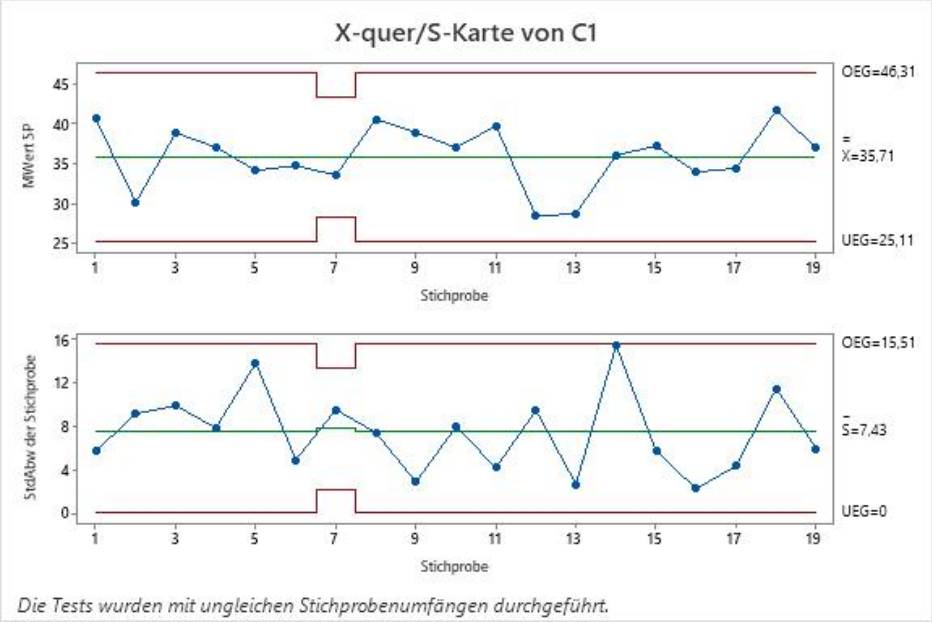


Figure 38: $\bar{x} - s$ - Karte mit unterschiedlichen Stichprobengrößen

Zu einer Regelkarte sind zusätzliche (Hypothesen-) Tests wählbar, die sog. Stabilitätskriterien. Diese Kriterien basieren ebenfalls auf der Annahme, dass die Daten einer normalverteilten Grundgesamtheit entstammen. Werden die Kriterien verletzt, dann führt das in der Ausgabe zu einem Fehler (rot markierte Mittelwert- oder Streuungspunkte).

Für den konkreten Fall der Beispieldaten hier ist kein Fehler vorhanden. Anhand des folgenden Beispiels ist das besser zu sehen:

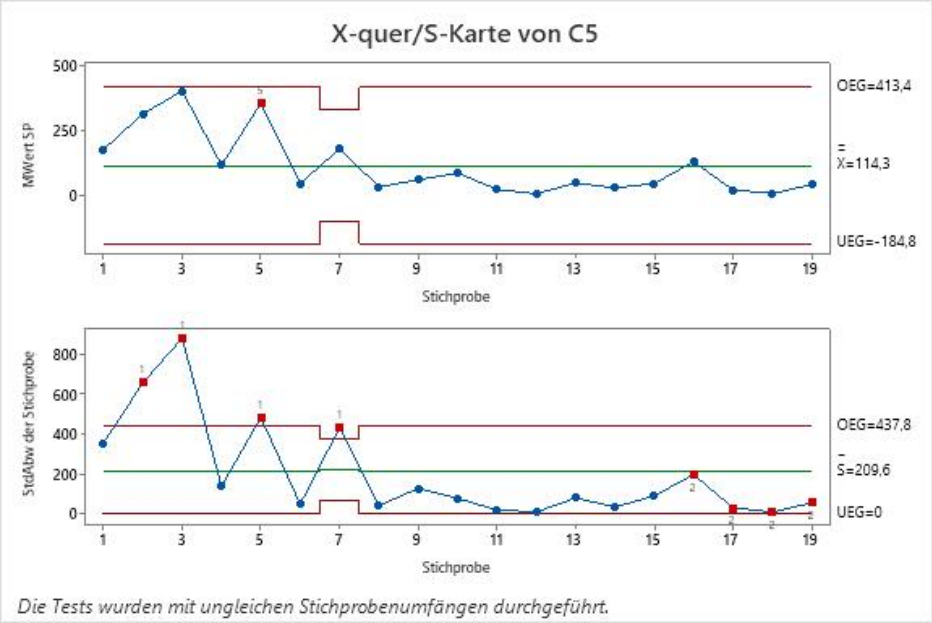


Figure 39: $\bar{x} - s$ - Karte mit Fehlern

Die Test-Fehler werden in Minitab noch in Klartext angegeben:

Testergebnisse für X-quer-Karte von C5

TEST 5. 2 von 3 Punkten mehr als 2 Standardabweichungen von der Mittellinie (auf einer Seite der Mittellinie).
Test fehlgeschlagen an den Punkten: 5

Testergebnisse für S-Karte von C5

TEST 1. Ein Punkt mehr als 3,00 Standardabweichungen von der Mittellinie.
Test fehlgeschlagen an den Punkten: 2; 3; 5; 7
TEST 2. 9 aufeinander folgende Punkte auf der gleichen Seite der Mittellinie.
Test fehlgeschlagen an den Punkten: 16; 17; 18; 19

Figure 40: Fehlerinformation zur X-S-Karten

Die Hypothesentests zu den Stabilitätskriterien sind unterschiedlich einzuordnen. Eine Überschreitung einer Eingriffsgrenze ist ein möglicher Fehler, der deutlich auf eine Prozessstörung hinweist; andere Fehler haben diese Deutlichkeit nicht. Die Fehler sind bzgl. dieser Deutlichkeit absteigend sortiert (d.h. nach Trennschärfe des Kriteriums). Das bedeutet, eine kleine Fehlernummer ist ein deutlicherer

Fehler, als eine Grosse.

Die Tests zu einer $\bar{x} - s$-Karte sind folgende:

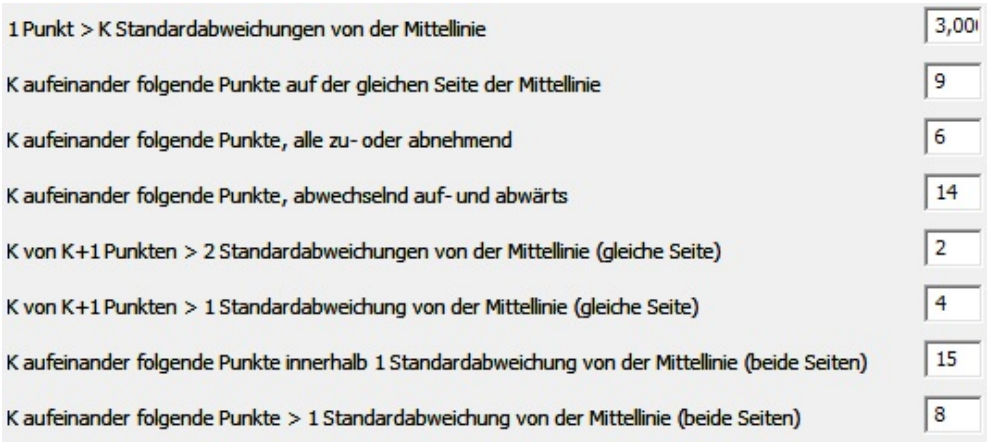


Figure 41: Tests zu den Stabilitätskriterien einer $\bar{x} - s$ - Karte

Anhand der Beschreibung erkannt man Tests für Runs, Trends oder ungleiche Verteilungen.

Die ersten vier Tests sind Mustererkennungen, die Muster sind: Ausreißer, Trend, und Runs (der letztgenannte ist zwei Mal vorhanden). Die letzten vier Tests sind Tests gegen Anteile von Messungen in bestimmten Bereichen der Karte, die Grundlage hierfür sind Wahrscheinlichkeiten bzgl. der Normalverteilung.

Die beiden letztgenannten Tests hier sind Tests gegen das sog. 'Middle-Third'-Kriterium. Dieser Test beruht darauf, dass bei einer Normalverteilung ca. 68%, also in etwa zwei Drittel, der Werte in dem Intervall $[\mu - \sigma, \mu + \sigma]$ liegen.

Vergleicht man die Fehler der Regelkarte mit dem XY-Plot über alle Daten (also ohne Verkürzung durch die Mittelwertberechnung je Stichprobe, siehe Abb. 19), dann kann man auf die Fehlerursachen schließen.

Jedoch sind die nicht alle Besonderheiten der einzelnen ursprünglichen Daten so zu erkennen (wie man z.B. durch Vergleich der $\bar{x}$-Spur mit an dem ersten Ausreißertest in Abb. 19 sieht: der Ausreißer führt in der Regelkarte nicht zu einem

Fehler).

Prinzipiell gilt: je größer die Stichprobe, desto 'unempfindlicher' (im Sinne von Fehlertolerant) ist die Regelkarte; fehlerhafte Effekte werden dann durch die Tests nicht mehr so häufig angezeigt.

Es kann sinnvoll sein, die Spur aufzuteilen. Z.B. wenn der Prozess verändert wurde, oder auch in dem Fall, wenn es einen Prozessabschnitt mit einigen Fehlern gab.

In letzterem Fall führt das typischerweise dazu, dass auch, wenn der Prozess aktuell wieder gut unter Kontrolle ist, einige Fehler durch die Tests angezeigt werden. So führt beispielsweise ein großer Ausreißer nach oben dazu, dass der Test, der bei einer zu großen Anzahl Datenpunkte unterhalb des Mittelwertes anschlägt, häufig auftritt (denn das arithmetische Mittel wird durch den Ausreißer nach oben verschoben, es sind dann überproportional viele Daten unterhalb der Mittelwertlinie).

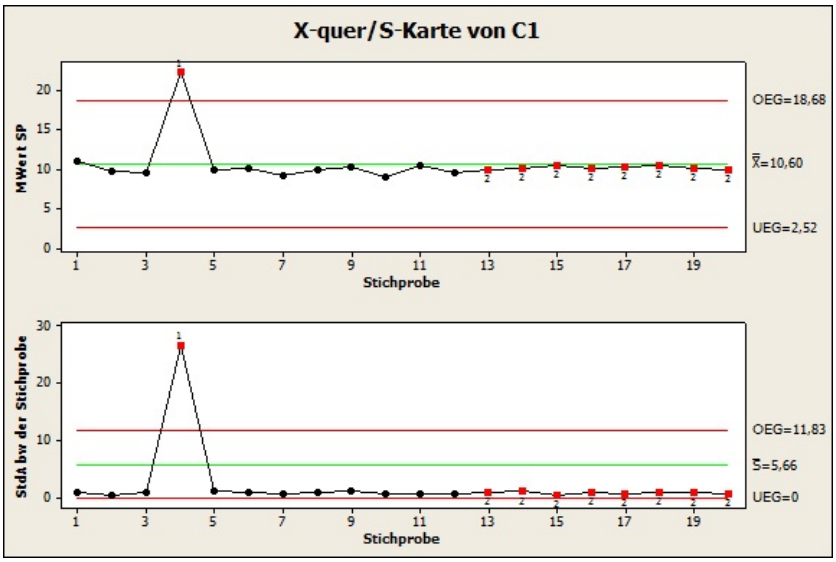


Figure 42: QRK mit Ausreißer provoziert Folgefehler

Die Fehlermeldung ist:

```
Testergebnisse für X-quer-Karte von C1

TEST 1. Ein Punkt mehr als 3,00 Standardabweichungen von der
Mittellinie.
Test fehlgeschlagen an den Punkten:  4

TEST 2. 9 aufeinander folgende Punkte auf der gleichen Seite der
Mittellinie.
Test fehlgeschlagen an den Punkten:  13; 14; 15; 16; 17; 18; 19; 20

TEST 7. 15 Punkte innerhalb 1 Standardabweichung von der Mittellinie
(ober- und unterhalb der Mittellinie).
Test fehlgeschlagen an den Punkten:  19; 20


Testergebnisse für S-Karte von C1

TEST 1. Ein Punkt mehr als 3,00 Standardabweichungen von der
Mittellinie.
Test fehlgeschlagen an den Punkten:  4

TEST 2. 9 aufeinander folgende Punkte auf der gleichen Seite der
Mittellinie.
Test fehlgeschlagen an den Punkten:  13; 14; 15; 16; 17;
                                     18; 19; 20
```

Es wäre also sinnvoll, den Ausreißer hier zu isolieren. Das kann mit Hilfe der Abschnitte geschehen.

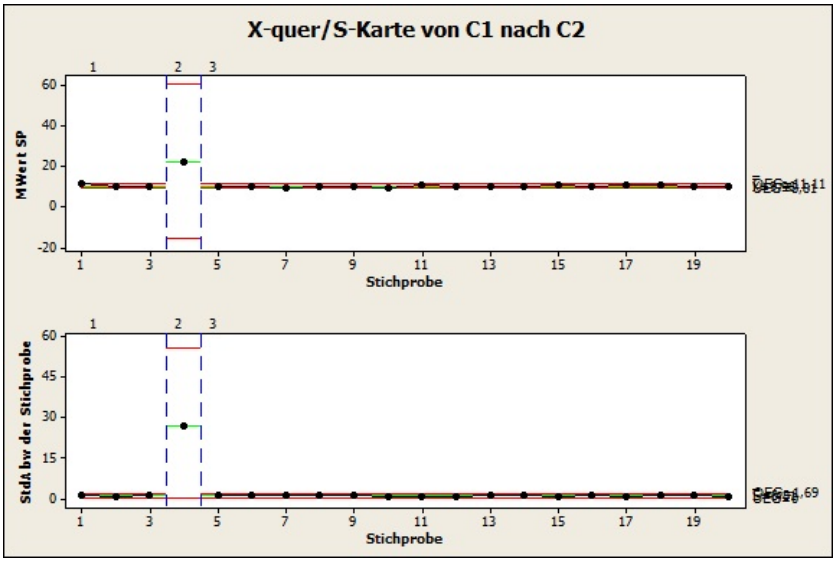


Figure 43: QRK mit Ausreißer und Abschnitten - ohne Folgefehler

Insofern ist eine Regelkarte über die Einzelwerte als Ersatz für den XY-Plot zu definieren, in die dann gewisse Tests (bezogen auf die Normalverteilung !) bereits integriert sind. Das ist die I-MR-Karte (I steht für 'Indivuelle Werte', also Einzelwerte, MR für 'Moving Range', also gleitender Mittelwert).

Für den Beispieldatensatz ergibt die I-MR-Karte:

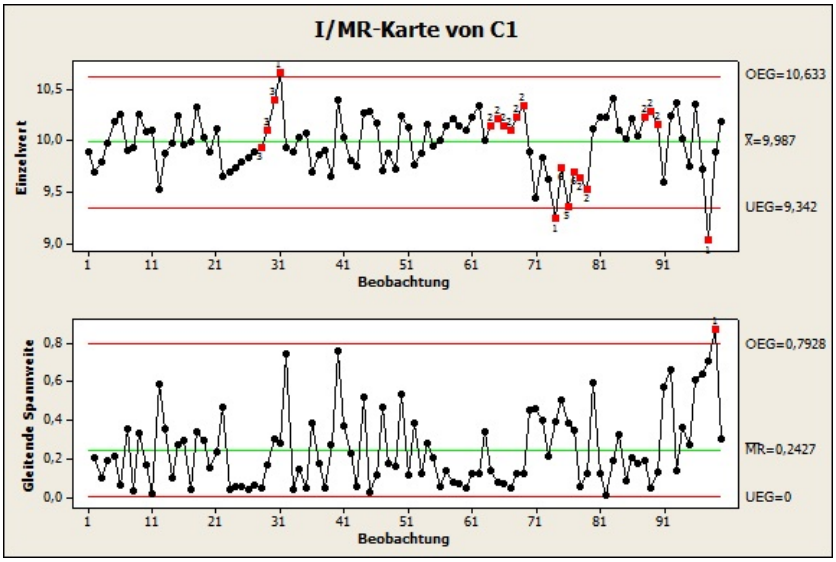


Figure 44: I-MR-Karte

Man erkennt deutlich mehr ausgewiesene Fehler (was zu erwarten war), deren Beschreibung ist:

```
Testergebnisse für I-Karte von C1

TEST 1. Ein Punkt mehr als 3,00 Standardabweichungen von der
Mittellinie.
Test fehlgeschlagen an den Punkten:  31; 74; 98

TEST 2. 9 aufeinander folgende Punkte auf der gleichen Seite der
Mittellinie.
Test fehlgeschlagen an den Punkten:  64; 65; 66; 67; 68; 69; 78;
                                     79; 88; 89; 90

TEST 3. 6 aufeinander folgende Punkte, alle zu- oder abnehmend.
Test fehlgeschlagen an den Punkten:  28; 29; 30; 31

TEST 5. 2 von 3 Punkten mehr als 2 Standardabweichungen von der
Mittellinie (auf einer Seite der Mittellinie).
Test fehlgeschlagen an den Punkten:  76
```

```
TEST 6. 4 von 5 Punkten mehr als 1 Standardabweichung von der
Mittellinie (auf einer Seite der Mittellinie).
Test fehlgeschlagen an den Punkten:  75; 76; 77; 78; 79


Testergebnisse für MR-Karte von C1

TEST 1. Ein Punkt mehr als 3,00 Standardabweichungen von der
Mittellinie.
Test fehlgeschlagen an den Punkten:  99
```

Die hier vorgestellten Regelkarten werden an anderer Stelle nochmal verwendet werden, diese werden dann in gewisse grafische Auswertungen automatisch ausgewiesen.

Zudem sind noch eine ganze Reihe weiterer Regelkarten für stetige Merkmale definiert worden, alle zu unterschiedlichen Zwecken. Darauf wird hier im Einzelnen nicht weiter eingegangen, es wird hier nur eine Liste und die Karte zu den Beispieldaten für eine eigenständige Interpretation angegeben:

Die I/MR - R/S-Karte:

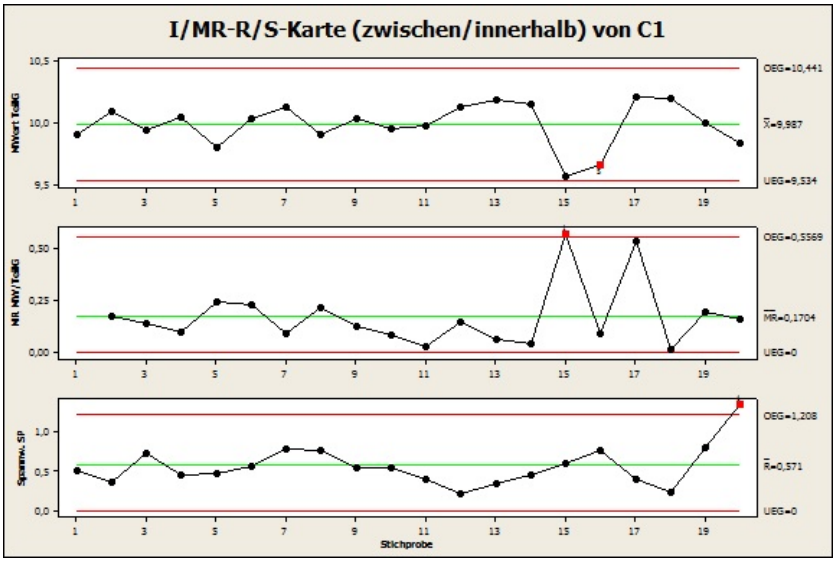


Figure 45: I/MR - R/S - Karte

Zeitlich gewichtete Regelkarten:

Die Gleitender Durchschnitt-Karte:

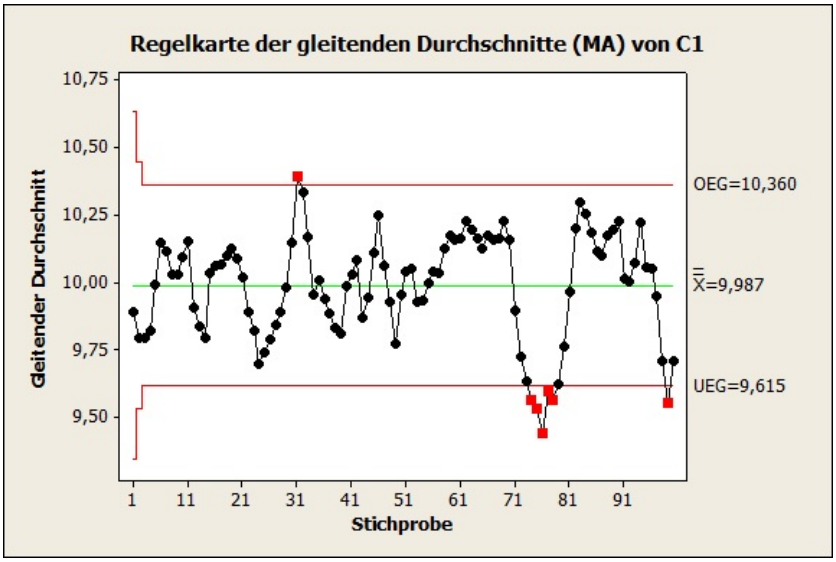


Figure 46: Gleitender Durchschnitt - Karte

Die EWMA-Karte (Exponential Weighted Moving Average):

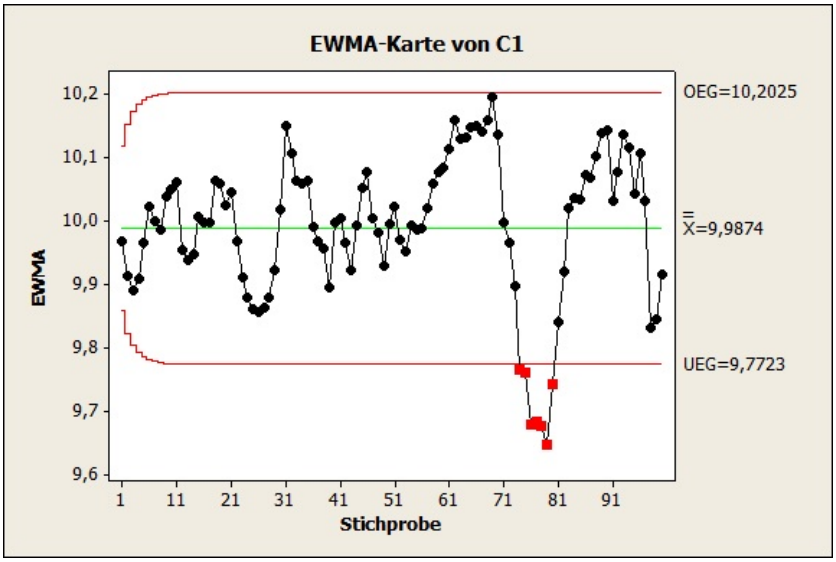


Figure 47: EWMA - Karte

Die CUSUM-Karte (Cumulated Sum, kumulierte Summe der Abweichungen vom Sollwert):

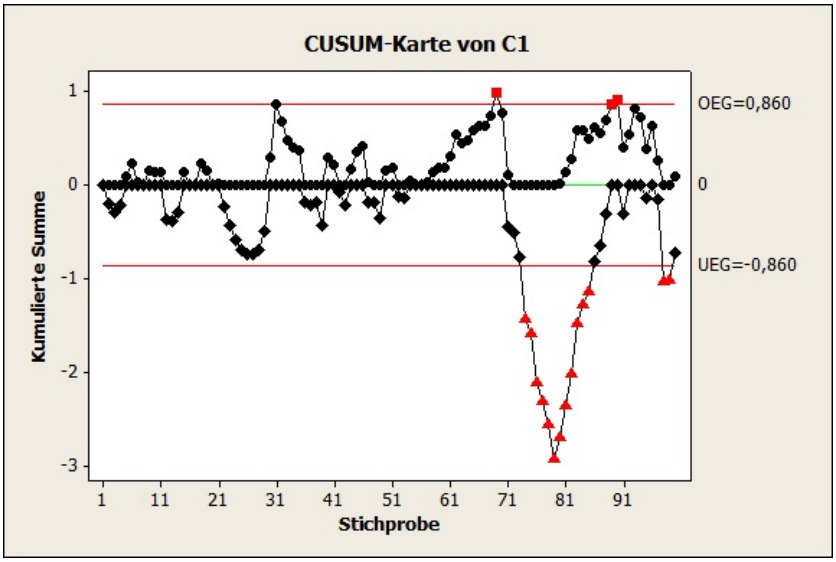


Figure 48: CUSUM - Karte

Die EWMA-Karte ist geeignet, langfristige Trends usw. zu finden. Die CUSUM-Karte hingegen ist geeignet, ein 'Weglaufen' des
Prozesses schneller als eine I/MR- oder $\bar{x} - s$-Karte zu erkennen.

7.2 Definition einer Qualitätsregelkarte für ein attributives Merkmal

Bei attributiven Merkmalen sind vor allem Anteile wichtig, so also auch Quoten; z.B. Fehlerquoten beobachten für einen Lieferanten dessen gelieferte Qualität.

Eine (relative) Quote (z.B. von fehlerhaften Teilen in einer Lieferung) sei mit p bezeichnet; ist N die Anzahl der zur Ermittlung der Quote verwendeten Teile, dann ist $p * N$ demnach die Anzahl der Teile, die der Quote entsprechen (also z.B. die absolute Anzahl der fehlerhaften Teile).

Diese Quote bzw. die Anzahl der fehlerhaften Teile wird nun als Prozessgrößen dargestellt:

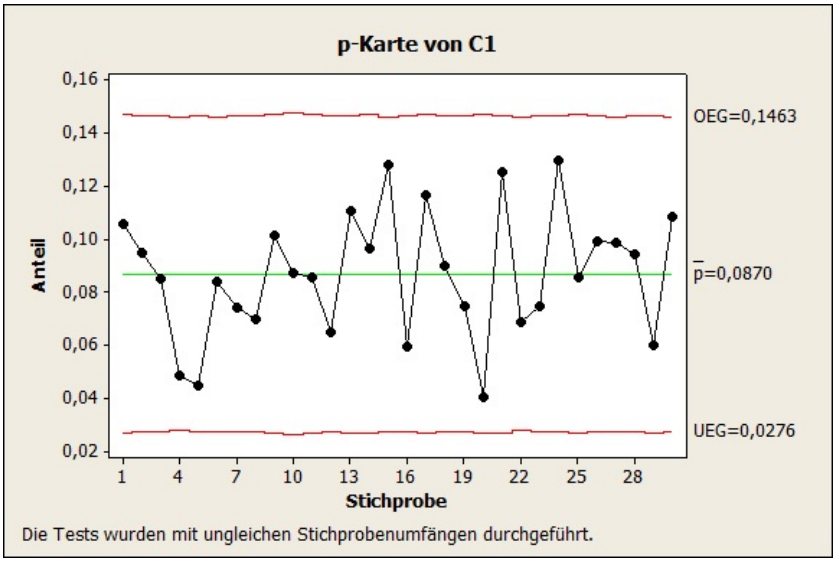


Figure 49: p - Karte für eine Quote p

Es sind unterschiedliche Stichprobengrößen verwendet worden, man beachte den Hinweis auf der Grafik.

In der Karte sind wiederum eine OEG und eine UEG zu erkennen. Diese werden berechnet als Konfidenzintervalle zur Binomialverteilung.

Auch zur p-Karte sind Stabilitätskriterien definiert, die zugehörigen (Hypothesen-) Tests beziehen sich dann auf die Binomialverteilung. Die Tests sind:

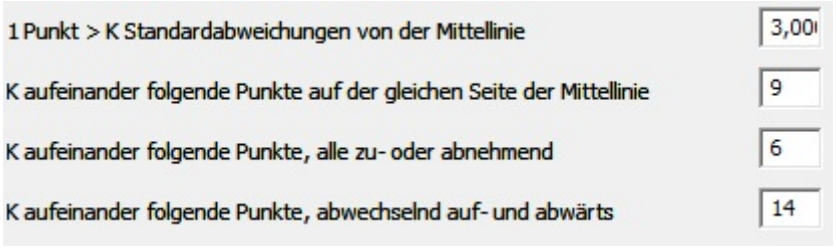


Figure 50: Tests zu den Stabilitätskriteren einer p - Karte

Auf einer Np-Karte werden dann die absoluten Fehler als Spur dargestellt:

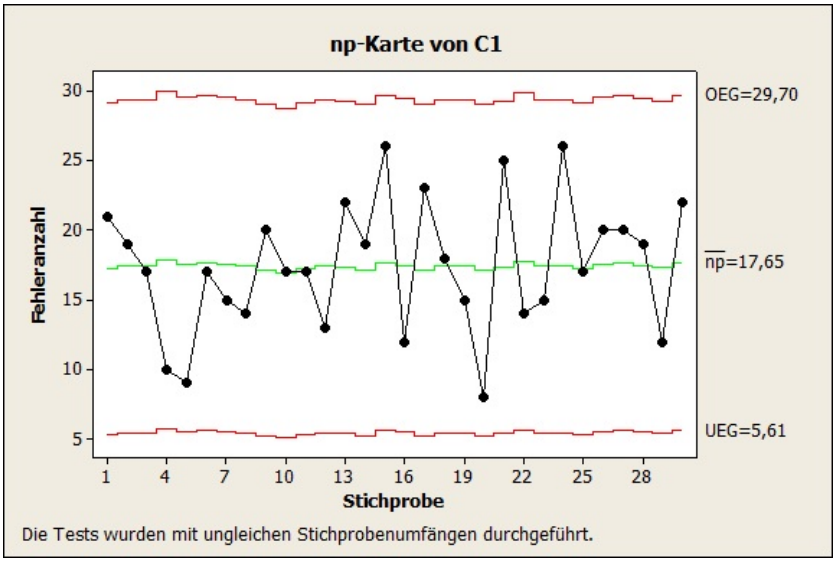


Figure 51: Np - Karte

Zu erkennen ist jetzt (im Unterschied zu einer $\bar{x} - s$- oder p-Karte), dass die Mittellinie nicht durchgezogen gerade ist; die mittlere Quote hängt bei dieser Karte eben von der Stichprobengröße ab.

Die Stabilitätskriterien und die zugehörigen Tests sind analog zur p-Karte.

Auch für Poisson-Verteilte Merkmale gibt es eine Regelkarte zur Überwachung, eine sog. u-Karte. Auf die wird hier nicht weiter eingegangen, die Arbeitsweise ist sehr ähnlich zur p-Karte (es werden lediglich andere Grenzen berechnet, eben auf Basis der Poisson-Verteilung.

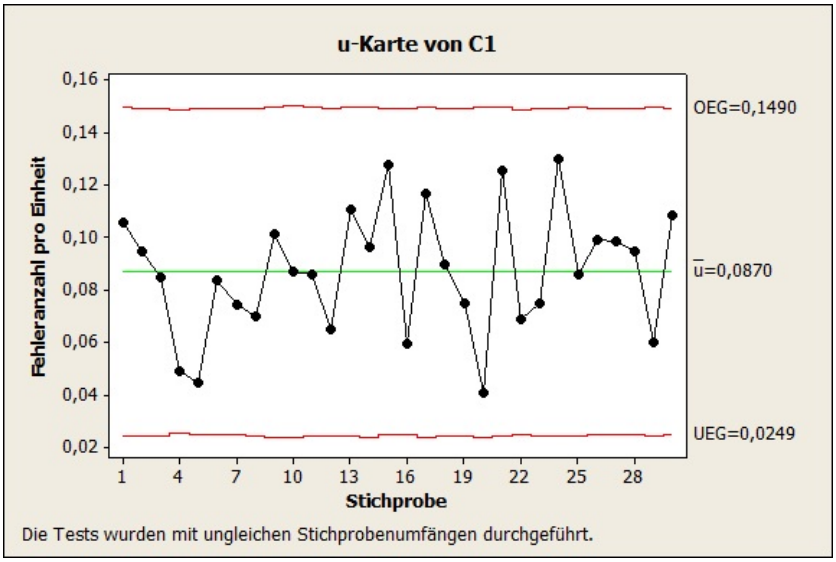


Figure 52: u - Karte

8 Annahmeprüfung

Jetzt verlassen wir die produktionsbegleitende Überprüfung der Qualität mittels SPC und betrachten, wie die Qualität beim Kunden überprüft werden kann.

Dem Kunden werden z.T. komplexe Produkte geliefert, bestehend aus diversen Komponenten. Diese werden im QM durch eine Vielzahl von Merkmalen beschrieben. Sobald eines der Merkmale die zulässigen Toleranzgrenzen verletzt, ist das Produkt für den Kunden nicht mehr brauchbar.

Eine einfache Abschätzung, wie groß die Ausschussquote für das Gesamtprodukt ist, sofern man Informationen über die Fehlerquoten der einzelnen Merkmale hat, ist im Anhang zu finden.

Auf Seite des Kunden wird nun eine Überprüfung der Gesamtfunktionalität der Produkte initiiert, die Annahmeprüfung. Diese bestehen im Wesentlichen aus dem Annahmetest.

Annahmetests testen also wiederum mittels Stichproben die einzelnen Merkmale und geben ein Urteil im Sinne von z.B. 'OK' oder 'Ausschuss' ab.

Dieses Merkmal ist nun ein binomiales Merkmal. Binomiale Merkmale werden im Idealfall durch die Binomialverteilung erfasst.

Wichtig hierbei (wie schon bereits in 7.2 angedeutet) ist die Quote für die Rate des Ausschusses. Diese Quote dient in der Praxis z.B. in Verträgen zwischen Lieferant und Produzent als Qualitätsanforderung.

Um eine solche Quote zu überprüfen, benötigen wir also Aussagen über die Wahrscheinlichkeit darüber, dass wenn eine bestimmte Fehlerzahl in der Stichprobe gefunden wurde, wie gross (geschätzt) dann die Quote in der gesamten Lieferung ist.

Für dieses Vorgehen wird nun ein Hypothesentest aufgebaut. Was dann wiederum den Mathematiker interessiert, sind die Fehler, die diesem Hypothesentest unterliegen.

In der Praxis muss für ein Prüfungsergebnis eines zugelieferten Loses eine eindeutige Entscheidung gefällt werden können. Das bedeutet, bei einer festgestellten Fehlerzahl für eine geprüfte Stichprobe für das zugelieferte Los muss unmittelbar eine Entscheidung gefällt werden können (die z.B. auch als Arbeitsanweisung der im Einkauf beschäftigten QM-Mitarbeiter dient).

Zu diesem Zweck wird nun eine max. zulässige Fehleranzahl (c) definiert, bis zu der das Los angenommen wird. Wird diese Fehleranzahl überschritten, dann wird das Los abgewiesen - ansonsten angenommen.

Das Bild des Prozesses zum Vorgehen bei einer solchen Prüfung ist:

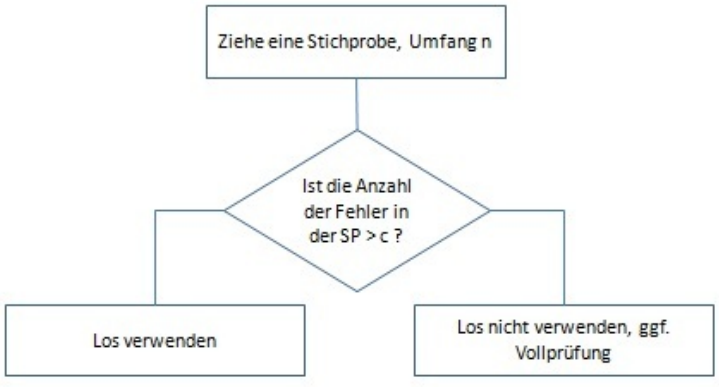


Figure 53: Prozessbild für den einstufigen Prüfplan

Die Frage ist nun: wie oft ist diese Entscheidung dann - erwartungsgemäß - falsch ?

Zunächst ist anzumerken, dass es zwei Fehler (wie in der Statistik üblich) gibt: die fälschliche Annahme, oder die fälschliche Rückweisung.

Und weiter: wie kann man ggf. diese Fehler auf ein bestimmtes (akzeptables) Maß reduzieren (Restrisiko) ? Welcher Aufwand ist dann gegeben ?

8.1 Einstufiger Prüfplan

Die Hypothesentestfehler, die möglich sind, sind:

- die Quote ist in der Realität OK, das Los wird aber abgelehnt
- die Quote ist in der Realität nicht OK, das Los wird aber akzeptiert

Diese Aussagen hängen nun von der Größe des Loses, d.h. des Stichprobenumfanges, und von der gefundenen Fehleranzahl im Los ab. Als Qualität des Tests sind dann die beiden Fehler, der erster Art α, und der zweiter Art β zu berechnen.

Bei den ersten Betrachtungen von Hypothesentests wird typischerweise das sog. Signifikanzniveau α betrachtet, jedoch zunächst nicht die Trennschärfe $1 - \beta$. Bei der Vorgehensweise hier müssen jedoch beide Fehler berücksichtigt werden. Zudem ist es in der Praxis so, dass die Annahme von fehlerhaften Teilen (weitaus) schlimmer ist, als die fehlerhafte Rückweisung. D.h. der Fehler 2. Art wird jetzt bestimmend.

Dazu wird die sog. Operationscharakteristik betrachtet. Diese Funktion ist im Anhang genauer erläutert (siehe Abschnitt 19.2).

Für die hier schon motivierte 'Gut-Schlecht-Prüfung' einer Massenproduktion als Annahmeprüfung beim Kunden wird folgendes vereinbart:

- ein Los ist 'gut' und wird angenommen vom Kunden, wenn höchstens 2% fehlerhafte Teile im Los enthalten sind
- ein Los ist 'schlecht' und wird abgelehnt (zurückgewiesen), wenn es mehr als 5% fehlerhafte Teile enthält
- für den Fehler erster Art, also eine Zurückweisung eines 'guten' Loses, wird eine Wahrscheinlichkeit von höchstens 5% gefordert
- für den Fehler zweiter Art, also eine Annahme eines 'schlechten' Loses, wird eine Wahrscheinlichkeit von max. 10% gefordert

Anm.: die Quoten 2% und 5%, die hier ein Los als 'gut' oder 'schlecht' definieren, heißen die annehmbare und die zurückzuweisende Qualitätslage. Der Bereich zwischen diesen Qualitätslagen heißt indifferente Qualitätslage.

Man hat nun den Stichprobenumfang und die max. zulässige Fehlerzahl für diese Stichprobe aufgrund dieser Festlegungen zu bestimmen.

Die Losgröße ist hier noch nicht eingebracht worden in die Betrachtungen. Da eine wie hier beschriebene Prüfung eine Stichprobenprüfung ohne zurücklegen ist, muss eigentlich die Hypergeometrische Verteilung zur Berechnung der zu bestimmenden Größen herangezogen werden. Für größere Losgrößen mit kleinen Fehlerquoten jedoch kann statt der Hypergeometrischen Verteilung die Binomialverteilung verwendet werden (was eine Stichprobenprüfung mit Zurücklegen

entspricht; bei großen Losen ist die Wahrscheinlichkeit, dass man mal ein Teil doppelt prüft, sehr gering). Und für sehr große Lose kann dann wiederum die Binomialverteilung durch die Normalverteilung ersetzt werden (Normal-Approximation).

Werkzeuge verwenden typischerweise die Binomialverteilung, 'auf Papier' kann man besser mit der Normalverteilung rechnen.

Das Beispiel oben führt in Mintab im Dialog für 'Annahmestichprobenprüfung nach Attributen' ...

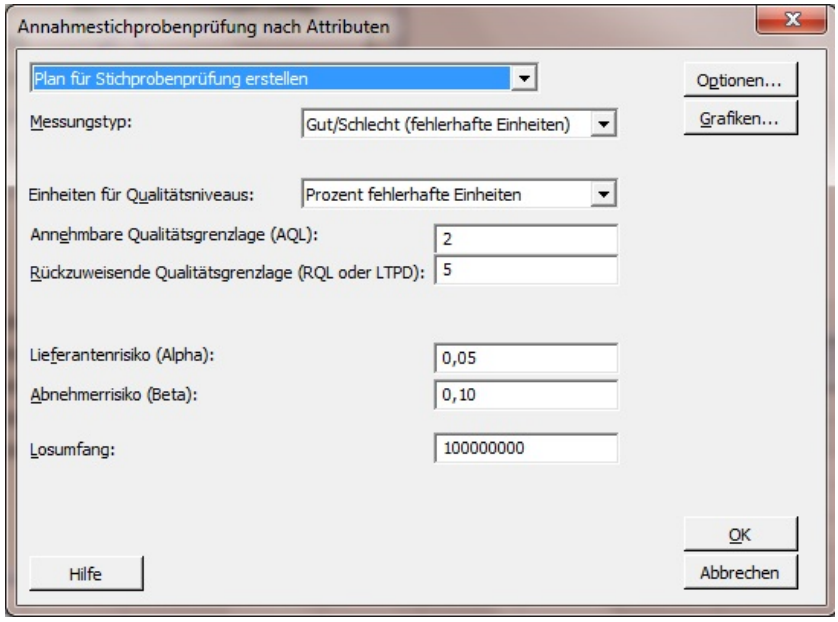


Figure 54: Eingabe Minitab Annahmeprüfung - Prüfplan

... zu folgendem Prüfplan:

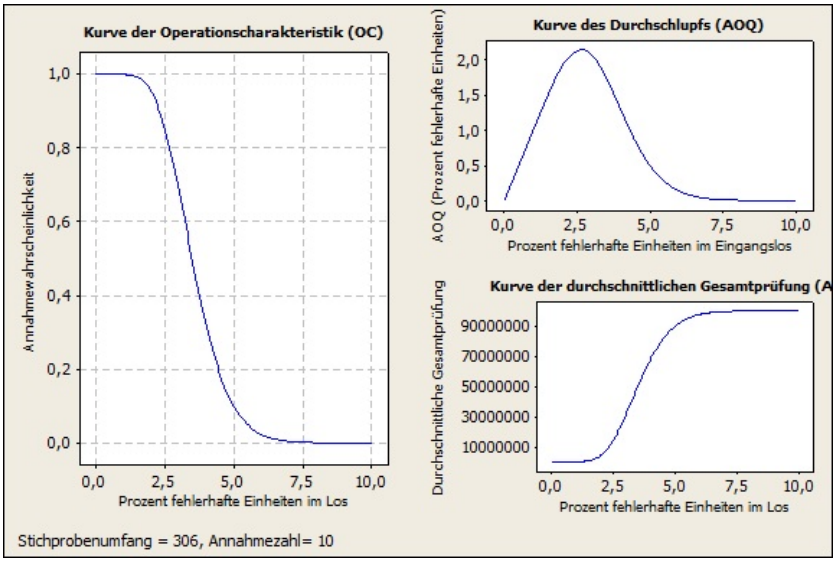


Figure 55: Minitab Annahmeprüfung - Prüfplan

D.h. um die statistischen Vorgaben zu erfüllen, ist als Stichprobengröße $n = 306$ und als max. zulässige Fehleranzahl in der Stichprobe $c = 10$ zu setzen.

Folgende Rechnung kann hierzu auch mit Hilfe der Normalverteilung durchgeführt werden:

Es müssen folgende Gleichungen erfüllt sein:

$$\begin{aligned} \alpha = 0.05 \quad &: \quad Pr(X < c|\pi = 0.02) = 0.95 \\ \beta = 0.10 \quad &: \quad Pr(X < c|\pi = 0.05) = 0.10 \end{aligned}$$

Wobei X die Anzahl der fehlerhaften Teile in der Probe ist. Der Schwellenwert c heißt Annahmekennzahl. Sei nun angenommen, dass Losmenge und auch Stichprobenumfang relativ groß sind, so dass man mit der Normalverteilung als Approximation der Hypergeometrischen Verteilung rechnen kann. Dann ist:

$$\begin{aligned} \text{Liefermengeist'gut'} \quad &: \quad X \sim N(n \cdot 0.02; n \cdot 0.02 \cdot 0.98) \\ \text{Liefermengeist'schlecht'} \quad &: \quad X \sim N(n \cdot 0.05; n \cdot 0.05 \cdot 0.95) \end{aligned}$$

Das ergibt zwei Gleichungen mit zwei Unbekannten.

Für $\alpha = 0.05$ folgt:

$$\Phi\left(\frac{c - n \cdot 0.02}{\sqrt{n \cdot 0.02 \cdot 0.98}}\right) = 0.95 \Rightarrow \frac{c - n \cdot 0.02}{\sqrt{n \cdot 0.0196}} = 1,645$$

Für $\beta = 0.10$ folgt:

$$\Phi\left(\frac{c - n \cdot 0.05}{\sqrt{n \cdot 0.05 \cdot 0.95}}\right) = 0.10 \Rightarrow \frac{c - n \cdot 0.05}{\sqrt{n \cdot 0.0475}} = -1,282$$

Die Lösung dieses Gleichungssystems ist $n = 288,66$ und $c = 9,68$, d.h. die Stichprobengröße ist etwas kleiner geworden durch die Normal-Approximation.

8.2 Durchschlupf

Wenn ein Los angenommen wird (und nur dann), können fehlerhafte Teile in die Produktion 'schlüpfen'. Das ist ein Risiko für den Kunden. Der Anteil dieser eigentlich unbrauchbaren Teile wird dann als Durchschlupf bezeichnet. Mit Hilfe des definierten Tests kann dieser Schlupf berechnet werden (genauer: dessen Erwartungswert).

Um diesen exakt berechnen zu können, muss man definieren, wie man mit den aus dem Los genommenen, für die Prüfung benötigten Anteil umgeht. Die folgenden drei Möglichkeiten werden hierfür betrachtet:

S1: der Prüfanteil der Stichprobe wird vollständig weggelassen

S2: aus dem Prüfanteil der Stichprobe werden die fehlerhaften Teile aussortiert

S3: aus dem Prüfanteil der Stichprobe werden die fehlerhaften Teile durch 'gute' Teile ersetzt

Zudem wird definiert, was passiert, wenn ein Los als 'schlecht' identifiziert wurde. Das Los kann dann folgendermaßen behandelt werden:

L1: das Los wird vollständig weggelassen

L2: das Los wird vollständig geprüft, die fehlerhaften Teile werden aussortiert

L3: das Los wird vollständig geprüft, die fehlerhaften Teile werden durch 'gute' Teile ersetzt

Der Durchschlupf ist ein Erwartungswert, der sich aufgrund der erwarteten Fehlerquote ergibt. Durchschlupf passiert nur, wenn ein Los angenommen wird. Dann ist die zu erwartende Quote mit Hilfe der Operationscharakteristik berechenbar. Der Durchschlupf wird dabei in Prozent angegeben, und zwar dann mit Hilfe des Quotient der Anzahlen der fehlerhaften Teile zur Gesamtanzahl der Teile, die im Los sind.

Der mittlere (d.h. der erwartete) Durchschlupf heißt auch 'Average Outgoing Quality', kurz: AOQ.

Es gilt folgende allgemeine Überlegung:

Es seien Y_1 die Anzahl der durchschlüpfenden fehlerhaften Einheiten, Y_2 die Losgröße, dann ist

$$AOQ = AOQ(p) = \frac{E(Y_1)}{E(Y_2)}$$

Durchschlupf existiert nur, wenn das Los angenommen wird, d.h.:

$$Y_1 = \begin{cases} 0 \\ (N-n)\cdot p \end{cases} \text{ falls } \begin{matrix} \text{AnzahlFehler} > c \\ \text{AnzahlFehler} \leq c \end{matrix}$$

Dann ist weiter:

$$P(Y_1 = 0) = 1 - OC(p), \quad P(Y_1 = (N-n)\cdot p) = OC(p)$$

Woraus folgt:

$$E(Y_1) = 0 \cdot (1 - OC(p)) + (N-n)\cdot p \cdot OC(p) = (N-n)\cdot p \cdot OC(p)$$

Y_2 ist abhängig davon, wie mit dem Los bzw. der Stichprobe verfahren wird. Wird z.B. die Stichprobe komplett nicht weiter verwendet und sobald die Anzahl

fehlerhafter Teile größer ist als c, dann wird das Los abgelehnt und nicht (also nichts aus dem Los) verwendet, dann folgt für diese Situation:

$$Y_2 = \begin{cases} 0 \\ (N-n) \end{cases} \text{ falls } \begin{matrix} \text{AnzahlFehler} > c \\ \text{AnzahlFehler} \leq c \end{matrix}$$

Nun muss noch der Erwartungswert von Y_2 berechnet werden:

$$E(Y_2) = 0 \cdot (1 - OC(p)) + (N-n) \cdot OC(p) = (N-n) \cdot (OC(p)$$

Somit hat man Zähler und Nenner für AOQ bestimmt, es folgt:

$$AOQ(p) = \frac{(N-n) \cdot p \cdot OC(p)}{(N-n) \cdot OC(p)} = p$$

Mit einer anderen Verfahrensweise für die Stichprobe oder das Los insg. bekommt man nun andere Ergebnisse, für die Verfahren S1, S2, S3 und L1, L2, L3 errechnet man folgende Matrix:

	L_1	L_2	L_3
S_1	$AOQ(p) = p$	$\frac{p \cdot OC(p)}{1-p(1-OC(p))}$	$p \cdot CO(p)$
S_2	$\frac{(N-n) \cdot p \cdot OC(p)}{(N-n) \cdot OC(p) + n \cdot (1-p)}$	$\frac{(N-n) \cdot p \cdot OC(p)}{N - np - (N-n) \cdot p \cdot (1-OC(p))}$	$\frac{(N-n) \cdot p \cdot OC(p)}{N - n \cdot p}$
S_3	$\frac{(N-n) \cdot p \cdot OC(p)}{(N-n) \cdot OC(p) + n}$	$\frac{(N-n) \cdot p \cdot OC(p)}{N - (N-n) \cdot p(1-OC(p))}$	$\frac{N_n}{N} \cdot p \cdot OC(p)$

8.3 Aufwand

Analog zum Durchschlupf kann auch der Prüfaufwand, also die Anzahl der erwarteten vorzunehmenden Prüfungen, berechnet werden.

Der erwartete Aufwand heißt auch Average Total Inspection.

Sei Y_3 die nach gegebenen Prüfplan zu prüfenden Anzahl von Teilen, dann ist:

$$ATI = E(Y_3), AFI := \frac{ATI}{N}$$

ATI bedeutet Average Fraction Inspected (der relative mittlere Prüfaufwand).

Auch der Aufwand für die durchzuführenden Prüfungen wird nach Verfahren für die Stichprobe bzw. das Los berechnet, es folgt:

	L_1	L_2	L_3
S_1	$ATI(p) = n$	$N - (N-n) \cdot OC(p)$	$n(1-p) + (N-n) \cdot \frac{(1-OC(p))}{1-p}$
S_2	n	$N - (N-n) \cdot OC(p)$	$n(1-p) + (N-n) \cdot \frac{(1-OC(p))}{1-p}$
S_3	$\frac{n}{(1-p)}$	$\frac{N-(N-n)(OC(p)+p-p \cdot OC(p))}{1-p}$	$\frac{n+(N-n) \cdot (1-OC(p))}{1-p}$

Der Aufwand ist bis auf wenige Ausnahmen stochastisch und dann direkt mit der $OC(p)$-Kurve verbunden.

8.4 Der Algorithmus von Guenther

Wie hier zu sehen ist, definieren die zu wählenden Parameter (AQL, α) und (RQL, β) jeweils Koordinatenpaare im Graphen der OC-Kurve. Die OC-Kurve wird dann im Idealfall genau durch diese Punkte hindurch gelegt, um die statistischen Anforderungen an den Annahmetest zu erfüllen (das entspricht der Berechnung für die Normalverteilung, vgl. 8.1), und liefert dann die beiden Ergebnisgrößen Stichprobenumfang und max. zulässige Fehleranzahl in der Stichprobe, n, c).

Um einen Prüfplan ohne Minitab auszurechnen kann folgender Algorithmus von Guenther verwendet werden:

1. Setze $c = 0, n = 1$
2. Berechne $\tilde{\beta} = OC(RQL)$
3. Falls $\beta \leq \tilde{\beta}$, dann: $n = n + 1$, gehe zu 2.
4. Berechne $1 - \tilde{\alpha} = OC(AQL)$
5. Falls $1 - \tilde{\alpha} \leq 1 - \alpha$, dann: $c = c + 1$, $n = c$, gehe zu 2.

Man kann zeigen: dieser Algorihthmus ist insofern optimal, da er zu den Vorgaben automatisch den Prüfplan mit kleinstem Aufwand errechnet.

Als Beispiel(-ergebnis) für einen Prüfplan ohne Zurücklegen sei angegeben:

Ist AQL = 5%, RQL = 10%, $\alpha = 5\%$, $\beta = 10\%$, dann folgt: $n = 29$, $c = 3$, $1 - \alpha = 97.60\%$, $\beta = 9.926\%$.

Offensichtlich kann der Prüfprozess abgekürzt werden, z.B. bei folgenden Gegebenheiten:

- sind $k < n$ Teile geprüft worden und die Anzahl der Fehler ist größer als c, dann wird das Los vorzeitig abgelehnt
- sind k Teile geprüft worden und es gilt: $n - k+$ Anzahl Fehler ist kleiner als c, dann wird das Los vorzeitig angenommen

8.5 Zweistufiger Prüfplan

Ziel ist es nun, den Prüfaufwand zu reduzieren. Die Vorgaben sollen in jedem Fall beibehalten werden, also AQL, RQL, α und β sollen nicht verändert werden (dürfen).

Da c zur Stichprobengröße gehört, kann nur die Stichprobengröße verändert werden.

Die Idee ist jetzt folgende:

Man zieht $n_1 < n$ Teile und prüft diese.

Ist die gefundene Fehleranzahl in dieser kleinen Stichprobe sehr klein ($\leq c_1$), so kann das Los (mit einer gewissen Wahrscheinlichkeit) als 'gut' identifiziert und dann frühzeitig angenommen werden.

Ist die gefundene Fehleranzahl in dieser kleinen Stichprobe sehr gross ($\geq d_1$), so kann das Los (mit einer gewissen Wahrscheinlichkeit) als 'schlecht' identifiziert

und dann frühzeitig abgelehnt werden.

Ist die gefundene Fehlerzahl zwischen diesen beiden Frühindikatorenwerten, also $c_1 <$ Anzahl Fehler $< d_1$, dann ist die Lage noch unklar, es wird dann eine zweite Stichprobe des Umfanges n_2 zur Klärung der Situation gezogen.

Für diese 'Nachprüfung' selbst gibt es dann keine weitere max. zulässige Fehlerzahl; bei der Entscheidung, ob das Los nach der zweiten Prüfung angenommen oder abgelehnt wird, ist die Gesamtzahl der Fehler zu berücksichtigen (und nicht nur die in der zweiten Prüfung gefundene Anzahl). Insofern ist eine Annahmezahl als Anzahl max. Fehler aus beiden Prüfungen definiert, diese sei wieder mit c bezeichnet. Es sei dann $c_2 := c - c_1$.

Folgende Relationen sind klar:

$$0 \leq c_1 < d_1 + 1 \leq c_2 < n_1 < n_1 + n_2$$

Zusätzlich gelten:

$$\frac{c_1}{n_1} < \frac{c}{n_1 + n_2}, \quad \frac{c+1}{n_1 + n_2} < \frac{d_1}{n_1}$$

Ob so konstruierte Prüfpläne tatsächlich dann einen geringeren Aufwand bei gleichbleibender statistischer Sicherheit haben, ist aufzuzeigen über die Operationscharakteristik und die Größen AOQ und ATI bzw. AFI.

Der Prozessablauf für eine zweistufige Prüfung ist:

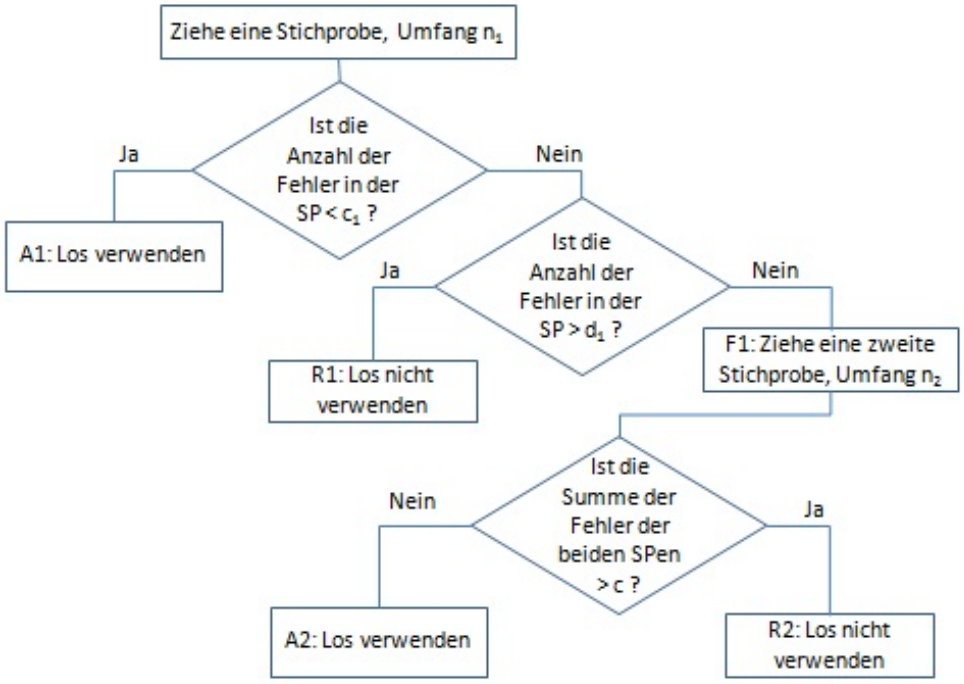


Figure 56: Zweistufiger Prüfplan

Zunächst gilt folgender Zusammenhang für die Wahrscheinlichkeiten des Erreichens der einzelnen Stufen A1, A2, R1, R2 und F1:

$$P(A1) + P(A2) + P(R1) + P(R2) = P(A1) + P(R1) + P(F1) = 1$$

Die folgende Rechnung ermittelt die Operationscharakteristik des zweistufigen Prüfplanes:

Die Operationscharakteristik gilt für die Annahme des Loses:

$$OC(p) = P(A1) + P(A2)$$

$P(A1)$ ist analog der Operationscharakteristik des einstufigen Prüfplanes anzusetzen (siehe Abschnitt 19.2).

Für $P(A2)$ folgt:

$$P(A2) = P(c_1 < \#F(n_1) < d_1 \wedge \#F(n_1 + n_2) \leq c)$$

Hierbei ist mit $\#F(n)$ die Anzahl der in der Stichprobe vom Umfang n gefundenen Fehler bezeichnet.

Die Ereignisse sind diskret (die Binomialverteilung wird hier angesetzt), die Fehler in den Teilen werden zudem als unabhängig angenommen, dann folgt:

$$\begin{aligned} P(A2) &= \sum_{j=c_1+j}^{d_1-1} P(\#F(n_1) = j \wedge \#F(n_2) \leq c_2 - j) \\ &= \sum_{j=c_1+j}^{d_1-1} P(\#F(n_1) = j) \cdot P(\#F(n_2) \leq c_2 - j) \end{aligned}$$

Die Wahrscheinlichkeiten P in dieser Formel können dann durch die Binomialverteilung $B(n; p)$ ersetzt werden (siehe 19.2), es folgt:

$$OC(p) = G_B(c_1) + \sum_{j=c_1+1}^{d_1-1} B(j; p) \cdot G_B(c_2 - j)$$

8.6 Mittlerer Aufwand eines zweistufigen Prüfplanes

Bei einem einfachen Prüfplan ist der Aufwand durch die Anzahl n der zu prüfenden Teile unmittelbar gegeben bzw. z.B. über den Algorithmus berechenbar. Bei einem zweistufigen Prüfplan ist der Aufwand nun stochastisch, es kann je nach Ergebnis der ersten Prüfung n_1 oder $n_1 + n_2$ sein.

Die Wahrscheinlichkeit, dass der Aufwand n_1 ist, ist gleich der Wahrscheinlichkeit, dass nicht die zweite Prüfung durchgeführt werden muss, ist also $1 - P(F1)$. Analog ist die Wahrscheinlichkeit für den Aufwand $n_1 + n_2$ dann gleich $P(F1)$.

Mit Hilfe dieser Überlegung kann nun der Erwartungswert für den Prüfungsaufwand berechnet werden. Dieser wird mit Average Sample Number, kurz: ASN, beze-

ichnet:

$$ASN = ASN(p) = n_1(1 - P(F1)) + (n_1 + n_2) \cdot P(F1) = n_1 + n_2 \cdot P(F1)$$

Analog können nun noch die Kennzahlen AOQ, ATI und AFI berechnet werden. Das wird hier nicht weiter ausgeführt.

8.7 Äquivalente Prüfpläne

Wenn nun auf die so beschriebene Art unterschiedliche Prüfpläne konstruiert werden können, dann stellt sich die Frage, welcher denn nun 'besser' ist. Zunächst ist das Wort 'besser' zu klären: 'besser' wäre in dem Sinn hier ein Prüfplan, wenn der (mittlere, d.h. erwartete) Aufwand geringer ist.

Dazu die folgenden Definitionen:

Zwei Prüfpläne heißen äquivalent, wenn die Operationscharakteristik beider Prüfpläne gleich ist.

Dass die Operationscharakteristik zweier Prüfpläne exakt gleich ist, ist selten der Fall; darum wird folgendes definiert:

Zwei Prüfpläne heißen approximativ äquivalent, wenn sie in den Punkten $(AQL, 1-\alpha)$ und (RQL, β) übereinstimmen.

Und schliesslich:

Ein Prüfplan PP1, der zu einem zweiten Prüfplan PP2 approximativ äquivalent ist, heißt effizienter, wenn der mittlere Prüfaufwand geringer ist.

Als Effizienzmass wird definiert:

$$\omega = \frac{ASNL}{n}, ASNL := \max ASN(p), \quad \text{max.Aufwand}$$

Beispiel:

Der einfache Prüfplan $n = 136$, $c = 7$ ist zu dem zweistufigen Prüfplan $n_1 = 50$, $c_1 = 1$, $d_1 = 5$, $n_2 = 100$, $c_2 = 7$ approximativ äquivalent. Für beide Prüfpläne ist $(AQL, 1 - \alpha) = (0.027, 0.95)$, $(RQL, \beta) = (0.085, 0.1)$.

8.8 Sequentieller Prüfplan

Die Idee durch eine Aufsplittung der Stichprobenprüfung in mehrere Teilprüfungen, um so den Aufwand zu reduzieren, führt in seiner extremsten Ausprägung dazu, dass jede Prüfstufe nur noch vom Umfang 1 ist. Dieses Verfahren ist dann der sog. sequentielle Prüfplan.

Die Ergebnisse der sequentiellen Prüfungen für die Teile liefern eine Bernouilli-Verteilung, also eine Sequenz aus 1 und 0. Zu prüfen ist nun, wann das Los angenommen oder abgelehnt werden kann bzw. ob weiter geprüft werden muss.

Es wird hierzu ein Hypothesentest aufgebaut mit den folgenden Hypothesen:

$$\begin{aligned} H0 &: p = p_0 \\ H1 &: p = p_1 \end{aligned}$$

(Die beiden Schranken p_0 und p_1 werden dann für die Annahmeprüfung durch AQL und RQL ersetzt.)

Die Stichprobe liefert eine Abfolge $x_1, \ldots, x_n$, wobei x_i entweder eine 1 oder eine 0 ist. Die 1 wird dabei und hier im Folgenden im Falle eines Fehlers angenommen.

Zu überprüfen ist nun, ob eine der beiden Quoten p_0 bzw. p_1 angenommen werden darf. Also etwas präziser: ob eine der beiden Quoten wahrscheinlich genug ist, um diese akzeptieren zu können.

Hierzu wird die Likelihood-Funktion als Ausdruck für diese Wahrscheinlichkeit definiert, die fehlerhaften bzw. OK-Teile sind ja bereits entsprechend kodiert:

$$L(p) := \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

Um die Wahrscheinlichkeit des Zutreffens von p_0 bzw. p_1 miteinander zu vergleichen wird der Quotient aus den beiden Wahrscheinlichkeiten gebildet.

$$LQ_n := \frac{L(p_1}{L(p_0)}$$

Dieser Quotient ist die Prüfgröße eines Hypothesentests (des sog. LQ-Test), der grundsätzlich wie folgt funktioniert:

- H_0 wird verworfen, wenn $LQ_n \ll 1$
- H_1 wird angenommen, wenn $LQ_n \gg 1$

Zum Test werden nun die Wahrscheinlichkeiten für den Fehler erster und zweiter Art betrachtet:

Sei α die Fehlerwahrscheinlichkeit erster Art, β die Fehlerwahrscheinlichkeit zweiter Art, und weiter:

$$A := \frac{\beta}{1-\alpha}, \quad B := \frac{1-\beta}{\alpha}$$

Mit Hilfe dieser Größen ist der Test dann entscheidbar:

- H_0 wird angenommen, wenn $LQ_n \leq A$
- H_1 wird angenommen, wenn $LQ_n \geq B$

Das bedeutet konkret:

- ist $LQ_n \geq B$, dann wird H_0 verworfen
- ist $LQ_n \leq A$, dann wird H_0 angenommen

- ist $LQ_n \in (A, B)$, dann wird eine weitere Stichprobe gezogen

Anm.: ist $\alpha + \beta < 1$, dann ist $A < B$.

Für die praktische Durchführung werden diese Relationen noch umgerechnet:

Sei nun k die Anzahl defekter Teile aus n Stichproben. Dann ist:

$$LQ_n = \frac{L(P_1}{L(P_0)} = \frac{\prod_{i=1}^{n} p_1^{x_i}(1-p_1)^{1-x_i}}{\prod_{i=1}^{n} p_0^{x_i}(1-p_0)^{1-x_i}} = \frac{p_1^k(1-p_1)^{n-k}}{p_0^k(1-p_0)^{n-k}} = \left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)^k \cdot \left(\frac{1-p_1}{1-p_0}\right)^n$$

Weitergezogen werden muss, wenn $A < LQ_n < B$; da $\ln$ eine monoton wachsende Funktion ist, folgt daraus:

$$\ln(A) < \ln(LQ_n) < \ln(B)$$

Und somit:

$$\ln(A) < k \cdot \ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right) + n \cdot \ln\left(\frac{1-p_1}{1-p_0}\right) < \ln(B)$$

Die interessante Größe hier ist k, denn die Anzahl der gefundenen Fehler steuert den Test und damit die Entscheidung; nach k wird nun umgestellt, es folgt:

$$\frac{\ln(A) - n \cdot \ln\left(\frac{1-p_1}{1-p_0}\right)}{\ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)} < k < \frac{\ln(B) - n \cdot \ln\left(\frac{1-p_1}{1-p_0}\right)}{\ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)}$$

Für einen konkreten Prüfplan sind p_0 durch AQL und p_1 durch RQL zu ersetzen.

Man definiert nun noch die in der obigen Formel durch die Vorgaben konstanten Anteile:

$$a := \frac{\ln(A)}{\ln\left(\frac{RQL(1-AQL)}{AQL(1-RQL)}\right)}, \quad b := \frac{\ln(B)}{\ln\left(\frac{RQL(1-AQL)}{AQL(1-RQL)}\right)}, \quad c := \frac{\ln\left(\frac{1-AQL}{1-RQL}\right)}{\ln\left(\frac{RQL(1-AQL)}{AQL(1-RQL)}\right)}$$

Der Test verläuft nun wie folgt:

Bestimme für eine Stichprobe vom Umfang n das Intervall:

$$I_n := (a + c \cdot n, b + c \cdot n)$$

Anm.: in Bezug auf n ist I_n ein Bereich, der durch zwei Geraden begrenzt wird, die Steigung der Geraden ist c, die absoluten Glieder der Geraden sind a und b.

Sei k die Anzahl der Defekte in der Stichprobe. Dann wird folgendermaßen entschieden:

- Ist $k < a+c\cdot n$, dann ist H_0 anzunehmen (d.h. das Los wird angenommen)
- Ist $k > b + c \cdot n$, dann ist H_1 anzunehmen (d.h. das Los wird abgelehnt)
- Ansonsten: ziehe erneut

Beispiel:

Zu den in 8.7 gegebenen approximativ äquivalenten Prüfplänen ist der sequentielle Prüfplan wie folgt zu definieren:

$$\begin{aligned} \alpha &= 0.05, \beta = 0.1 \\ A &= \frac{\beta}{1-\alpha} = \frac{0.1}{1-0.05} = 0.105263158 \\ B &= \frac{1-\beta}{\alpha} = 18 \\ a &= -1.863228903, b = 2.392148456, c = 0.050865943 \end{aligned}$$

Das Intervall I_n zu diesem Prüfplan ist dann folgendermaßen:

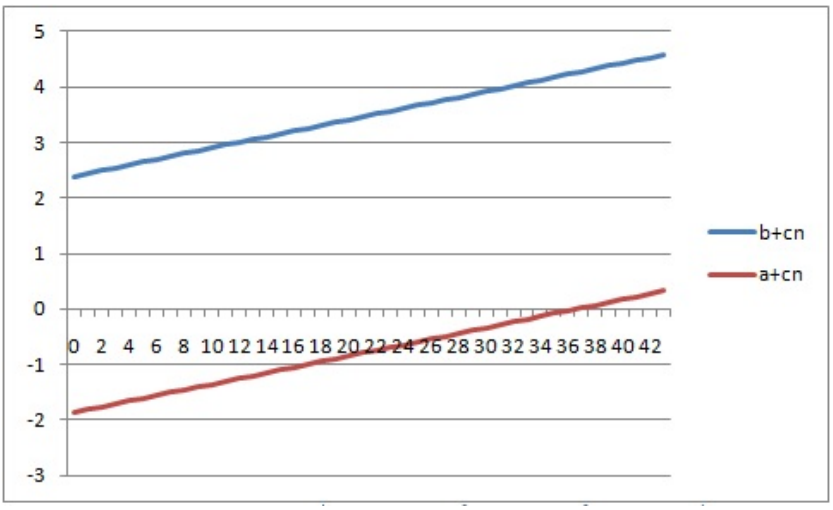


Figure 57: Intervall I_n zum sequentiellen Prüfplan

Mit Hilfe dieser Werte kann nun die Frage beantwortet werden, wann das Los schnellstmöglich angenommen bzw. abgelehnt werden kann.

Eine Annahme ist schnellstens dann möglich, wenn keine Fehler gefunden werden. Dann wird das Intervall I_n nach unten verlassen und H_0 wird angenommen.

Die Gleichung zur Berechnung lautet:

$$a + c \cdot n = 0 \Rightarrow n = -\frac{a}{c} = 36,63018515$$

Analog für den Fall der schnellstmöglichen Ablehnung des Loses, dann werden in jeder gezogenen Stichprobe Defekte gefunden, die Gleichung dazu lautet:

$$b + c \cdot n = n \Rightarrow n = \frac{b}{1-c} = 2,520348351$$

Nach frühestens 37 Prüfungen kann also ein Los angenommen werden, nach frühestens 3 Prüfungen kann ein Los abgelehnt werden.

9 Regressionen für stetige Antwortvariablen

Regressionen sind im Wesentlichen mathematische Ursache-Wirkungs-Analysen, oder allgemeiner, Zusammenhanganalysen. Diese werden in den Standard-Vorlesungen der Statistik detailliert erläutert; hier werden die Verfahren nur nochmal übersichtlich dargestellt und es werden Verfahren vorgestellt, die auf Regressionsanalysen aufbauen.

Regressionen 'erklären' Zusammenhänge, d.h. Ziel ist es, ein gewisses Verhalten, z.B. Werte bei Zielgrößen, auf Einflüsse zurückzuführen. Insb. sollen so die Zielgrößen 'angesteuert', d.h. im Prozess erreicht werden können. Es wird dann berechnet, mit welchen Einstellungen das Ziel erreicht werden kann (Zielgrößenoptimierung). Diese Zielsetzung wird hier im weiteren bei gezielten Untersuchungen des Qualitätsmanagements, der Messsystemanalyse und dem Design of Experiments, verfolgt und erläutert.

Zu berücksichtigen ist noch, dass bei so durch Optimierung gefunden Einstellungen die Zielwerte nicht exakt erreicht werden, sondern abweichen und 'schwanken' (sich sozusagen in einem gewissen Masse nicht erklärbar zufällig verhalten).

Ziel ist es also immer, einen funktionalen Zusammenhang (funktional bedeutet: eine abhängige zu jeder unabhängigen Variablen) und die Abweichung davon bestmöglich und plausibel zu berechnen. Geometrisch gesprochen wird also zu einem Vektor $\vec{x}$ dann ein 'bestmögliches' y gesucht (die Formulierung ist auch für einen Vektor als y möglich, darauf wird hier jedoch zunächst verzichtet). Hinzu kommt dann ein Term, der die Abweichung von dieser idealen Formel beschreibt. Das ist das sog. Residuum als Beschreibung der Differenz zwischen der Regressionsformel und den tatsächlichen Messwerten:

$$y = f(\vec{x}) + \epsilon(\vec{x})$$

Anm.: i.d.R. wird das Residuum als Verteilung unabhängig von der Variablen $\vec{x}$ angegeben, also berechnet bzw. geschätzt. Grundlage ist dann im wesentlichen ein Verteilungstest (vgl. Abschn. 5.4.1), der das Residuum als stochastische Funktion bestimmt. Als einfachster Ansatz wird dann die Normalverteilung gesetzt und die Regression ist solange gültig, solange es keinen Widerspruch (in Form einer Signifikanz) zu dieser Annahme gibt.

Im Abschnitt 2.2 ist auf die unterschiedlichen Merkmalstypen eingegangen worden, um Qualität von Produkten vollumfänglich prinzipiell zu beschreiben. Insofern muss ein Analyst, der sich mit QM-Problematiken aller Art auseinandersetzt, mit allen Merkmalskombinationen für y und die $\vec{x}$ 'rechnen', also umgehen können.

Eine Ursache-Wirkungs-Analyse kann also Zusammenhänge zwischen Merkmalen aufzeigen. Diese Merkmale können unterschiedlich ausgeprägt sein. Analysen sind denkbar, um Gründe für die Produktion von Ausschuss o.ä. zu finden.

Da es für die in Frage kommenden Unabhängigen-Vektorkomponenten von $\vec{x}$ und auch für ein y je vier Merkmalsausprägungen gibt, sind im einfachsten Fall insg. 16 Kombinationen möglich. Diese Detaillierung wird jedoch nicht benötigt. Die Darstellung hier wird nun auf die zwei Oberkategorien 'attributiv' und 'metrisch' verkürzt dargestellt. Das folgende Schaubild soll nun aufzeigen, welche Analysen (Algorithmen, Methoden, Verfahren) in einer gegebenen Situation auszuwählen sind:

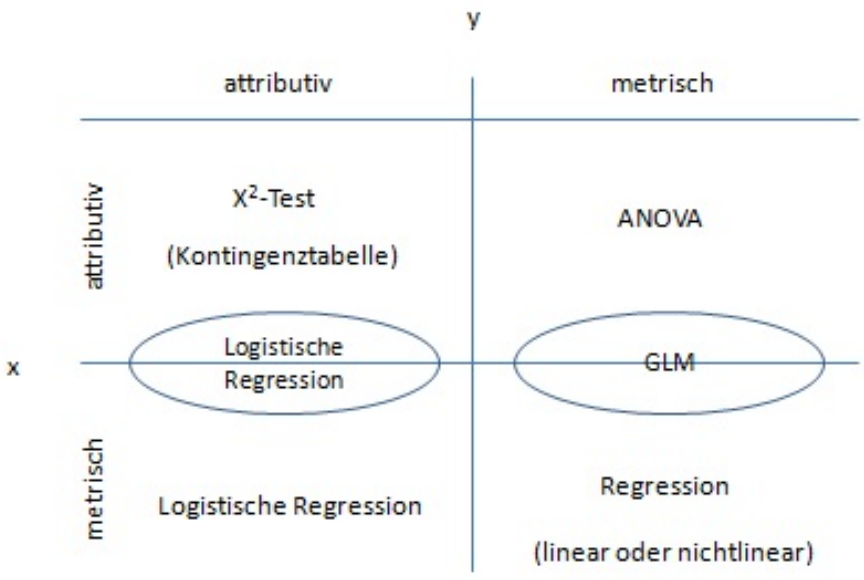


Figure 58: Regressionen - 4+2-Schema

Die Verfahren in den Ellipsen sind anzuwählen, wenn die $\vec{x}$-Merkmale gemischt sind, es also mind. ein attributives und mind. ein metrisches Merkmal vorkommt. Die Ausgabe ist dann zweigeteilt für die jeweils verbundenen Methoden.

9.1 Varianzanalyse, ANOVA

9.1.1 Die Grundidee der Varianzanalyse

Ziel der (univariaten) Varianzanalyse oder engl.: Analysis of Variance (kurz: ANOVA) ist es, die Varianz der Zielgröße (y, stetig) bzgl. einzelner Faktoren ($\vec{x}$, nominal) zu erklären. Das geschieht dadurch, dass ein multipler t-Test durchgeführt. Die Formulierung der zugehörigen Hypothesen ist:

H_0: die Mittelwerte der Messungen zu den einzelnen Faktoren sind nicht signifikant unterschiedlich

H_1: mind. ein Mittelwert ist zu mind. einem anderen Mittelwerten der Messungen zu den einzelnen Faktoren signifikant unterschiedlich

Die Grundlage der Analyse ist der sog. Verschiebungssatz; in der abstrakten Form wird er folgendermaßen geschrieben:

$$Var(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

oder in der Formulierung, wie sie für konkrete Daten besser geeignet ist:

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

Für die Formulierung mit Faktoren ist die Variable x mit einem Lauf- und einem Faktor-Index zu versehen. Dann folgt:

$$\begin{aligned} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{\bar{x}})^2 &= \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{\bar{x}})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \\ SST &= SSB + SSE \end{aligned}$$

Anm.: dividiert man diese letzte Gleichung durch SST, dann steht links eine 1, oder anders formuliert: 100%. Die Anteile SSB/SST und SSE/SST sind dann als prozentuale Aufteilung der 100% Streuung auf die innere und die äußere Streuung zu lesen.

Hierbei ist i ist der Index für einen Faktor, j ist der laufende Index der Messungen je Faktor, $\bar{x}_i$ der Mittelwert der Daten zum Faktor i, $\bar{\bar{x}}$ der Mittelwert über alle Daten, unabhängig von den Faktoren. SST bedeutet 'Sum of Squares Total', SSB Sum of Squares Between (Variation zwischen den Gruppen) und SSE Sum of Squares Error (Within, Variation innerhalb der Gruppen).

Für den Hypothesentest der ANOVA werden folgende Werte berechnet:

$$MS_W = \frac{SSE}{N-k}, \quad MS_B = \frac{SSB}{k-1}$$

Der Quotient $\frac{MS_B}{MS_W}$ ist nun ein Maß dafür, wie weit die Mittelwerte nun voneinander entfernt sind.

Ist die Variation zwischen den Gruppen erheblich größer als die Variation innerhalb der Gruppen, dann sind die Lagen der Mittelwerte der Gruppen deutlich (signifikant ...) voneinander verschieden; und dann ist der Quotient F sehr groß

Die Nullhypothese wird nun verworfen, wenn F größer ist als das $1-\alpha$-Quantil einer F-Verteilung mit den Freiheitsgraden $(k-1)$ und $(N-k)$ (einseitiger Test).

Die Voraussetzungen für die ANOVA-Analyse sind:

- die Daten für jeden Faktor folgen einer Normalverteilung
- Varianzhomogenität (die Varianzen sind annähernd gleich groß)

Sind diese Voraussetzungen nicht erfüllt, dann muss zur Analyse ein nichtparametrisches Verfahren Anwendung finden.

Anm.: in der ANOVA und allen darauf basierenden Verfahren (z.B. GLM) wird typischerweise die Entscheidung bzw. Berechnung des p-Wertes anhand der in-

dividuellen Varianzen der Faktoren getroffen, sondern anhand der sog. 'zusammengefassten Standardabweichung'. Diese ist nichts anderes als der Mittelwert der Standardabweichungen der Werte zu den Faktoren. Durch diese Vereinfachung können Analyseergebnisse ggf. verfälscht werden.

Die Voraussetzung der Normalverteilung wird auch anhand des Residuums geprüft (also eigentlich 'zu spät', nämlich nach der Berechnung des Modells). So wird dann nachträglich das berechnete Modell als sinnvoll bestätigt - oder eben verworfen. Die Residuenwerte müssen also konkret gesprochen einen Test auf Normalverteilung bestehen.

9.1.2 Anwendungsbeispiel ANOVA

Gegeben seien nun Daten der folgenden Gestalt:

Faktor	Zielgrösse
1	18,95
1	12,62
1	11,94
1	14,42
2	10,06
2	7,19
2	7,03
2	14,66
3	10,92
3	13,28
3	14,52
3	12,51
4	10,46
4	21,40
4	18,10
4	22,50

Figure 59: Daten für die ANOVA

Zunächst der Test auf Normalverteilung für die Faktoren. In Minitab kann das am einfachsten mit einem 'mehrfachen Wahrscheinlichkeitsnetz' erfolgen:

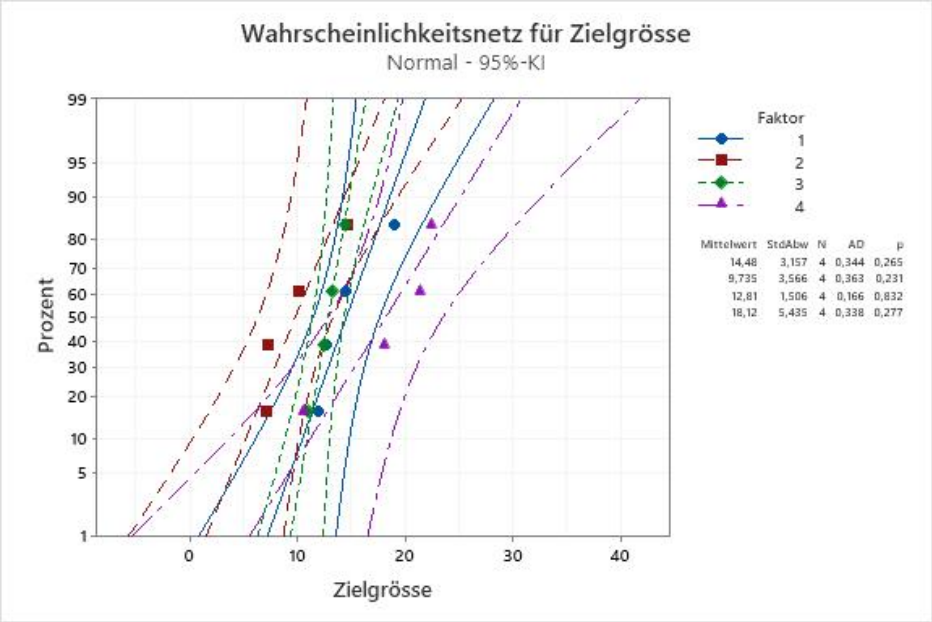
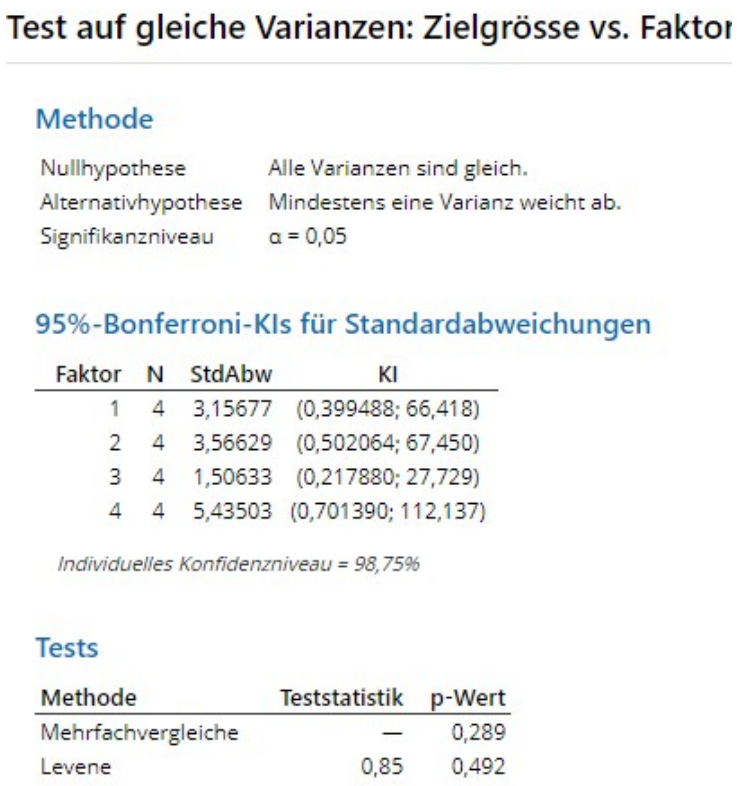


Figure 60: Test auf NV für Faktoren in ANOVA

Alle p-Werte sind größer als 5%, somit ist der Test für alle Faktoren bestanden (siehe Tabelle in der Grafik).

Der Test auf Homogenität der Varianzen liefert:



Test auf gleiche Varianzen: Zielgrösse vs. Faktor

Methode

Nullhypothese	Alle Varianzen sind gleich.
Alternativhypothese	Mindestens eine Varianz weicht ab.
Signifikanzniveau	α = 0,05

95%-Bonferroni-KIs für Standardabweichungen

Faktor	N	StdAbw	KI
1	4	3,15677	(0,399488; 66,418)
2	4	3,56629	(0,502064; 67,450)
3	4	1,50633	(0,217880; 27,729)
4	4	5,43503	(0,701390; 112,137)

Individuelles Konfidenzniveau = 98,75%

Tests

Methode	Teststatistik	p-Wert
Mehrfachvergleiche	—	0,289
Levene	0,85	0,492

Figure 61: F-Test ANOVA Output

D.h. auch die Voraussetzung der Varianz-Homogenität ist gegeben.

Anm.: in diesem Beispiel ist die Stichprobengröße für jeden Faktor lediglich 4. Das ist für einen Test auf Verteilung also auch für einen Test auf gleiche Varianzen eigentlich viel zu klein ... siehe dazu auch oben.

Die ANOVA-Analyse in Minitab ist somit 'erlaubt', diese ergibt folgenden Output:

Einfache ANOVA: Zielgrösse vs. Faktor

Methode

Nullhypothese	Alle Mittelwerte sind gleich.
Alternativhypothese	Nicht alle Mittelwerte sind gleich.
Signifikanzniveau	α = 0,05

Für die Analyse wurde von gleichen Varianzen ausgegangen.

Faktorinformationen

Faktor	Stufen	Werte
Faktor	4	1; 2; 3; 4

Varianzanalyse

Quelle	DF	Kor SS	Kor MS	F-Wert	p-Wert
Faktor	3	146,4	48,79	3,58	0,047
Fehler	12	163,5	13,62		
Gesamt	15	309,9			

Zusammenfassung des Modells

S	R-Qd	R-Qd(kor)	R-Qd(prog)
3,69094	47,24%	34,05%	6,20%

Mittelwerte

Faktor	N	Mittelwert	StdAbw	95%-KI
1	4	14,48	3,16	(10,46; 18,50)
2	4	9,73	3,57	(5,71; 13,76)
3	4	12,807	1,506	(8,787; 16,828)
4	4	18,12	5,44	(14,09; 22,14)

Zusammengefasste StdAbw = 3,69094

Figure 62: Einfache ANOVA

Der p-Wert des F-Tests zeigt eine Signifikanz auf; das bedeutet, die Null-Hypothese (alle Mittelwerte sind 'gleich') muss verworfen werden. Insofern sind mindestens zwei Mittelwerte deutlich verschieden voneinander.

Die einfache Textplotgrafik deutet die Lagen und die Konfidenzintervalle der Mittelwerte auf. Die unterschiedlichen Mittelwertlagen sind zu erkennen. Auch die folgende Grafik mit den Fehlerbalken zeigt das auf:

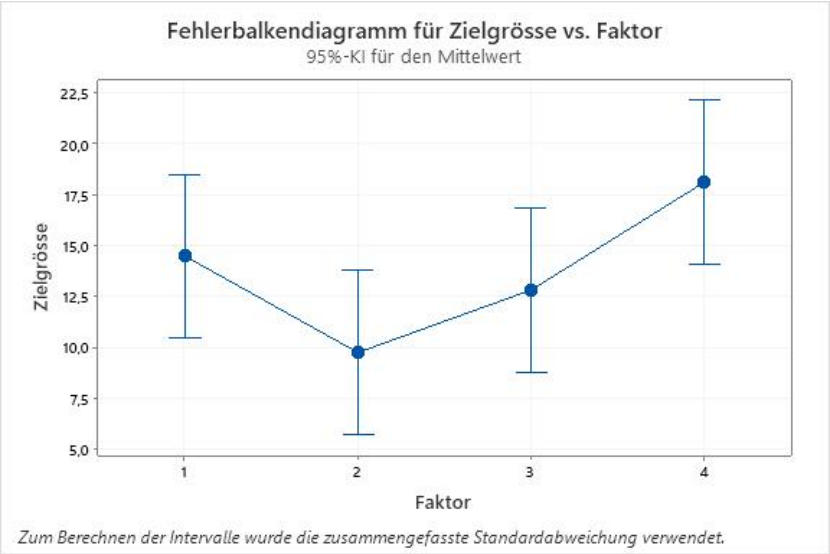


Figure 63: Einfache ANOVA - Fehlerbalken

Die Residuen und die Anpassungen der Werte an das Modell (Anm.: die Anpassungswerte sind die Mittelwerte der Zielgröße je Faktor) sind:

RESI2	ANPASS2
4,4675	14,4825
-1,8625	14,4825
-2,5425	14,4825
-0,0625	14,4825
0,3250	9,7350
-2,5450	9,7350
-2,7050	9,7350
4,9250	9,7350
-1,8875	12,8075
0,4725	12,8075
1,7125	12,8075
-0,2975	12,8075
-7,6550	18,1150
3,2850	18,1150
-0,0150	18,1150
4,3850	18,1150

Figure 64: Residuum und Anpassungen ANOVA

Der Test auf Normalverteilung für das Residuum ergibt folgendes Ergebnis:

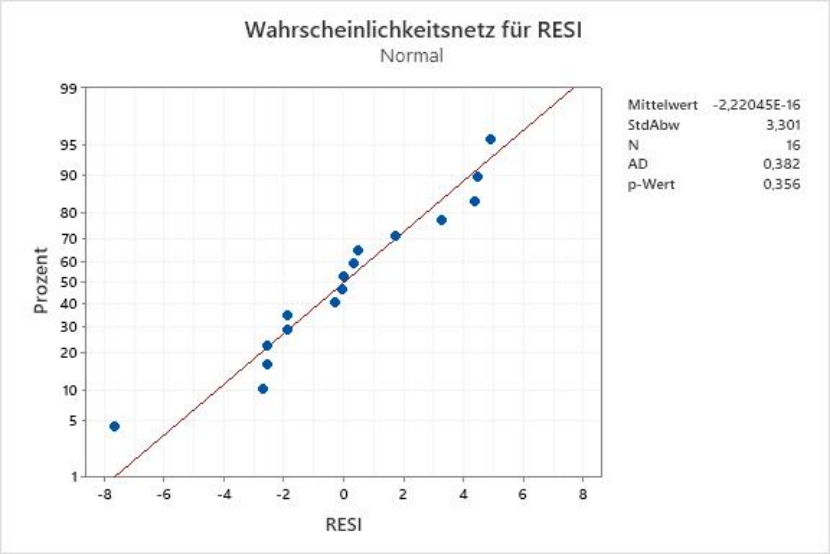


Figure 65: Verteilungsanalyse ANOVA Residuum Gesamt

Der p-Wert der Verteilungsanalyse ist größer als 0.05, insofern ist die Annahme,

das Residuum ist normalverteilt, zulässig. Die Rechnung und das Ergebnis der ANOVA ist dementsprechend gültig.

Zur Analyse des Residuums dient auch die folgende grafische Übersicht:

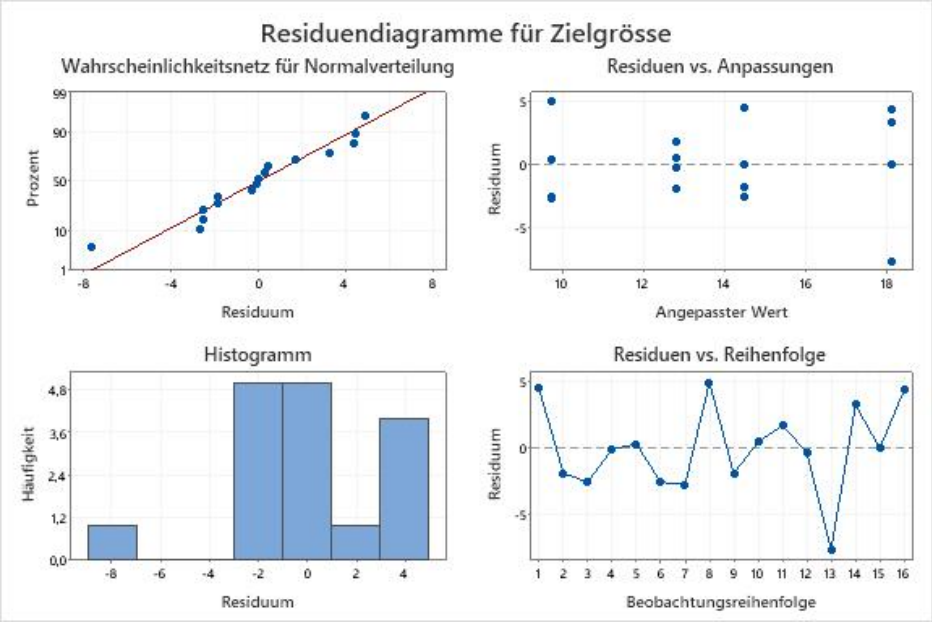


Figure 66: 4 in 1 Diagramm Residuum ANOVA

Der Boxplot gibt Hinweise auf Lage und Verteilung der Messwerte gleichzeitig:

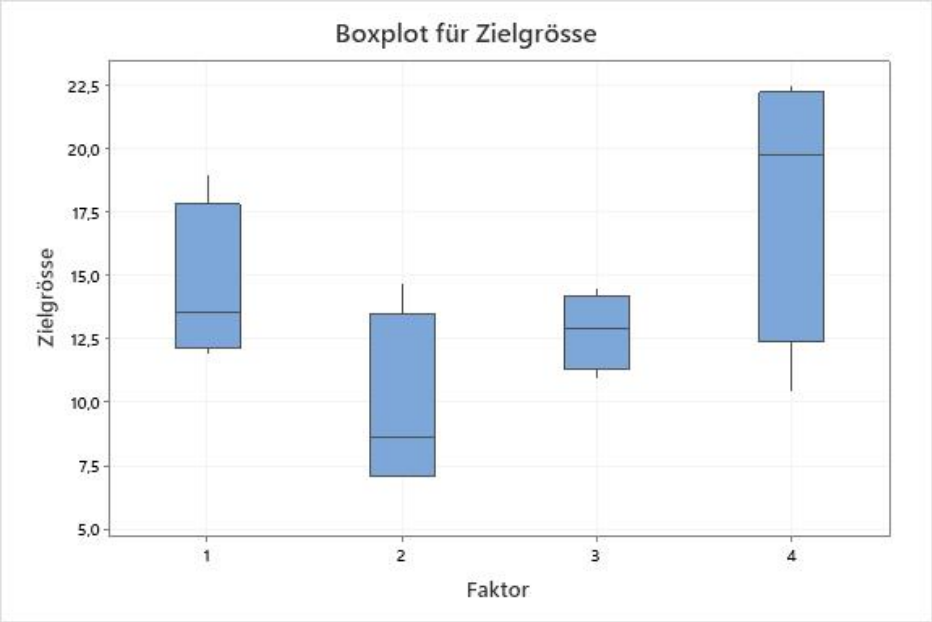


Figure 67: Boxplot ANOVA

Der Intervallplot zeit nochmal separat die Mittelwertverschiebung und die Konfidenzintervalle der Mittelwerte auf:

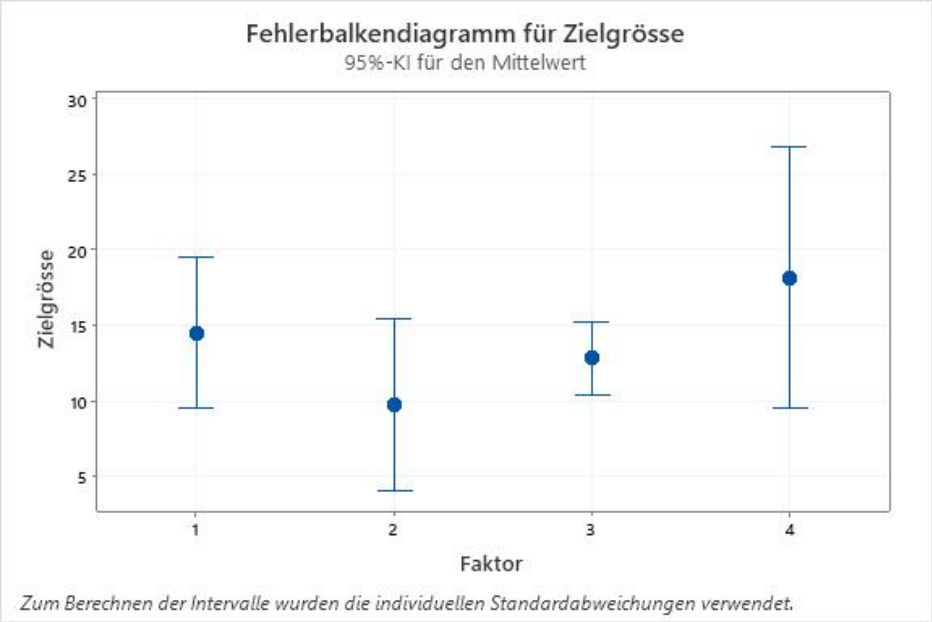


Figure 68: Intervallplot ANOVA

Anhand der Grafik mit den Fehlerbalken (d.h. der Konfidenzintervalle) ist zu sehen, dass das eigentliche Ergebnis signifikant unterschiedlicher Mittelwerte je Faktor aufgrund der kleinen Stichprobe und gleichzeitig der zusammengefassten Standardabweichung als Berechnungsgrundlage die tatsächliche Situation schlecht bewertet.

9.1.3 Trennschärfe der ANOVA

Zum Hypothesentest der ANOVA ist auch die Trennschärfe berechenbar. Diese hängt hier von folgenden Werten ab:

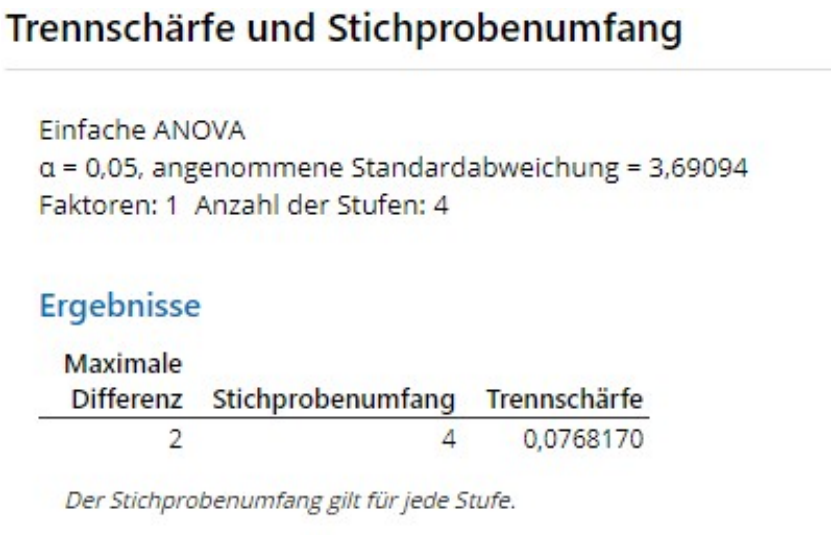
- Anzahl der Faktorstufen
- Stichprobenumfänge (diese müssen für jeden Faktor gleich sein; ein sog. balanciertes Schema)
- Max. Differenz zwischen den Mittelwerten, die festgestellt werden soll

- Standardabweichung

Für das Fallbeispiel hier gilt:

- Anzahl der Faktorstufen: 4
- Stichprobenumfänge: 4
- Max. Differenz zwischen den Mittelwerten, die festgestellt werden soll: wird auf 2 gesetzt
- Standardabweichung (siehe Output): $S = 3,69094$

Das Ergebnis der Berechnung ist:



Trennschärfe und Stichprobenumfang

Einfache ANOVA
α = 0,05, angenommene Standardabweichung = 3,69094
Faktoren: 1 Anzahl der Stufen: 4

Ergebnisse

Maximale Differenz	Stichprobenumfang	Trennschärfe
2	4	0,0768170

Der Stichprobenumfang gilt für jede Stufe.

Figure 69: Trennschärfe ANOVA

Die Trennschärfe ist also sehr gering (als akzeptabler und praxisnaher Wert wird 80% angesehen). Das besagt, dass die Wahrscheinlichkeit für einen Fehler zweiter sehr Art hoch ist. Wenn also ein Unterschied in der Realität bzgl. der Mittelwerte vorhanden ist, dann wird dieser nur zu 7% auch durch den Test gefunden.

Dennoch wird im Test der p-Wert mit $4,7\%$ berechnet, d.h. es wird ein signifikanter Unterschied zwischen den Mittelwerten gefunden. Die Wahrscheinlichkeit, dieses so zu finden, ist für die zu bewertende Situation sehr gering ! Insofern ist die Stichprobengröße in der Praxis bereits ausreichend, um diesen Unterschied zu finden.

Wollte man theoretisch sicher gehen, und berechnet die Stichprobengröße, die notwendig ist, um die 80% Trennschärfe auch zu erreichen, dann bekommt man:

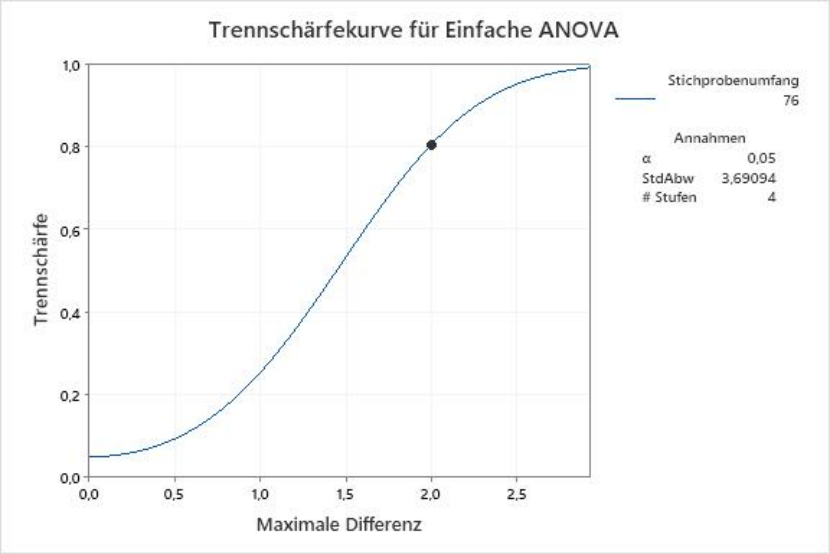


Figure 70: Trennschärfe ANOVA - theoretisch

D.h. die theoretisch geforderte Stichprobengröße ist 76 - je Faktor, also der Gesamtaufwand ist 304 Versuche.

Dennoch muss man die restlichen 288 theoretisch geforderten Versuche nicht durchführen. Die Argumentation ist: erhöht man die Stichprobengröße, dann verkleinern sich die Konfidenzintervalle. Da diese bereits für die Stichprobengröße 4 ausreichend klein sind, um eine Signifikanz auszuweisen, wäre dieses Ergebnis mit (weitaus) mehr Aufwand lediglich noch deutlicher. Insofern reicht die kleine Stichprobe aus, denn der Hypothesentest stellt nur einen Unterschied fest - und berechnet nicht, wie groß dieser ist. Das Ergebnis würde sich also bei dem erheblichen Mehraufwand lediglich bestätigen.

Diese Interpretation ist wichtig insb. für Verfahren, die versuchen mit bereits recht kleinen Stichproben zu Aussagen zu kommen. Lediglich kleine Stichproben zu planen ist im übrigen der Sinn der statistischen Versuchsplanung, die das Ziel hat, mit kleinen Stichproben bereits Signifikanzen zu finden.

Aus diesem Beispiel ist zudem erkennbar, wie man in der Praxis vorgehen sollte: man testet erst eine (relativ kleine) Anzahl von Tests; das insb. auch, um eine erste Schätzung für die Mittelwerte und die Standardabweichungen zu bekommen. Anschließend kann man die Stichprobengröße planen. Man beendet jedoch die Tests, wenn entweder der p-Wert kleiner 5% wird, oder die Anzahl berechneter Stichproben erreicht ist. Ersteres führt zu einer H_1-Bewertung, letzteres zu H_0.

9.2 Mehrfaktorielle ANOVA

Die Idee der einfaktoriellen ANOVA, mittels der Varianzen ggf. Mittelwertsunterschiede mit einem entsprechenden Hypothesentest aufzuzeigen, kann auf mehr als einen Faktor erweitert werden. Dadurch ist es möglich, eine Zielgröße bzgl. mehrerer Faktoren in einer Berechnung zu erklären und ein entsprechendes Modell aufzustellen.

Als abstraktes Modell wird nun eine Erweiterung von (14) angesetzt:

$$y = b + a_1 * x_1 + \cdots + a_m * x_m + \epsilon \tag{14}$$

Die x_1 bis x_m sind die m Faktoren, die nun die Antwortvariable y erklären sollen.

Die Ideen der einfachen ANOVA werden nun auf diese Situation übertragen; somit sind folgende Größen auch für die hier gegebene Situation definiert:

- Koeffizienten
- Effekte
- Anpassungsgrad bzw. Bestimmtheitsmaß bzw. Erklärungsgrad
- Gruppierungen (hier nun auch für Faktorkombinationen)

- Residuum (ein Residuum für die Modellgleichung)
- Standardabweichung (eine)

Das Modell wird analog der einfachen ANOVA anhand der Stichprobenmittelwerten relativ zum Gesamtmittelwert gebildet.

Als Beispiel betrachtet man eine Antwortvariable $Y =$ 'Gehalt', es wird eine Abhängigkeit vermutet von den 3 Faktoren $x_1 =$ 'Region', $x_2 =$ 'Geschlecht' und $x_3 =$ 'Familienstand'.

Anm.: das Gehalt ist vermutlich nicht normalverteilt; das sei hier für ein anschauliches Beispiel mal (fiktiv) angenommen.

$$\left.\begin{array}{l} Gehalt \\ Geschlecht \\ Familienstand \end{array}\right\} \rightarrow Gehalt$$

Es seien zu den hier angegebenen Faktoren jeweils lediglich 2 Werte gegeben, also für 'Region' $\{R_1, R_2\}$, für 'Geschlecht' $\{G_1, G_2\}$ und für 'Familienstand' $\{F_1, F_2\}$ Es seien weiter nun folgende Koeffizienten (korrespondierend zu den 3 Faktoren hier im Beispiel) angenommen:

$$b = 4000, \quad x_{R_1} = -500, \quad x_{G_1} = +200, \quad x_{F_1} = +0,5 \tag{15}$$

Da jeder Faktor in diesem Beispiel lediglich 2 mögliche Werte besitzt, ergeben sich die Faktoren der restlichen Faktorwerte wie folgt:

$$b = 4000, \quad x_{R_2} = +500, \quad x_{G_2} = -200, \quad x_{F_2} = -0,5 \tag{16}$$

Eine Prognose für z.B. x_{R_1}, x_{G_2} und x_{F_1} wird dann berechnet durch:

$$y = 4000 + (-500) + (-200) + (+0,5) + \epsilon \tag{17}$$

Der Effekt bei einem Wechsel von x_{R_1} auf x_{R_2} ist dann $500 - (-500) = 1000$. Das bedeutet, in R_2 ist das Durchschnittsgehalt 1000 höher, als in Region R_1.

Ob die einzelnen Faktoren nun die Mittelwerte signifikant beeinflussen (also ob die Mittelwertunterschiede zwischen den Faktorwerten mindestens in einem Faktorwertpaar signifikant unterschiedlich sind), wird nun mit Hilfe eines Hypthesentests ausgewiesen.

$$\begin{aligned} Nullhypothese\ H_0 &: \text{alle Mittelwerte zu den Faktoreinstellungen sind gleich} \\ Gegenhypothese\ H_1 &: \text{es existiert mind. ein signifikanter Mittelwertunterschied zwischen zwei Faktoreinstellungen des betrachteten Faktors.} \end{aligned} \tag{18}$$

Beispiel: man betrachte die Abhängigkeit des Gehaltes von Angestellten in Abhängigkeit von der Region (Nord/Süd), vom Alter (Jung/Alt), von der Religion (christlich / nicht christlich) und dem Geschlecht (M/W). Hier die ersten Zeilen der Datenerfassung des Beispiels, es liegen insg. 112 Datensätze vor.

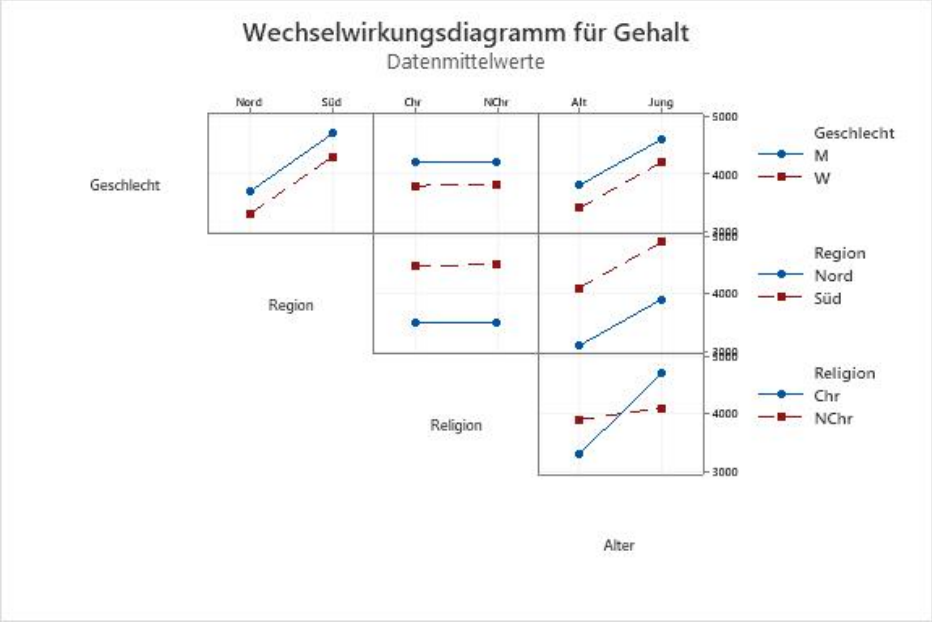


Figure 71: Beispieldatenerfassung für das GLM-Beispiel

Die Varianzanalyse für diese 4 Faktoren liefert nun folgende Tabelle:

Varianzanalyse

Quelle	DF	Kor SS	Kor MS	F-Wert	p-Wert
Geschlecht	1	4366197	4366197	46,01	0,000
Region	1	27565522	27565522	290,48	0,000
Religion	1	30	30	0,00	0,986
Alter	1	17737360	17737360	186,91	0,000
Fehler	107	10153950	94897		
Fehlende Anpassung	11	9939279	903571	404,07	0,000
Reiner Fehler	96	214672	2236		
Gesamt	111	59823060			

Die Faktoren Geschlecht, Region und Alter sind signifikant; d.h. die Gehaltsunterschiede sind bzgl. dieser Faktoren z.T. sehr unterschiedlich. Die Mittelwerte und insb. die Unterschiede können wie folgt mit dem sog. Haupteffektdiagramm deutlich gemacht werden:

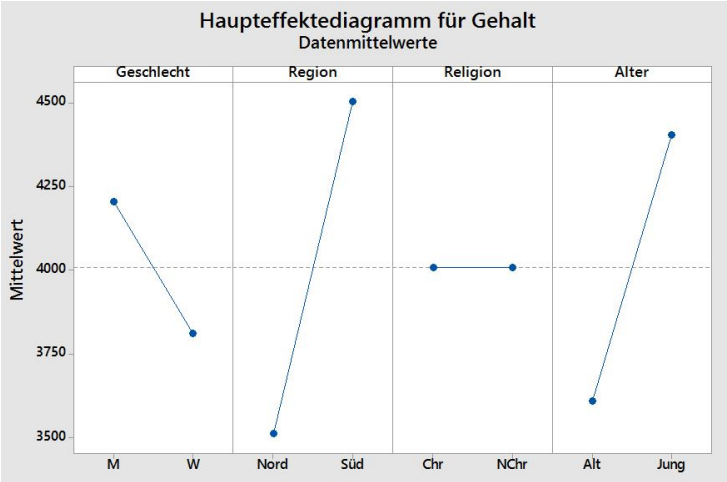


Figure 72: Haupteffekte der 4 Faktoren des Beispiels

Anhand dieser Darstellung kann man die mittleren Zahlen der Gehälter je Faktoreinstellung ablesen.

Da der Faktor Religion die Gehälter (im Mittel) nicht beeinflusst, wird dieser Faktor nun aus der Modellbildung herausgenommen, anschließend wird das Modell nochmal gerechnet; das führt dann zu folgender Varianzanalysetabelle:

Varianzanalyse

Quelle	DF	Kor SS	Kor MS	F-Wert	p-Wert
Geschlecht	1	4366197	4366197	46,44	0,000
Region	1	27565522	27565522	293,19	0,000
Alter	1	17737360	17737360	188,66	0,000
Fehler	108	10153981	94018		
Fehlende Anpassung	4	2371	593	0,01	1,000
Reiner Fehler	104	10151609	97612		
Gesamt	111	59823060			

Das Modell selbst lautet wie folgt:

Koeffizienten

Term	Koef	SE Koef	t-Wert	p-Wert	VIF
Konstante	4007,4	29,0	138,31	0,000	
Geschlecht					
M	197,4	29,0	6,81	0,000	1,00
W	-197,4	29,0	-6,81	0,000	*
Region					
Nord	-496,1	29,0	-17,12	0,000	1,00
Süd	496,1	29,0	17,12	0,000	*
Alter					
Alt	-398,0	29,0	-13,74	0,000	1,00
Jung	398,0	29,0	13,74	0,000	*

Regressionsgleichung

Gehalt = 4007,4 + 197,4 Geschlecht_M -197,4 Geschlecht_W -496,1 Region_Nord
+496,1 Region_Süd -398,0 Alter_Alt +398,0 Alter_Jung

Die Streuung und der Anpassungsgrad berechnen sich zu:

S	R-Qd	R-Qd(kor)	R-Qd(prog)
306,624	83,03%	82,56%	81,75%

9.2.1 Prinzip der Sparsamkeit, Modell-Robustheit und Prognose

Diese Reduktion des Modells auf ein möglichst einfaches Modell mit eben möglichst wenig - und dann ausschließlich signifikanten - Effekten heiß auch das Prinzip der Sparsamkeit . Dieses Prinzip ist zu beachten und durchzuführen, damit man ein möglichst robustes Modell bekommt.

Robust bedeutet in diesem Zusammenhang folgendes Phänomen: liegen genau so viele Messwerte vor wie Terme im Modell sind, dann erhält man ein Bestimmtheitsmaß von 100%. Das klingt zwar schön, wird aber durch einen weiteren Messwert i.d.R. sofort Makulatur. Der Anpassungsgrad wird dann nicht mehr 100% sein; somit ist das Modell nicht robust ggü. weiteren Messwerten.

Sind nun (erheblich) weniger Terme im Modell, als man Messwerte hat, wird allein durch den Glättungseffekt der vielen Meßwerte die berechneten Mittelwerte sich auch bei weiteren Messungen nicht sonderlich verschieben. Die Mittelwerte sind somit stabil gegen weitere Messergebnisse, d.h. das Modell ist robust.

Der Grad der Robustheit wird durch den Wert $R^2(prog)$ zum Ausdruck gebracht. Für einen Anpassungsgrad von 100%, der sich bei gleicher Anzahl von Parametern und Messwerten (automatisch) ergibt, ist der prognostizierte R^2 jedoch sehr klein. Der Unterschied zwischen R^2 und $R^2(prog)$ zeigt also auf, wie stabil das Modell ist. Mehr noch: die Stabilität wird durch die sog. Freiheitsgrade bestimmt. je mehr Freiheitsgrade vorhanden sind, desto mehr Messungen hat man. Wie man von den Konfidenzintervallen her weiß: je mehr Messungen, desto kleiner die Konfidenzintervalle. Der großer Unterschied zwischen R^2 und $R^2(prog)$ weist also auf eine zu kleine Stichprobengröße hin.

Es gilt die Faustregel, dass ein Unterschied zwischen R^2 und $R^2(prog)$ von max. 10% ein Modell genügend robust ausweist.

Die errechneten Koeffizienten dienen nun der Prognose. Das Modell liefert zum Beispiel für Geschlecht $=$ 'M', Region $=$ 'Nord' und Alter $=$ 'Jung':

```
Variable    Einstellung
Geschlecht            M
Region            Nord
Alter             Jung
```

```
SE
Anpassung  Anpassung        95%-KI              95%-PI
4106,70    57,9465  (3991,84; 4221,56)  (3488,16; 4725,24)
```

Anm.: das Prognoseintervall berechnet sich durch ungefähr $\pm 2 * S = \pm 600$; das ergibt einen guten Schätzwert für die Prognose der Werte, im Fall unseres Beispiels hier des Gehaltes.

9.2.2 Wechselwirkungen

Die hier angegebene Prognose ist nur dann sinnvoll, wenn die Faktoren voneinander unabhängig sind. Also wenn z.B. grundsätzlich in einer Region R_2 unabhängig vom Geschlecht und Alter stets signifikant mehr oder weniger im Durchschnitt verdient wird. Das ist in der Realität nicht immer gegeben - und statistisch besteht nun die Aufgabe, solche Fälle, also solche Abhängigkeiten auch aufzudecken, zu bewerten und der Prognose zuzuführen.

Eine solche Abhängigkeit kann z.B. sein, dass zwar in Region R_1 die Personen mit G_1 mehr verdienen, in R_2 ist es aber gerade umgekehrt: da verdienen Personen mit G_2 mehr.

Ein solcher Effekt heißt eine Wechselwirkung .

Zu jeder Wechselwirkung wird nun durch die Varianzanalyse analog wie zu den bisher betrachteten Faktoren ein Koeffizient berechnet. Dieser wird mittels eines Hypothesentests auf gleich 0 getestet. Die Argumentation ist nun wie gehabt: ist der zu der Wechselwirkung gerechnete p-Wert $> 5\%$, dann ist die Wechselwirkung im Modell nicht signifikant - und dann kann diese aus dem Modell entfernt werden. Ist das nicht der Fall, also ist der p-Wert $< 5\%$, dann liegt eine signifikante Wechselwirkung vor. Welcher Art die Wechselwirkung ist, kann dann anhand der Koeffizienten nur schwer gesehen werden. Hierzu gibt es dann die sog. Wechselwirkungsdiagramme, die die Art und Weise der Wechselwirkung aufzeigen.

Anm.: die Effekte, die nur von einem Faktor und dann unabhängig von allen anderen betrachteten Faktoren signifikant sind, heißen dann Haupteffekte .

9.2.3 Beispiel:

Wir betrachten die Datenbasis nun zudem mit Berechnung aller möglichen Wechselwirkungen (in dem Beispiel hier sind Wechselwirkung zwischen 2, 3 oder 4 Faktoren möglich !). Die Varianzanalysetabelle sieht dann wie folgt aus:

Varianzanalyse

Quelle	DF	Kor SS	Kor MS	F-Wert	p-Wert
Geschlecht	1	4366197	4366197	1952,54	0,000
Region	1	27565522	27565522	12327,14	0,000
Religion	1	30	30	0,01	0,908
Alter	1	17737360	17737360	7932,05	0,000
Geschlecht*Region	1	760	760	0,34	0,561
Geschlecht*Religion	1	6612	6612	2,96	0,089
Geschlecht*Alter	1	127	127	0,06	0,812
Region*Religion	1	167	167	0,07	0,785
Region*Alter	1	1150	1150	0,51	0,475
Religion*Alter	1	9923874	9923874	4437,90	0,000
Geschlecht*Region*Religion	1	1282	1282	0,57	0,451
Geschlecht*Region*Alter	1	334	334	0,15	0,700
Geschlecht*Religion*Alter	1	488	488	0,22	0,641
Region*Religion*Alter	1	305	305	0,14	0,713
Geschlecht*Region*Religion*Alter	1	4179	4179	1,87	0,175
Fehler	96	214672	2236		
Gesamt	111	59823060			

Figure 73:

Anhand der p-Werte erkennt man umgehend, dass Alter * Religion eine signifikante Wechselwirkung ist. Zu sehen ist das wiederum sehr gut im sog. Wechselwirkungsdiagramm :

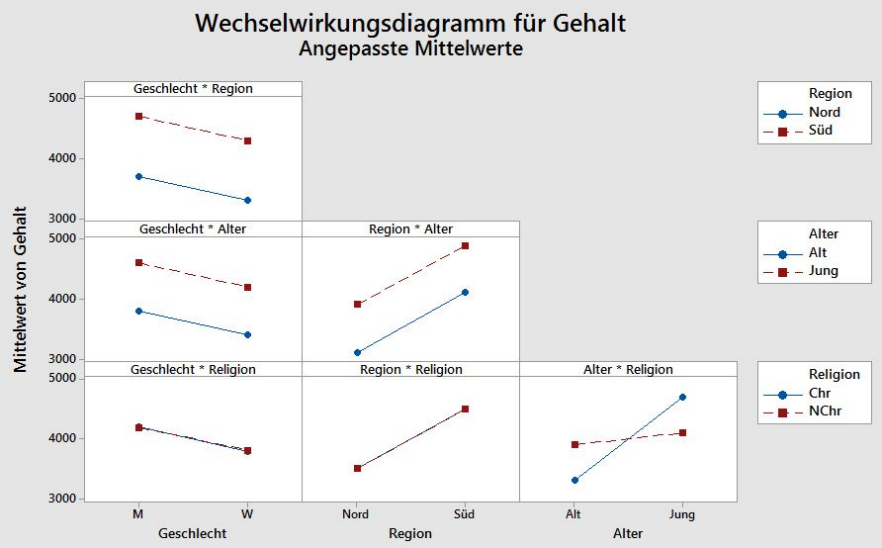


Figure 74: 2fach-Wechselwirkungen der 4 Faktoren des Beispiels

Hier in unserem Beispiel verdienen Nichtchristen, egal ob jung oder alt, im Mittel nahezu gleich viel. Bei den Christen verdienen die Jungen jedoch deutlich mehr als die Älteren (die zudem weniger verdienen, also die Nicht-Christen).

Das Prinzip der Sparsamkeit wird nun angewendet, die nicht-signifikanten Terme werden eliminiert.

Die prinzipielle Vorgehensweise zur Eliminierung ist, dass nicht signifikante Terme der Hierarchie folgend aus dem Modell genommen werden. Die Hierarchie eines Terms ist nichts anderes, als die Anzahl der Faktoren, aus dem der Term besteht. So sind Haupteffekte in der Hierarchiestufe 1, Zweifachwechselwirkungen in der Stufe 2, usw.

Das Modell wird klassisch (manuell) nach dem Prinzip der Sparsamkeit reduziert, in dem in der Gruppe der höchsten Hierarchiestufe der Term mit dem größten p-Wert eliminiert wird.

Anm.: Achtung: nach jeder Eliminierung verändert sich das Modell; d.h. die p-Werte und somit deren Signifikanzbewertung können sich mit jeder Reduktion verändern, typischerweise werden die p-Werte kleiner (da sich die Anzahl der Stichproben auf weniger Terme verteilt und somit die Konfidenzintervalle kleiner

werden). Es kann halt passieren, dass ein Term, der im Modell ohne Reduktion nicht signifikant war, nach ein paar Reduktionen signifikant wird. Insofern muss die Reduktion schrittweise vorgenommen werden. Daraus erklärt sich der Begriff, der im Minitab-Menü auf der zugehörigen Schaltfläche erscheint:

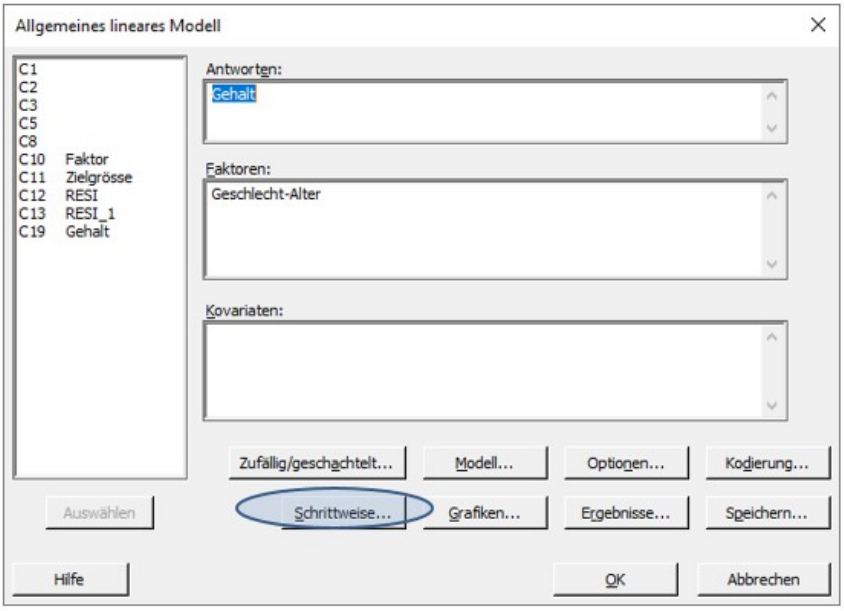


Figure 75: Aktivieren des Prinzips der Sparsamkeit

In Minitab stehen dann 4 Methoden zur Verfügung:

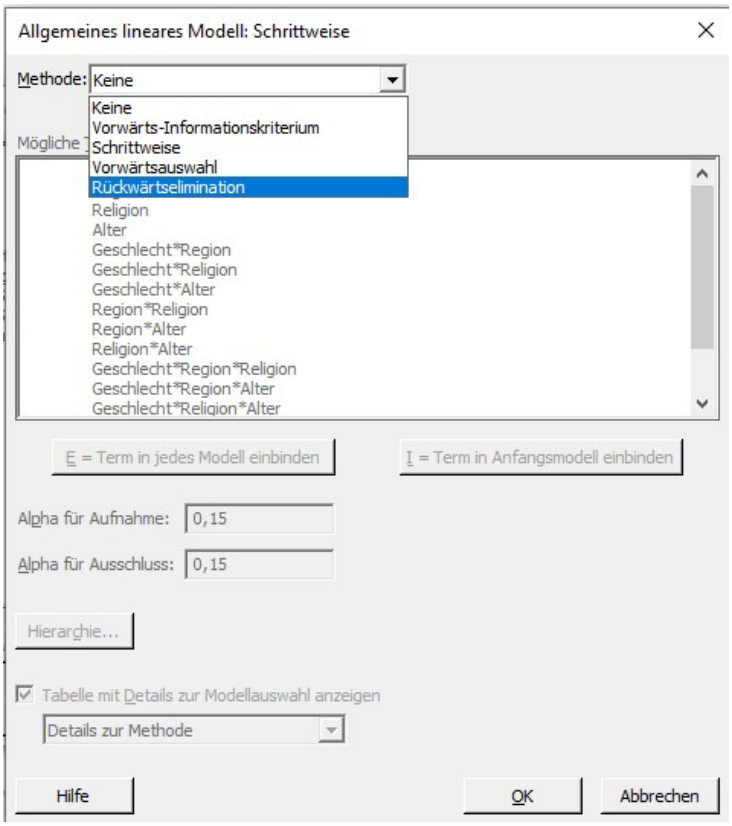


Figure 76: PdS-Methoden

Auf die genauen Vorgehensweisen und deren Vor- und Nachteile wird hier nicht weiter eingegangen. Die klassische Vorgehensweise ist für unsere Zwecke völlig ausreichend, zudem eine Methode, die signifikante Terme auch sicher findet. Wie oben erläutert, wird klassisch die Reduktion von Modellen rückwärts vorgenommen, die 'Rückwärtselimination' entspricht diesem Vorgehen.

Die Berechnung des Modells unter Verwendung der Rückwärtselimination führt nun zu folgendem Modell:

Allgemeines lineares Modell: Gehalt vs. Geschlecht; Region; Religion; Alter

Methode

Faktorkodierung (-1; 0; +1)

Rückwärtselimination von Termen

α für Ausschluss = 0,1

Faktorinformationen

Faktor	Typ	Stufen	Werte
Geschlecht	Fest	2	M; W
Region	Fest	2	Nord; Süd
Religion	Fest	2	Chr; NChr
Alter	Fest	2	Alt; Jung

Varianzanalyse

Quelle	DF	Seq SS	Beitrag	Kor SS	Kor MS	F-Wert	p-Wert
Geschlecht	1	4366197	7,30%	4366197	4366197	2051,57	0,000
Region	1	27565522	46,08%	27565522	27565522	12952,35	0,000
Religion	1	30	0,00%	30	30	0,01	0,905
Alter	1	17737360	29,65%	17737360	17737360	8334,34	0,000
Geschlecht*Religion	1	6612	0,01%	6612	6612	3,11	0,081
Religion*Alter	1	9923874	16,59%	9923874	9923874	4662,98	0,000
Fehler	105	223464	0,37%	223464	2128		
Fehlende Anpassung	9	8792	0,01%	8792	977	0,44	0,912
Reiner Fehler	96	214672	0,36%	214672	2236		
Gesamt	111	59823060	100,00%				

Zusammenfassung des Modells

S	R-Qd	R-Qd(kor)	PRESS	R-Qd(prog)	AICc	BIC
46,1327	99,63%	99,61%	254252	99,57%	1186,27	1206,62

Figure 77: GLM-4F-PdS1

Koeffizienten

Term	Koef	SE Koef	95%-KI	t-Wert	p-Wert	VIF
Konstante	4007,40	4,36	(3998,76; 4016,04)	919,31	0,000	
Geschlecht						
M	197,44	4,36	(188,80; 206,09)	45,29	0,000	1,00
W	-197,44	4,36	(-206,09; -188,80)	-45,29	0,000	*
Region						
Nord	-496,11	4,36	(-504,75; -487,46)	-113,81	0,000	1,00
Süd	496,11	4,36	(487,46; 504,75)	113,81	0,000	*
Religion						
Chr	-0,52	4,36	(-9,16; 8,12)	-0,12	0,905	1,00
NChr	0,52	4,36	(-8,12; 9,16)	0,12	0,905	*
Alter						
Alt	-397,96	4,36	(-406,60; -389,31)	-91,29	0,000	1,00
Jung	397,96	4,36	(389,31; 406,60)	91,29	0,000	*
Geschlecht*Religion						
M Chr	7,68	4,36	(-0,96; 16,33)	1,76	0,081	1,00
M NChr	-7,68	4,36	(-16,33; 0,96)	-1,76	0,081	*
W Chr	-7,68	4,36	(-16,33; 0,96)	-1,76	0,081	*
W NChr	7,68	4,36	(-0,96; 16,33)	1,76	0,081	*
Religion*Alter						
Chr Alt	-297,67	4,36	(-306,31; -289,02)	-68,29	0,000	1,00
Chr Jung	297,67	4,36	(289,02; 306,31)	68,29	0,000	*
NChr Alt	297,67	4,36	(289,02; 306,31)	68,29	0,000	*
NChr Jung	-297,67	4,36	(-306,31; -289,02)	-68,29	0,000	*

Regressionsgleichung

Gehalt = 4007,40 + 197,44 Geschlecht_M - 197,44 Geschlecht_W - 496,11 Region_Nord
+ 496,11 Region_Süd - 0,52 Religion_Chr + 0,52 Religion_NChr - 397,96 Alter_Alt
+ 397,96 Alter_Jung + 7,68 Geschlecht*Religion_M Chr - 7,68 Geschlecht*Religion_M
NChr - 7,68 Geschlecht*Religion_W Chr + 7,68 Geschlecht*Religion_W NChr
- 297,67 Religion*Alter_Chr Alt + 297,67 Religion*Alter_Chr Jung
+ 297,67 Religion*Alter_NChr Alt - 297,67 Religion*Alter_NChr Jung

Figure 78: GLM-4F-PdS2.jpg

Anpassungen und Bewertung für ungewöhnliche Beobachtungen

Beob	Gehalt	Anpassung	SE Anpassung	95%-KI	Resid	Std. Resid	Entf Resid
26	4684,0	4593,5	11,5	(4570,6; 4616,4)	90,5	2,03	2,06
31	3876,4	4012,5	11,5	(3989,6; 4035,4)	-136,1	-3,05	-3,18
78	2913,3	3020,3	11,5	(2997,4; 3043,1)	-106,9	-2,39	-2,45
96	3319,3	3422,3	11,5	(3399,5; 3445,2)	-103,0	-2,31	-2,36
104	4089,0	4214,0	11,5	(4191,1; 4236,8)	-125,0	-2,80	-2,89

Beob	HOCH	Cook-D	DFITS	
26	0,0625	0,04	0,531057	R
31	0,0625	0,09	-0,820061	R
78	0,0625	0,05	-0,632659	R
96	0,0625	0,05	-0,608350	R
104	0,0625	0,07	-0,747242	R

R Großes Residuum

Figure 79: GLM-4F-PdS3.jpg

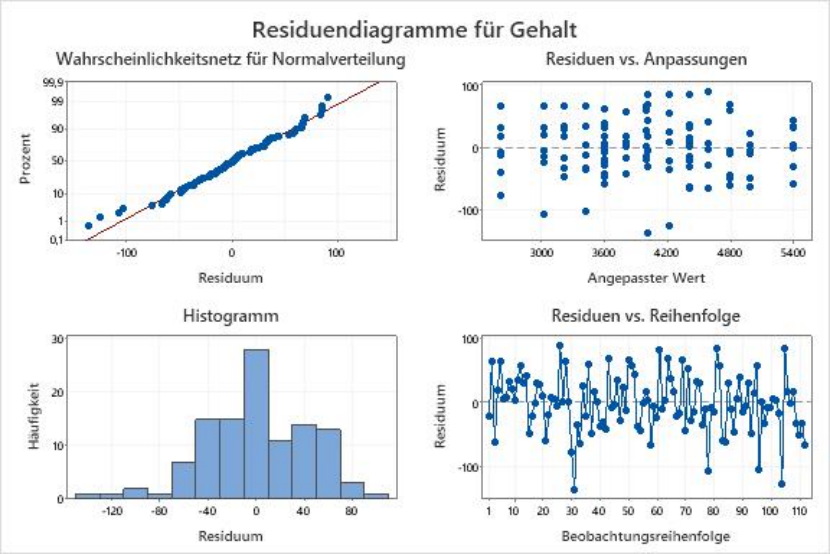


Figure 80: GLM-4F-PdS4.jpg

Ein wichtiger Aspekt erscheint in der Tabelle der Varianzanalyse: die Beitrags-Prozente. Hiermit kann nun die Frage nach der Wichtigkeit der Terme im Modell beantwortet werden. Die Wichtigkeit kann dabei als Sensitivität der Änderung der Zielgröße unter Änderung des jeweiligen Terms gesehen werden. Zu beachten ist hierbei die Skala, die die Terme in der Stichprobe hatten (die Sensitivität wird

relativ zum Definitionsbereich ausgewertet).

Im Beispiel hier ist die Region der wichtigste Einfluss auf ein Gehalt, vor dem Alter. Die Wechselwirkung von Alter mit Religion ist der schwächste Einfluss nach dieser Analyse auf das Gehalt.

Anm.: es kann sein, dass in einer Modellbildung ein Term nach p-Wert zwar signifikant ist, dass jedoch der Beitrags%-Wert relativ klein ist. Man wäre dann geneigt, diesen Term nach dem Prinzip der Sparsamkeit ebenfalls (also nachträglich) aus dem Modell zu nehmen. Ein brauchbares Kriterium für die Praxis, um einen solchen Term zu eliminieren, wäre, wenn der Betrags%-Wert in etwa der Hälfte des Beitrages des 'Unerklärlichen', also des Residuums, entspricht. Dieser Beitrags%-Wert des Residuums heißt in Minitab 'Fehler' und besteht aus dem Beitrag für fehlende Anpassung (also ob ein Wechselwirkungs- oder auch ein nichtlinearer Term im Modell ggf. fehlt) und dem dem reinen Fehler der zufälligen Restunsicherheit der Messungen.

9.2.4 Residuum-Analyse, VIF-Werte, ungewöhnliche Beobachtungen, AIC und BIC

Die Residuendiagramme dienen ausschließlich der (qualitativen,. anhand der Grafiken) Überprüfung , ob die Residuendaten einer Normalverteilung folgen. Dieses ist hier relativ gut der Fall, wie das Wahrscheinlichkeitsnetz als auf das Histogramm aufzeigen. Die Abfolge der Messungen in der Zeit (unten rechts) zeigt zudem ebenfalls keine (offensichtlichen) Muster auf.

Eine explizite Analyse des Residuums ist möglich unter Verwendung von 'Speichern' in Minitab:

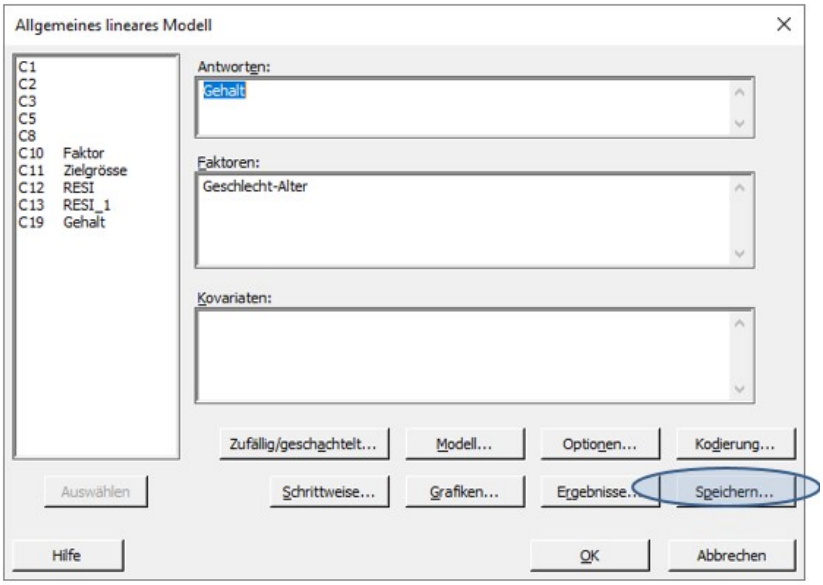


Figure 81: GLM-Speichern des Residuums

In Minitab wird dann eine Spalte erzeugt namens 'RESI'; der Test auf Normalverteilung liefert hierfür:

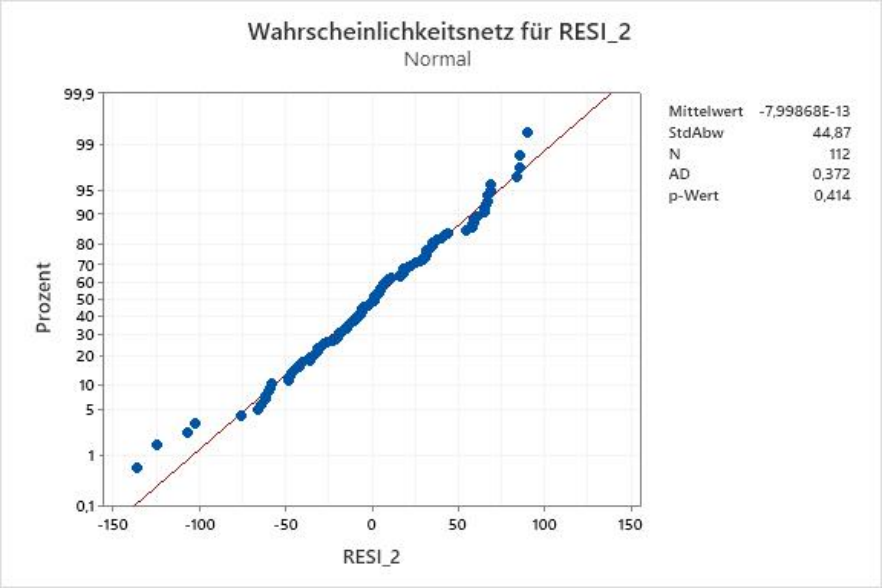


Figure 82: GLM-Res-NV

Der p-Wert des AD-Tests ist hier $0,414$, also deutlich größer als 5%. Insofern liegt hier kein KO-Kriterium für das Modell vor - wie anhand der Grafiken bereits vermutet.

Ein weiteres sehr wichtiges Kriterium für die Bewertung des Modells sind die in der Tabelle der Koeffizienten vorhandenen VIF-Werte. VIF bedeutet 'Variance Inflation Faktor'. Dieser Faktor misst einen lineare Zusammenhang zwischen den Faktoren. Ein Modell wird ungenauer in der Prognose, wenn die VIF-Werte groß werden. Als Faustregel gilt: ein VIF-Wert bis 5 ist für ein Modell OK.

Anm.: der hier angesprochene lineare Zusammenhang ist von einem bilateralen Zusammenhang, wie er beim Korrelationskoeffizienten berechnet wird, zu unterscheiden. Lineare Zusammenhänge sind allgemeinere (lineare) Abhängigkeiten zwischen ggf. mehreren Faktoren, die mit z.B. Korrelationskoeffizienten von Spearmen nicht entdeckt werden können.

Die Größen AIC und BIC bedeuten Akaike Information Criterium und Bayes Information Criterium. Diese messen, inwiefern ein Modell mit mehr Faktoren ggü. einen Modell mit weniger Faktoren tatsächlich ein besseres Modell darstellt. Besser bedeutet, dass das bessere Modell mehr Informationen, also eine bessere

Anpassung und damit weniger Unsicherheiten enthält. Man erinnere sich an die Auffassung, dass Stichproben, z.B. von einem komplexen System, Informationsgewinnung über dieses System ist; dieses wird nun auch übertragen auf Aussagekraft von Modellen. Auf die genauen Formeln wird hier verzichtet; für Experten hier aber der Hinweis, dass die aus der Physik als Kennzahl über die Information (also die Ordnung) in einem System definierte Größe der Entropie als Vorbild dient. Diese Kriterien können auch für das Prinzip der Sparsamkeit verwendet werden. Dann geht es nicht mehr darum, signifikante Terme in einem Modell zu behalten, sondern Terme, die auch eine größtmögliche Informationseffizienz im Modell besitzen.

Die Tabelle der ungewöhnlichen Beobachtungen schließlich kennzeichnet zwei unterschiedliche Auffälligkeiten: zum einen, wenn ein Messergebnis nicht gut zum Modell passt (d.h. das Residuum dieses Wertes ist recht groß), oder wenn das Modell sehr stark nur von einzelnen Messungen abhängt.

Anhand dieser Regressionsgleichung ist nun eine Prognose möglich. Für die Faktorwerte 'M', 'Nord', 'Jung' und 'Chr', die in Minitab wie folgt eingestellt werden können:

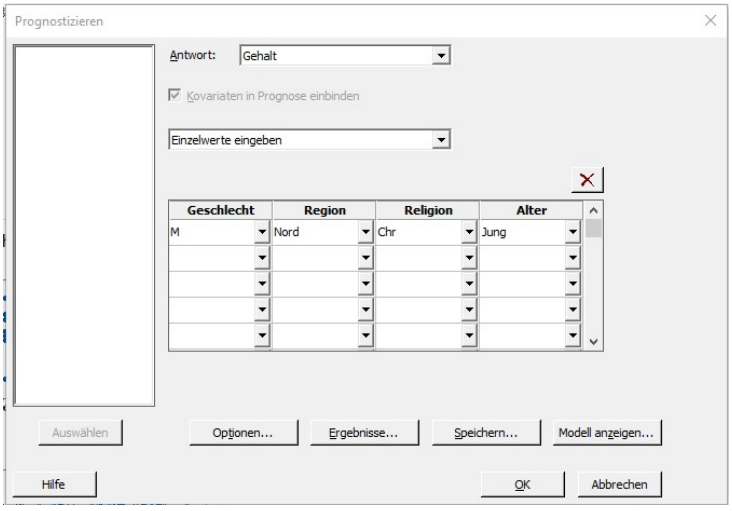


Figure 83:

folgt:

Informationen zum allgemeinen linearen Modell

Terme
Geschlecht Region Religion Alter Geschlecht*Religion Religion*Alter

Einstellungen

Variable	Einstellung
Geschlecht	M
Region	Nord
Religion	Chr
Alter	Jung

Prognose

Anpassung	SE Anpassung	95%-KI	95%-PI
4411,53	11,5332	(4388,66; 4434,40)	(4317,24; 4505,81)

Figure 84:

Das Prognoseintervall (PI) ist die Prognose der Gehälter für diese Einstellungen, die dann zwischen 4.317,24 und 4.505.81 liegt.

Die Mittelwerte können nun wieder in Gruppen eingeteilt werden mit Hilfe der Kontrastberechnung.

Das Residuum wird wiederum als normalverteilt angenommen; die Streuung ist im Beispiel mit 46,3738 angegeben, was je Faktoreinstellung bei einer Prognose zu einer Gehaltsschwankung (95%) von fast 100 führt.

10 Lineare Regression

Bei der linearen Regression wird nun eine metrische Antwortvariable in Relation zu einer oder mehreren metrischen Zielgrößen gesetzt.

Zunächst ist die Frage, ob es überhaupt einen prinzipiellen linearen (!) Zusammenhang zwischen den Variablen gibt. Diese Frage beantwortet der sog. Korre-

lationskoeffizient .

Der Korrelationskoeffizient nach Pearson misst den Grad eines linearen Zusammenhangs. Dieser Zusammenhang kann positiv oder auch negativ sein, in dem Sinn, dass bei einer Abhängigkeit von x und y dann ein linearer Zusammenhang besteht: $y = a * x + b + \epsilon$ (siehe auch die Modellierung der ANOVA). Ist der Zusammenhang negativ, dann ist das Vorzeichen von a negativ; das bedeutet, bei anwachsendem x wird y kleiner.

Berechnet man den Korrelationskoeffizienten zwischen 'Alter' und 'Gehalt' anhand der Beispieldaten, dann bekommt man folgendes Resultat:

```
Korrelation nach Pearson für Alter und Gehalt = 0,176
p-Wert = 0,164
```

Dieser Wert zeigt, dass es keinen guten Zusammenhang zwischen Alter und Gehalt gibt. Das wird auch deutlich am sog. Streudiagramm, das die gemessenen Werte (x, y) als Koordinaten darstellt.

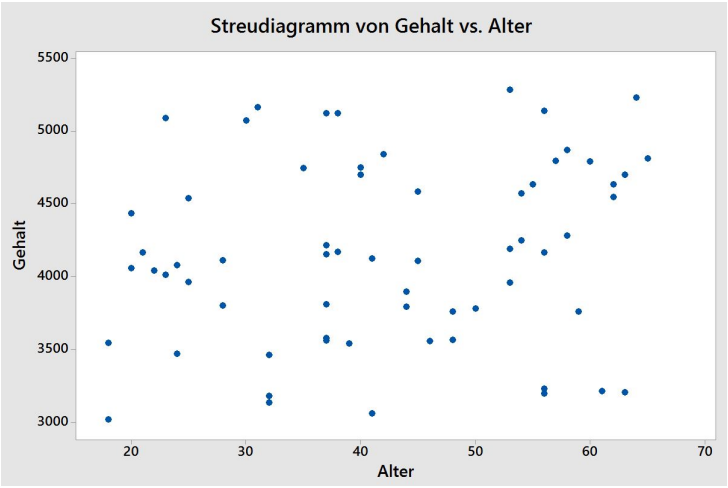


Figure 85: Streudiagramm für das Beispiel Alter vs. Gehalt

Ein hoher Zusammenhang würde in etwa zu solch einem Streudiagramm führen:

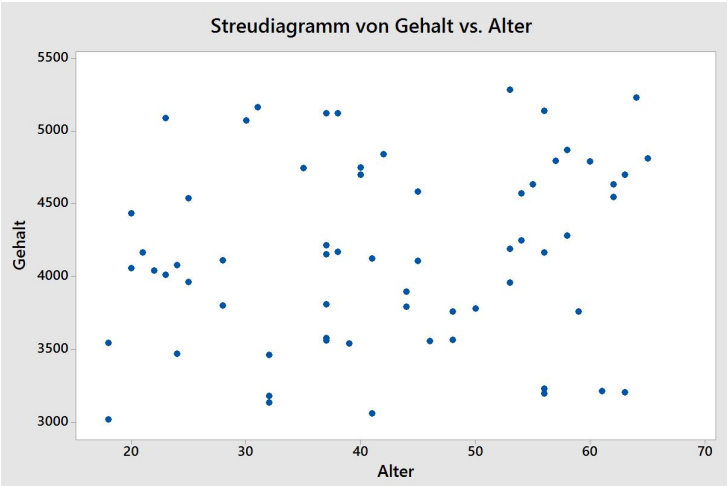


Figure 86: Streudiagramm für das Beispiel hohe Korrelation

```
Korrelation nach Pearson für C7 und Gehalt = 0,993
p-Wert = 0,000
```

Anm.: der p-Wert ist wieder das Ergebnis eines Hypothesentests, H_0 ist: Korrelationskoeffizient $= 0$.

Der Korrelationskoeffizient ist eine relativ schwache Größe, sie berücksichtigt z.B. nicht nichtlineare Zusammenhänge, oder bei mehr Variablen einen nicht-linearen (also z.B. einen quadratischen oder kubischen) Zusammenhang.

Jedoch ist bei einem hohen Korrelationskoeffizienten schon mal nahezu gesichert, dass es einen starken linearen Zusammen zwischen den betrachteten Größen gibt - und sich somit auch eine lineare Regression, wie jetzt vorgestellt, lohnt.

Die Lineare Regression berechnet nun mit Hilfe der Methode der kleinsten Quadrate eine bestmöglich zu den Daten passende Gerade (eben ein lineares Modell $y_R = a * x + b$). Diese Methode minimiert die Differenzen in den y-Werten zwischen der gesuchten Geraden und den Datenpunkten. Die dann berechnete Geraden stellt somit einen Mittelwert bzgl. der Differenzen dar, und ist somit analog der ANOVA oder des GLM die Prognose-Größe (im Sinne des Erwartungswertes). Die Streuung der Werte der Prognose wird ebenfalls analog der Varianzanalyse

als normalverteiltes Residuum dargestellt, dass sich auf die Prognosewerte aufaddiert.

Nun zunächst zu den zwei Beispielfällen 'Alter' vs. 'Gehalt' bzw. die zum Gehalt hoch korrelierte Größe 'C7'.

Die berechnete Gerade ergibt sich für den ersten Fall wie folgt:

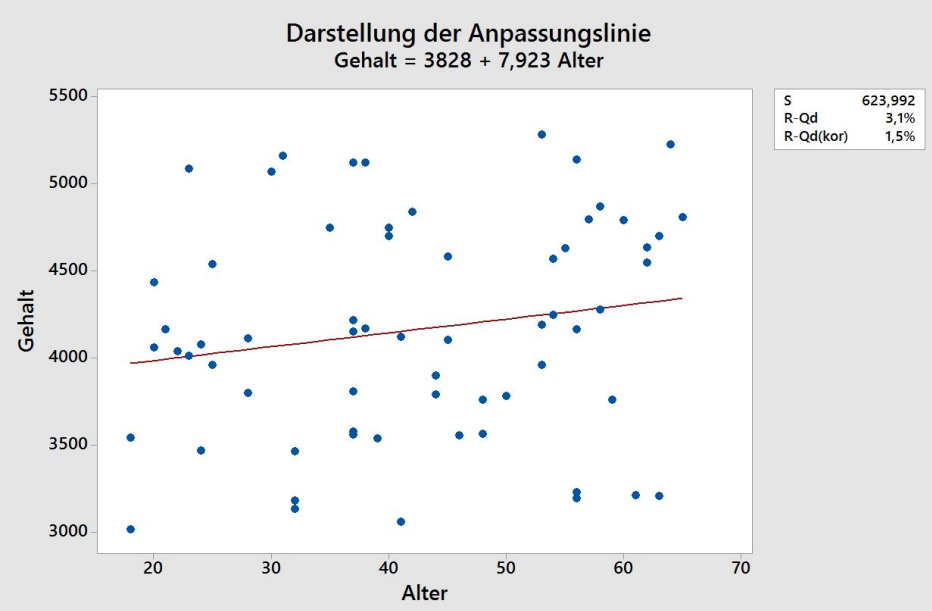


Figure 87: Regressionsgeraden für schwach korrelierte Größen

Für den zweiten Fall ergibt sich:

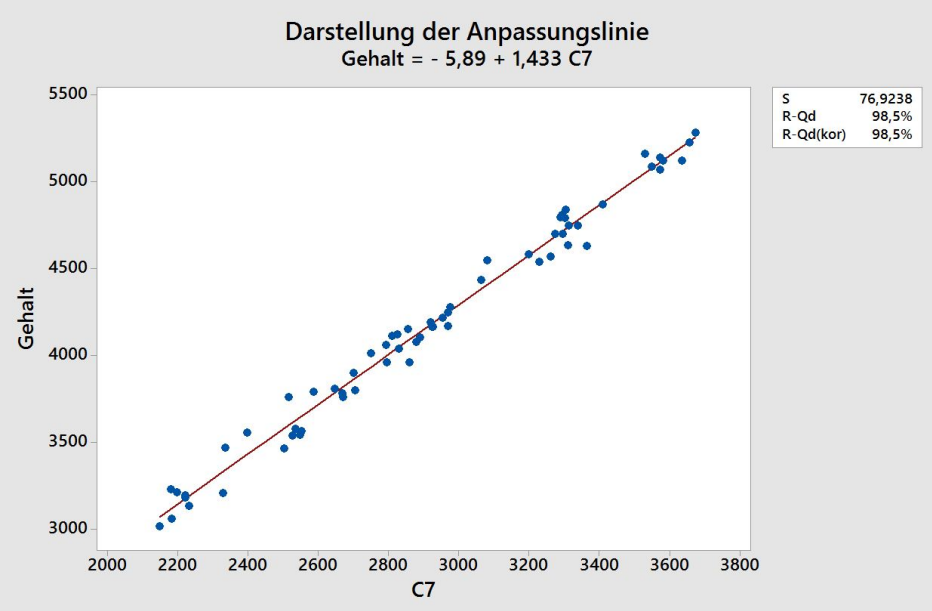


Figure 88: Regressionsgeraden für stark korrelierte Größen

Die Größe R^2 (entspricht dem Quadrat des Korrelationskoeffizienten ...) ist nun der Anpassungsgrad oder das Bestimmtheitsmaß für die gerechnete Regression. Im ersten Fall ist dieses mit 3,1% sehr klein (und weit entfernt von den auch hier geforderten 80%). Im zweiten Fall ist die Bestimmtheit oder Anpassung mit 98,5% sehr hoch, mit diesem Modell ist dann eine Prognose oder Zielgrößenbestimmung erlaubt.

Dieser Sachverhalt wird zudem noch durch die sehr unterschiedlichen Standardabweichungswerte deutlich: fast 624 vs. fast 77.

das Modell berechnet sich für den zweiten Fall wie folgt:

```
Varianzanalyse

Quelle       DF    Kor SS    Kor MS   F-Wert  p-Wert
Regression    1  24545950  24545950  4148,18   0,000
C7            1  24545950  24545950  4148,18   0,000
Fehler       62    366871      5917
Gesamt       63  24912821
```

```
Zusammenfassung des Modells

S    R-Qd  R-Qd(kor)  R-Qd(prog)
76,9238  98,53%     98,50%      98,43%


Koeffizienten

Term         Koef  SE Koef  t-Wert  p-Wert   VIF
Konstante    -5,9     65,4   -0,09   0,929
C7         1,4329   0,0222   64,41   0,000  1,00


Regressionsgleichung

Gehalt = -5,9 + 1,4329 C7
```

Für den zweiten Fall kann man nun eine Prognose wagen, so ist z.B. für $x = 3000$ die Prognose wie folgt:

```
Regressionsgleichung

Gehalt = -5,9 + 1,4329 C7


Variable  Einstellung
C7              3000


SE
Anpassung  Anpassung        95%-KI              95%-PI
4292,80    9,82791  (4273,15; 4312,44)  (4137,78; 4447,82)
```

Hier ist die Funktionsweise des Residuums zu erkennen: zu dem Prognosewert $-5,9+1,4329*3000$ wird nun ein 95%-Konfidenzintervall mit nahezu (also grob

geschätzt) $\pm 2 * S = \pm 2 * 77 = \pm 154$ hinzuaddiert.

Anm.: das Residuum ist nicht von x abhängig, also wird prinzipiell das hier angegebene Intervall zu jeder beliebigen Prognose an jedem x-Wert hinzuaddiert. Diese Vereinfachung kann problematisch sein.

Die Umkehrung, also die Vorgabe eines y-Wertes und die Berechnung des zugehörigen x-Wertes, ist ebenfalls möglich. Hier als Beispiel sei nun der y-Wert 4292,8 vorgegeben und der entsprechend der Geraden zugehörige x-Wert gesucht. Die Umstellung der Geradenformel liefert:

$$y = b + a * x \Rightarrow x = \frac{1}{a} * y - \frac{b}{a} \tag{19}$$

Die Berechnung mit einem Tool liefert:

Antwort	Ziel	Untergrenze	Soll	Obergrenze	Gewichtung	Bedeutung
Gehalt	Soll	4200	4292,8	4500	1	1

Lösung

Gehalt Lösung	Zusammengesetzte C7	Anpassung	Erwünschtheit
1	2995,61	4286,51	0,932233

Mehrere Antwortprognosen

Variable	Einstellung
C7	2995,61

SE

Antwort	Anpassung	Anpassung	95%-KI	95%-PI
Gehalt	4286,51	9,81	(4266,90; 4306,12)	(4131,50; 4441,52)

Anm.: die kleine Differenz von 2995,61 auf 3000 ist aufgrund von Rundungsfehlern entstanden.

11 Anwendung ANOVA: Gage R& R

Als eine Anwendung der ANOVA wird nun die sog. Messsystemanalyse betrachtet.

Messungen führen in der Praxis selbst für gleiche Objekte zu unterschiedlichen Messergebnissen. Das bedeutet, die Messwerte besitzen eine (ungewollte) Streuung, die nicht durch den Herstellungs- oder Verwendungsprozess bedingt ist, sondern durch das Messsystem. Die Gründe dafür können vielfältig sein, der folgende Streuungsbaum gibt Hinweise dazu:

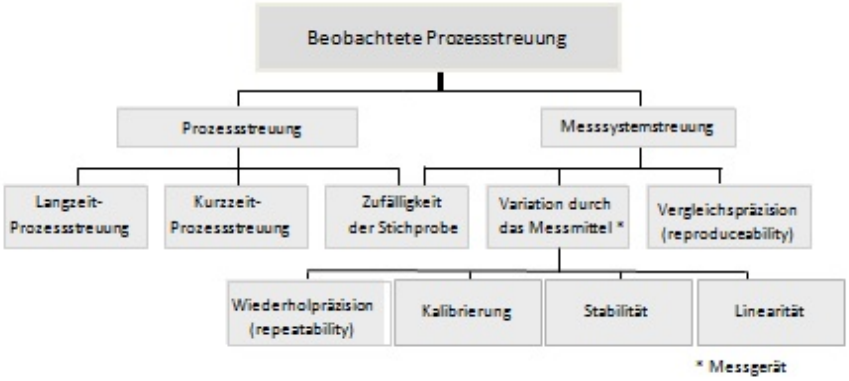


Figure 89: Einflüsse auf Messwerte - Gründe für eine Streuung

Betrachtet man das Zustandekommen von Messwerten in der Praxis, dann sind elementar zwei Einflussfaktoren als wesentlich identifizierbar:

- der Prüfer
- das Messgerät

Diese Einflüsse sind attributiv, also Faktoren.

Die Streuung der Messwerte soll nun sozusagen den jeweiligen Verursachern zugewiesen werden können. Die Methode dazu ist bereits diskutiert worden (vgl. Abschn. 9.1.1), die Gesamtstreuung (100%) kann nun auf die jeweiligen Einflussfaktoren mittels der Bestimmung von innerem und äußerem Streuungsanteil aufgeteilt werden. Diese Streuungsanteile müssen lediglich den Verursachern zugeordnet werden. Zu beachten ist, dass die Gesamtstreuung als Streuung der Messwerte die Streuung in den Messgrößen der gemessenen Teile beinhaltet. Folgendes Bild stellt diese Streuungsaufteilung dar:

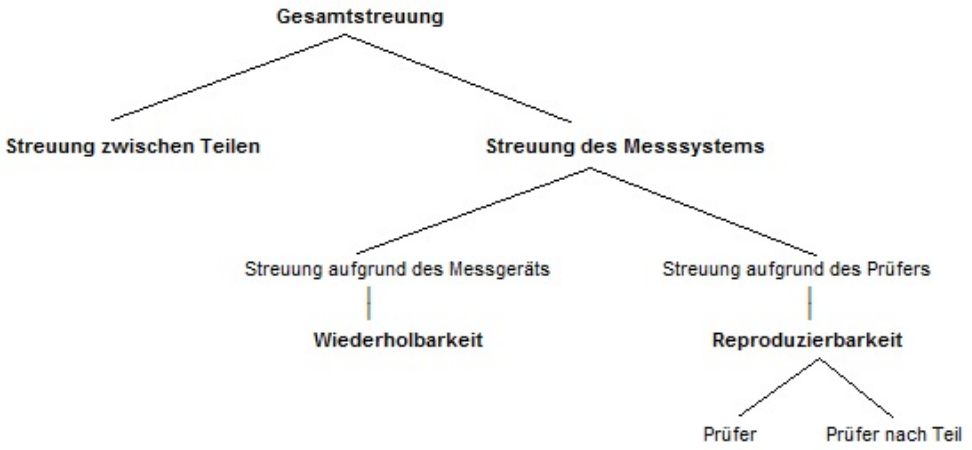


Figure 90: Streuungsaufteilung bei der Messsystemanalyse mittels ANOVA

Durch diese Streuungsaufteilung werden zwei Begriffe definiert, die Wiederholbarkeit (engl.: Repeatability) und die Reproduzierbarkeit (engl.: Reproducability). Diese beiden 'R'-Wörter geben der Messsystemanalyse im engl. den Namen: Gage R& R.

Ein ideales Messsystem würde 100% Streuung den Teilen zuweisen (es sind ja unterschiedliche Teile mit unterschiedlichen Messgrößen gemessen worden, das würde hierdurch dann ausgedrückt), und jeweils 0% Streuung für die Prüfer und das Messgerät. Im Hypothesentest der ANOVA wäre dann der p-Wert für die Teile $= 0\%$, und für die Prüfer und Teile, sowie die Wechselwirkung zwischen diesen $= 100\%$.

Die Wiederholbarkeit wird gemessen, indem ein Prüfer mit dem selben Gerät ein Teil mehrfach bemisst. Dieser Wert stellt also die durch das Messgerät allein

verursachte Streuung dar.

Die Reproduzierbarkeit wird gemessen, indem mehrere Prüfer ein Teil mit einem Gerät (mehrfach) bemessen.

Die Streuung 'Prüfer nach Teil' ist die Wechselwirkung zwischen Prüfer und Teil. Es kann sein, dass ein Prüfer ganz bestimmte Teile anders als die anderen Teile misst. Dieser Effekt wird durch diese Analyse ausgewiesen.

In der Industrie sind unterschiedliche Standards definiert worden, um Messsysteme zu prüfen:

- MSA Typ 1: ein bekanntes Teil wird (von einem Prüfer) 25 mal gemessen
- MSA Typ 2: mind. 10 Teile werden von mind. 3 Prüfern mind. 2 Mal gemessen
- MSA Typ 3: mind. 25 Teile werden von 1 Prüfer mind. 2 Mal gemessen

Eine besondere Problematik ist gegeben, wenn die Teile nicht mehrfach gemessen werden können; das ist z.B. der Fall bei zerstörenden Prüfungen. Dann müssen für jede einzelne Prüfung unterschiedliche Teile herangezogen werden. Dann ist vorher zu klären in wie weit die für eine quasi-wiederholte Messung geeignet sind, z.B. durch Verwendung von Teilen eines Produktionsloses (Charge).

Im Falle von wiederholbaren Messungen spricht man von gekreuzter MSA, im Falle nicht wiederholbarer Messungen von einer geschachtelten MSA (das ergibt sich aus den Linien eines Messzuweisungsplanes für die Prüfer zu den Teilen; im Falle der wiederholbaren Messungen kreuzen sich diese Linien. Wenn Sie die Annahme der Homogenität der Teile treffen können, hängt die Entscheidung zwischen der gekreuzten und der geschachtelten Messsystemanalyse für zerstörende Prüfungen davon ab, wie der Messprozess eingerichtet ist. Wenn alle Prüfer Messwerte für Teile aus sämtlichen Chargen erfassen, verwenden Sie die gekreuzte oder die erweiterte Messsystemanalyse. Wenn eine Charge jeweils von nur einem einzigen Prüfer getestet wird, verwenden Sie die geschachtelte oder die erweiterte Messsystemanalyse. In allen Fällen, in denen Prüfer ihnen eindeutig zugewiesene Teile messen, liegt ein geschachteltes Design vor. Wenn Sie neben Prüfer und Teil weitere Faktoren einbinden müssen oder ein nicht

balanciertes Design, feste Faktoren bzw. eine Mischung von gekreuzten und geschachtelten Faktoren vorliegen, müssen Sie die erweiterte Messsystemanalyse ausführen.

11.0.1 Beispiel zur MSA

Als Beispiel seien Daten von jeweils 3 Testmessungen zu 10 Teilen von 3 Messungen (MSA Typ 2) gegeben (hier sind nur die ersten Datenzeilen angegeben):

Teil	Prüfer	Messwert
1	A	0,29
1	A	0,41
1	A	0,64
2	A	-0,56
2	A	-0,68
2	A	-0,58
3	A	1,34
3	A	1,17
3	A	1,27
4	A	0,47
4	A	0,50
4	A	0,64

Figure 91: Daten MSA Typ 2

Die ANOVA-Analyse gibt folgenden Output:

```
Messsystemanalyse - ANOVA-Methode

Zweifache ANOVA-Tabelle mit Wechselwirkungen

Quelle            DF       SS       MS        F      p
Teil               9  88,3619  9,81799  492,291  0,000
Prüfer             2   3,1673  1,58363   79,406  0,000
Teil * Prüfer     18   0,3590  0,01994    0,434  0,974
Wiederholbarkeit  60   2,7589  0,04598
Gesamt            89  94,6471

Alpha für Ausschluss des Wechselwirkungsterms = 0,25

Zweifache ANOVA-Tabelle ohne Wechselwirkungen
```

```
Quelle            DF       SS       MS        F      p
Teil               9  88,3619  9,81799  245,614  0,000
Prüfer             2   3,1673  1,58363   39,617  0,000
Wiederholbarkeit  78   3,1179  0,03997
Gesamt            89  94,6471

R&R (gesamt)        %Beitrag  (der Quelle      VarKomp  VarKomp)
R&R (gesamt)          0,09143      7,76
Wiederholbarkeit      0,03997      3,39
Reproduzierbarkeit    0,05146      4,37
Prüfer                0,05146      4,37
Zwischen den Teilen   1,08645     92,24
Gesamtstreuung        1,17788    100,00

Streu.s in

Unters.  %Streuung s in Untersuchung
Quelle                StdAbw        (6*StdAbw)         (%SU)
R&R (gesamt)          0,30237        1,81423           27,86
Wiederholbarkeit      0,19993        1,19960           18,42
Reproduzierbarkeit    0,22684        1,36103           20,90
Prüfer                0,22684        1,36103           20,90
Zwischen den Teilen   1,04233        6,25396           96,04
Gesamtstreuung        1,08530        6,51180          100,00

Anzahl der eindeutigen Kategorien = 4
```

Der p-Wert für die Teile ist 0%; die (gewisse) Teile werden also auch als unterschiedlich gemessen.

Der p-Wert für Prüfer ist ebenfalls 0%. Das bedeutet, dass die Prüfer gleiche Teile unterschiedlich gemessen haben, und zwar signifikant unterschiedlich. Das ist problematisch.

Der p-Wert für die Wechselwirkung zwischen Prüfern und Teilen ist mit 97,4% sehr hoch. Insofern haben die Prüfer alle Teile gleichmäßig gemessen, es wurden

keine Teile besonders (signifikant) gut oder schlecht gemessen.

Im Programm ist ein Ausschlusskriterium definiert, dass nicht-signifikante Wechselwirkungen für die Analyse ausschließt und diese dann ohne die Wechselwirkung wiederholt.

Die Aufteilung der Streuung nach Methodik der ANOVA ist der Spalte '%Beitrag (der VarKomp)' zu entnehmen. Auf die Streuungen begründet durch die Prüfer und das Messgerät entfallen insg. lediglich 7,76%, die Streuung in den Teilen wird dann konsequenterweise (Summe der Streuungen = 100%) zu 92,24%.

11.0.2 Kriterien der AIAG

Die 'Automotive Industry Interaction Group' (kurz: AIAG) hat sich auf folgende Bewertungen für Messsysteme geeinigt:

Ist der Wert in der Spalte '%Beitrag' ...

- kleiner als 1%, dann ist das Messsystem akzeptabel
- zwischen 1% und 9%, dann ist das Messsystem in Abhängigkeit von der Anwendung, der Kosten des Messgeräts, den Reparaturkosten und anderen Faktoren ggf. akzeptabel
- größer als 9%, dann ist das Messsystem nicht akzeptabel und sollte verbessert werden

Zusätzlich werden die einzelnen Anteile der Streuung quadriert betrachtet (Achtung: die Summe der Quadrate ist größer als 100%, das Quadrat einer Summe ist größer als die Summer der Quadrate !):

Ist der Wert in der Spalte '%Streuung in Untersuchung (%Toleranz, %Prozess)' (kurz: '%SU') ...

- kleiner als 10%, dann ist das Messsystem akzeptabel
- zwischen 10% und 30%, dann ist das Messsystem in Abhängigkeit von der Anwendung, der Kosten des Messgeräts, den Reparaturkosten und anderen Faktoren ggf. akzeptabel

- größer als 30%, dann ist das Messsystem nicht akzeptabel und sollte verbessert werden

Im Beispiel hier sind die Werte für '%Beitrag' 7.76 und für '%SU' $27,86$, es handelt sich also um ein Messsystem, dass 'in Abhängigkeit von der Anwendung, der Kosten des Messgeräts, den Reparaturkosten und anderen Faktoren ggf. akzeptabel' ist.

Des weiteren wird eine 'Anzahl Kategorien' angegeben. Die Anzahl der Kategorien ist die Anzahl Klassen von Teilen, die durch das Messsystem unterscheidbar sind. Konkret ist diese Anzahl die der nicht überlappenden Konfidenzintervalle, die die Spannweite der Produktstreuung umfasst. Diese können auch als die Anzahl von Gruppen in den Prozessdaten interpretiert werden, die das Messsystem unterscheiden kann.

Die Berechnung ist zunächst Division der Standardabweichung der 'Teile' durch die Standardabweichung der 'R&R gesamt'; dieser Wert wird dann multipliziert mit 1.41 und die Nachkommastellen werden abgeschnitten (Entier-Funktion oder Gauss-Klammer, engl.: floor-function).

$$\text{Anzahl Kategorien} = \left\lfloor \frac{1,04233}{0,30237} * 1.41 \right\rfloor = \lfloor 4,82608063 \rfloor = 4$$

Sofern die Einteilung der Teile in gewisse Kategorien erforderlich ist, kann zudem die Fehlerwahrscheinlichkeit angegeben werden, mit der das Messsystem eine falsche Klassifikation angibt. Das geschieht anhand der gemeinsamen und der bedingten Wahrscheinlichkeiten wie folgt:

Die Wahrscheinlichkeit, dass das Teil 'schlecht' ist, es wird aber durch die Messung akzeptiert, ist:

$$P(X \notin R, Y \in R) = F_Y(b) - F_Y(a) - F_{(X,Y)}(b,b) + F_{(X,Y)}(a,b) + F_{(X,Y)}(b,a) - F_{(X,Y)}(a,a)$$

Die Wahrscheinlichkeit, dass das Teil 'gut' ist, es wird aber durch die Messung nicht akzeptiert, ist:

$$P(X \in R, Y \notin R) = F_X(b) - F_X(a) - F_{(X,Y)}(b,b) + F_{(X,Y)}(a,b) + F_{(X,Y)}(b,a) - F_{(X,Y)}(a,a)$$

Die bedingten Wahrscheinlichkeiten für die Fehlklassifikationen 'Teil ist schlecht, wird aber akzeptiert' und 'Teil ist gut, wird aber abgewiesen' sind dann:

$$P(Y \in R | X \notin R) = \frac{P(X \notin R, Y \in R)}{P(X \notin R)}, \quad P(Y \notin R | X \in R) = \frac{P(X \in R, Y \notin R)}{P(X \in R)}$$

Hierbei ist:

$$P(X \in R) = F_X(b) - F_X(a), \quad P(X \notin R) = 1 - P(X \in R)$$

$F(X,Y)$ ist die Verteilungsfunktion der bivariaten Normalverteilung (der Zufallsvektor ist entsprechend (X,Y)) mit Mittelwert $\vec{\mu} = (\theta, \theta)^T$ und der Varianz-Kovarianz-Matrix:

$$\Sigma = \begin{pmatrix} \sigma^2_{Teile} & \sigma^2_{Teile} \\ \sigma^2_{Teile} & \sigma^2_{total} \end{pmatrix}$$

$F_{(X)}$ und $F_{(Y)}$ sind die zugehörigen Randverteilungen, d.h.:

$$\begin{aligned} F_{(X,Y)}(x,y) &= \frac{1}{2\pi\sqrt{\det(\Sigma)}} \int_{-\infty}^{y} \int_{-\infty}^{x} e^{\frac{1}{2}\vec{\omega}^T \Sigma^{-1} \vec{\omega}} du_1 du_2, \quad \vec{\omega} = (u_1 - \theta, u_2 - \theta)^T \\ F_{(X)}(x) &= \frac{1}{\sqrt{2\pi\sigma^2_{Teile}}} \int_{-\infty}^{x} e^{\frac{-(u-\theta)^2}{2\sigma^2_{Teile}}} du \\ F_{(Y)}(y) &= \frac{1}{\sqrt{2\pi\sigma^2_{total}}} \int_{-\infty}^{y} e^{\frac{-(u-\theta)^2}{2\sigma^2_{total}}} du \end{aligned}$$

Als Unterstützung der Analyse wird das Ergebnis zudem grafisch ausgegeben:

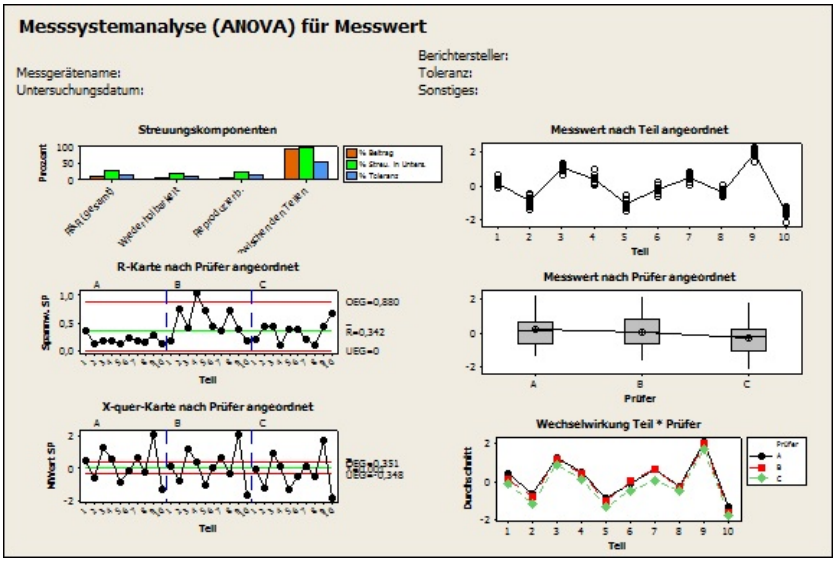


Figure 92: Grafik MSA Typ 2

In der Grafik sind folgende Auffälligkeiten festzustellen:

- die erste Teilgrafik stellt die Streuungsaufteilung als Säulendiagramm dar. Hier ist abzuschätzen, dass das Messsystem nach den Vorgaben der AIAG aktuell nur bedingt brauchbar ist.
- die R-Karte zeigt, dass Prüfer 'B' eine recht hohe Streuung in den Teilen gemessen hat. Das bedeutet, dieser Prüfer hat die gleichen Teile sehr unterschiedlich gemessen. Dass die Teile wesentlich besser wiederholt gemessen werden können, zeigen die relativ guten Streuungswerte der Prüfer 'A' und 'C'. Das Messsystem kann also wesentlich verbessert werden, wenn Prüfer 'B' die Wiederholbarkeit verbessert.
- die fünfte und sechste Grafik ('Messwert nach Prüfer angeordnet', 'Wechselwirkung Teil * Prüfer') zeigen, dass Prüfer 'C' systematisch kleinere Werte gemessen hat, als die beiden anderen Prüfer. Das führt zu schlechten Streuungswerten in der Reproduzierbarkeit. Zu klären ist, ob Prüfer 'C' zu kleine, oder die beiden anderen Prüfer zu große Werte messen. Hier ist ebenfalls Potenzial zur Verbesserung des Messsystems.

Anm.: die Anzahl der Kategorien hängt sehr vom Spektrum der Messwerte der Teile ab. Vergrößert man dieses Spektrum (nimmt also z.B. ein Teil mit höheren Messwertdaten hinzu), dann erhöht sich dadurch die Anzahl der Kategorien. Insofern ist das Messwertspektrum in Bezug auf den tatsächlichen Bedarf zu wählen, und nicht nach den Messsystemmöglichkeiten. Oder deutlich: ein Kategorienfehler eines Messsystems kann durch eine solche Erhöhung des Messwertspektrums verschleiert werden.

11.0.3 Linearität und Bias

Die Linearität ist ein Maß für die Gleichheit der Messqualität der Teile im Messspektrum. Unter Bias versteht man eine systematische Messwertverschiebung in den Messungen im Messwertspektrum. Für den Bias müssen also Referenzwerte für die Teile angegeben werden.

Zur Linearität wird eine lineare Regression für die Messwerte durchgeführt. Die Referenzwerte werden dann von den Anpassungswerten abgezogen. Dadurch wird der Bias bestimmt. Die Linearität ist dann eine Regression der Differenzwerte; es wird dann ein Hypothesentest auf 'Steigung = 0' durchgeführt, um auf eine lineare Verschiebung zu testen.

Der Beispieldatensatz ist wie folgt aufgebaut:

Teil	Referenz	Messwert
1	2	2,7
1	2	2,5
1	2	2,4
1	2	2,5
1	2	2,7
1	2	2,3
1	2	2,5
1	2	2,5
1	2	2,4
1	2	2,4
1	2	2,6
1	2	2,4

Figure 93: MSA: Linearität und Bias

Die Daten werden als Einzelwertdiagramm angegeben:

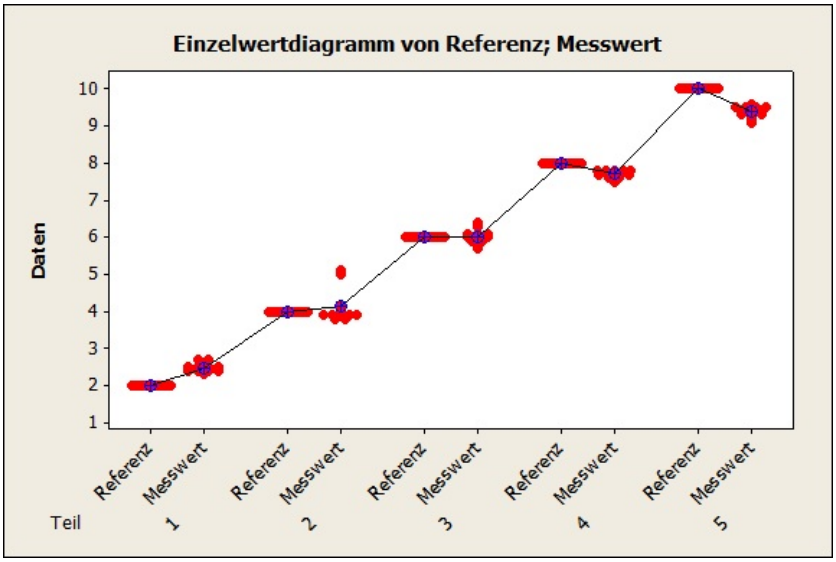


Figure 94: MSA: Linearität und Bias - Einzelwertdiagramm

Es sind schon aufgrund dieser einfachen Grafik gewisse Abweichungen sichtbar.

Die Ausgabe der Linearitäts- und Abweichungsanalyse ist dann:

```
Toleranzintervall: Messwert

Methode

Toleranzintervalltyp                      Beidseitig
Konfidenzniveau                           95%
Prozent der Grundgesamtheit in Intervall  95%

Statistik

Variable   N  Mittelwert  StdAbw
Messwert  90       0,001   1,031

95%-Toleranzintervall, mit Normalverteilungsmethode

Variable  Untergrenze  Obergrenze
```

```
Messwert       -2,321       2,324

95%-Toleranzintervall, mit verteilungsfreier Methode

Erreichte
Variable  Untergrenze  Obergrenze  Konfidenz
Messwert       -2,160       2,260       94,3
```

Eine grafische Übersicht wird ebenfalls ausgegeben:

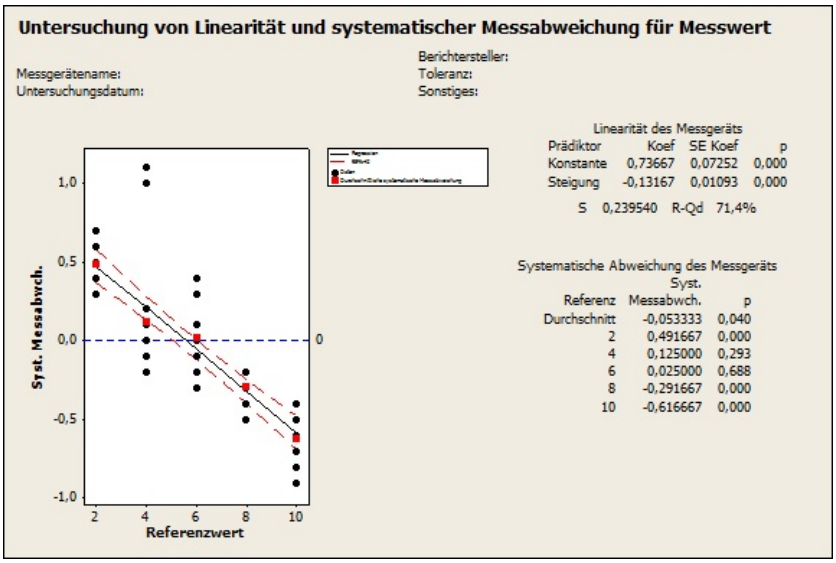


Figure 95: MSA: Linearität und Bias - Grafik

Das Ergebnis ist folgendermaßen zu interpretieren:

- die Grafik zeigt eine systematische Verschiebung der Messwerte; die kleinen Teile werden zu groß, die großen Teile zu klein gemessen.
- dieser Effekt ist signifikant (p-Wert Steigung $= 0\%$), die Steigung selbst ist -0.13167.

- auch sind die Messwerte gegenüber den Referenzwerten systematisch signifikant verschoben, der p-Wert ist ebenfalls 0%, die Verschiebung ist 0.73667.
- die systematische Messwertabweichung ist für die Punkte 4 und 6 nicht signifikant.
- das Konfidenzintervall ist das der Regression. Man sieht in der Grafik diverse Messungen außerhalb dieses Intervalles.

11.0.4 Exkurs: Messsystemanalyse für attributive Daten

Eine Messsystemanalyse ist auch für attributive Daten, also z.B. 'OK / Nicht OK'-Zuweisungen oder Aussortierungen, möglich. Hier wird unterschieden, ob die Messungen automatisch durchgeführt werden, oder ob Prüfer mit ihren subjektiven Einschätzungen oder Meinungen als 'Messgeräte' dienen. Die erste Methode heißt analytisch, die zweite Prüferübereinstimmung.

Es werden dann Bias, Linearität sowie Wiederholbarkeit und Reproduzierbarkeit bestimmt.

Für eine analytische Methode ist folgendes zu beachten:

Es sind mindestens acht Teile erforderlich, damit eine Messsystemanalyse für attributive Daten durchgeführt werden kann. Das Teil mit dem kleinsten Referenzwert muss null Annahmen und das Teil mit dem größten Referenzwert muss die maximale Anzahl der möglichen Annahmen aufweisen. Bei der AIAG-Methode zum Prüfen der systematischen Messabweichung müssen mindestens sechs Teile mit mehr als null Annahmen und weniger als 20, der maximalen Anzahl von Annahmen, vorhanden sein. Für die Regressionsmethode können mehr als sechs Teile zwischen den zwei extremen Referenzwerten vorhanden sein.

Wenn Sie die untere Toleranzgrenze angeben, muss das Teil mit dem kleinsten Referenzwert 0 Annahmen und das Teil mit dem höchsten Referenzwert die maximale Anzahl der möglichen Annahmen aufweisen. Für die untere Toleranzgrenze gilt somit: Mit steigenden Referenzwerten nimmt die Anzahl der Annahmen zu. Wenn Sie die obere Toleranzgrenze angeben, muss das Teil mit dem kleinsten Referenzwert die maximale Anzahl der möglichen Annahmen und das Teil mit dem höchsten Referenzwert 0 Annahmen aufweisen. Für die obere Toleranzgrenze gilt

somit: Mit steigenden Referenzwerten nimmt die Anzahl der Annahmen ab.

Sie können die Anzahl der Versuche als Konstante oder Datenspalte eingeben. Wenn die Anzahl der Versuche nicht für alle Teile identisch ist, müssen Sie eine Spalte mit der Anzahl der Versuche für jedes Teil eingeben.

Die Anzahl der Versuche für jedes Teil muss größer als oder gleich 15 sein. Wenn Sie die AIAG-Methode (Standardeinstellung) zum Prüfen der systematischen Messabweichung des Messsystems verwenden, müssen Sie über mindestens 20 Versuche pro Teil verfügen.

Als Beispiel sei folgender Fall betrachtet:

Ein Fahrzeughersteller möchte die systematische Messabweichung und Wiederholbarkeit eines automatisierten Attributmesssystems messen. Das System verfügt über eine untere Toleranzgrenze von -0,020 und eine obere Toleranzgrenze von 0,020. Der Hersteller ließ das Messsystem 20-mal von 10 Teilen durchlaufen. Die Teile besitzen Referenzwerte in Intervallen von 0,005 von -0,050 bis -0,005.

Folgende Daten seien nun gegeben:

Teilenummer	Referenz	Annahmen
1	-0,050	0
2	-0,045	1
3	-0,040	2
4	-0,035	5
5	-0,030	8
6	-0,025	12
7	-0,020	17
8	-0,015	20
9	-0,010	20
10	-0,005	20

Figure 96: MSA: attributive Daten - analytisch

Die MSA liefert dann folgendes Ergebnis:

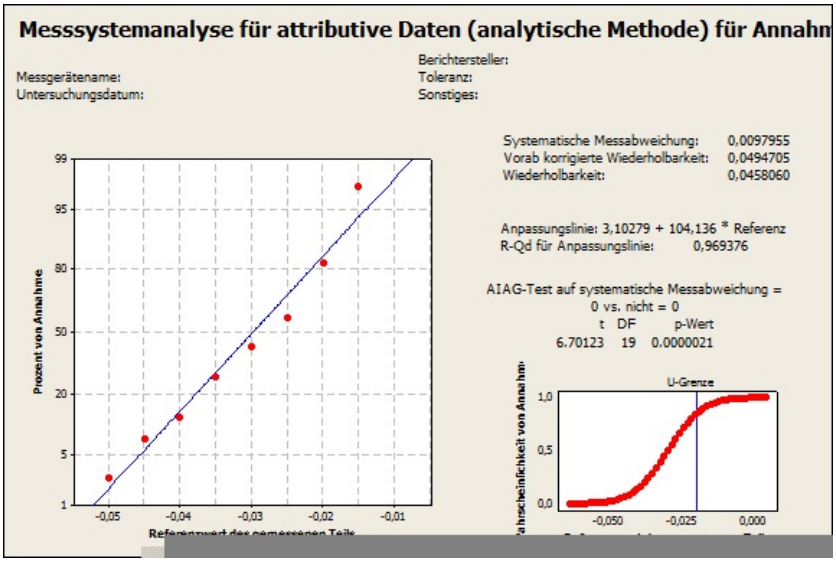


Figure 97: MSA: attributive Daten - analytisch, Ergebnis

Die systematische Messabweichung in der Messsystemanalyse für attributive Daten (analytisch) beträgt $0,0097955$ und die korrigierte Wiederholbarkeit $0,0458060$. Der Test auf systematische Messabweichung deutet an, dass sich die systematische Messabweichung signifikant von 0 (p-Wert $= 0\%$) unterscheidet. Das bedeutet, es liegt für die Daten hier höchstwahrscheinlich im attributiven Messsystem eine systematische Abweichung vor.

Für eine Analyse einer Prüferübereinstimmung sei folgendes Fallbeispiel betrachtet:

Ein Unternehmen im Bildungsbereich, das Tests abhält, schult fünf neue Prüfer für den schriftlichen Teil der standardisierten Prüfung für Aufsätze in der zwölften Klasse. Dabei muss die Fähigkeit der Prüfer untersucht werden, dass sie Aufsätze in Übereinstimmung mit den Standards einstufen können. Jeder Prüfer hat fünfzehn Aufsätze auf einer Fünf-Punkte-Skala (-2, -1, 0, 1, 2) bewertet.

Die Daten zu diesem Beispiel sind:

Prüfer	Stichprobe	Einstufung	Attribut
Schmidt	1	2	2
Mayer	1	2	2
Hoffmann	1	2	2
Dreger	1	1	2
Henschel	1	2	2
Schmidt	2	-1	-1
Mayer	2	-1	-1
Hoffmann	2	-1	-1
Dreger	2	-2	-1
Henschel	2	-1	-1
Schmidt	3	1	0
Mayer	3	0	0
Hoffmann	3	0	0

Figure 98: MSA: attributive Daten - Prüferübereinstimmung, Daten

Die Ausgabe des Ergebnisses der Analyse ist rein grafisch:

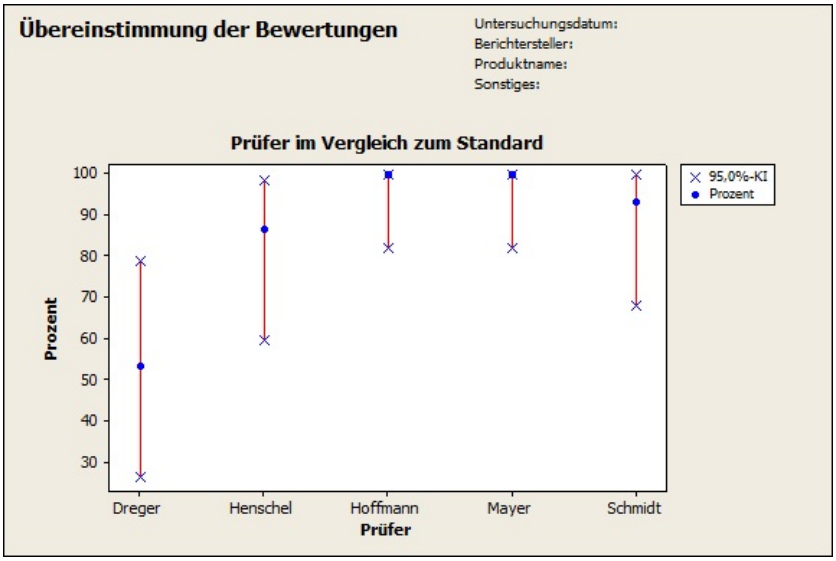


Figure 99: MSA: attributive Daten - Prüferübereinstimmung, Ergebnis

12 GLM

Das verallgemeinerte lineare Modell wird verwendet, um univariate Varianzanalysen mit balancierten und nicht balancierten Designs, Kovarianzanalysen und

Regression für jede Antwortvariable auszuführen.

Die Berechnungen werden mit Hilfe eines Regressionsverfahrens durchgeführt. Eine Designmatrix mit 'vollem Rang' wird aus den Faktoren und Kovariaten (also attributiven und metrischen Regressoren) gebildet, und für jede Antwortvariable wird eine Regression in den Spalten der Designmatrix durchgeführt.

Sie müssen ein hierarchisches Modell angeben. Wenn ein hierarchisches Modell einen signifikanten Wechselwirkungsterm enthält, müssen sämtliche Wechselwirkungen niedriger Ordnung sowie alle Haupteffekte, die die Wechselwirkung umfasst, im Modell erscheinen.

Faktoren können gekreuzt oder geschachtelt (vgl. Abschn. 11) sowie fest oder zufällig sein. Kovariaten können miteinander oder mit Faktoren gekreuzt sein oder innerhalb von Faktoren geschachtelt sein.

Quadratische oder andere nicht-lineare Regressionen werden durch Umrechnung der Variablenwerte erreicht.

Das Ziel der GLM-Berechnung ist es, signifikante (und nur solche) Koeffizienten als Vorfaktoren zu den Faktoren bzw. Kovariaten zu ermitteln. Diese beschreiben dann das Modell. Es wird zudem ein Anpassungsgrad ermittelt.

Das Vorgehen wird an einem Fallbeispiel. Die Daten dazu sind wie folgt:

Reifenhersteller	Profiltiefe	ABS	Bremsweg
GT	10,0	An	20,4
GT	1,5	Aus	24,7
MX	1,5	An	23,2
MX	10,0	An	25,3
LS	10,0	Aus	27,1
GT	10,0	An	20,4
LS	1,5	An	25,3
LS	1,5	Aus	30,8
LS	10,0	Aus	28,0
MX	10,0	Aus	29,3
MX	10,0	Aus	27,4
GT	10,0	Aus	26,5
LS	1,5	An	25,0
GT	1,5	Aus	25,9
MX	10,0	An	21,3
MX	1,5	An	23,2
GT	10,0	Aus	28,4
MX	1,5	Aus	28,0
LS	1,5	Aus	28,0
LS	10,0	An	22,3
GT	1,5	An	22,0
MX	1,5	Aus	27,4
GT	1,5	An	19,8
LS	10,0	An	25,3

Figure 100: GLM: Daten

Als Intervallplot sehen die Daten wie folgt aus:

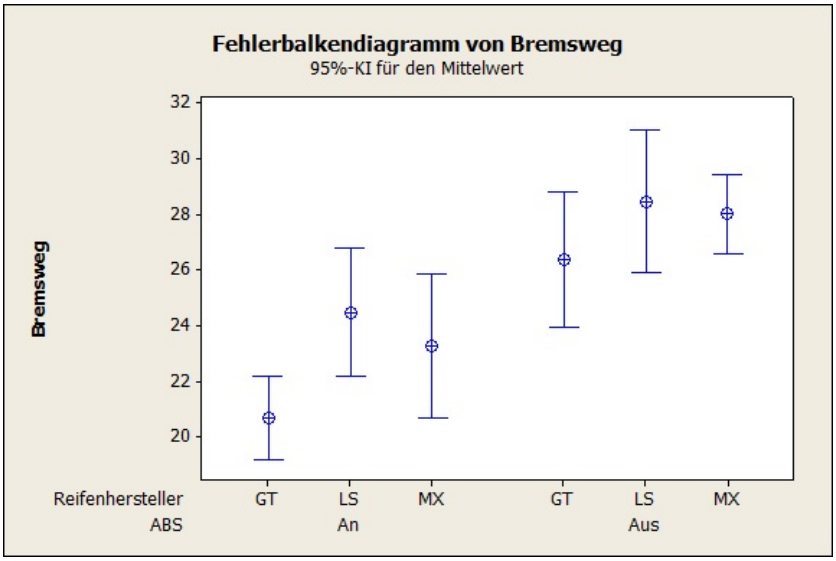


Figure 101: GLM: Intervallplot der Daten

Es soll herausgefunden werden, welche der angegebenen Faktoren einen Einfluss auf den Bremsweg haben bzw. ob es signifikante Wechselwirkungen gibt.

Die Analyse für diesen Datensatz liefert als Ergebnis:

Allgemeines lineares Modell: Bremsweg vs. ABS; Profiltiefe; Reifenherste

Faktor	Typ	Stufen	Werte
ABS	fest	2	An; Aus
Profiltiefe	fest	2	1,5; 10,0
Reifenhersteller	fest	3	GT; LS; MX

Varianzanalyse für Bremsweg unter Verwendung von korrigierter SS für Tests

Quelle	DF	Seq SS	Kor SS	Kor MS	F	p
ABS	1	140,167	140,167	140,167	70,67	0,000
Profiltiefe	1	0,107	0,107	0,107	0,05	0,821
Reifenhersteller	2	37,316	37,316	18,658	9,41	0,003
ABS*Profiltiefe	1	1,215	1,215	1,215	0,61	0,449
ABS*Reifenhersteller	2	2,986	2,986	1,493	0,75	0,492
Profiltiefe*Reifenhersteller	2	6,656	6,656	3,328	1,68	0,228
ABS*Profiltiefe*Reifenhersteller	2	2,573	2,573	1,286	0,65	0,540
Fehler	12	23,800	23,800	1,983		
Gesamt	23	214,818				

S = 1,40831 R-Qd = 88,92% R-Qd(kor) = 78,77%

Term	Koef	SE Koef	t	p
Konstante	25,2083	0,2875	87,69	0,000
ABS				
An	-2,4167	0,2875	-8,41	0,000
Profiltiefe				
1,5	0,0667	0,2875	0,23	0,821
Reifenherste				
GT	-1,6958	0,4065	-4,17	0,001
LS	1,2667	0,4065	3,12	0,009
ABS*Profiltiefe				

```
An   1,5                        0,2250   0,2875   0,78  0,449
ABS*Reifenherste
An  GT                         -0,4458   0,4065  -1,10  0,294
An  LS                          0,4167   0,4065   1,02  0,326
Profiltiefe*Reifenherste
1,5        GT                 -0,4792   0,4065  -1,18  0,261
1,5        LS                  0,7333   0,4065   1,80  0,096
ABS*Profiltiefe*Reifenherste
An   1,5        GT              0,4375   0,4065   1,08  0,303
An   1,5        LS             -0,3500   0,4065  -0,86  0,406

Ungewöhnliche Beobachtungen für Bremsweg

SE              Std.
Beob  Bremsweg  Anpassung  Anpassung  Residuum  Resid.
4   25,3000    23,3000     0,9958    2,0000    2,01 R
15   21,3000    23,3000     0,9958   -2,0000   -2,01 R

R kennzeichnet eine Beobachtung mit einem großen standardisierten Residuum.
```

Die Analyse zeigt, dass lediglich zwei Faktoren signifikant 'Bremsweg' beeinflussen, 'ABS' und 'Reifentyp'.

Diese Terme werden aus dem Modell genommen.

Anmerkungen:

- Die Herausnahme von Termen erfolgt eigentlich Schritt für Schritt. Man beginnt mit den nicht-signifikanten Termen höchster Wechselwirkungsordnung und reduziert so das Modell bis auf die wesentlichsten Faktoren bzw. Kovariaten.
- es kann passieren, dass ein Modell nach der ersten Rechnung keine signifikanten Terme enthält. Dann verfährt man dennoch wie eben beschrieben. Es kann dann sein, dass nach der Reduktion um die Terme höherer Wechselwirkungen das Modell dann signifikante Terme 'bekommt'.

Das GLM wird anschließend nochmal gerechnet (die Reduktion ist hier verkürzt wiedergegeben):

```
Allgemeines lineares Modell: Bremsweg vs. ABS; Reifenhersteller

Faktor            Typ   Stufen  Werte
ABS               fest       2  An; Aus
Reifenhersteller  fest       3  GT; LS; MX


Varianzanalyse für Bremsweg unter Verwendung von korrigierter SS für Tests

Quelle            DF   Seq SS   Kor SS   Kor MS      F      p
ABS                1  140,167  140,167  140,167  75,08  0,000
Reifenhersteller   2   37,316   37,316   18,658   9,99  0,001
Fehler            20   37,336   37,336    1,867
Gesamt            23  214,818


S = 1,36631   R-Qd = 82,62%   R-Qd(kor) = 80,01%


Term             Koef  SE Koef      t      p
Konstante     25,2083   0,2789  90,39  0,000
ABS
An            -2,4167   0,2789  -8,67  0,000
Reifenherste
GT            -1,6958   0,3944  -4,30  0,000
LS             1,2667   0,3944   3,21  0,004
```

Die Interpretation des Ergebnisses ist wie folgt:

- Es sind zwei Terme signifikant, 'ABS' und 'Reifenhersteller'
- der Anpassungsgrad des Modells ist mit 82% sehr gut; insofern sind Modellberechnungen und -interpretationen wie die Folgenden möglich.

- die Konstante ist $25,21$. Das ist der über alle Einstellungen mittlere Bremsweg (Mittelwert über alle Daten)
- ist 'ABS' 'an', dann verkürzt sich der Bremsweg ausgehend vom Mittel (also der Konstante) um 2.4
- die Summe aller Koeffizienten muss Null ergeben. Insofern verlängert sich 'Bremsweg' um 2.4, wenn 'ABS' 'aus' ist. Dieser Wert wird - leider - im Output unterdrückt.
- der Reifentyp 'GT' verkürzt 'Bremsweg' um 1.7, 'LS' verlängert um 1.2, folglich verlängert der Typ 'MX' den Bremsweg um ca. 0.5
- möchte man einen Bremsweg von (im Mittel) 23.2 erreichen, so kann man das mit den Einstellungen 'ABS' an und dem Reifentyp 'MX' (die Rechnung ist: $25.2 - 2.4 + 0.4$)
- möchte man einen Bremsweg von unter 25 erreichen, dann muss 'ABS' 'an' sein.

Die einzelnen Bremswege im Mittel können folgender Tabelle, die die mittleren Werte für die einzelnen Faktorkombinationen ausweist, entnommen werden.

25,2083		
	-2,4167	2,4167
-1,6958	21,0958	25,9292
1,2667	24,0583	28,8917
0,4291	23,2207	28,0541

Solche Zielgrößeneinstellungen sind nur gültig, wenn das errechnete Modell gültig ist, also nur signifikante Terme enthält und einen genügend hohen Anpassungsgrad besitzt.

Es ist noch das Residuum auf Normalverteilung zu testen (es gelten für das GLM die gleichen Annahmen wie für die ANOVA, vgl. Abschn. 9.1.1).

Zunächst ist eine Residuumanalyse noch grafisch sinnvoll:

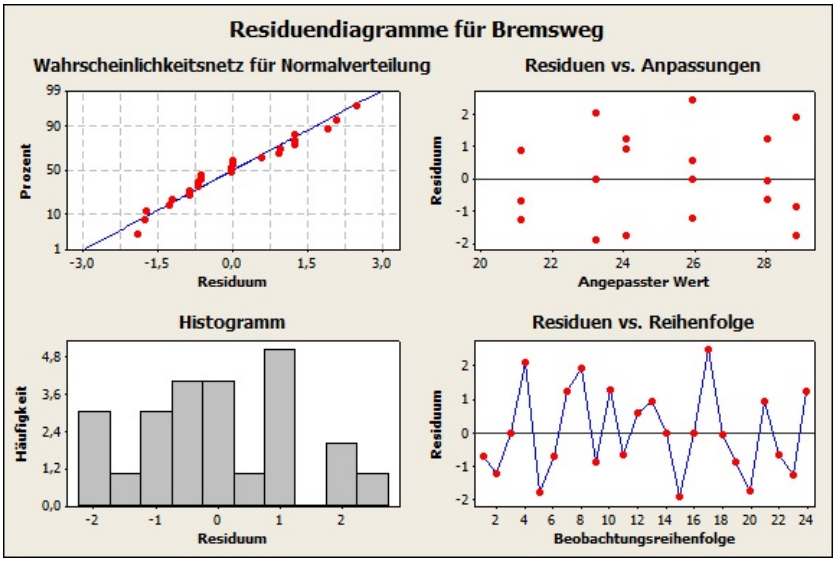


Figure 102: GLM: Residuum-Grafiken

Der Test für das Residuum auf Normalverteilung ist positiv:

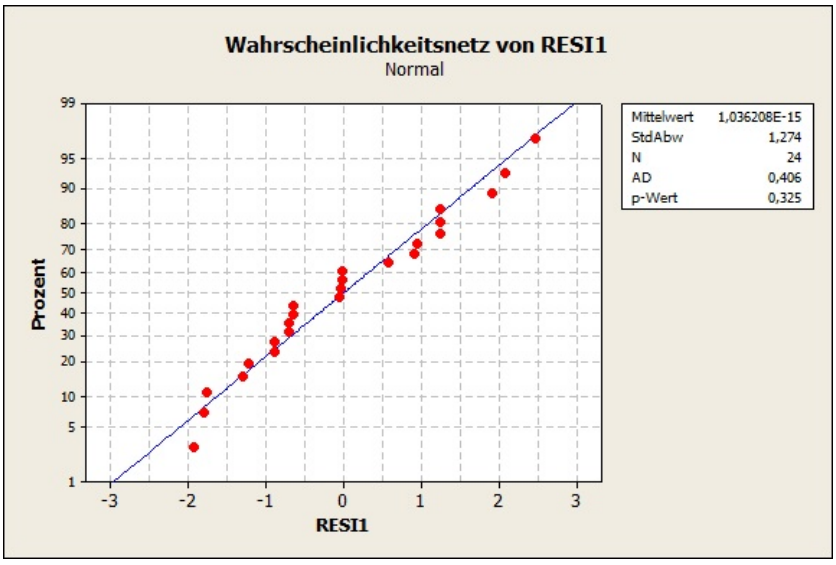


Figure 103: GLM: Test Residuum auf Normalverteilung

13 Anwendung GLM: Design of Experiments

Ziel von Experimenten ist es, Zusammenhänge - im Sinne von Gesetzmäßigkeiten - herauszufinden. Der einfachst mögliche Zusammenhang zwischen zwei steigen Variablen wäre dann ein linearer. Diesen kann man mit Hilfe einer linearen Regression berechnen.

Nun ist zudem gefordert, dass ein solcher Zusammenhang mit möglichst wenig Versuchen und dennoch statistisch möglichst sicher (also konkret: mit hoher Signifikanz und Trennschärfe) aufgezeigt wird.

13.1 Full Factorial Designs

Dazu betrachte man zunächst den einfachen Fall zweier Variablen x und y. Um einen linearen Zusammenhang zwischen x und y zu ermitteln benötigt man zwei Versuche, für zwei unterschiedliche Einstellungen von x, z.B. x_U und x_O.

Durch die beiden so ermittelten Punkte im Koordinatensystem (x_U, y_U) und (x_O, y_O) kann man genau eine Gerade legen; diese wäre dann als das lineare Modell zu interpretieren.

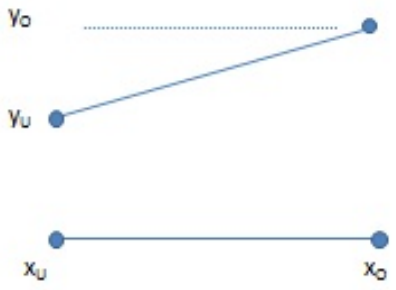


Figure 104: Min. Anzahl Messungen 1 Faktor, lineares Modell

Es fehlt lediglich der Zufall. Es ist so kein Residuum berechnet, und somit hat man so kein Wissen über den Anpassungsgrad der so gefundenen Funktion.

Der Anpassungsgrad wird über die Abweichung der realen Messwerte von der theoretischen, bestmöglich angepassten Geraden, berechnet. Diese Abweichung wird wiederum über die Standardabweichung gemessen. Die Standardabweichung muss homogen für alle (beide) x-Punkte sein, das ist ebenfalls zu prüfen; zudem muss das Schema balanciert sein. Insofern benötigt man für jeden x-Punkt mind. zwei Messungen.

Zur Bestimmung der Geraden und des Anpassungsgrades sind dann also insg. 4 Messungen erforderlich.

Die x-Werte werden dabei nicht verändert, sondern fest eingestellt. Diese sind auch für den Fall, dass x kontinuierlich ist, lediglich Faktoren.

Kommt nun noch ein zweiter Faktor (x_2, der erste Faktor sei nun x_1) hinzu, dann hat man die Situation von 2 unabhängigen Dimensionen. Um ein lineares Modell zu errechnen, genügen dann 3 Punkte (durch drei Punkte kann man eine Ebene legen). Die Variationen des zweiten Faktors sollten jedoch für jede Einstellung des ersten Faktors - zunächst - durchgeführt werden. Das ergibt dann insg. 4 Punkte: $(x_{1_U}, x_(2_U)$, $(x_{1_U}, x_(2_O)$, $(x_{1_O}, x_(2_U)$, $(x_{1_O}, x_(2_O)$.

Nach dieser Überlegung hätte man sozusagen einen Punkt zu viel. Jedoch gibt es den Effekt der Wechselwirkung (vgl. die Ausführungen in 12). Mit Hilfe dieses zusätzlichen Punktes kann also eine Wechselwirkung der Form $x_1 * x_2$ berechnet werden (konkret wird zu diesem Produkt ein Koeffizient bestimmt).

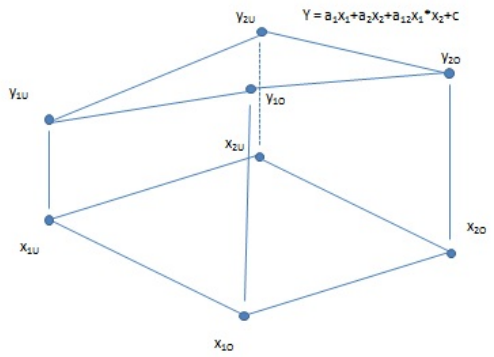


Figure 105: 2-Faktoren-Schema

Zu erkennen ist anhand dieser Erläuterungen, dass die Anzahl der Punkte mit der Anzahl der Koeffizienten übereinstimmt.

Bei drei Faktoren x_1, x_2 und x_3 erhält man insg. acht Punkte (für jede der beiden Faktoreinstellungen von x_3 sind ja die vier Kombinationen bzgl. der Faktoren x_1 und x_2, nämlich $(x_{1_U}, x_(2_U)$, $(x_{1_U}, x_(2_O)$, $(x_{1_O}, x_(2_U)$, $(x_{1_O}, x_(2_O)$ definiert. So erkennt man die Gesetzmäßigkeit zwischen der Anzahl der Faktoren mit zwei Ausprägungen und der Anzahl Punkte:

- die Anzahl der (Mess-) Punkte im Schema ist gleich 2 hoch der Anzahl der Faktoren

Das Schema der Punkte, an denen gemessen wird, wird ins geometrische übertragen. Für 3 Faktoren ergibt sich gerade ein Würfel mit $2^3 = 8$ Ecken und somit mit 8 Koeffizienten; das eindeutig bestimmbare lineare Modell lautet dann:

$$y = a_1x_1 + a_2x_2 + a_3x_3 + a_{12}x_1 * x_2 + a_{23}x_2 * x_3 + a_{13}x_1 * x_3 + a_{123}x_1 * x_2 * x_3 + c$$

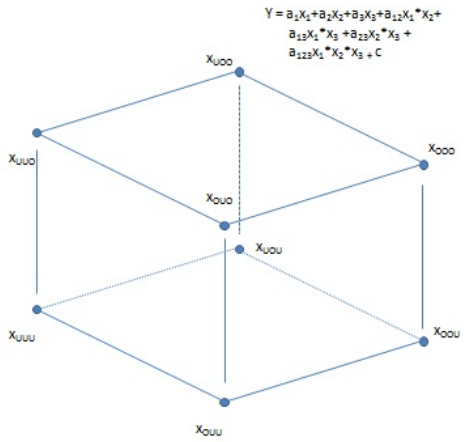


Figure 106: 3-Faktoren-Schema

Dieses Schema benötigt dann mindestens 16 Messungen (mind. zwei Messungen je Punkt), um auch genügend statistische Informationen zu ermitteln. Ein solches Schema (Design) heißt auch ein Full Factorial Design.

Die Grundlage für das Experimentieren ist die Variation in den x-Werten, die dann fixiert werden. Die Variation in den Einstellungen wird also minimiert.

Der einfachste Ansatz wäre es, einen t-Test für alle Einstellungen durchzurechnen, oder eine ANOVA. Jedoch damit erhält man noch kein Modell. Ziel des Designs hier ist es, ein Modell abzuleiten und dann ggf. eine Zielgrößenoptimierung (evtl. auch für mehrere Zielgrößen, und diese dann mit gegebenem Zielkonflikt) durchzuführen. Für ein Modell gelten die Regeln des GLM (vgl. Abschn. 12).

Der t-Test-Gedanke provoziert folgenden Hinweis: die gesuchte Funktion ist linear, also stetig. Wählt man die Punkte x_U und x_O bei einem 1-Faktor-Design sehr dicht nebeneinander, dann wird ggf. keine Signifikanz bzgl. des Unterschieds der Mittelwerte y_U und y_O nachgewiesen werden können. Insofern muss man mit der Wahl der Distanz der Punkte zueinander relativ großzügig sein.

In der Versuchsplanung werden die Faktoren i.d.R. einfach mit großen Buchstaben bezeichnet, also A, B, C. Wechselwirkungen sind als Produkt von Faktoren gegeben durch $AB, ACD, ABEF$.

Die Eckpunkte in einem vollfaktoriellen Versuchsplan werden geometrisch als die Ecken des Einheitswürfels aufgefasst - und entsprechend kodiert. In einem 3-faktoriellen Design wäre $(-1, 1, -1)$ dann der Punkt (x_U, y_O, Z_U). Weiter abkürzend verwendet man dann nur noch das Vorzeichen. Aus $(-1, 1, -1)$ wird dann $(-, +, -)$.

13.2 Reduziertes Design, Teilfaktorieller Versuchsplan

Eine Frage wirft sich auf in Bezug, wenn Wechselwirkungsterme ggf. nicht in Betracht gezogen werden sollen: kann man dann das Design so reduzieren, dass dennoch alle Haupteffekte aufgezeigt werden können ?

Wenn nur die Koeffizienten der Haupteffekte berechnet werden sollen, dann braucht man für ein 3-faktorielles Design statt der oben angegebenen 8 lediglich 4 Punkte. Eine weitere Überlegung fordert: zu jedem Faktor muss eine Variation im Design enthalten sein. Wenn man beispielsweise in dem Würfel-Design für 3 Faktoren die vier unteren oder vorderen Punkte weglässt, dann fehlt eine komplette Variation, in diesen Fällen für x_3 bzw. x_2.

Das bedeutet, von jedem Koordinaten-Typ, im Schema oben 'U' bzw. 'O', muss zu jedem Faktor eine Ausprägung vorhanden sein.

Das führt auf zwei Möglichkeiten; diese sind aus der nachfolgenden Zeichnung herauszulesen:

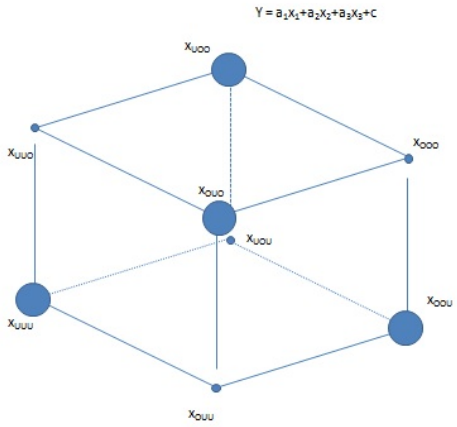


Figure 107: Reduziertes 3-Faktoren-Schema

Entweder die vier jetzt dick markierten Punkte bilden das reduzierte Design, oder die anderen vier klein gebliebenen Punkte werden für die Experimente verwendet.

Mit diesem reduzierten Schema können nun keine Wechselwirkungen nachgewiesen werden.

Bei einem so reduzierten Design spricht man dann von einem teil-faktoriellen Versuchsplan

13.3 Saturierung, Blockvariablen und Blöcke, Zentralpunkte, Auflösung, Aliasstruktur, Vermengung, Faltung

Jeder Punkt in einem Design ist Stellvertreter für einen Term im Regressionsmodell. Können alle Terme des (linearen) Ansatzes berechnet werden, dann spricht man von einem saturierten (gesättigten) Modell.

Sofern man ein Modell benötigt, das weitere Terme berechnet (z.B. für Kovariaten, also zufällige Variablen, die nicht steuerbar sind, dennoch aber im Modell berechnet werden sollen), dann benötigt man konsequenter Weise mehr Punkte.

Die Distanz zwischen den Punkten ist entscheidend für das Erkennen von Effekten. Als zusätzliche Punkte sind dann solche zu wählen, die zu den Standard-

Punkten größtmöglichen Abstand besitzen. Bei einem Würfel wäre das der Mittelpunkt.

Bei attributiven Variablen sind nun Zwischenstellen nicht möglich (sofern man Mischungen ausschließt). Dann ist ein solcher Mittelpunkt im Würfel für die Experimente nicht wählbar. Es ist aber ein Mittelpunkt für jedes attributive Merkmal wählbar.

Man spricht bei einem solchen attributiven Merkmal nun von einer Blockvariablen (gewisse Kanten des Würfels sind nicht metrisch, dadurch werden gewisse Flächen 'geblockt'), bzw. einem zu jedem Attribut zugehörigen Block.

Je Block ist dann ein Zentraler Punkt sinnvoll definierbar.

Es gilt:

- besitzt das Modell keine Blöcke und keine zentralen Punkte, dann ist die Anzahl der berechenbaren Terme im Modell gleich der Anzahl der (eindeutigen) Eckpunkte -1
- besitzt das Modell zentrale Punkte, jedoch keine Blöcke, dann ist die Anzahl der berechenbaren Terme im Modell gleich der Anzahl der (eindeutigen) Eckpunkte

Effekte, die nicht unabhängig geschätzt werden können, werden vermengt genannt.

Vermengungen treten auf, wenn Sie einen teil-faktoriellen Versuchsplan verwenden, da nicht für alle Kombinationen von Faktorstufen ein Durchlauf ausgeführt wird. Wenn Faktor A beispielsweise mit der Drei-Faktor-Wechselwirkung BCD vermengt ist, entspricht der geschätzte Effekt für A der Summe des Effekts von A und des Effekts von BCD.

Solche Effekte weisen eine sog. Aliasstruktur auf. Die Aliasstruktur beschreibt die Vermengung innerhalb eines Versuchsplans.

Als Angabe, wann bei einem Design gewissen Wechselwirkungen berechnet werden können, dient die sog. Auflösung der Versuchspläne.

Die Auflösung beschreibt also das Ausmaß, in dem Effekte in einem teil-faktoriellen Versuchsplan eine Aliasstruktur mit anderen Effekten aufweisen (Vermengung).

Grundsätzlich wird die Verwendung eines teil-faktoriellen Versuchsplans mit der für die erforderliche Zerlegung höchstmöglichen Auflösung angestrebt. Beispielsweise sollten Sie in der Regel lieber einen Versuchsplan auswählen, in dem Haupteffekte mit Drei-Faktor-Wechselwirkungen (Auflösung IV) vermengt sind, als einen Versuchsplan, in dem Haupteffekte mit Zwei-Faktor-Wechselwirkungen (Auflösung III) vermengt sind.

Versuchspläne mit Auflösung III, IV und V sind am weitesten verbreitet. Diese sind wie folgt definiert:

Auflösung III: die Haupteffekte weisen keine Aliasstruktur zu anderen Haupteffekten, jedoch eine Aliasstruktur zu Zwei-Faktor-Wechselwirkungen auf.

Auflösung IV: Haupteffekte weisen keine Aliasstruktur zu anderen Haupteffekten oder Zwei-Faktor-Wechselwirkungen auf, einige Zwei-Faktor-Wechselwirkungen weisen jedoch eine Aliasstruktur zu anderen Zwei-Faktor-Wechselwirkungen auf, und Haupteffekte weisen eine Aliasstruktur mit Drei-Faktor-Wechselwirkungen auf.

Auflösung V: Haupteffekte oder Zwei-Faktor-Wechselwirkungen weisen keine Aliasstruktur zu anderen Haupteffekten oder Zwei-Faktor-Wechselwirkungen auf, Zwei-Faktor-Wechselwirkungen weisen jedoch eine Aliasstruktur zu Drei-Faktor-Wechselwirkungen auf, und Haupteffekte weisen eine Aliasstruktur mit Vier-Faktor-Wechselwirkungen auf.

Durch sog. Falten kann eine Vermengung von Faktoren und deren Wechselwirkungen verringert werden. Vermengungen treten in teil-faktoriellen Versuchsplänen auf, in denen mindestens ein Effekt nicht getrennt von den anderen Effekten geschätzt werden kann (Aliasstruktur).

Der Schlüssel zur Aliasstruktur liegt im sog. Identitätsausdruck (hier: $I + ABCDE$).

Man betrachte zum Beispiel die Identität plus eine Fünffach-Wechselwirkung $I + ABCDE$. Wenn bestimmt werden soll, welche Effekte vermengt sind, multiplizieren Sie den betreffenden Term mit dem Identitätsausdruck, und streichen

die quadratischen und ggf. weitere höhere Terme.

So ermitteln Sie z. B. die Terme, mit denen BC vermengt ist:

$$(BC)(I + ABCDE) = BC + AB2C2DE = BC + ADE$$

BC und ADE sind also miteinander vermengt.

Durch das sog. Falten kann eine Vermengung von Faktoren und deren Wechselwirkungen verringert werden.

Vermengungen treten in teil-faktoriellen Versuchsplänen auf, in denen mindestens ein Effekt nicht getrennt von den anderen Effekten geschätzt werden kann. Die nicht gesondert behandelbaren Effekte weisen dann eine sog. Aliasstruktur auf.

Versuchspläne mit der Auflösung IV können durch das Falten von Versuchsplänen der Auflösung III erstellt werden.

Angenommen, es wird für den Faktor A gefaltet. Dann weisen alle Zwei-Faktor-Wechselwirkungen von A keine anderen Haupteffekte bzw. Zwei-Faktor-Wechselwirkungen auf.

Wenn für alle Faktoren gefaltet wird, dann weisen alle Haupteffekte weder andere Haupteffekte noch Zwei-Faktor-Wechselwirkungen auf.

Beim Falten eines Versuchsplans verkürzt sich normalerweise die Aliasstruktur, da gefaltet weniger Terme miteinander vermengt werden. Insbesondere werden beim Falten für alle Faktoren alle Ausdrücke in der Aliasstruktur mit ungerader Buchstabenanzahl weggelassen. Beim Falten für einen Faktor werden alle Ausdrücke mit diesem Faktor in der Aliasstruktur ausgelassen.

Angenommen, es liegt ein Versuchsplan mit fünf Faktoren vor. Dieser weist folgende Aliasstruktur für die nicht gefalteten bzw. gefalteten Versuchspläne auf (in beiden Fällen wurde einmal für alle Faktoren und einmal nur für Faktor A gefaltet):

$$I + ABD + ACE + BCDE$$

Der gefaltete Versuchsplan ist dann:

$$I + BCDE$$

13.4 Split-Plot-Design

Das Wesen des Experimentierens hier ist, dass Merkmalsausprägungen variiert werden. In der Praxis kann es der Fall sein, dass ein Merkmal nur schwer geändert werden kann, sei es aus technischen oder auch ökonomischen Gründen.

Für einen solchen Fall ist das sog. Split-Plot-Design angegeben worden; diese Designs sind zuerst in der Landwirtschaft eingesetzt worden.

13.5 Praktisches Erstellen eines Full-Factorial-DoE

Für das Design eines zweistufigen vollfaktoriellen Versuchsplanes mit 3 Faktoren sind jeweils 2 Werte $(+, -)$ anzugeben, um den Plan zu erstellen. Die folgende Beispieltabelle weist alle Informationen auf, die dann notwendig sind:

Faktor A	Sprühdruck	2 bar	5 bar
Faktor B	Temperatur	30 °C	55 °C
Faktor C	Grundierung	G234A	G400A

Figure 108: DOE Faktor-Tabelle

Für 2 Durchläufe wird dann das entsprechende Datenblatt zu:

StdRfolge	Durchlaufreihenfolge	Zentralpunkt	Blöcke	Sprühdruck	Temperatur	Grundierung
8	1	1	1	5	55	G400A
12	2	1	1	5	55	G234A
4	3	1	1	5	55	G234A
5	4	1	1	2	30	G400A
13	5	1	1	2	30	G400A
10	6	1	1	5	30	G234A
16	7	1	1	5	55	G400A
14	8	1	1	5	30	G400A
1	9	1	1	2	30	G234A
3	10	1	1	2	55	G234A
2	11	1	1	5	30	G234A
7	12	1	1	2	55	G400A
9	13	1	1	2	30	G234A
11	14	1	1	2	55	G234A
15	15	1	1	2	55	G400A
6	16	1	1	5	30	G400A

Figure 109: DOE Datenblatt teil-faktorieller Versuchsplan

Die zugehörige Textausgabe ist:

```
Vollfaktorieller Versuchsplan

Faktoren:     3   Basisversuchsplan:       3; 8
Durchläufe:  16   Replikationen:              2
Blöcke:       1   Zentralpunkte (gesamt):     0


Alle Terme ohne Aliasstruktur.




Versuchsplantabelle (randomisiert)

Durchlauf  A  B  C
1  +  +  +
2  +  +  -
3  +  +  -
4  -  -  +
5  -  -  +
6  +  -  -
7  +  +  +
8  +  -  +
9  -  -  -
```

```
10  -  +  -
11  +  -  -
12  -  +  +
13  -  -  -
14  -  +  -
15  -  +  +
16  +  -  +
```

Ein teilfaktorieller Versuchsplan analog der Vorgehensweise in 13.2 ist:

Die Datenblatt-Ausgabe für einen teil-faktoriellen Versuchsplan ist:

StdRfolge	Durchlaufreihenfolge	Zentralpunkt	Blöcke	Sprühdruck	Temperatur	Grundierung
8	1	1	1	5	55	G400A
2	2	1	1	5	30	G234A
4	3	1	1	5	55	G400A
3	4	1	1	2	55	G234A
7	5	1	1	2	55	G234A
5	6	1	1	2	30	G400A
6	7	1	1	5	30	G234A
1	8	1	1	2	30	G400A

Figure 110: DOE Datenblatt teil-faktorieller Versuchsplan

Die Textausgabe für einen teil-faktoriellen Versuchsplan enthält Angaben zur Aliasstruktur und ist nun:

```
Teilfaktorieller Versuchsplan

Faktoren:    3   Basisversuchsplan:       3; 4   Auflösung:  III
Durchläufe:  8   Replikationen:              2   Fraktion:   1/2
Blöcke:      1   Zentralpunkte (gesamt):     0

* HINWEIS * Einige Haupteffekte sind mit Zwei-Faktor-Wechselwirkungen vermengt.


Versuchsplangeneratoren: C = AB


Aliasstruktur:  I = ABC
```

```
Aliasstruktur

I + ABC

A + BC
B + AC
C + AB


Versuchsplantabelle (randomisiert)

Durchlauf  A  B  C
1  +  +  +
2  +  -  -
3  +  +  +
4  -  +  -
```

Als fiktives Modell werden mit Hilfe normalverteilter Zufallszahlen ϵ mit Mittelwert 0 und Standardabweichung 3 folgende Zuweisungen (bezogen auf die Eckenkodierungen mit -1- und $+1$-Werten) verwendet:

$$\begin{aligned} Werte &= 200 + 5 * Grundierung - 16 * Spruehdruck * Temperatur + \epsilon \\ Kosten &= 100 - 7 * Grundierung + 6 * Spruehdruck + 8 * Temperatur + \epsilon \end{aligned}$$

Es resultiert:

StdRfolge	Durchlaufreihenfolge	Zentralpunkt	Blocke	Sprühdruck	Temperatur	Grundierung	Werte	Kosten
8	1	1	1	5	55	G400A	191,158	109,158
12	2	1	1	5	55	G234A	178,251	120,251
4	3	1	1	5	55	G234A	175,931	117,931
5	4	1	1	2	30	G400A	194,992	84,992
13	5	1	1	2	30	G400A	191,240	81,240
10	6	1	1	5	30	G234A	207,412	101,412
16	7	1	1	5	55	G400A	189,223	107,223
14	8	1	1	5	30	G400A	224,344	94,344
1	9	1	1	2	30	G234A	180,746	94,746
3	10	1	1	2	55	G234A	213,773	111,773
2	11	1	1	5	30	G234A	209,089	103,089
7	12	1	1	2	55	G400A	221,871	95,871
9	13	1	1	2	30	G234A	180,058	94,058
11	14	1	1	2	55	G234A	211,264	109,264
15	15	1	1	2	55	G400A	222,889	96,889
6	16	1	1	5	30	G400A	218,541	88,541

Figure 111: DOE Datenblatt voll-faktorieller Versuchsplan mit Daten

Die Varianzanalysetabelle für Werte ist:

Faktorielle Anpassung: Werte vs. Sprühdruck; Temperatur; Grundierung

Geschätzte Effekte und Koeffizienten für Werte (kodierte Einheiten)

Term	Effekt	Koef	SE Koef	t	p
Konstante		200,67	0,5134	390,84	0,000
Sprühdruck	-2,86	-1,43	0,5134	-2,79	0,024
Temperatur	-0,26	-0,13	0,5134	-0,25	0,808
Grundierung	12,22	6,11	0,5134	11,90	0,000
Sprühdruck*Temperatur	-30,95	-15,47	0,5134	-30,14	0,000
Sprühdruck*Grundierung	0,93	0,46	0,5134	0,90	0,392
Temperatur*Grundierung	-0,74	-0,37	0,5134	-0,72	0,494
Sprühdruck*Temperatur*Grundierung	0,69	0,34	0,5134	0,67	0,521

S = 2,05377 PRESS = 134,975
R-Qd = 99,25% R-Qd(prog) = 97,00% R-Qd(kor) = 98,59%

Varianzanalyse für Werte (kodierte Einheiten)

Quelle	DF	Seq SS	Kor SS	Kor MS	F
Haupteffekte	3	629,96	629,96	209,99	49,78
Sprühdruck	1	32,73	32,73	32,73	7,76
Temperatur	1	0,27	0,27	0,27	0,06
Grundierung	1	596,96	596,96	596,96	141,53
2-Faktor-Wechselwirkungen	3	3836,74	3836,74	1278,91	303,21
Sprühdruck*Temperatur	1	3831,12	3831,12	3831,12	908,28
Sprühdruck*Grundierung	1	3,45	3,45	3,45	0,82
Temperatur*Grundierung	1	2,17	2,17	2,17	0,51
3-Faktor-Wechselwirkungen	1	1,90	1,90	1,90	0,45
Sprühdruck*Temperatur*Grundierung	1	1,90	1,90	1,90	0,45
Residuenfehler	8	33,74	33,74	4,22	
Reiner Fehler	8	33,74	33,74	4,22	
Gesamt	15	4502,35			

Quelle	p
Haupteffekte	0,000
Sprühdruck	0,024
Temperatur	0,808
Grundierung	0,000
2-Faktor-Wechselwirkungen	0,000
Sprühdruck*Temperatur	0,000
Sprühdruck*Grundierung	0,392
Temperatur*Grundierung	0,494
3-Faktor-Wechselwirkungen	0,521
Sprühdruck*Temperatur*Grundierung	0,521
Residuenfehler	
Reiner Fehler	
Gesamt	

Geschätzte Koeffizienten für Werte unter Verwendung von Daten in nicht kodierten Einheiten

Term	Koef
Konstante	81,6887
Sprühdruck	34,1209
Temperatur	2,87818
Grundierung	9,01306
Sprühdruck*Temperatur	-0,825280
Sprühdruck*Grundierung	-0,47226
Temperatur*Grundierung	-0,093857
Sprühdruck*Temperatur*Grundierung	0,0184000

Die Varianzanalyse für Kosten ist:

Faktorielle Anpassung: Kosten vs. Sprühdruck; Temperatur; Grundierung

Geschätzte Effekte und Koeffizienten für Kosten (kodierte Einheiten)

Term	Effekt	Koef	SE Koef	t	p
Konstante		100,674	0,5134	196,08	0,000
Sprühdruck	9,139	4,570	0,5134	8,90	0,000
Temperatur	15,742	7,871	0,5134	15,33	0,000
Grundierung	-11,784	-5,892	0,5134	-11,48	0,000
Sprühdruck*Temperatur	1,052	0,526	0,5134	1,02	0,336
Sprühdruck*Grundierung	0,929	0,465	0,5134	0,90	0,392
Temperatur*Grundierung	-0,736	-0,368	0,5134	-0,72	0,494
Sprühdruck*Temperatur*Grundierung	0,690	0,345	0,5134	0,67	0,521

S = 2,05377 PRESS = 134,975
R-Qd = 98,25% R-Qd(prog) = 92,99% R-Qd(kor) = 96,72%

Varianzanalyse für Kosten (kodierte Einheiten)

Quelle	DF	Seq SS	Kor SS	Kor MS	F
Haupteffekte	3	1880,81	1880,81	626,938	148,63
Sprühdruck	1	334,12	334,12	334,118	79,21
Temperatur	1	991,29	991,29	991,285	235,01
Grundierung	1	555,41	555,41	555,411	131,68
2-Faktor-Wechselwirkungen	3	10,05	10,05	3,350	0,79
Sprühdruck*Temperatur	1	4,43	4,43	4,427	1,05
Sprühdruck*Grundierung	1	3,45	3,45	3,454	0,82
Temperatur*Grundierung	1	2,17	2,17	2,169	0,51
3-Faktor-Wechselwirkungen	1	1,90	1,90	1,904	0,45
Sprühdruck*Temperatur*Grundierung	1	1,90	1,90	1,904	0,45
Residuenfehler	8	33,74	33,74	4,218	
Reiner Fehler	8	33,74	33,74	4,218	
Gesamt	15	1926,51			

Quelle	p
Haupteffekte	0,000
Sprühdruck	0,000
Temperatur	0,000
Grundierung	0,000
2-Faktor-Wechselwirkungen	0,531

Sprühdruck*Temperatur	0,336
Sprühdruck*Grundierung	0,392
Temperatur*Grundierung	0,494
3-Faktor-Wechselwirkungen	0,521
Sprühdruck*Temperatur*Grundierung	0,521
Residuenfehler	
Reiner Fehler	
Gesamt	

Geschätzte Koeffizienten für Kosten unter Verwendung von Daten in nicht kodierten Einheiten

Term	Koef
Konstante	67,4220
Sprühdruck	1,85424
Temperatur	0,531509
Grundierung	-2,98694
Sprühdruck*Temperatur	0,0280529
Sprühdruck*Grundierung	-0,47226
Temperatur*Grundierung	-0,093857
Sprühdruck*Temperatur*Grundierung	0,0184000

Die hier errechneten Modelle enthalten nicht-signifikante Terme. Diese werden eliminiert (siehe dazu auch Abschn. (12), die Eliminierung gilt für beide Modelle gleichzeitig !), es folgt für 'Werte':

Faktorielle Anpassung: Werte vs. Sprühdruck; Temperatur; Grundierung

Geschätzte Effekte und Koeffizienten für Werte (kodierte Einheiten)

Term	Effekt	Koef	SE Koef	t	p
Konstante		200,67	0,4842	414,40	0,000
Sprühdruck	-2,86	-1,43	0,4842	-2,95	0,013
Temperatur	-0,26	-0,13	0,4842	-0,27	0,795
Grundierung	12,22	6,11	0,4842	12,61	0,000

Sprühdruck*Temperatur	-30,95	-15,47	0,4842	-31,95	0,000

S = 1,93699 PRESS = 87,3176
R-Qd = 99,08% R-Qd(prog) = 98,06% R-Qd(kor) = 98,75%

Varianzanalyse für Werte (kodierte Einheiten)

Quelle	DF	Seq SS	Kor SS	Kor MS	F	p
Haupteffekte	3	629,96	629,96	209,99	55,97	0,000
Sprühdruck	1	32,73	32,73	32,73	8,72	0,013
Temperatur	1	0,27	0,27	0,27	0,07	0,795
Grundierung	1	596,96	596,96	596,96	159,11	0,000
2-Faktor-Wechselwirkungen	1	3831,12	3831,12	3831,12	1021,11	0,000
Sprühdruck*Temperatur	1	3831,12	3831,12	3831,12	1021,11	0,000
Residuenfehler	11	41,27	41,27	3,75		
Fehlende Anpassung	3	7,53	7,53	2,51	0,59	0,636
Reiner Fehler	8	33,74	33,74	4,22		
Gesamt	15	4502,35				

Ungewöhnliche Beobachtungen für Werte

Beob	StdRfolge	Werte	Anpassung	SE Anpassung	Residuum	Std. Resid.
8	14	224,344	220,955	1,083	3,389	2,11R

R kennzeichnet eine Beobachtung mit einem großen standardisierten Residuum.

Geschätzte Koeffizienten für Werte unter Verwendung von Daten in nicht kodierten Einheiten

Term	Koef
Konstante	81,6887
Sprühdruck	34,1209
Temperatur	2,87818

Grundierung	6,10821
Sprühdruck*Temperatur	-0,825280

Und für 'Kosten':

Faktorielle Anpassung: Kosten vs. Sprühdruck; Temperatur; Grundierung

Geschätzte Effekte und Koeffizienten für Kosten (kodierte Einheiten)

Term	Effekt	Koef	SE Koef	t	p
Konstante		100,674	0,4842	207,90	0,000
Sprühdruck	9,139	4,570	0,4842	9,44	0,000
Temperatur	15,742	7,871	0,4842	16,25	0,000
Grundierung	-11,784	-5,892	0,4842	-12,17	0,000
Sprühdruck*Temperatur	1,052	0,526	0,4842	1,09	0,301

S = 1,93699 PRESS = 87,3176
R-Qd = 97,86% R-Qd(prog) = 95,47% R-Qd(kor) = 97,08%

Varianzanalyse für Kosten (kodierte Einheiten)

Quelle	DF	Seq SS	Kor SS	Kor MS	F	p
Haupteffekte	3	1880,81	1880,81	626,938	167,10	0,000
Sprühdruck	1	334,12	334,12	334,118	89,05	0,000
Temperatur	1	991,29	991,29	991,285	264,21	0,000
Grundierung	1	555,41	555,41	555,411	148,03	0,000
2-Faktor-Wechselwirkungen	1	4,43	4,43	4,427	1,18	0,301
Sprühdruck*Temperatur	1	4,43	4,43	4,427	1,18	0,301
Residuenfehler	11	41,27	41,27	3,752		
Fehlende Anpassung	3	7,53	7,53	2,509	0,59	0,636
Reiner Fehler	8	33,74	33,74	4,218		
Gesamt	15	1926,51				

Ungewöhnliche Beobachtungen für Kosten

Beob	StdRfolge	Kosten	SE Anpassung	Std. Anpassung	Residuum	Resid.
8	14	94,344	90,955	1,083	3,389	2,11R

R kennzeichnet eine Beobachtung mit einem großen standardisierten Residuum.

Geschätzte Koeffizienten unter Verwendung von Daten in nicht kodierten Einheiten

Term	Koef
Konstante	67,4220
Sprühdruck	1,85424
Temperatur	0,531509
Grundierung	-5,89179
Sprühdruck*Temperatur	0,0280529

Anmerkungen:

- die Eliminierung gilt für beide Modelle gleichzeitig; das hat zur Folge, dass für 'Werte' der nicht-signifikante Term für 'Temperatur' und für 'Kosten' der ebenfalls nicht-signifikante Term für die Wechselwirkung 'Sprühdruck * Temperatur' enthalten bleiben muss (da diese Terme im jeweils anderen Modell signifikant sind).
- die Koeffizienten in der ANOVA-Tabelle beziehen sich auf die kodierten Einheiten (also auf die Eckpunkt-Koordinaten des Einheitswürfels, notiert mit -1- und $+1$-Werten). Um das Modell z.B. auch für technische Zielgrößenoptimierungen zu verwenden, müssen die Koordinaten in die technischen Dimensionen umgerechnet werden. Die umgerechneten Koeffizienten werden in der Tabelle 'Geschätzte Koeffizienten für Kosten unter Verwendung von Daten in nicht kodierten Einheiten' angegeben.

Als weitere Hilfsmittel zur Analyse werden noch zwei Grafiken als Unterstützung ausgegeben, das sog. Pareto-Diagramm der Effekte:

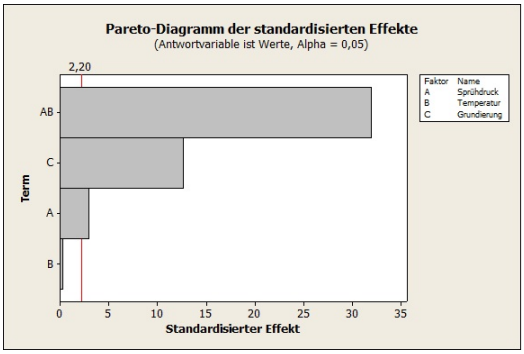


Figure 112: DOE Pareto-Diagramm der Effekte für Werte

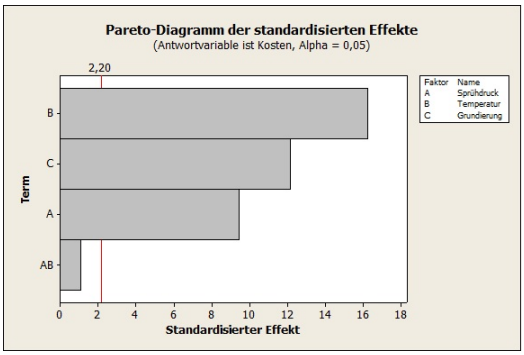


Figure 113: DOE Pareto-Diagramm der Effekte für Kosten

Anm.: die Linie im Pareto-Diagramm ist eine Signifikanzlinie und zeigt an, welche Terme signifikant sind.

Sowie eine Residuumanalyse (das Residuum muss normalverteilt sein ...):

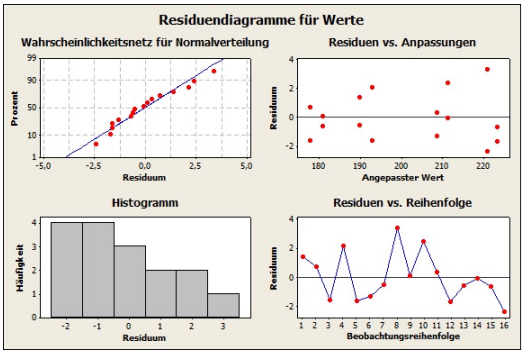


Figure 114: DOE Residuendiagramm für Werte

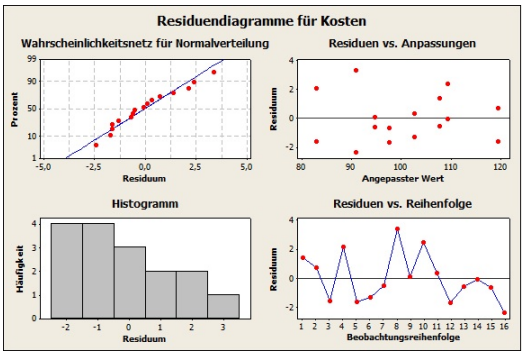


Figure 115: DOE Residuendiagramm für Kosten

Anm.: da für 'Werte' und für 'Kosten' die gleichen Zufallszahlen verwendet wurden, wird hier für beide Variablen das gleiche Residuum berechnet.

Der Test auf Normalverteilung für das Residuum ist im übrigen positiv:

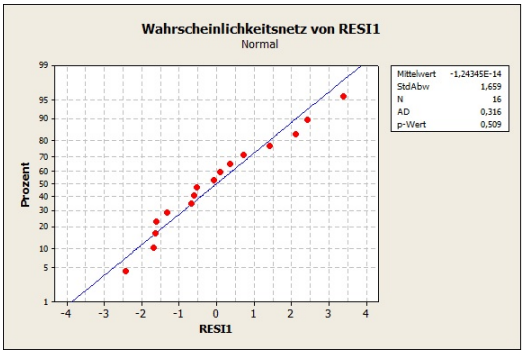


Figure 116: DOE Residuum: Test auf Normalverteilung

Die Haupteffekte werden im Haupteffekte-Diagrammen grafisch dargestellt; analog die Wechselwirkungen.

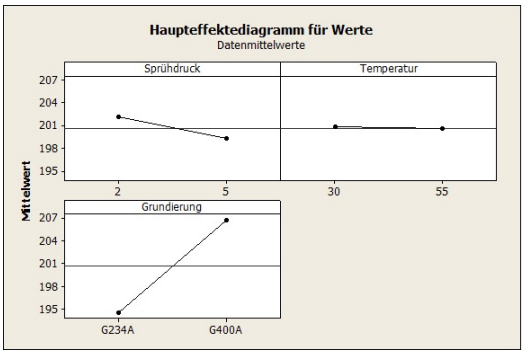


Figure 117: DOE Haupteffektediagramm Werte

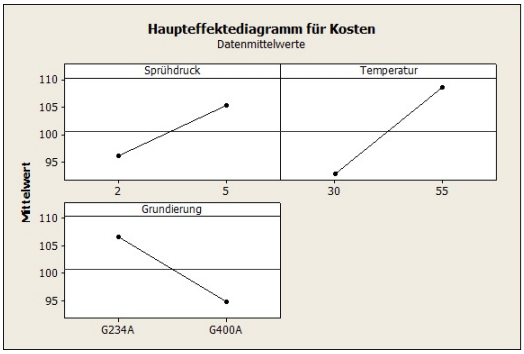


Figure 118: DOE Haupteffektediagramm Kosten

Anm.: je steiler die Gerade, desto stärker ist der Haupteffekt. Für 'Werte' ist zu erkennen, dass die Temperatur praktisch keine Wirkung auf das Ergebnis der Experimente hatte.

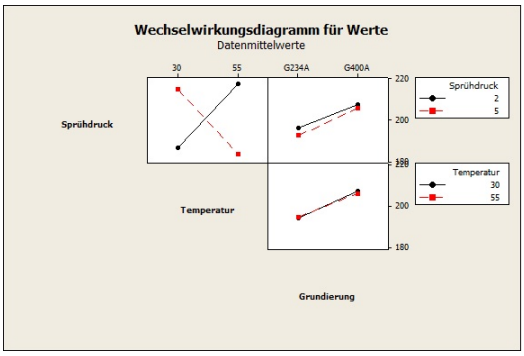


Figure 119: DOE Wechselwirkungsdiagramm Werte

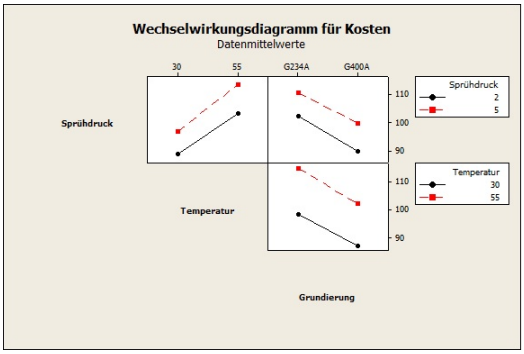


Figure 120: DOE Wechselwirkungsdiagramm Kosten

Es ist lediglich nur eine Wechselwirkung in den Modellen zu finden, in 'Werte' für 'Sprühdruck * Temperatur'. Nur für diese Wechselwirkung kreuzen sich die Linien im Diagramm.

Die Mittelwerte der Messungen werden im Würfeldiagramm aufgezeigt:

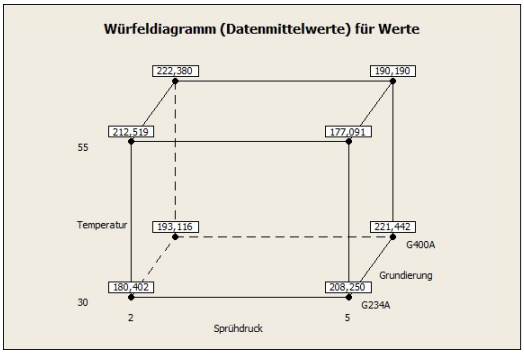


Figure 121: DOE Würfeldiagramm Werte

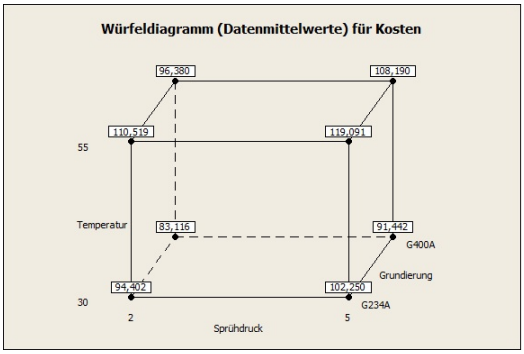


Figure 122: DOE Würfeldiagramm Kosten

Das Würfeldiagramm kann auch genutzt werden, um prinzipielle Einstellungen (Faustregeln) zu finden. So ist z.B. im Würfeldiagramm für 'Werte' zu erkennen, dass die hinteren 4 Ecken größere Mittelwerte in den Messungen hervorgebracht haben, also die vorderen 4. Insofern ist die Grundierung 'G400A' zu wählen, wenn 'Werte' möglichst groß sein soll. Weitere solcher einfachen Einschätzungen sind dann nicht zu erkennen.

13.6 Optimierung

Die im vorigen Abschnitt errechneten Daten können nun für eine Zielgrößenoptimierung verwendet werden. Ziel ist es, für 'Werte' und 'Kosten' eine optimale Einstellung so zu finden, dass gewisse Mindestforderungen an beide Größen vorgegeben werden können, gleichzeitig bei einem Zielkonflikt je nach Zielsetzung die Größen dann erreicht, maximiert oder minimiert werden sollen.

Im hier angegebenen Beispiel sollen:

- die Werte für 'Werte' maximal werden, mindestens 200 betragen, akzeptabel für einen Kompromiss soll 120 sein
- die Kosten jedoch minimiert werden und höchstens 100 betragen, akzeptabel für einen Kompromiss sind 90

In Minitab steht dafür folgender Dialog zur Verfügung:

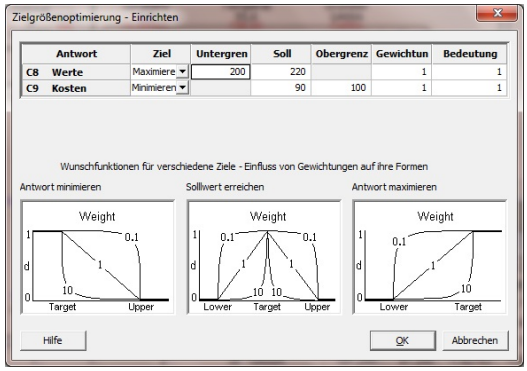


Figure 123: DOE Optimierung Dialog - Einstellungen

Die Mechanismen für 'Bedeutung' und 'Gewichtung' für einen Fall der multikriteriellen Optimierung sind noch zu erläutern:

Die Gewichtung gibt an, wie vom Wunsch abgewichen werden kann, sofern dieses erforderlich ist. Als Gewichtung ist eine Zahl zwischen 0.1 und 10 wählbar. Für den Wert 0.1 wird am wenigsten auf das erreichen des Wertes Wert gelegt. Die Bewertung erfolgt mittels einer Transformation und ist in den Hilfegrafiken des Dialoges angedeutet. Bewertet wird der Abstand vom Wunschwert, exponentiell gewichtet.

Nach dem Berechnen der individuellen Erwünschtheiten für die Antwortvariablen werden diese kombiniert, um ein Maß der zusammengesetzten Erwünschtheit bzw. der Gesamterwünschtheit des Systems mit mehreren Antwortvariablen zu erhalten. Dieses Maß der zusammengesetzten Erwünschtheit ist das gewichtete geometrische Mittel der individuellen Erwünschtheiten für die Antwortvariablen. Anschließend kann durch das Maximieren der zusammengesetzten Erwünschtheit die optimale Lösung bestimmt werden (d. h. die optimalen Betriebsbedingungen).

Das Ergebnis für die o.g. Einstellungen ist nun:

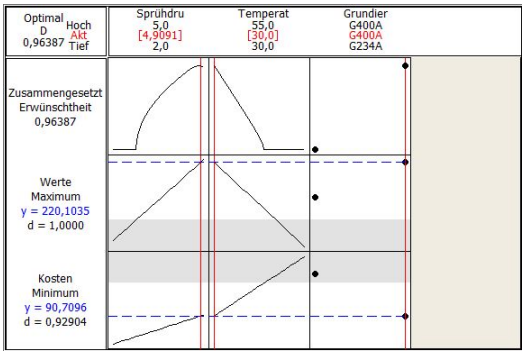


Figure 124: DOE Optimierung Ergebnis

Die Variable d gibt an, inwieweit die Vorgaben für 'Werte' und 'Kosten' erreicht werden konnten. Für 'Werte' sind die Vorgaben 100% erreicht worden ($d = 1$), für 'Kosten' lediglich zu fast 93%.

Setzt man nun die 'Bedeutung' im Dialog 123 für 'Kosten' auf 10, so folgt ein anderes Ergebnis durch diese neue Bewertung:

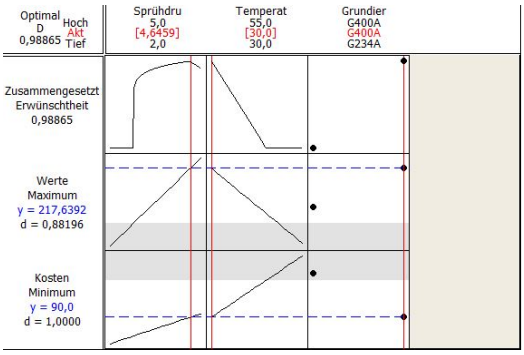


Figure 125: DOE Optimierung Ergebnis mit geänderter Bedeutung

Weitere Hilfsmittel sind das Kontur- und das Wirkungsflächendiagramm, um sich die Situation zu veranschaulichen:

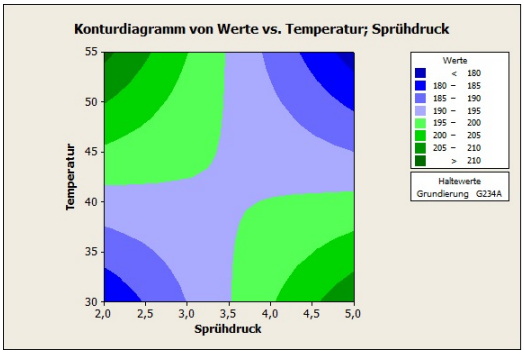


Figure 126: DOE Kontur-Diagramm

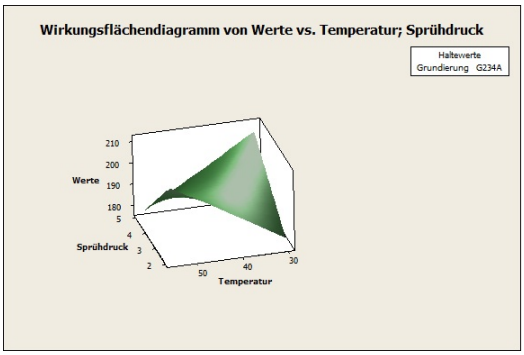


Figure 127: DOE Wirkungsflächen-Diagramm

Um den Zielkonflikt bzw. den Kompromiss zu beurteilen gibt es noch ein überlagertes Konturdiagramm:

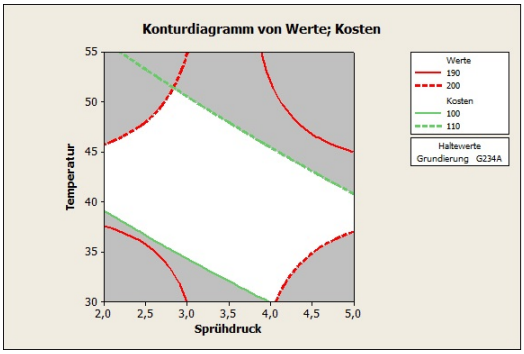


Figure 128: DOE überlagertes Konturdiagramm

14 Lebensdauer- und Zuverlässigkeitsanalyse

In der Praxis ist jeder Hersteller dazu angehalten, über die Haltbarkeit (auch unter Belastungen) seiner Produkte etwas zu wissen - oder zu versprechen. Die Kunden der Produkte müssen sich nicht selten auf eine gewisse Haltbarkeit verlassen können, das beginnt bei der Haltbarkeit von Lebensmitteln bis hin zu Stahlseilen an einer Brücke oder die Achsen in einem Kraftfahrzeug.

Sofern es lediglich um eine aktuelle auszuhaltende Last geht, ist die Messung der Zuverlässigkeit noch recht einfach; ganz anders sieht es aus, wenn die Produkte über Jahre ihren Dienst nicht versagen sollen. Aussagen für eine Zuverlässigkeit über 20 Jahre sind keine Seltenheit. Man kann erahnen, dass (aber nicht nur) für solche Fälle gewisse technische Tricks angewendet werden müssen, um z.B. langlebige Oberflächen in (wesentlich) kürzerer Zeit schlussendlich zu beschädigen. Die Mathematik muss in ihren Schätzungen der Haltbarkeit (oder andersherum: eines Ausfalls oder Schadens) diesen Zusammenhang natürlich berücksichtigen.

Die nun vorgestellten mathematischen Methoden haben das Ziel die Vorgehensweisen und zugehörigen Datenformate der Praxis entsprechend (vollständig) zu berücksichtigen.

Vorab sei noch erwähnt, dass Zuverlässigkeit und Lebensdauer sich prinzipiell nicht unterscheiden. Historisch sind Analysen gegen die Zeit zunächst entwickelt worden, das den Begriff der Lebensdauer(-Analyse) begründet. Jedoch sind

Analysen für eine Zuverlässigkeit, z.B. die Haltbarkeit von Drähten unter Zuglast mit den gleichen Verfahren und Methoden durchführbar. Hier wird im weiteren nicht mehr unterscheiden, was konkret die Unabhängige ist: Zeit oder Zugkraft, oder was aus technischer Sicht betrachtet werden muss.

14.1 Lebensdauerdiagramme

Zunächst ist es sinnvoll, die Datenerfassung aus Experimenten zur Zuverlässigkeits- oder Lebensdauerprüfung vorzustellen. Man betrachte also eine Datenerfassung von einem Experiment, hier in der Zeit:



Figure 129: Ausfalldiagramm ohne Zensur

Eine Problematik ergibt sich sofort: in diesem Bild, also bei diesen Daten, wird davon ausgegangen, dass alle Prüflinge auch ausgefallen sind. In der Praxis, z.B. bei noch laufenden oder rollierenden Versuchen, bei denen ausgefallene Bauteile umgehend durch neue ersetzt werden, muss dieses Nicht-ausfallen entsprechend berücksichtigt werden. Wann dann diese nichtausgefallenen Bauteile ausfallen, ist halt nicht bekannt. Man spricht hier in Bezug auf die Ausfallzeit von einer Zensur der Daten.

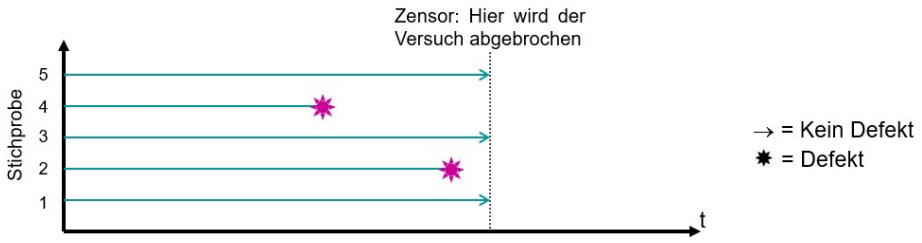


Figure 130: Ausfalldiagramm mit Zensur

Sofern man keine zensierten Daten hat, ist eine Histogrammdarstellung sinnvoll, um ein Gespür für das Ausfallverhalten in der Zeit zu bekommen. Für das Beispiel hier ist dieses:

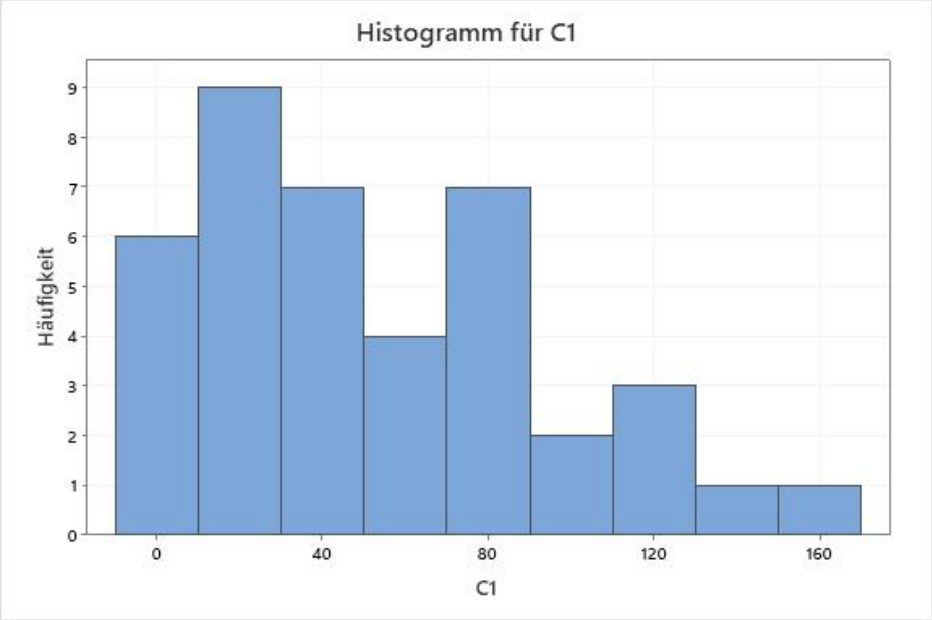


Figure 131: Histogramm von Ausfalldaten

14.2 Zensur

Die Ausfalldaten können nun gemischte Informationen beinhalten, also Ausfälle und Nicht-Ausfälle beinhalten. Diese Situation ist nicht nur in Tests zu finden, praktisch nicht vermieden werden kann dieses bei Felddaten und rollierenden

Tests.

Wichtig ist, die nicht-ausgefallenen Bauteile nicht einfach nicht zu berücksichtigen in der Analyse. Ob ein Bauteil ausgefallen ist oder nicht, das ist für die Analyse wichtig; diese Zensur wird in den Daten entsprechend kodiert.

14.2.1 Allgemeine und Intervall-Zensur

Die folgenden Angaben sind für eine Allgemeine oder eine Intervall-Zensur notwendig:

- Start der Messung
- Ende der Messung
- Häufigkeit der Ausfälle zwischen Start und Ende

Beispiel:

Start	Ende	Häufigkeit
0	1	1
1	2	3
2	3	6
3	5	10
5	10	12
1	*	12
2	*	15
3	*	30
5	*	45

Figure 132: Beispiel: Intervallzensur

Intervalle können also unterschiedliche Breite besitzen, sie müssen lediglich aufsteigend sortiert sein.

Bei dieser Zensur ist lediglich bekannt, wie viele Bauteile in einem Zeitraum (Intervall mit Start und Ende) ausgefallen sind. Die Bauteile, die noch nicht ausgefallen sind, haben in der rechten Zeitspalte des Intervalles ein Sternchen. Also $1,', 12$ bedeutet, dass 12 Bauteile, für die vor einer Zeiteinheit der Test begann, noch intakt sind.

14.2.2 Rechtszensur

Diese Zensur für exakte Zeitpunktmessungen vorgesehen und einfacher; es wird mit Hilfe einer Indexspalte angegeben, ob das Bauteil ausgefallen ist, oder nicht. Der Index kann in Minitab frei gewählt werden, hier im Beispiel seien die Eins-Einträge die Ausfälle, 0 bedeutet, dass das Bauteil noch funktioniert.

Ausfallzeit	Zensor
6,031	1
73,908	1
85,955	1
97,291	1
39,067	1
13,218	1
13,722	1
8,875	0
14,279	0

Figure 133: Beispiel: Rechtszensur

14.3 Ausfallverhalten

14.3.1 Verteilungsfreie Schätzungen

Kaplan-Meier:

Der Kaplan-Meier-Schätzer (KME), auch bezeichnet als Produkt-Limit-Schätzer (PLE), schätzt die Überlebenszeit für Stichproben. Die Überlebensfunktion S

ist:

$$S(x) = \prod_{i:x_i \leq x} (1 - h_i) \tag{20}$$

Ein Schätzer ist dann:

$$\hat{S}(x) = \prod_{i:x_i \leq x} (1 - \hat{h}_i). \tag{21}$$

Die h_i bzw. $\hat{h}_i$ hier in den Formeln sind die Hazard-Rate, also die Funktion, die die Wahrscheinlichkeit eines Ausfalls auf den noch intakten Bauteilen schätzt:

$$\hat{h}_i = \frac{d_i}{n_i} \qquad i = 1, 2, ..., k \tag{22}$$

d_i Anzahl von Defekten bis zur Zeit x_i.

n_i Anzahl der Einheiten unter Risiko (d.h. die noch ausfallen können) zur Zeit x_i.

Der Schätzer für die Überleben wird dann zu:

$$\hat{S}(x) = \prod_{i:x_i \leq x} \left(1 - \frac{d_i}{n_i}\right) \qquad i = 1, 2, ..., k. \tag{23}$$

Der Kaplan-Meier-Schätzer kann ebenfalls mit rechtszensierten Daten schätzen.

Liegen exakte Ausfallzeitpunkte für die Ausfälle vor, dann ist die Methode von Kaplan-Meier eine Schätzmethode, die parameterfrei, d.h. nicht verteilungsgebunden, die Schätzungen durchführen kann. Hier ein Ergebnis mit Minitab:

Verteilungsfreie Schätzwerte

Merkmale der Variablen

Mittelwert (MTTF)	Standardfehler	Normales 95,0%-KI Untergrenze	Obergrenze	Q1	Median	Q3	IQR
41,7022	9,16355	23,7420	59,6625	8,87501	20,2562	64,3400	55,4650

Kaplan-Meier-Schätzwerte

Zeit	Anzahl gefährdeter Einheiten	Anzahl der Ausfälle	Überlebenswahrscheinlichkeit	Standardfehler	Normales 95,0%-KI Untergrenze
4,334	20	1	0,95	0,048734	0,854483
5,746	19	1	0,90	0,067082	0,768522
6,031	18	1	0,85	0,079844	0,693509
7,678	17	1	0,80	0,089443	0,624695
8,875	16	1	0,75	0,096825	0,560227
13,218	15	1	0,70	0,102470	0,499163
13,722	14	1	0,65	0,106654	0,440963
14,279	13	1	0,60	0,109545	0,385297
17,320	12	1	0,55	0,111243	0,331968
20,256	11	1	0,50	0,111803	0,280869
34,510	10	1	0,45	0,111243	0,231968
37,922	9	1	0,40	0,109545	0,185297
39,067	8	1	0,35	0,106654	0,140963
50,747	7	1	0,30	0,102470	0,099163
64,340	6	1	0,25	0,096825	0,060227
73,908	5	1	0,20	0,089443	0,024695
77,724	4	1	0,15	0,079844	0,000000
85,955	3	1	0,10	0,067082	0,000000
97,291	2	1	0,05	0,048734	0,000000
161,121	1	1	0,00	0,000000	0,000000

Figure 134: Ausfallschätzungen nach Kaplan-Meier

Für beliebig zensierte Daten ist ebenfalls eine Schätzung möglich, hierfür gibt es in Minitab zwei Methoden:
Turnbull:

Aktuarmethode

Die Einzelheiten der Methoden hier zu erläutern würden den Rahmen sprengen, zumal die Algorithmik auf einigen Schritten und Optimierungen basiert. Insofern hier lediglich eine Beispielausgabe:

14.3.2 Verteilungsbasierte Schätzungen

Die nicht-parametrische Schätzung hat den Nachteil, dass lediglich in den Ausfallzeitpunkten die empirische Ausfallwahrscheinlichkeit, eben durch die Summation der Ausfälle bis zu diesem Zeitpunkt, angegeben wird. Eine Berechnung der Wahrscheinlichkeiten von Ausfällen auch in den Zwischenzeitpunkten muss approximativ vorgenommen werden.

Besser wäre es, praktisch für alle Zeitpunkte eine Ausfallschätzung zu bekommen. Hierfür wird eine Wahrscheinlichkeitsfunktion benötigt, die das gewünschte leistet. Jedoch muss diese bestimmt werden. Da nach der Bestimmung sämtliche Schätzungen lediglich noch auf der Funktion beruhen (die tatsächlichen Ausfälle werden dann nicht mehr in Erwägung gezogen), ist eine Schätzung einer solchen Funktion sorgfältig vorzunehmen.

In der Praxis werden drei verschiedene Defektmechanismen anhand von Verteilungen unterschieden:

- Verschleiß durch (Langzeit Ab-) nutzung
- Abreißphänomene (z.B. beim Zerreißen eines Drahtes)
- Verschleißfreie Abnutzung, dafür Belastung z.B. durch Temperatur usw., z.B. bei elektronischen Geräten

Für diese Phänomene sind 3 Verteilungen typisch, der Reihe nach sind dies: Weibull-Verteilung, Kleinste Extremwert-Verteilung und die Lognormal-Verteilung.

Identifizieren der Ausfallverteilung:

Anders als in der klassischen Prozessanalyse geschehen die Ausfälle keineswegs fokussiert, also lediglich oder wesentlich in einem bestimmten Zeitraum. So gibt es Frühausfälle, genauso wie sehr langlebige Produkte. Insofern kann man schon erahnen, dass eine Beschreibung der Ausfälle nicht gut mit einer Normalverteilung erfolgen kann. Die Frage nach der Verteilung muss also nun ganz allgemein gestellt werden.

Sofern man mit einer Verteilung arbeiten möchte, ist aus meiner Sicht ein Abgleich der nicht-verteilungsbasierten (d.h. nicht-parameterischen) Schätzungen mit

den verteilungsbasierten durchzuführen. Die Ausfallschätzungen sollten halt für alle Zeiten sehr ähnlich sein, sonst ist das gewählte Verteilungsmodell (!) falsch .

Die Analyse der Verteilung geschieht mit der Methode von Anderson-Darling (siehe dazu auch den Abschnitt in der Prozessfähigkeitsanalyse). Minitab bietet jedoch im Menü zu 'Zuverlässigkeit/Lebensdauer' ein auf diese Analysen zugeschnittene Vorgehensweise an (so kann z.B. eine Zensur hier berücksichtigt werden, was bei der Prozessanalyse ja nicht notwendig ist):

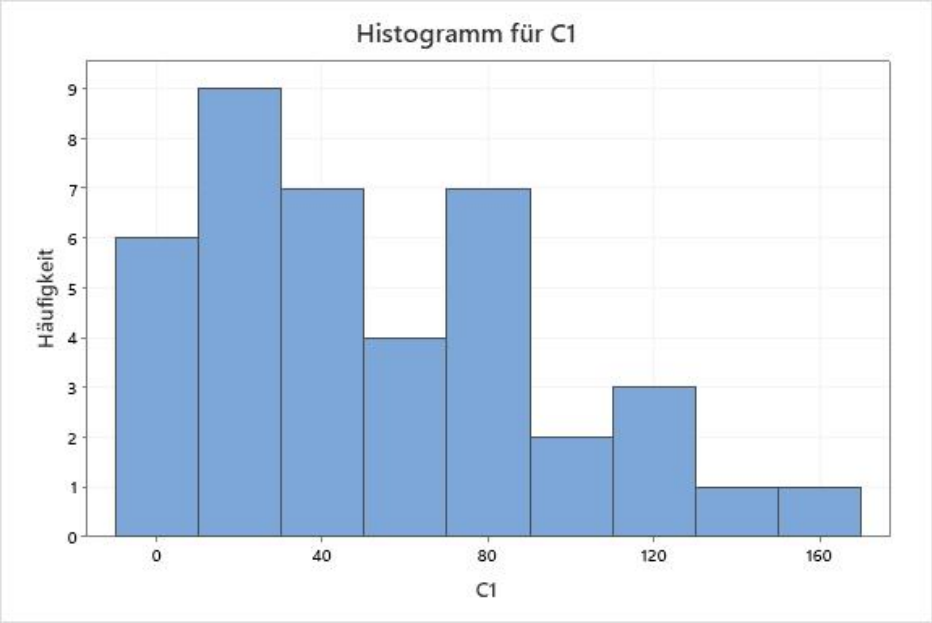


Figure 135: Histogramm der Ausfalldaten

In diesem Fall hier ist die entsprechende Tabelle der AD-Werte wie folgt:

Güte der Anpassung

Verteilung	Anderson-Darling (kor)
Weibull	0,986
Lognormal	0,996
Exponential	0,961
Loglogistisch	0,980
Weibull mit 3 Parametern	0,905
Lognormal mit 3 Parametern	0,977
Exponential mit 2 Parametern	0,988
Loglogistisch mit 3 Parametern	0,936
Kleinster Extremwert	2,173
Normal	1,671
Logistisch	1,676

Figure 136: AD-Werte für die Bestimmung der passenden Ausfallverteilung

Die Auswahl der Verteilung geschieht wieder in 2 Schritten: zu den vier oder weniger besten Verteilungen nach AD-Ergebnis betrachte man final, um die Verteilung bestmöglich passend auszuwählen, die Wahrscheinlichkeitsnetze. Durch partielle Auswahl im Minitab-Menü ist das einfach möglich:

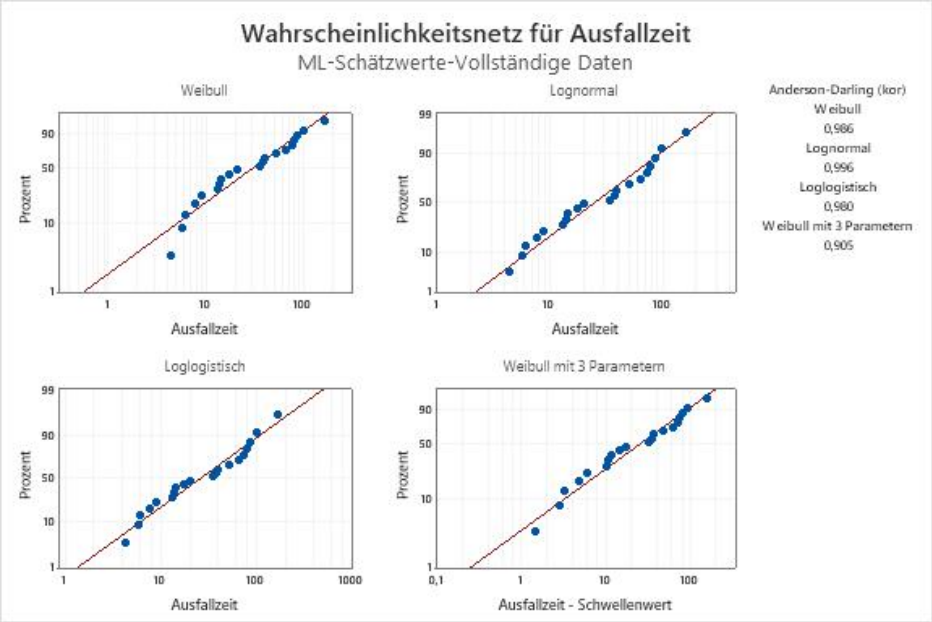


Figure 137: Bestimmung der bestpassenden Ausfallverteilung

Aufgrund der besten hier zu sehenden - qualitativ - Anpassung (nach WS-Netz) fällt die Wahl auf die Lognormal-Verteilung.

Verteilungsübersicht und Plausibilisierung des Ausfallverhaltens

Anm.: die 3-parametrischen Verteilungen verschieben die Ausfallverteilungen, die typischerweise ab $t = 0$ Wahrscheinlichkeiten > 0 besitzen, auf der t-Achse. Der dritte Parameter ist ein sog. Threshold. Ist der Threshold ¿ 0, dann wird demnach in der Zeit von $t = 0$ bis zum Threshold-Wert mit (sicher ... WS $= 0$) keinem Ausfall gerechnet (sog. Time of no Risk). Ist der Threshold negativ, so ist im Zeitpunkt $t = 0$ eine positive Wahrscheinlichkeit gegebene, was bedeutet, dass es Bauteile gibt, die bereits vor Nutzung defekt sind. Aus diesen beiden Gründen werden normalerweise 3-parametrische Verteilungen bei der Analyse ausgeschlossen. Lediglich bei Felddatenanalysen kann ein Threshold akzeptiert werden, wenn z.B. der Hersteller Bauteile bis zum Verbauen einige Zeit im Lager liegen lässt. Das dem so ist, kann die Analyse durch einen positive Threshold-Parameter aufzeigen.

Als nächstes wird die so gefundene Verteilung aus Sicht der Ausfälle, aber auch

bzgl. der Ausfallrate, der Ausfällt je Zeiteinheit, betrachtet:

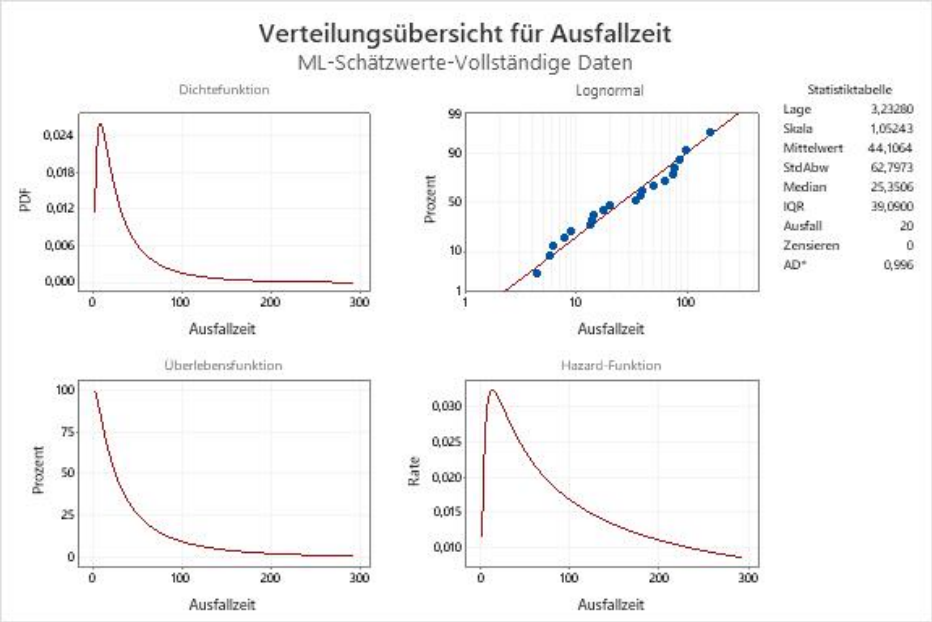


Figure 138: Übersicht über die ausgewählte Ausfallverteilung

Die sog. Hazard-Funktion (wird auch Hades-Funktion genannt) beschreibt, wie die aktuell noch nicht ausgefallenen Objekte dann zukünftig geschätzt ausfallen werden. Zu diesem Zweck wird die Ausfallwahrscheinlichkeit (also die Dichte der Ausfallverteilung) relativ zu den aktuell Überlebenden gesehen. Das geschieht durch den folgenden Quotienten:

$$H(t) = \frac{f(t)}{1 - F(t)} \tag{24}$$

wobei f die Dichte der Ausfallwahrscheinlichkeit und $1 - F(t)$ die Anzahl der noch nicht ausgefallenen ist.

Anm.: abkürzend wird auch geschrieben:

$$R(t) = 1 - F(T) \tag{25}$$

mit $R =$ Zuverlässigkeit (engl.: Reliability).

Anm.: diese Grafiken zeigen auf, wie die Bauteile ausfallen. Für den Analysten ist es wichtig, dieses Verhalten mit der Praxis zu vergleichen. Hierzu werden typischerweise Experten befragt, ob das hier zu sehende Ausfallverhalten bekannt - oder erwartet ist.

Schätzungen der Ausfallanteile für exakte Ausfalldaten

Hat man sich festgelegt auf eine Verteilung, dann kann man nun Ausfallanteile ggü. einer Zeit (z.B. der Garantiezeit) oder als Charakteristischen Wert unter Vorgabe von Prozentsätzen von Defekten die Zeitpunkte schätzen, wann diese erreicht sein werden. In Minitab sind die Schätzungen in der Perzentiltabelle oder als separate Wertetabelle angegeben:

Parameterschätzwerte

Parameter	Schätzwert	Standardfehler	Untere 95,0%-Grenze bei Normalverteilung
Lage	3,23280	0,235330	2,84572
Skala	1,05243	0,166404	0,811418

Log-Likelihood = -94,057

Güte der Anpassung

Anderson-Darling (korrigiert)
0,996

Merkmale der Verteilung

	Schätzwert	Standardfehler	Untere 95,0%-Grenze bei Normalverteilung
Mittelwert (MTTF)	44,1064	12,9383	27,2240
Standardabweichung	62,7973	31,1492	27,7717
Median	25,3506	5,96576	17,2139
Erstes Quartil (Q1)	12,4653	3,25002	8,11808
Drittes Quartil (Q3)	51,5553	13,4418	33,5756
Interquartilbereich (IQR)	39,0900	11,6728	23,9194

Figure 139: Schätzung Teil 1

Perzentiltabelle

Prozent	Perzentil	Standardfehler	Untere 95,0%-Grenze bei Normalverteilung
1	2,19127	0,992715	1,04010
2	2,91939	1,21137	1,47530
2,5	3,22224	1,29593	1,66287
3	3,50222	1,37138	1,83917
4	4,01616	1,50404	2,16915
5	4,48935	1,62051	2,47928
6	4,93578	1,72623	2,77665
7	5,36360	1,82432	3,06538
8	5,77803	1,91677	3,34815
9	6,18267	2,00491	3,62683
10	6,58015	2,08974	3,90276
20	10,4547	2,86303	6,66327
30	14,5983	3,66401	9,66071
40	19,4174	4,64226	13,1041
50	25,3506	5,96576	17,2139
60	33,0966	7,91264	22,3357
70	44,0223	11,0491	29,1326
80	61,4700	16,8336	39,1776
90	97,6651	31,0167	57,9262
91	103,944	33,7069	60,9748
92	111,223	36,8965	64,4498
93	119,817	40,7535	68,4774
94	130,203	45,5368	73,2460
95	143,150	51,6726	79,0559
96	160,016	59,9257	86,4255
97	183,498	71,8534	96,3628
98	220,132	91,3415	111,243
99	293,278	132,864	139,205

Figure 140: Schätzung Teil 2

Tabelle der Überlebenswahrscheinlichkeiten

Zeit	Wahrscheinlichkeit	Untere 95,0%-Grenze bei Normalverteilung
12	0,761345	0,617534

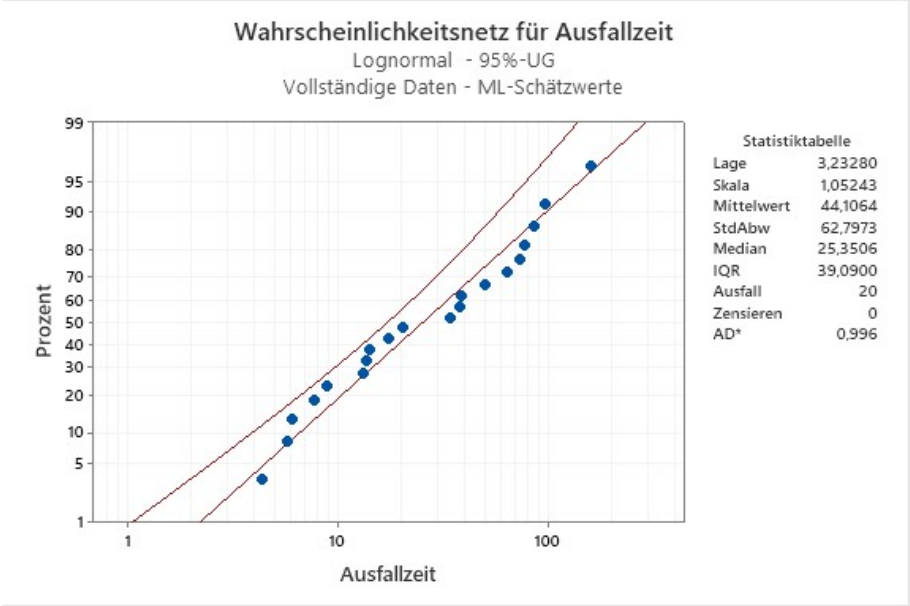


Figure 141: Schätzung Teil 3

Anm.: die Schätzungen sind im Sinne einer Risikoeinschätzung sog. konservativ vorzunehmen. Typischerweise reicht hierfür eine einseitige Schätzung in Bezug auf die Konfidenzintervalle - so wie hier zu sehen.

Schätzungen der Ausfallanteile für intervallzensierte Daten:

Für intervallzensierte Daten ist die oben vorgestellte Vorgehensweise völlig analog möglich. Hier sei lediglich noch auf die Brücke von intervall- und rechtszensierten Daten hingewiesen: in intervallzensierte Daten können exakte Ausfalldaten eingebaut werden, indem Start und Ende gleich gesetzt werden.

14.3.3 WeiBayes-Schätzung

Liegen nur wenige Ausfälle vor, dann sind die klassischen Schätzungen oft nicht gut geeignet, die statistische Unsicherheit in dieser Situation ist es möglich Ausfallschätzungen vorzunehmen, wenn ein Form- (Weibull) oder Skalenparameter (andere Verteilung) bekannt ist. Man spricht hier von historischem oder Vorwissen. In der Statistik wird Vorwissen mit Hilfe der Vorgehensweisen von Bayes betrieben, insofern ergibt sich hier die Bezeichnung WeiBayes.

Die folgende Situation stellt eine solche Situation mit sehr wenig Ausfällen dar:

Start	Ende	Häufigkeit
0	1	0
1	2	0
2	3	2
3	5	1
5	10	2
1	*	121
2	*	312
3	*	523
5	*	713

Figure 142: Daten für die Ausfallschätzung mit WeiBayes

In Minitab ist im Dialogfeld der Schätzung einfach in das entsprechende Feld ein (in dem Fall hier) Weibull-Formparameter einzutragen:

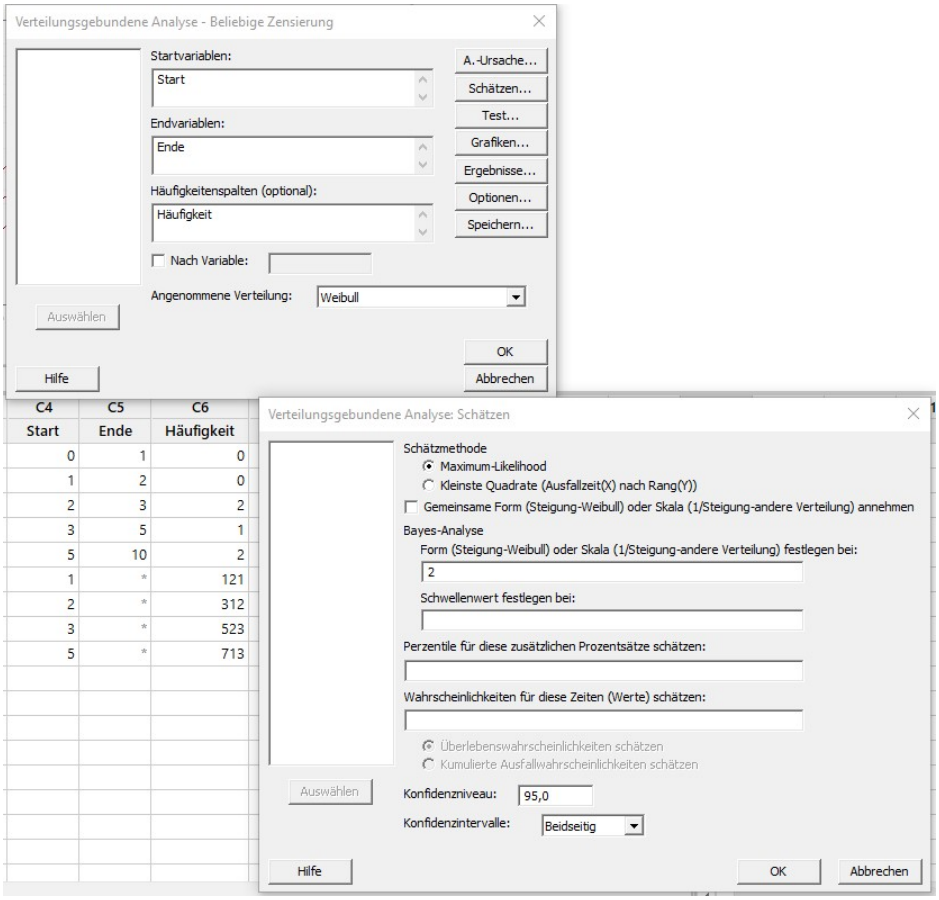


Figure 143: WeiBayes in MTB

Das Ergebnis der Schätzung ist dann:

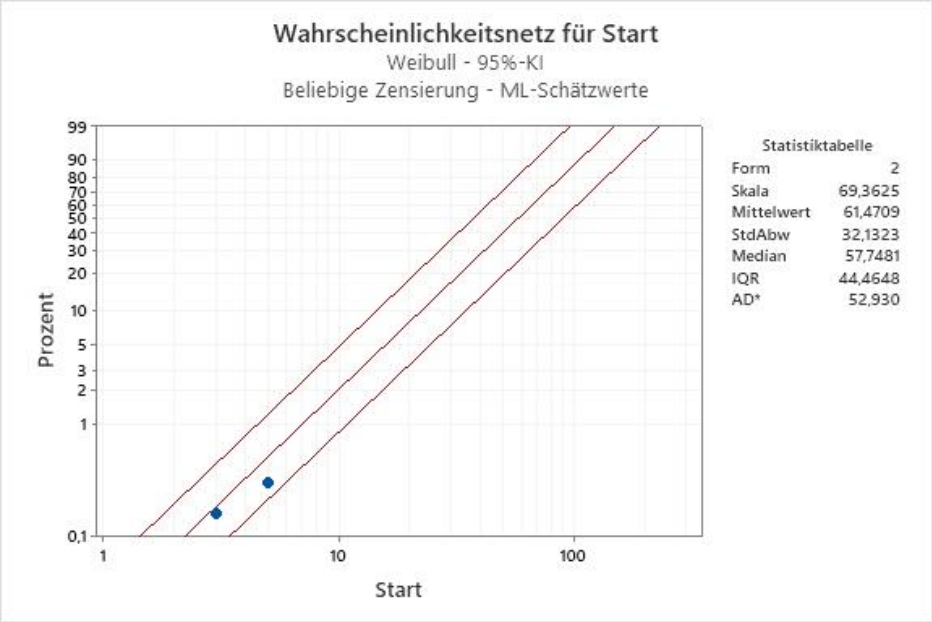


Figure 144: Ausfallschätzung mit WeiBayes

14.3.4 Maximum-Likelihood-Schätzung für Daten mit Zensur

Eine Maximum-Likelihood-Schätzung ist aus den Statistik-Vorlesungen bekannt; die Formel ohne Zensur für exakte Ausfalldaten wäre:

$$L = \prod_{i=1}^{n} f(t_i; \alpha, \beta) \tag{26}$$

wobei f die zugrunde gelegte Ausfallwahrschenilichkeitsdichte und α und β die Parameter dieser Dichte sind. Die Maximierung von L nach den Parametern ist die Aufgabe der Maximum-Likelihood-Methode.

Für zensierte Daten wird folgende Funktion verwendet:

$$L = \prod_{i=1}^{n} f(t_i; \alpha, \beta) * \prod_{j=1}^{m} (1 - F(t_j; \alpha, \beta)) \tag{27}$$

Die Ausfalldaten sind die Zeitpunkte t_i von 1 bis n, die zensierten Daten sind die Zeiten t_j von 1 bis m. Der Faktor $1 - F$ stellt das Wissen des Nichtausfalls ind er Stichprobe dar, F ist die Verteilungsfunktion der Ausfälle, $1 - F$ also die Verteilung des intakten Bauteils.

14.3.5 Parametertests

Weiter Entscheidend für die Schätzung z.B. der Ausfallperzentilen bei Defekten sind die Parameter, die für die jeweils zugrunde gelegte Verteilung geschätzt werden. Möchte man nun z.B. aufzeigen, dass es bzgl. des Ausfallverhaltens an verschiedenen Standorten oder Produkten ein signifikanter Unterschied ergibt, dann muss man sozusagen die komplette Verteilung (oder besser die wesentlichen Teile) als unterschiedlich identifizieren. Ein Test auf /un-)gleiche Parameter muss nun auch für den nicht-normalen Verteilungsfall, der bereits in den ersten Kapiteln für die Normalverteilung aufgezeigt wurde, also konkret: der t-Test für einen Mittelwerts- und der F-Test für einen Standardabweichungsunterschied, her.

In Minitab ist zu diesem Zweck ein Chi-Quadrat-Test vorhanden. Eine Beispielausgabe hier zeigt ein Ergebnis eines solchen Tests auf:

Hier zunächst die Datensituation:

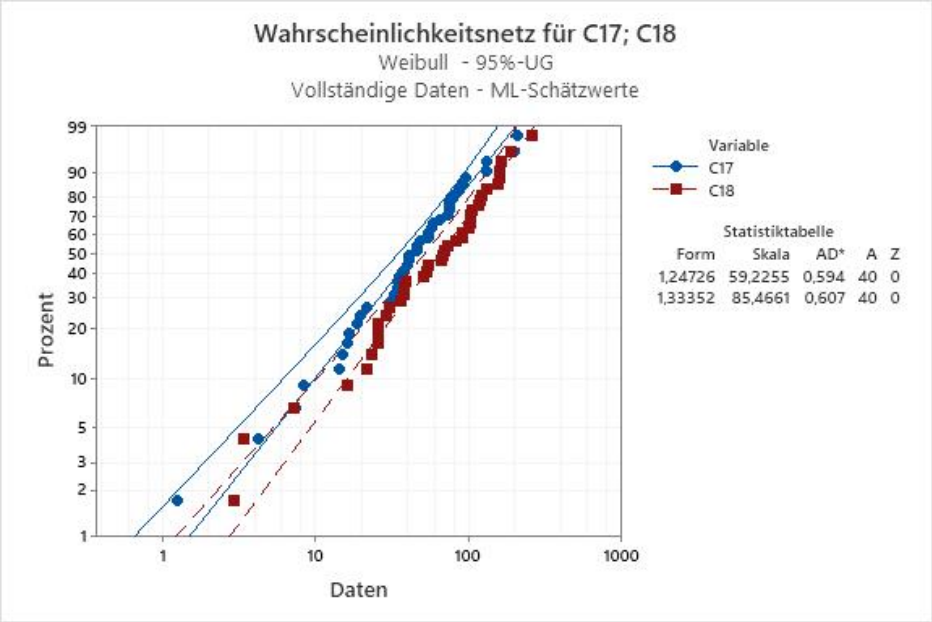


Figure 145: Zwei Weibull-Vert.: Form = 1,5, Skalen = 70 und 90

Das Testergebnis lautet:

Verteilungsanalyse: C17; C18

Test auf gleiche Form- und Skalenparameter

Chi-Quadrat	DF	p
4,09673	2	0,129

Test auf gleiche Formparameter

Chi-Quadrat	DF	p
0,146012	1	0,702

Test auf gleiche Skalenparameter

Chi-Quadrat	DF	p
4,03239	1	0,045

Figure 146: Testergebnis Parametertest Weibull

Die Interpretation ist nicht ganz einfach und wird dem Leser überlassen ...

14.4 Felddaten- und Garantieanalyse

Produziter Produkte werden schlussendlich den die Kunden geliefert, die dann z.B. als sog. OEM (Original Equipment Manufacturer, Erstausrüster) diese zugelieferten Bauteile in deren Produkten verbauen. Die Defekte, die z.B. in der Garantiezeit auftreten, sind dann zu analysieren, ob es z.B. auffällige Häufigkeiten wie Trends gibt.

Da sich Produkte und auch Fertigungsprozesse zudem permanent ändern, ist eine genaue zeitliche Zuordnung von Produktionszeitraum und Ausfall wichtig.

In der Praxis wird dieses durch eine relativ einfache Datenstruktur geleistet, die hiermit vorgestellt wird. Die Ausfälle werden dem Produktionszeitraum (typischerweise einem Monat) und der Gesamtzahl der Produktion zugeordnet durch folgende Dreiecksmatrix:

Versand	Monat1	Monat2	Monat3	Monat4	Monat5	Monat6	Monat7	Monat8	Monat9	Monat10	Monat11	Monat12
1000	0	0	0	0	0	0	0	0	0	0	0	0
1000		0	0	0	0	0	0	0	0	0	0	0
1000			0	0	1	4	3	1	1	2	4	5
1000				1	1	0	1	0	0	1	2	1
1000					0	0	0	1	3	1	4	2
1000						2	1	2	1	3	1	2
1000							1	0	1	1	2	1
1000								0	2	1	2	1
1000									0	0	1	0
1000										0	1	1
1000											1	0
1000												2

Figure 147: Garantiedatenstruktur

Die Auslieferungszahlen sind hier in der Spalte Versand zu finden. Zu diesen 12 Monaten sind die Rückläufer in den Monats-Spalten notiert.

Diese Daten werden nun in eine Intervallzensurgestalt umgerechnet. Zu fragen ist: Wie viele Bauteile sind im ersten Monat nach Produktion reklamiert worden. Diese Anzahl steht in den Diagonalen, also für den ersten Monat nach Auslieferung sind jeweils $1+2+1+1+2=7$ Bauteile defekt gewesen. Für den zweiten Monat sind es $1+1+2+1=5$.

Die Frage nach den noch intakten Bauteilen ergibt sich am einfachsten durch die Betrachtung der jüngsten Zahlen. Das wäre hier der Monat 12. In diesem Monat sind 1000 Bauteile produziert worden, und es wurden gleich in diesem Monat 2 Bauteile defekt gewesen (das entspricht der Zahl in der Spitze der Dreiecksmatrix). Somit sind noch 998 Bauteile aus Monat 12 intakt.

Für den Monat 11, also 2 Monate seit Produktionsbeginn dieser Bauteile, ist lediglich 1 Bauteil defekt gewesen. Das entspricht der Summe in Zeile 11 der Dreiecksmatrix.

In Minitab können nun diese Summen ausgerechnet werden und in das Schema einer Intervallzensur übertragen werden. Das Ergebnis hierfür ist:

Startzeit	Endzeit	Häufigkeiten
0	1	7
1	2	5
2	3	7
3	4	10
4	5	12
5	6	4
6	7	8
7	8	6
8	9	5
9	10	5
10	11	0
11	12	0
1	*	998
2	*	999
3	*	998
4	*	999
5	*	994
6	*	994
7	*	988
8	*	989
9	*	993
10	*	979
11	*	1000
12	*	1000

Figure 148: Garantiedaten in Intervallzensur

Die Analyse, die folgt, ist klar: Identifizierung der Verteilung, Schätzen der Ausfallperzentile, usw. Eine zusätzliche Funktionalität ist in Minitab noch vorhanden, indem die Ausfallzahlen prognostiziert werden können insb. unter Berücksichtigung der Garantiezeit. Bauteile, die nicht in der Garantiezeit ausfallen, sind betriebswirtschaftlich im Sinne eines Regresses des Herstellers nicht mehr relevant. Es

muss die Zahl der 'Bauteile in Risk' berechnet werden, um die Anzahl der Ausfälle mit der jeweiligen Ausfallwahrscheinlichkeit zu schätzen.

Produktionszeitplan

Zukünftiger Zeitraum	1	2	3	4	5	6	7	8	9	10	11	12
Produktionsmenge	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Tabelle der prognostizierten Anzahl und Kosten der Ausfälle

Zukünftiger Zeitraum	Potenzielle Anzahl der Ausfälle	Prognostizierte Anzahl der Ausfälle	95%-Poisson-KI Untergrenze	95%-Poisson-KI Obergrenze	Prognostizierte Kosten der Ausfälle
1	12931	13,107	7,000	22,366	2621,5
2	13931	27,493	18,193	39,868	5498,6
3	14931	43,180	31,272	58,127	8636,0
4	15931	60,189	45,952	77,445	12037,8
5	16931	78,542	62,137	97,949	15708,3
6	17931	98,256	79,792	119,712	19651,2
7	18931	119,351	98,901	142,784	23870,2
8	19931	141,844	119,462	167,201	28368,8
9	20931	165,751	141,478	192,994	33150,3
10	21931	191,089	164,955	220,188	38217,8
11	22931	217,873	189,901	248,804	43574,5
12	23931	246,116	216,326	278,862	49223,2

Figure 149: Garantieprognose - Zahlen

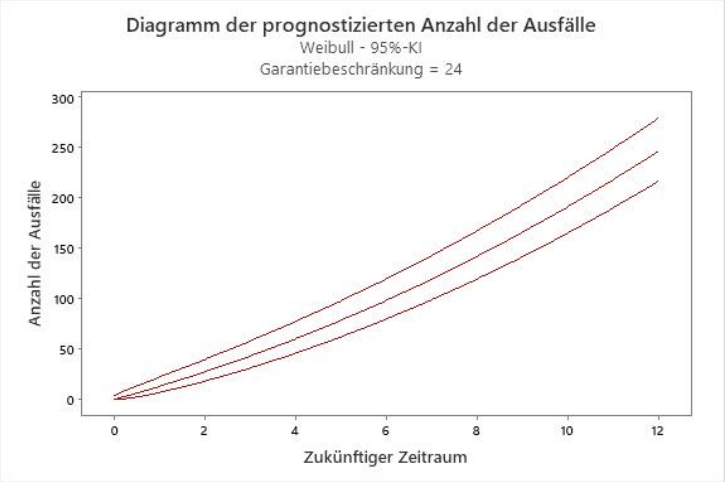


Figure 150: Garantieprognose - Grafik

14.5 Testplanung

Aus statistischer Sicht ist es wichtig, Tests auch so planen zu können, dass die Aussagen daraus Anforderungen an Trennschärfe genügen können (wie das auch bei den Tests basierend auf der Normalverteilung überlegt wurde).

Es werden in der Praxis grundsätzlich zwei Vorgehen unterschieden: die Schätzung der Parameter der Verteilung zum dem Zweck, dass (prinzipiell) jede Perzentile bzw. Prozentsatz geschätzt werden können. Zum anderen ein Nachweistest, mit dem typischerweise eine vorgegebene Mindestzuverlässigkeit, die die Bauteile erfüllen müssen, aufgezeigt werden soll.

Letztere Tests sind einfacher und erfordern weniger statistische Anforderungen, da lediglich eine Mindestanforderung gezeigt werden muss. Wie zuverlässig die Bauteile genau sind, ist nicht erforderlich zu zeigen - das führt auch in der Statistik zu weniger Anforderungen an die Stichproben.

14.5.1 Schätzungstests

Zunächst der Schätzungstest: bei einer Schätzung eines Parameters wird der wahre wert dieses Parameters durch ein Konfidenzintervall zum Ausdruck gebracht. Sind also $\hat{\alpha}$ und $\hat{\beta}$ die Parameterschätzungen, dann sind die wahren Parameter in Intervallen:

$$\alpha \in (\alpha_u, \alpha_o), \quad \beta \in (\beta_u, \beta_o) \tag{28}$$

Anm.: die Konfidenzintervalle sind nicht mehr symmetrisch und können i.d.R. nicht mehr einfach durch eine geschlossene Formel mit $\hat{\alpha}$ bzw. $\hat{\beta}$ ausgedrückt werden. Es gibt zudem ggf. viele verschiedene Formeln für diese Intervalle; welche die beste Formel ist, ist hier nicht zu klären, hier wird opportunistisch gearbeitet und das Ergebnis von Minitab wird übernommen ...

Die Annahme der Konfidenzintervalle für die Parameter der Verteilung führt letztlich zu einer Konfidenz der Wahrscheinlichkeitsverteilung; dieses ist hier zu sehen:

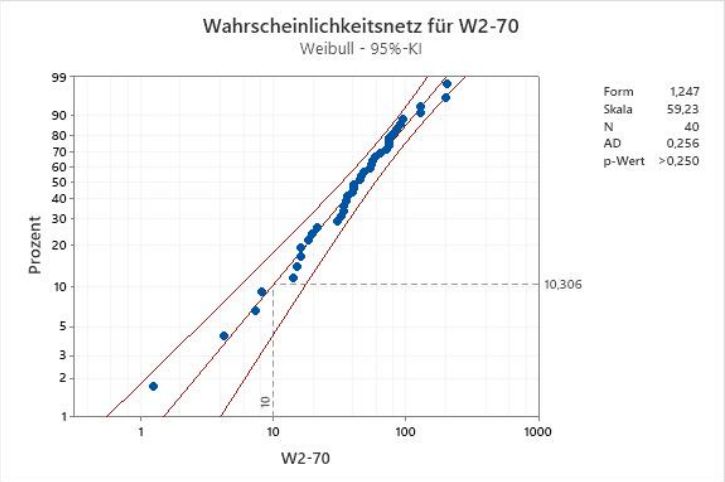


Figure 151: Verteilung mit Konfidenzintervallen

Für die hier zu sehende Perzentile 10 gibt es bzgl. der Schätzung des zugehörigen Ausfallprozentwertes nun ein Konfidenzintervall.

Umgekehrt kann nun auch der Prozentwert vorgegeben werden, das führt dann zu einer Konfidenz des Perzentils.

Hier können diese Intervalle mit jeweils ca. $\pm 5\%$ für den Prozentsatz bzw. ± 5 für den X-Achsenwert aus der Grafik heraus geschätzt werden.

Beides ist für eine Trennschärfe wichtig uns muss entsprechend verwendet werden können. Die Breite des Konfidenzintervalles ist also zu wählen (wird dann im weiteren Präzision genannt) in Form einer Prozent- oder einer Wertedifferenz. Aus dieser Vorgabe heraus berechnet sich dann die Stichprobengröße.

Hier der Minitab-Dialog mit den Angaben zu unseren Daten (d.h. Verteilung = Weibull, Form = 1,247, Skala = 59,23):

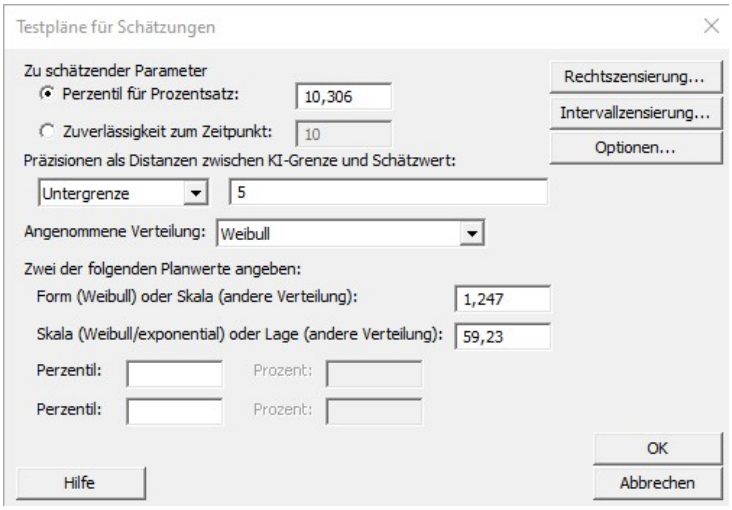


Figure 152: MTB-Dialog zur Testplanung - Schätzung

Anm.: die Schätzung nach Perzentile bzw. Wert wird mit der Radio-Schaltfläche eingestellt.

Die Ergebnisse für die Schätzungen lauten:

Testpläne für Schätzungen

Unzensierte Daten
Geschätzter Parameter: 10,306tes Perzentil
Berechneter Planschätzwert = 9,99730
Ziel-Konfidenzniveau = 95%
Präzision in Form eines einseitigen Konfidenzintervalls, das die Untergrenze für den Parameter darstellt.

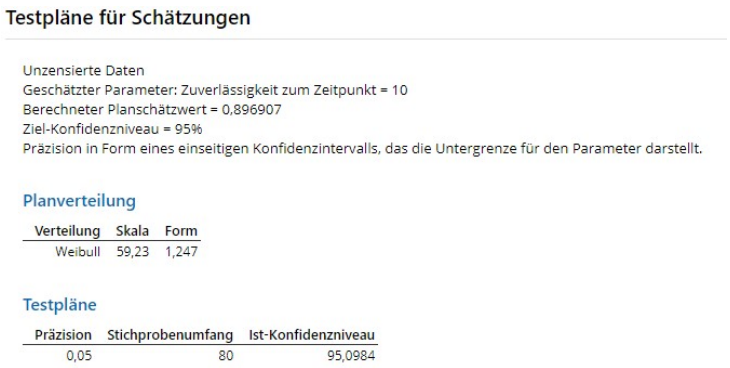
Planverteilung

Verteilung	Skala	Form
Weibull	59,23	1,247

Testpläne

Präzision	Stichprobenumfang	Ist-Konfidenzniveau
5	19	95,0153

Figure 153: Testplanung - Schätzung über Prozente



Testpläne für Schätzungen

Unzensierte Daten
Geschätzter Parameter: Zuverlässigkeit zum Zeitpunkt = 10
Berechneter Planschätzwert = 0,896907
Ziel-Konfidenzniveau = 95%
Präzision in Form eines einseitigen Konfidenzintervalls, das die Untergrenze für den Parameter darstellt.

Planverteilung

Verteilung	Skala	Form
Weibull	59,23	1,247

Testpläne

Präzision	Stichprobenumfang	Ist-Konfidenzniveau
0,05	80	95,0984

Figure 154: Testplanung - Schätzung über Werte

14.5.2 Nachweistests

Bei einem Nachweistest wird, wie oben erwähnt, eine Mindestanforderung an die Produkte getestet. Die Abnehmer der Produkte haben also ein Kriterium im statistischen Sinn definiert wie etwa: bis zu einer Zyklenzeit von 10000 dürfen max. 0,5% der Bauteile ausfallen.

Die statistische Testplanung sieht für einen solchen Nachweise typischerweise keine Ausfälle vor. Betrachtet man die Formel für das Überleben der Bauteile anhand der Verteilung, also $R(t)$, ist hier gefordert, dass kein Bauteil defekt wird vor 10000 Zyklen. Kein Bauteil bedeutet in Wahrscheinlichkeit gedacht, dass alle Bauteile überleben. Insofern ergibt sich folgende Gleichung:

$$R(t) = \alpha^{\frac{1}{N}} \tag{29}$$

wobei N die Anzahl der zu testenden Bauteile ist.

Umstellen liefert die für das statistische Bestehen notwendige Testzeit:

$$t = R^{-1}(\alpha^{\frac{1}{N}}) \tag{30}$$

Gibt man die Zeit des Nachweistests vor, dann erfolgt die Rechnung durch auflösen nach N.

Sind Ausfälle zugelassen, dann

$$R^N + \quad N(1-R)R^{N-1} + \quad \ldots \quad + \binom{N}{k}(1-R)^k R^{N-k} + \quad \ldots \quad + \binom{N}{m}(1-R)^m R^{N-m} = \quad \alpha$$

Figure 155: Testplanung: Gleichung für erlaubte Ausfälle

Da keine Ausfälle erwartet werden, kann nicht anhand einer Ausfallzahl geschätzt werden. Insofern ist für die Testplanung eine Minimalkenntnis (diese kann auch vernünftig angenommen sein) vorzugeben. Im Falle einer Weibullverteilung ist das die Form, für andere Verteilungen die Skala.

Der Dialog in Minitab zu dem Fallbeispiel oben hat z.B. für 6 Probebauteile folgende Gestalt:

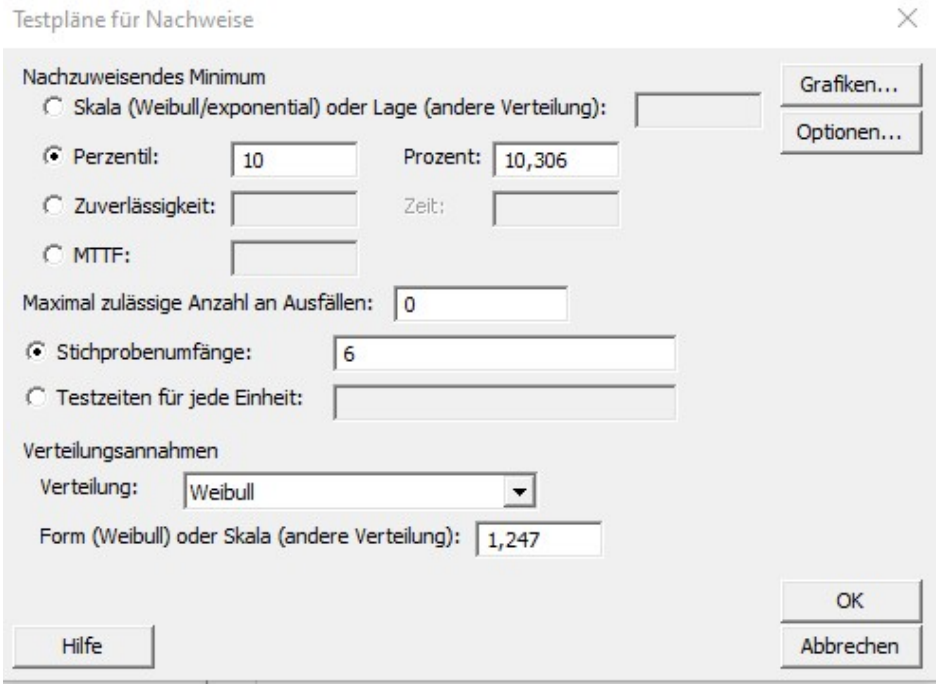


Figure 156: MTB Dialog für Nachweistestplanung

Das Ergebnis ist nichts weiter als die Zeit, die diese 6 Bauteile im Test ohne Ausfall schaffen müssen:



Testpläne für Nachweise

Testplan für Zuverlässigkeit
Verteilung: Weibull, Form = 1,247
Ziel-Perzentil = 10, Ist-Konfidenzniveau = 95%

Testpläne

Ausfälle im Test	Stichprobenumfang	Testzeit
0	6	33,9437

Figure 157: MTB-Ergebnis für Nachweistestplanung

14.5.3 Ausfallschätzung unter Vorwissen: Beyer/Lauster

Ist noch zu erarbeiten ...

14.6 Ausfallursachenanalyse

Besteht ein Produkt aus unterschiedlichen Komponenten, dann können diese Komponenten einzeln gesehen nach ihrer eigenen Verteilung ausfallen. Hier wird im weiteren ein System solcher Komponenten mit Ausfall betrachtet so, dass wenn eine Komponente ausfällt, dass dann das Gesamte System ausgefallen ist (Reihenschaltung der Komponenten).

Wenn man also zu einem Systemausfall den Grund für den Ausfall, also z.B. die ausgefallene Komponente, notiert, dann kann man mathematisch durch eine Faltung den Gesamtausfall beschreiben.

Also Beispiel hierzu ein Minitab-Datensatz:

C38	C39-T
Ausf	Komp
135,299	B
126,101	B
184,738	C
25,427	C
83,572	A
63,686	C
124,016	B
133,142	B
13,051	C
153,031	B
137,410	B
157,591	B
177,210	B
112,724	C
56,005	A
52,471	A
69,638	A

Figure 158: MTB-Daten für Ausfallursachenanalyse

In der Analyse besitzt jede Komponente nun ihre eigene Ausfallverteilung:

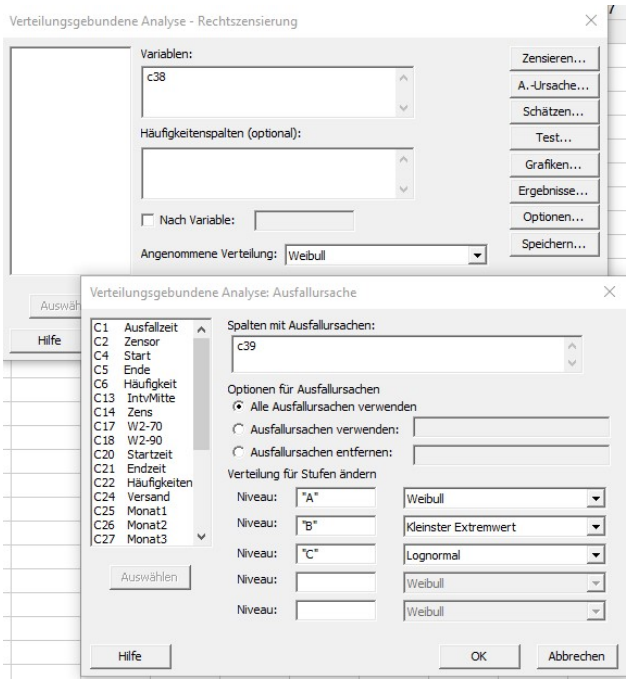


Figure 159: MTB-Dialoge für Ausfallursachenanalyse

Das Ergebnis ist eine zusammengesetzte Verteilung:

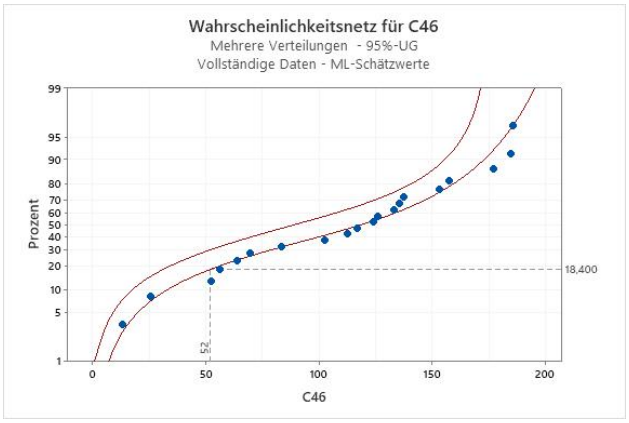


Figure 160: Ausfallursachenanalyse - Gesamtverteilung

Die einzelnen Verteilungen sind dabei die folgenden:

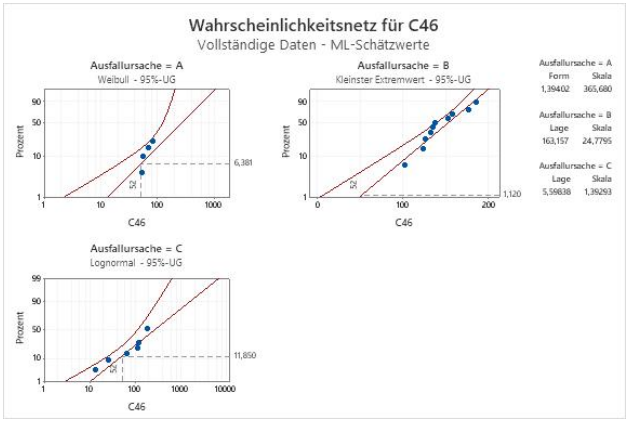


Figure 161: Ausfallursachenanalyse - einzelne Verteilungen

Um zu zeigen, dass die Wahrscheinlichkeiten tatsächlich passen, berechnen wir für eine feste Perzentile (z.B. 52) die Prozentsätze nach der totalen (Ausfall-) Wahrscheinlichkeit.

$$P_{ges} = 1 - (1 - P_1) * (1 - P_2) * (1 - P_3) \tag{31}$$

Mit konkreten Werten in Excel:

Ausf%	1-Ausf%	Produkt	1-Prod
6,38%	93,62%	81,60%	18,40%
1,12%	98,88%		
11,85%	88,15%		

Figure 162: Ausfallursachenanalyse - totale Wahrscheinlichkeit

14.7 Regression mit Lebensdauerdaten

Eine Regression gegen Lebensdauerdaten im Sinne der Erforschung von Ursache und Wirkung ist ebenfalls möglich. Die Vorgehensweisen sind gegenüber der linearen Regression sehr ähnlich, jedoch ist zu beachten, dass die Verteilungsfunktion der Ausfallwahrscheinlichkeiten nicht-normal ist. Das führt dazu, dass insb. das Residuum speziell betrachtet werden muss.

Bei einer linearen Regression wird als Regressionsgleichung eine lineare Funktion angegebene, die die Erwartungswerte in Bezug auf die unabhängigen Variablen berechnet. Dieses gilt für die Regression mit Lebensdauerdaten ebenfalls nicht. Die Daten beinhalten neben der Ausfallzeit nun noch gewisse Einstellungen, mit denen die Produkte hergestellt wurden oder denen sie unterlagen bis zum oder beim Ausfall. Ein Beispiel:

Start	Ende	Konstruktion	Gewicht
132	144	Standard	10,0
144	156	Standard	10,0
156	168	Standard	10,0
192	204	Standard	10,0
192	204	Standard	10,0
204	216	Standard	10,0
240	*	Standard	10,0
240	*	Standard	10,0
168	180	Standard	7,5
216	228	Standard	7,5
216	228	Standard	7,5
240	*	Standard	7,5
240	*	Standard	7,5
240	*	Standard	7,5
240	*	Standard	7,5

Figure 163: Ausfallanalyse Regression - Daten

Anm.: die Ausprägungen für 'Konstruktion' sind 'Standard' und 'Neu', die für

Gewicht $5, 7, 5 und 10$.

Die Regressionsrechnung wird in Minitab wie folgt eingestellt:

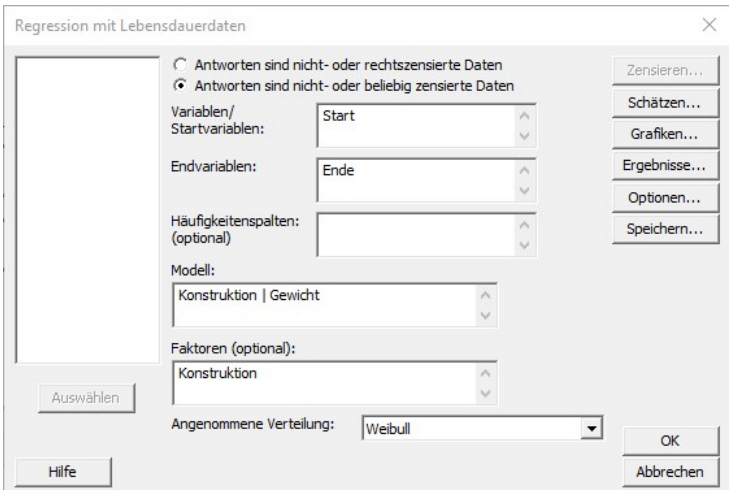


Figure 164: Ausfallanalyse Regression - MTB Dialog

Anm.: der senkrechte Strich im Modell-Feld ist die Kodierung in Minitab, Wechselwirkungen zu berechnen.

Das Ergebnis ist dann wie folgt:

Regressionstabelle

					Normales 95,0%-KI	
Prädiktor	Koef	Standardfehler	z	p	Untergrenze	Obergrenze
Schnittpunkt mit y-Achse	6,58886	0,229838	28,67	0,000	6,13839	7,03934
Konstruktion						
Standard	-0,500346	0,328752	-1,52	0,128	-1,14469	0,143996
Gewicht	-0,0450129	0,0264009	-1,70	0,088	-0,0967578	0,0067319
Konstruktion*Gewicht						
Standard	-0,0245658	0,0383084	-0,64	0,521	-0,0996490	0,0505173
Form	5,84180	1,08761			4,05577	8,41436

Figure 165: Ausfallanalyse Regression - erste Analyse

Die Vorgehensweise ist völlig analog zur Analyse eines GLM. Die Wechselwirkung ist nicht signifikant und muss aus dem Modell entfernt werden, final ergibt sich folgendes Modell:

Regression mit Lebensdauerdaten: Start vs. Konstruktion; Gewicht

Zensierungsinformation	Anzahl
Rechtszensierter Wert	25
Intervallzensierter Wert	23

Schätzmethode: Maximum-Likelihood

Verteilung: Weibull

Beziehung zu Beschleunigungsvariable(n): Linear

Regressionstabelle

					Normales 95,0%-KI	
Prädiktor	Koef	Standardfehler	z	p	Untergrenze	Obergrenze
Schnittpunkt mit y-Achse	6,68731	0,193766	34,51	0,000	6,30754	7,06709
Konstruktion						
Standard	-0,705643	0,0725597	-9,72	0,000	-0,847857	-0,563428
Gewicht	-0,0565899	0,0212396	-2,66	0,008	-0,0982187	-0,0149611
Form	5,79286	1,07980			4,02001	8,34755

Figure 166: Ausfallanalyse Regression - finale Analyse

Die Koeffizienten werden in folgendes Modell eingesetzt:

$$y(p) = e^{6{,}68731 - 0{,}705643 - 0{,}0565899 * Gewicht} + 5,79286 * \phi^{-1}(p) \tag{32}$$

für die Konstruktion 'Standard', für Konstruktion 'Neu' ergibt sich:

$$y(p) = e^{6{,}68731 - 0{,}0565899 * Gewicht} + 5,79286 * \phi^{-1}(p) \tag{33}$$

wobei das Residuum $\phi^-1(p)$ die Perzentile zu einem Prozentsatz liefert für die Kleinste Extremwertverteilung.

Das Residuum hat folgende Verteilungsfunktion für die Verteilungen der Ausfälle in der Regression:

- Weibull $\rightarrow$ Kleinste Extremwertverteilung
- Lognormal $\rightarrow$ Normal-Verteilung
- Kleinste Extremwertverteilung –¿ Kleinste Extremwertverteilung

Anm.: die Verteilung des Residuums muss eine Verteilung sein, die auch negative Werte zulässt ...

14.8 Beschleunigte Lebensdauertest

Die hier vorgestellte Analyse der Regression ist für stetige Variablen sofort als Regressionsrechnung für eine Beschleunigungsvariable zu interpretieren. In der Praxis werden mit beschleunigten Versuchen die Lebensdauertests z.T. erheblich verkürzt. Typischerweise verwendete Beschleunigungsvariablen sind Temperatur (insb. bei elektronischen Bauteilen), Druck (z.B. Anpressdruck) für mechanische Bauteile (Bremsen, Achsen). Jede Form der Überhöhung, die die Normallast der Bauteile im tatsächlichen Einsatz normalerweise nicht zu erleiden haben, ist eine Beschleunigungsvariable, die in den Versuchen in Stufen eingestellt wird.

Als Beschleunigungen mit attributiven Faktoren dienen z.B. Lastprofile, also komplexere Überlastungen, z.B. durch stark variierende Temperaturprofile in Kombination mit mechanischen Überforderungen. Solche Versuche sollen Praxisüberlastungen in Kombination darstellen und so in den Tests dienen.

Die Beschleunigungsrechnung stellt lediglich eine Erweiterung ggü. der oben vorgestellten Regression dar in dem Sinne, dass auch nichtlineare Beschleunigungsgesetze als Regressionsgesetze zugelassen sind. In Minitab stehen folgende Regressionen zur Verfügung:

- Arrhenius-Gesetz
- Power-Log-Gesetz (Ln(Power))
- Inverse Temperatur

Am bekanntesten in der Beschleunigungsanalyse ist das Arrhenius-Gesetz:

$$y(p) = e^{\beta_0 + \beta_1 * \left(\frac{11.604{,}53}{T\ (in\ Grad\ Celsius} + 273{,}16 \right)} + \sigma * \phi^{-1}(p) \qquad (34)$$

Das Inverse Temperatur-Gesetz entspricht dem Arrhenius-Gesetz, nur ist die Boltzmann-Konstante nicht vorhanden, d.h.

$$y(p) = e^{\beta_0+\beta_1*\left(\frac{1}{T(in\ Grad\ Celsius)}+273{,}16\right)} + \sigma * \phi^{-1}(p) \tag{35}$$

Das Ln(Potenz)-Gesetz lautet:

$$y(p) = e^{\beta_0+\beta_1*(\ln(Beschleunigungsvariable))} + \sigma * \phi^{-1}(p) \tag{36}$$

Beispiel: gegeben sei der folgende Datensatz, die Beschleunigungsvariable ist die Temperatur mit den Einstellungen 125, 150 und 175 Grad Celsius.

Um zu überprüfen, ob eine sinnvolle Beschleunigung hier durch die Temperatur erfolgt, ist als erstes zu prüfen, ob das Ausfall-Verhalten der Bauteile auch tatsächlich beschleunigt wird. Dafür betrachtet man die Wahrscheinlichkeitsnetze:

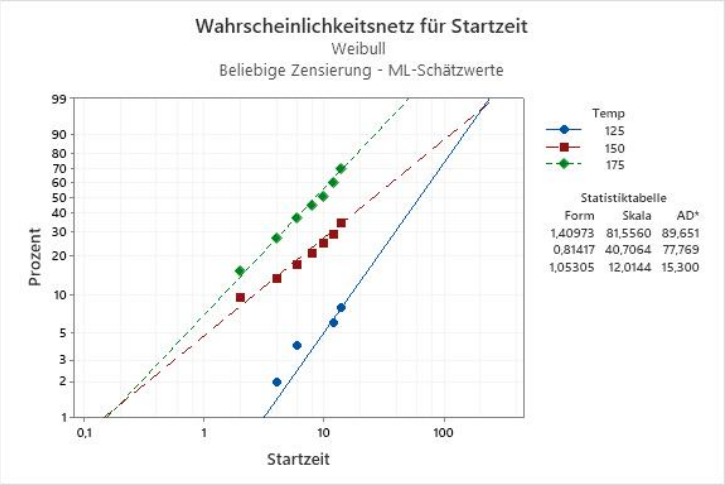


Figure 167: Wahrscheinlichkeitsnetze bei Beschleunigung

Zu sehen ist, dass die WS-Netze halbwegs parallel verlaufen und für jede höhere Stufe der Beschleunigung die Ausfälle auch tatsächlich früher erfolgen. Bestimmend für die Steilheit der Linien ist bei der Weibullverteilung die Form. Diese sollte bei einer idealen Beschleunigung gleich sein, d.h. die WS-Netze wären parallel. Dieses wird nun getestet.

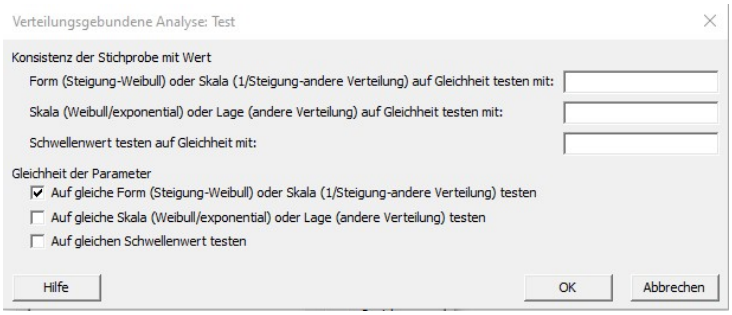


Figure 168: MTB-Dialog für Test gleiche Formparameter

Das Ergebnis ist:

Test auf gleiche Formparameter

Chi-Quadrat	DF	p
1,16213	2	0,559

Figure 169: Testergebnis gleiche Formparameter

Anm.: in den vorliegenden Versuchen liegen für die niedrigste Stufe sehr wenig Ausfälle vor. Dieses führt u.a. dazu, dass hier kein signifikanter Unterschied ausgewiesen wird.

Nun kann das Beschleunigungsgesetz aus der Stichprobe heraus bestimmt werden. Der Dialog in MTB dafür ist:

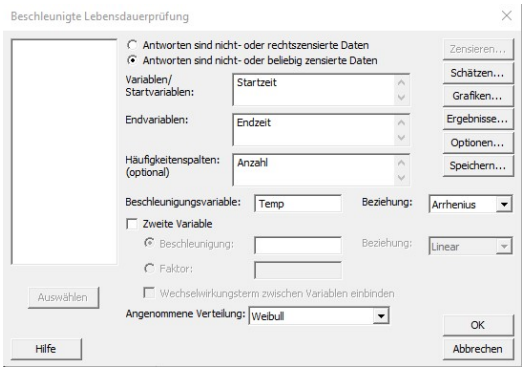


Figure 170: Dialog in Minitab für die Beschleunigungsrechnung

Das Ergebnis ist:

Beziehung zu Beschleunigungsvariable(n): Arrhenius

Regressionstabelle

					Normales 95,0%-KI	
Prädiktor	Koef	Standardfehler	z	p	Untergrenze	Obergrenze
Schnittpunkt mit y-Achse	-17,0990	4,13633	-4,13	0,000	-25,2061	-8,99195
Temp	0,755405	0,157076	4,81	0,000	0,447542	1,06327
Form	0,996225	0,136187			0,762071	1,30232

Figure 171: Ergebnis Beschleunigungsrechnung

Anm.: beim Arrhenius-Gesetz heißt die Konstante auch die sog. Aktivierungsenergie.

Am folgenden Diagramm ist zu erkennen, wie nun das Beschleunigungsgesetz ermittelt wird:

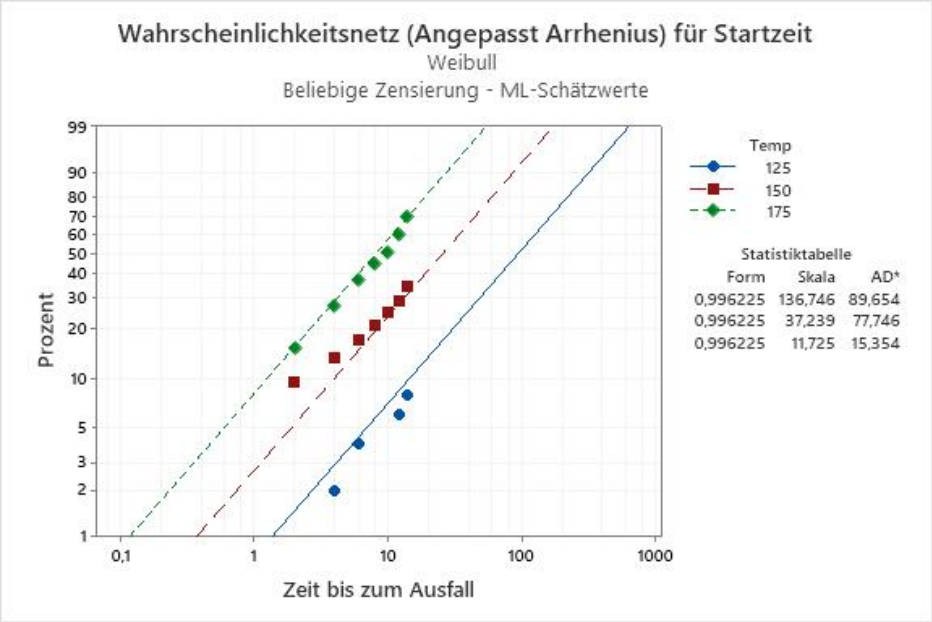


Figure 172: Beschleunigung - Wahrscheinlichkeitsnetz

Zu erkennen ist, das für jede Beschleunigungsstufe die Form der zugrundelegten Weibull-Verteilung gleich ist; der Formparameter wurde insg. gewichtet (nach Anzahl Ausfällen) gemittelt und so für die insg. Beschleunigung gesetzt.

Die Daten folgen recht gut den Wahrscheinlichkeitsnetzlinien, insofern ist - qualitativ gesehen - das Beschleunigungsgesetz akzeptabel.

Die Wirkung der Beschleunigung ist z.B. auf dem sog. Relationsdiagramm zu erkennen. Hier werden die Perzentile für 10, 50 (Median) und 90% ggü. der Temperaturveränderung vs. Startzeit als Kriterium eines Ausfalls aufgezeigt.

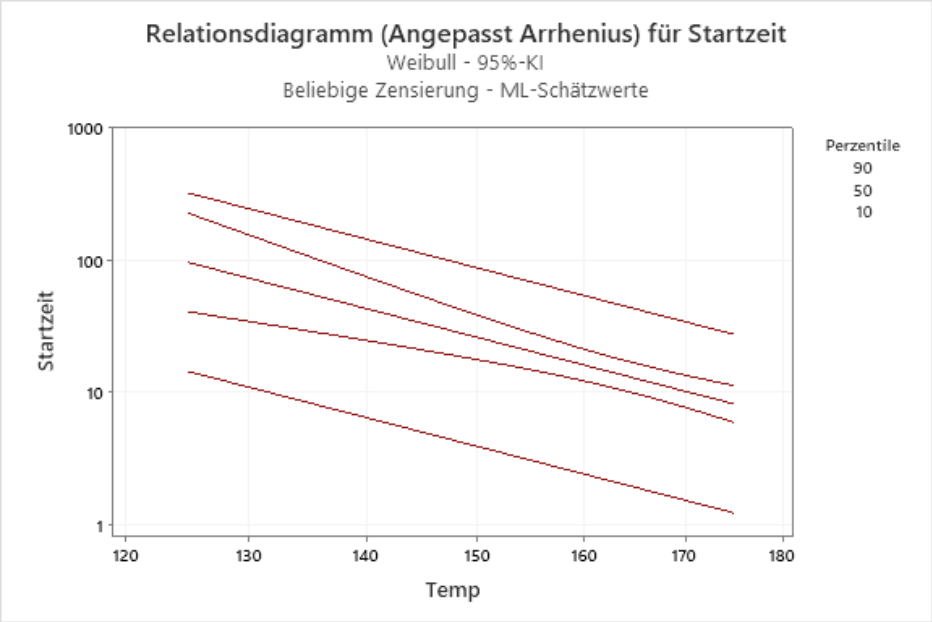


Figure 173: Relationsdiagramm der Beschleunigung

Zu beachten ist insb. die logarithmische Skala der Ausfallzeit (Startzeit). Bei höchster Beschleunigung von 175 Grad Celsius fallen demnach also geschätzt 50% der Bauteile bereits nach 10 Stunden aus, bei 125 Grad Celsius erst nach 100 Stunden.

Nun wird dieses Ergebnis verwendet, um mittels einer Extrapolation zu schätzen, wann die Bauteile unter Normallast ausfallen. Diese sei z.N: 60 Grad Celsius. Ähnlich wie bei der Regression muss nun ein neuer Prädiktorwert für die Regressionsgleichung der Beschleunigung angegeben werden. Der Dialog in Minitab hier ist:

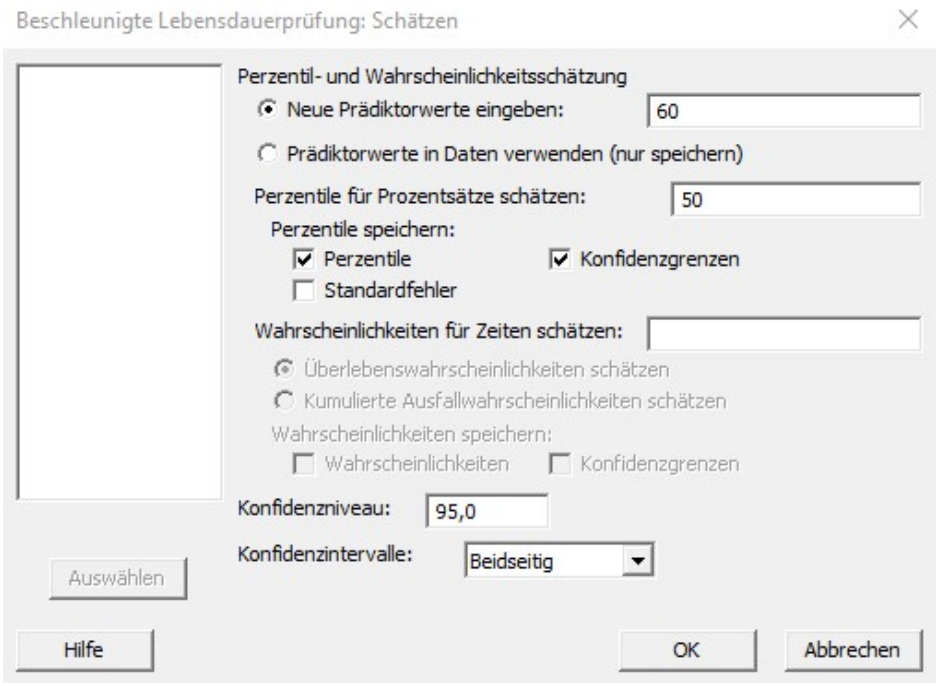


Figure 174: Dialog in MTB für die Extrapolation auf Designwertgröße

Die Normallast wird auch als Designwert bezeichnet. Um für diesen die Ausfallzeiten zu schätzen inkl. Konfidenzintervalle hier der Minitab-Dialog dazu:

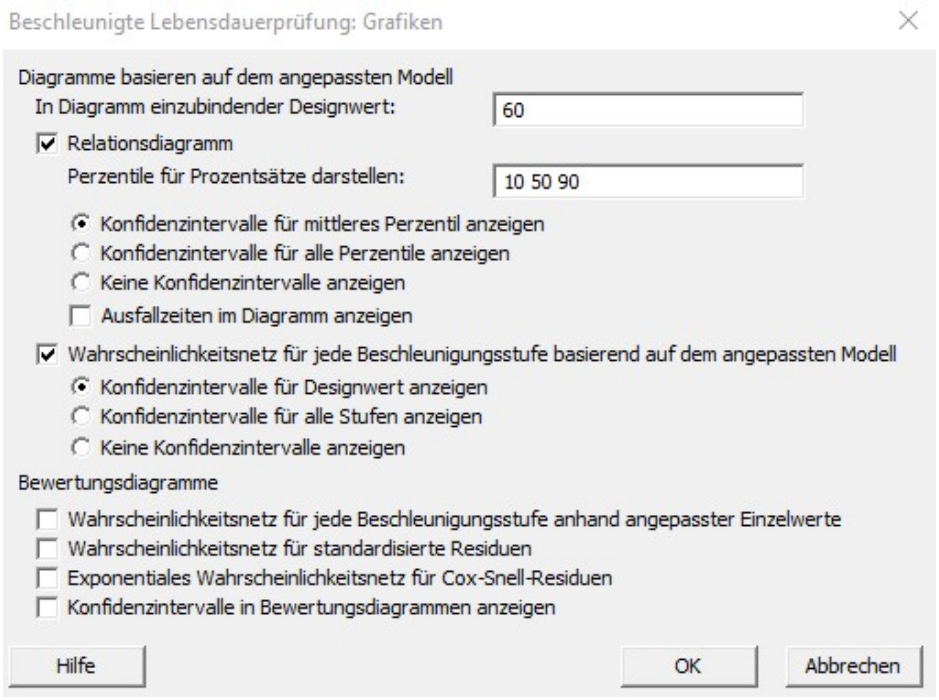


Figure 175: Dialog in MTB für die Angabe des Designwertes

Das Ergebnis ist dann etwas ernüchternd in Bezug auf die statistische Sicherheit:

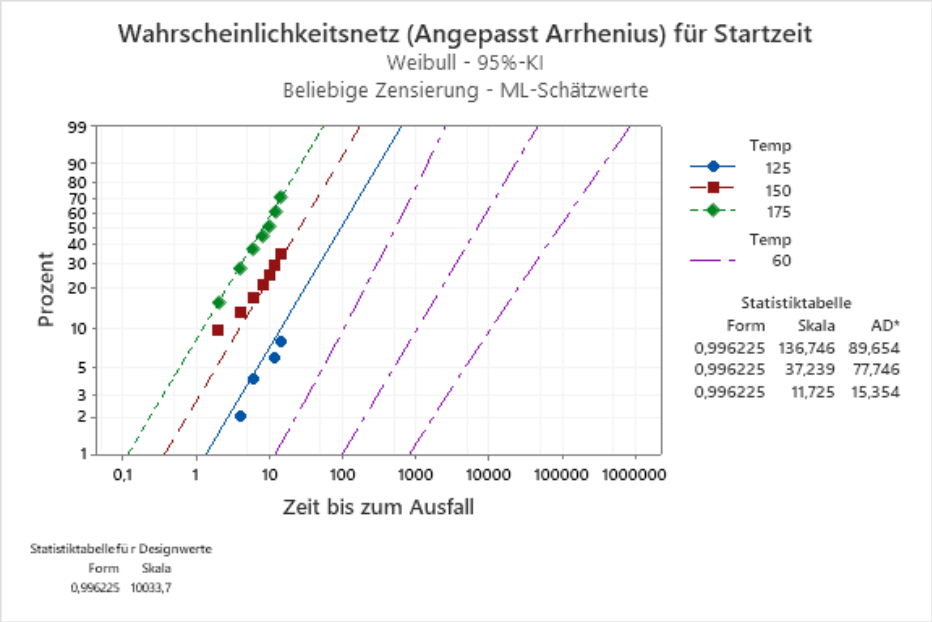


Figure 176: Ergebnis der Beschleunigungsrechnung für den sog. Designwert

Man darf der optischen Täuschung nicht unterliegen: die Konfidenzintervalle sind senkrecht in die Grafik einzutragen:

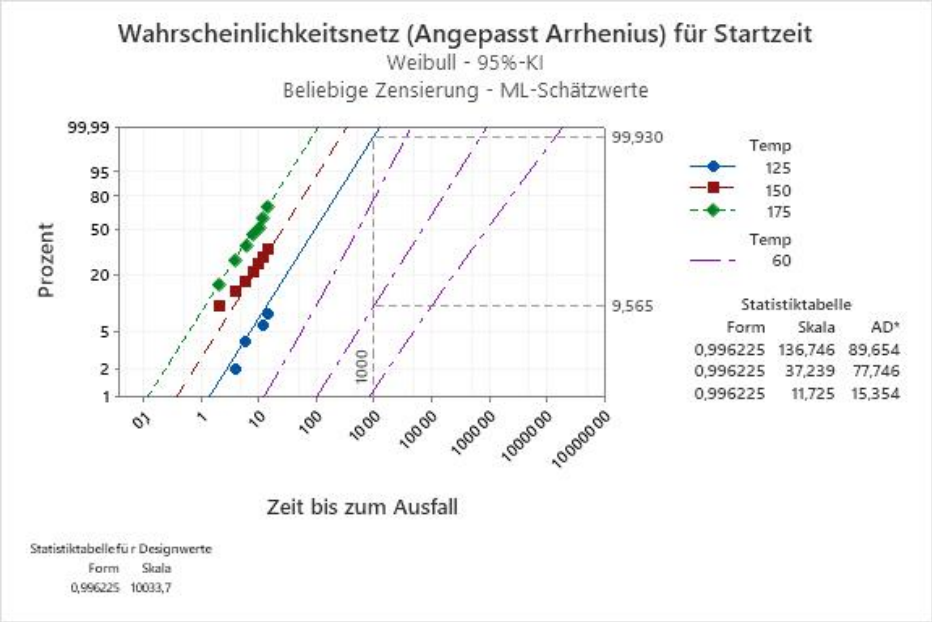


Figure 177: Ergebnis der Beschleunigungsrechnung für den sog. Designwert mit Konfidenz

Eine Konfidenz zwischen 9,6 und 99,9% bedeutet, dass man mit nahezu jedes Ausfallverhalten rechnen muss.

Die Gründe hierfür sind:

- Verwendung der Intervallzensur
- hohe Anzahl zensierter Werte
- Beschleunigungsgesetz ist nur akzeptabel, aber nicht sehr gut erkennbar (liefert zusätzliche Unsicherheiten für die Prognose, die zudem sehr stark hebelt)
- Design- und erste Testbeschleunigungsstufe sind relativ weit voneinander entfernt.

Solche Situationen sollte man in der Praxis dementsprechend vermeiden, da die Aussagen insb. durch die Extrapolation zu schwach werden.

Die relativ großen Konfidenzintervalle sind bei Praxisprüfungen stets ein Problem. Das ist ein Grund, warum Nachweistest unter Beschleunigung nicht zulässig sind. Jedoch um in der Entwicklung von Produkten in kurzer Zeit aufzuzeigen, dass bestimmte Veränderungen oder neue Materialien die Lebensdauer verändern (nach Wunsch erhöhen ...), sind solche Versuche sehr gut geeignet.

14.8.1 Testplanung für beschleunigte Lebensdauertests

Auch beschleunigte Lebensdauerversuche können geplant werden. Nachweise, wie oben geschrieben, entfallen. Jedoch für Schätzungen sind Planungen möglich. Hierfür muss man nun die Stichprobengrößen für die Beschleunigungsstufen (Stressstufen genannt) ermitteln. Um es vorweg zu nehmen, die Anzahl der Versuche muss aus statistischer Sicht nicht in jeder Stufe gleich groß sein ... Details hierzu in folgendem Fallbeispiel. Für die eben anvisierte Beschleunigung werden nun aus Sicht der Statistik sinnvolle Stichprobengrößen ermittelt.

Die Testplanung soll jetzt nicht mehr intervallzensiert sein (das ist auch möglich, jedoch aus obigen Gründen nicht zu empfehlen ...). Im Gegenteil wird für die beschleunigten Tests keine Zensur mehr vorgesehen.

Der Dialog in Minitab ist zweigeteilt, zunächst die Angaben über die Schätzsituation und das Vorwissen (Formparameter). Die Werte sind der Beschleunigungsanalyse (s.o.) zu entnehmen (und ggf. gerundet), entsprechen aber dem Beispiel.



Figure 178: Testplanung unter Beschleunigung

Im Dialog 'Stress' sind noch die jetzt vorzusehenden Stressstufen anzugeben:

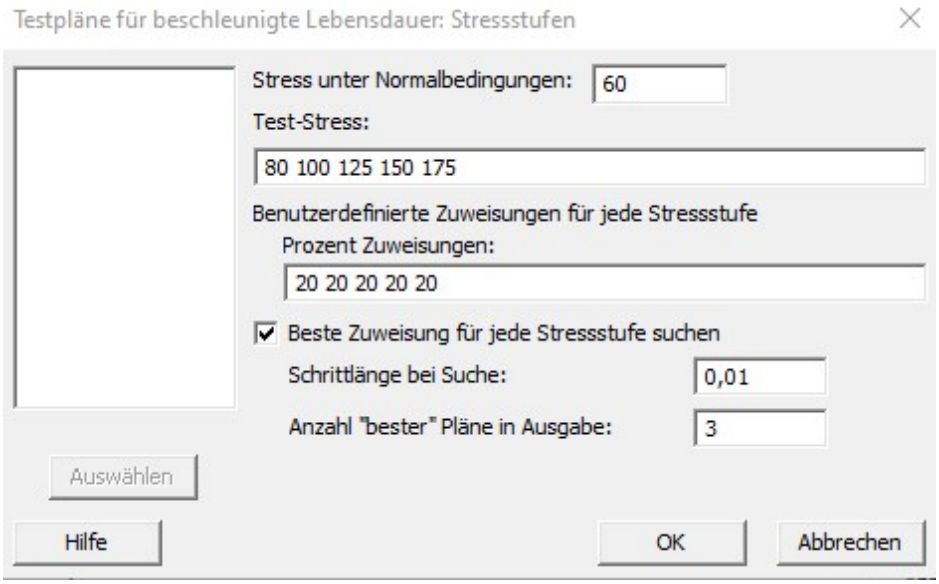


Figure 179: Testplanung unter Beschleunigung - Teil 2

Das Ergebnis dieser Planung ist:

Benutzerdefinierter Plan

Test-Stress	Prozentsatz der Ausfälle	Prozentuale Zuw	Stichproben-Einheiten	Erwartete Anzahl der Ausfälle
80	100	20	250	250
100	100	20	250	250
125	100	20	249	249
150	100	20	249	249
175	100	20	249	249

Erforderlicher Stichprobenumfang = 1247, Ist-Konfidenzniveau= 95,0022%

1.-bester „optimaler" Zuweisungstestplan

Test-Stress	Prozentsatz der Ausfälle	Prozentuale Zuw	Stichproben-Einheiten	Erwartete Anzahl der Ausfälle
80	100	55,0000	593	593
100	100	19,4063	209	209
125	100	1,4063	15	15
150	100	4,4063	48	48
175	100	19,7812	213	213

Erforderlicher Stichprobenumfang = 1078, Ist-Konfidenzniveau= 95,0003%

Figure 180: Testplanung unter Beschleunigung - Ergebnis

Zu sehen sind die Testpläne, die wie vorgesehen zunächst gleiche Anzahl Versuche je Stressstufe haben sollten. Benötigt werden hierzu insg. 1247 Bauteile. Der optimale Testplan nach besserer Verteilung der Stichproben auf die Stressstufen benötigt für die gleiche statistische Sicherheit noch 1078 Bauteile.

Die Anzahlen auf die Stufen ist typisch für eine Approximation, insb. für Extrapolationen. Auf der Stressstufe, die nächsten am Designwert liegt, sind am meisten Bauteile zugeteilt; danach folgt die Stufe, die am weitesten vom Designwert entfernt ist. Diese beiden Punkte stabilisieren eine Approximationsgerade am besten. Die anderen Stressstufen sind dann noch für eine weitere Stabilisierung geplant.

14.9 Probit-Analyse

Die sog. Probit-Analyse hat als Antwortvariable nicht mehr Zeiten oder Zeitintervalle, sondern ggü. einem verwendeten Stress, der auf Bauteilen lastet, eine Anzahl von Ausfällen. Diese Form der Regression ist bereits bekannt, es handelt sich dann bei 'Ausfall/nicht Ausfall'-Kodierung um eine logistische binäre Regression. In der Probit-Analyse für die Zuverlässigkeit bzw. Lebensdauer stehen jedoch die Ausfallverteilungen für Ausfallmechanismen zur Verfügung, um gegen diese eine entsprechende Analyse zu rechnen.

Das folgende Beispiel zeigt eine typische Datensituation: die Ausfälle sind ggü. einer Anzahl Versuche vorsummiert.

Stress	N	Brüche
500	15	1
550	15	1
600	10	2
650	10	3
700	9	5
750	10	8
800	10	8
850	10	9

Figure 181: Probit - Daten für Fallbeispiel

Der Dialog für die Probit-Analyse ist in Minitab wie folgt auszufüllen (als Verteilung sei hier einfach die Normalverteilung gewählt, im insb. die Verbindung zur logistischen Regression explizit aufzuzeigen):

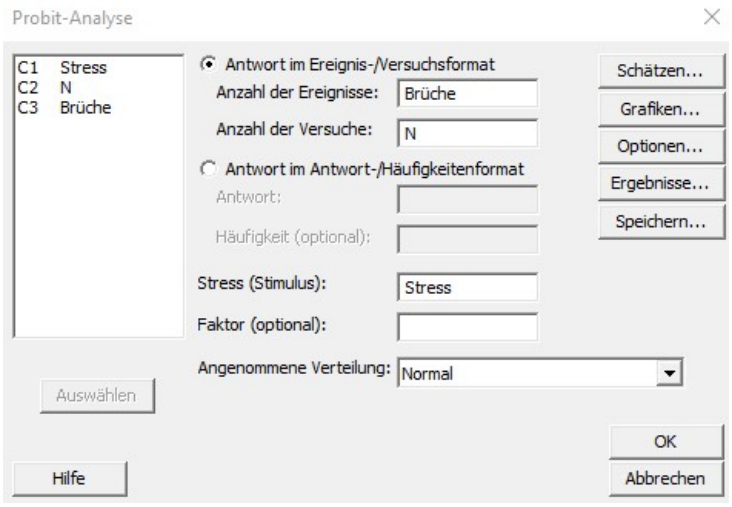


Figure 182: MTB Probitanalyse Dialog 1

Die Schätzwerte sind noch einzutragen:

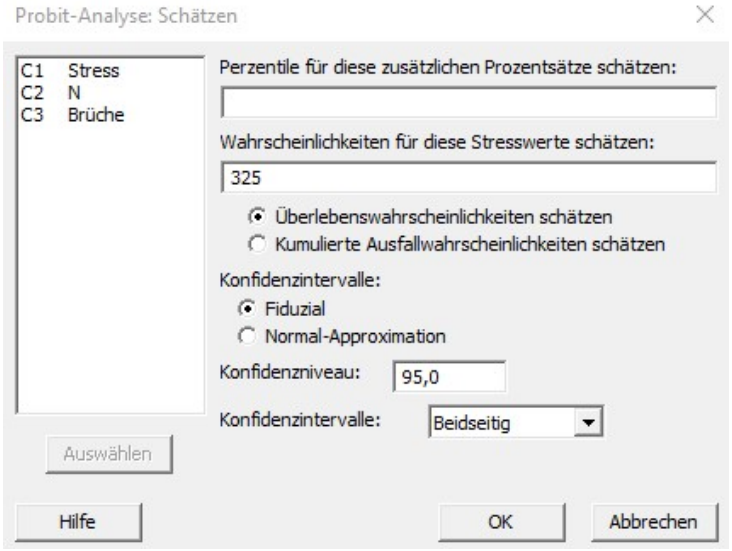


Figure 183: MTB Probitanalyse Dialog 2

Das Ergebnis der Rechnung ist:

Regressionstabelle

Variable	Koef	Standardfehler	z	p
Konstante	-6,20376	1,06565	-5,82	0,000
Stress	0,0089596	0,0015615	5,74	0,000
Natürlich				
Antwort	0			

Figure 184: Probitanalyse Ergebnis Regressionstabelle

Das aus der Berechnung abgeleitete Modell ist:

$$P(Bruch) = \Phi(-6,20376 + 0,0089596 * Geschwindigkeit) \qquad (37)$$

Für z.B. eine Geschwindigkeit von 1000 ist $P(Bruch) = \Phi(-6,20376+8,9596) = \Phi(2,75584) = 0,99797$. D.h. es ergibt sich eine Wahrscheinlichkeit von 99,797%, dass bei einer Geschwindigkeit von 1000 ein Bruch erfolgt.

Für eine Geschwindigkeit von 325 (s.o.) folgt:

Tabelle der Überlebenswahrscheinlichkeiten

		95,0%-Fiduzial-KI	
Stress	Wahrscheinlichkeit	Untergrenze	Obergrenze
325	0,999502	0,992822	1,00000

Figure 185: Probitanalyse Ergebnis Regressionstabelle

Anm.: in der Rechnung oben ist die Ausfallwahrscheinlichkeit angegeben, in der Probitschätzung von Minitab die Überlebenswahrscheinlichkeit.

14.9.1 Probit-Analyse als logistische Regression

Diese Datensituation kann in Minitab auch mit der logistischen Regression auf die folgende Weise gerechnet werden:

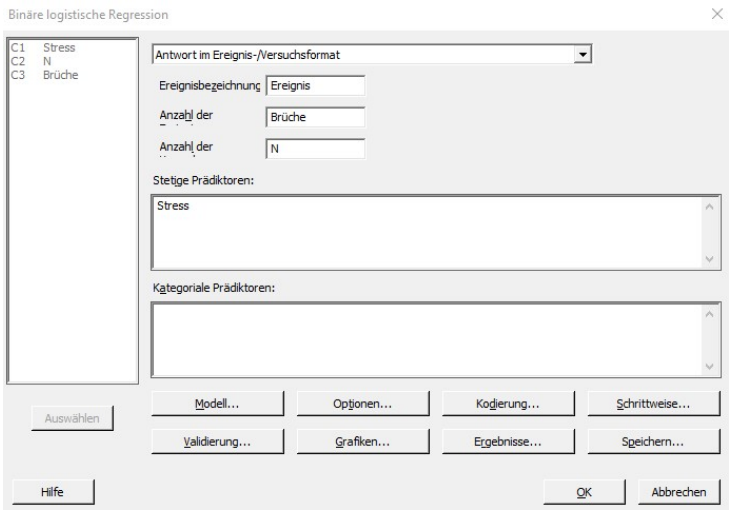


Figure 186: Probitanalyse als logistische Regression im Dialog

In den Optionen ist zudem noch die Normalverteilung zu wählen (im Standard wird die logistische Regression verwendet), das Ergebnis ist analog dem zur Probit-Analyse:

Regressionsgleichung

p(Ereignis) = Φ(Y')

Y' = -6,20 + 0,00896 Stress

Φ = kumulative Verteilungsfunktion der Standardnormalverteilung

Koeffizienten

Term	Koef	SE Koef	z-Wert	p-Wert	VIF
Konstante	-6,20	1,07	-5,82	0,000	
Stress	0,00896	0,00156	5,74	0,000	1,00

Figure 187: Ergebnis der logistischen Regression

15 Transformationen

Unter Transformation wird hier verstanden, dass versucht wird, Daten, die keiner der zur Verfügung stehenden Verteilungsfunktionen genügend gut folgen, in eine der Standard-Normalverteilung folgenden Datenset zu transformieren. Die bekanntesten Transformationen sind die Box-Cox-Transformation

$$y_i^* = \begin{cases} \frac{(y_i+c)^\lambda - 1}{\lambda} & \text{für} \quad \lambda \neq 0 \\ \ln(y_i + c) & \text{für} \quad \lambda = 0 \end{cases} \tag{38}$$

Der Parameter λ kann durch eine Optimierung berechnet werden.

Einen Schritt weiter geht die Johnson-Transformation:

15.0.1 Johnson-Transformation

Die Johnson-Transformation verwendet folgende Funktionen (Funktionsfamilie):

Johnson-Familie	Transformationsfunktion	Spannweite
S_B	$\gamma + \eta \ln [(x - \varepsilon) / (\lambda + \varepsilon - x)]$	$\eta, \lambda > 0, -\infty < \gamma < \infty, -\infty < \varepsilon < \infty, \varepsilon < x < \varepsilon + \lambda$
S_L	$\gamma + \eta \ln (x - \varepsilon)$	$\eta > 0, -\infty < \gamma < \infty, -\infty < \varepsilon < \infty, \varepsilon < x$
S_U	$\gamma + \eta \operatorname{Sinh}^{-1} [(x - \varepsilon) / \lambda]$, wobei $\operatorname{Sinh}^{-1}(x) = \ln [x + \operatorname{sqrt} (1 + x^2)]$	$\eta, \lambda > 0, -\infty < \gamma < \infty, -\infty < \varepsilon < \infty, -\infty < x < \infty$

Figure 188: Funktionsfamilie der Johnson-Transformation

Der Algorithmus verläuft dann wie folgt:

- Es werden nahezu alle geeigneten Transformationsfunktionen der Johnson-Familie berücksichtigt
- Die Parameter in der Funktion werden geschätzt
- Die Daten werden unter Verwendung der Transformationsfunktion transformiert

- Für die transformierten Daten werden die Anderson-Darling-Statistik und der entsprechende p-Wert berechnet
- Es wird die Transformationsfunktion mit dem größten p-Wert ausgewählt, der größer als der im Dialogfeld Transformieren angegebene p-Wert ist (Standardwert 0,10). Andernfalls ist keine Transformation angemessen.

15.0.2 Anwendung auf Lebensdauerdaten

Als Beispiel verwenden wir die Lebensdauerdaten:

Ausfallzeit
20,791
39,980
68,733
179,802
62,226
57,644
204,216
120,015
56,860
120,420
231,849

Figure 189: Ausfalldaten für Johnson-Transformation

Die Ausgabe der Johnson-Transformation für diese Daten ist:

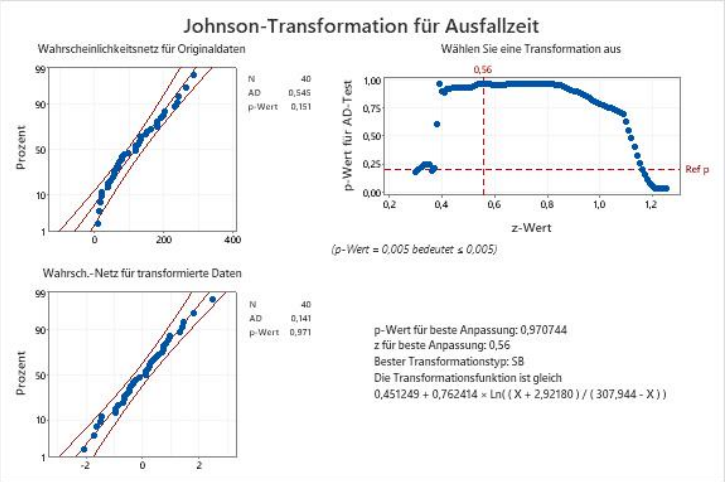


Figure 190: Ergebnis für die Johnson-Transformation

Die explizite Formel der JT lautet also:

$$X^* = 0,451249 + 0,762414 * \ln((X + 2,92180)/(307,944 - X)) \qquad (39)$$

Gibt man diese Formel in den Minitab-Rechner - minimal angepasst - ein,

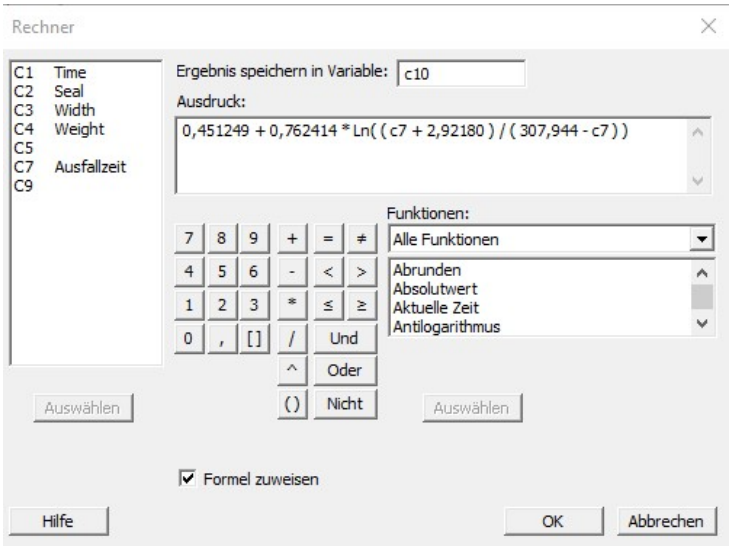


Figure 191: Rechner explizit für die Johnson-Transformation

dann wird hierdurch die Transformation bestätigt:

C9	C10
JT aus MTB	JT aus Rechner
-1,45020	-1,45020
-0,94545	-0,94545
-0,46782	-0,46782
0,72178	0,72178
-0,56088	-0,56088
-0,63056	-0,63056
0,97854	0,97854
0,12769	0,12769
-0,64289	-0,64289

Figure 192: Vgl. JT Minitab - JT Rechner

Ausgehend von den transformierten Daten wird nun die 10%-Perzentile berechnet:

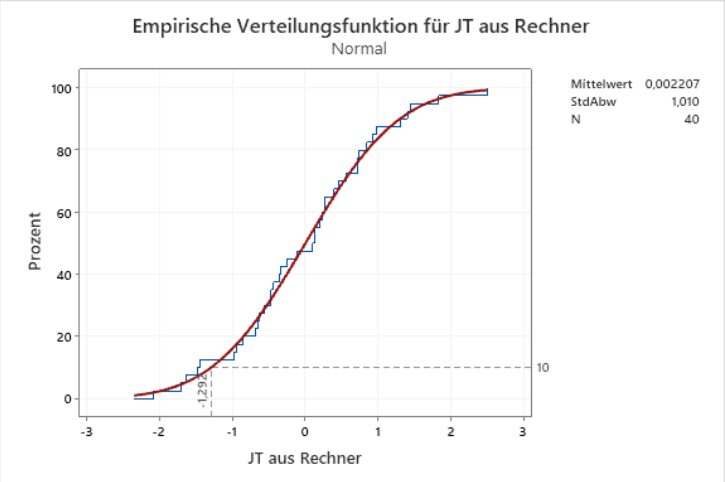


Figure 193: 10%-Perzentile der JT-transformierten Daten

Nun zur Rücktransformation: diese lautet für die hier angezeigte JT (Minitab berechnet diese Formel nicht, das ist manuell zu erledigen; jedoch hat Johnson für seine Transformationsfamilie nur solche Funktionen ausgewählt, für die die Rücktransformation auch existiert und relativ leicht durchführbar ist):

$$X = (307,944 + 2,92180)/(-1 - \exp((X^* - 0,451249)/0,762414)) + 307,944 \tag{40}$$

Im Rechner diese Formel eingetragen und bzgl. der transformierten Daten zeigt die Korrektheit:

Ausfallzeit		JT aus MTB	JT aus Rechner		Rücktrafo
20,791		-1,45020	-1,45020		20,791
39,980		-0,94545	-0,94545		39,980
68,733		-0,46782	-0,46782		68,733
179,802		0,72178	0,72178		179,802
62,226		-0,56088	-0,56088		62,226
57,644		-0,63056	-0,63056		57,644
204,216		0,97854	0,97854		204,216
120,015		0,12769	0,12769		120,015
56,860		-0,64289	-0,64289		56,860
120,420		0,13183	0,13184		120,420
231,849		1,31020	1,31021		231,849

Figure 194: Überprüfung der Rücktransformationsformel

Die 10%-Perzentile rücktransformiert ergibt:

10%-Perz	Rücktrafo10%
-1,292	25,7551

Figure 195: 10%-Perzentile der JT-transformierten Daten rücktransformiert

Der eigentliche Schätzwert von 30,8 wird nur annähernd wiedergegeben.

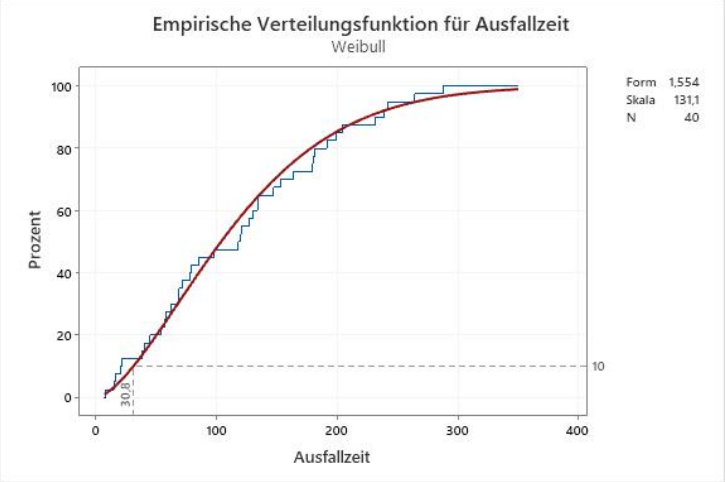


Figure 196: 10%-Perzentile der Weibull-Verteilung

Transformationen sind also mit einer gewissen Vorsicht anzuwenden. Immerhin ist es erstaunlich, dass die Johnsontransformation auch für recht unstete Daten recht gute Ergebnisse erzielt. Die Anwendung einer Transformation sollte aber eine Art Ultima Ratio sein, da die Fehlerfortschleppung durch Anpassung, Transformation und Rücktransformation recht groß sein kann.

16 Anwendung von Multivariate Analysen im QM

16.1 Hauptkomponentenanalyse

16.1.1 Einleitung

Die Hauptkomponentenanalyse wird verwendet, um in einem (sehr) großen Datensatz (z.B. ¿ 100 Variablen ...) eine (wesentlich) kleinere Anzahl nicht korrelierter Variablen zu identifizieren, die dann als sog. 'Hauptkomponenten' bezeichnet werden.

Mit der Hauptkomponentenanalyse erstellen Sie dann neue Variablen (Hauptkomponenten), bei denen es sich um lineare Kombinationen der beobachteten Variablen aus der Stichprobe handelt. Das Ziel der Hauptkomponentenanalyse besteht darin, die größtmögliche Varianz mit der kleinstmöglichen Anzahl von Hauptkomponenten zu erklären.

Die Hauptkomponentenanalyse ist oftmals selbst nicht zielführend, sondern wird als Schritt in einer Abfolge von weiteren Analysen verwendet. Sie können die Hauptkomponentenanalyse z. B. verwenden, bevor Sie eine Regressionsanalyse durchführen, um Multikollinearität zu vermeiden oder die Anzahl der Prädiktoren im Verhältnis zur Anzahl der Beobachtungen (stark) zu reduzieren. Somit ist das Verfahren eine Dimensionsreduktion basierend auf der Idee, Multikollinearität gänzlich auszuschalten.

Als Transformation in einen neue Variablenraum (in Minitab: Werte) sind Analyseergebnisse basierend auf den neuen Variablen jedoch wieder in die reale Welt zurück zu transformieren, damit diese dort genutzt werden können. Die Rücktransformation ist kein einfach zu lösendes Problem und ist in Minitab nicht ohne weiteres möglich. Für die Rücktransformation muss eine inverse Matrix erstellt werden, die Hauptkomponentenwerte sind dann mit dieser Matrix in den realen Bereich zu transformieren.

16.1.2 Das Prinzip der Methode

Bei der Hauptkomponentenanalyse werden zunächst die orthogonalen Eigenvektoren der Korrelations- oder Kovarianzmatrix der Variablen berechnet. Die Matrix der Hauptkomponenten ist dann das Produkt der Eigenvektormatrix und der Matrix der ursprünglichen (unabhängigen) Variablen.

Die erste Hauptkomponente erklärt den größten Prozentsatz der Gesamtstreuung in den Daten, die zweite den zweitgrößten Prozentsatz usw. Das Ziel der Hauptkomponentenanalyse ist, den größtmöglichen Anteil der Streuung mit einer kleinstmöglichen Anzahl von Komponenten zu erklären.

Die entstehenden Hauptkomponentenvektoren sind unabhägig, d.h. die Korrelation dieser Vektoren zueinander ist 0.

Die Koeffizienten der Hauptkomponenten sind eindeutig bis auf das Vorzeichen, sofern die Eigenwerte eindeutig und nicht null sind. Wenn ein Eigenwert eine höhere Vielflachheit besitzt, ist der durch alle Hauptkomponentenvektoren mit demselben Eigenwert aufgespannte Raum eindeutig, die einzelnen Vektoren sind es jedoch nicht. Deshalb stimmen manchmal die von der Software berechneten Werte nicht mit z.B. Beispielen aus der Literatur oder mit den Ergebnissen anderer Software überein.

Verfügt die Kovarianzmatrix über einen Rang $r < p$, wobei p die Anzahl der Variablen ist, dann sind $p–r$ Eigenwerte gleich null. Die Eigenvektoren, die diesen Eigenwerten entsprechen, sind ggf. nicht eindeutig. Dies kann geschehen, wenn die Anzahl an Beobachtungen kleiner als p ist oder eine (exakte) Multikollinearität vorliegt.

Eigenvektoren, die sich aus den Koeffizienten für die einzelnen Variablen zusammensetzen, sind die Gewichtungen für die einzelnen Variablen zur Berechnung der Werte (sog. Scores) der Hauptkomponenten. Die Eigenvektoren werden als Spalten der orthogonalen Matrix in der Spektralzerlegung der Kovarianz- oder Korrelationsmatrix S bzw. R berechnet. Da R symmetrisch ist, existiert dann eine orthogonale Matrix V, so dass $V^T * R * V = D$ ist, wobei D eine Diagonalmatrix ist. Analog gilt dann auch $R = VDV^T$. Die Diagonalelemente von D sind dann die Eigenwerte. Die Eigenvektoren wiederum sind die Spalten von V.

Durch die Diagonalisierung werden die Varianzen durch die Eigenwerte, also die Diagonalelemente von D ausgedrückt, die Gesamtstreuung ist dann die Summe der Eigenwerte. Der erklärende Varianzanteil einer Variablen wird dann relativ durch

$$\frac{\lambda_k}{\sum_{i=1}^{N} \lambda_i} \tag{41}$$

dargestellt, wobei N die Anzahl der Variablen ist. So wird die Varianzsumme der ersten n Eigenwerte zu:

$$\frac{\sum_{j=1}^{n} \lambda_j}{\sum_{i=1}^{N} \lambda_i} \tag{42}$$

Als sinnvolle Anzahl auszuwählender Eigenvektoren sei das Kaiser-Kriterium erwähnt, das besagt: wähle die Anzahl Eigenvektoren als Basis für die Komplexitätsreduktion mittels Hauptkomponentenanalyse, für die gitl:

$$\frac{\lambda_k}{\sum_{i=1}^{N} \lambda_i} > 1 \tag{43}$$

Zudem wird für die Variablen die sog. Mahalanobis-Distanz berechnet. Diese ist ein Maß für den Abstand zwischen einem Punkt im multivariaten Raum und dem Gesamtmittelwert (dem Zentroiden) des Raumes. Die Formel für dieses Maß lautet:

$$Y_i = \sqrt{(Y_i - \bar{Y})^T S^{-1} (Y_i - \bar{Y})} \tag{44}$$

Eine Ausreißererkennung kann anhand der folgenden Formel durchgeführt werden:

$$\sqrt{p * F^{-1}(0.95; p, n-p-1)} \tag{45}$$

wobei p = Anzahl der Variablen und n = Anzahl Zeilen ist.

16.1.3 Beispiel

Hier die Ergebnisse einer Hauptkomponentenanalyse:

Eigenwertanalyse der Korrelationsmatrix

Eigenwert	3,2952	2,1433	0,6565	0,4256	0,2965	0,1326	0,0504
Anteil	0,471	0,306	0,094	0,061	0,042	0,019	0,007
Kumulativ	0,471	0,777	0,871	0,932	0,974	0,993	1,000

Eigenvektoren

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Ausbildung	0,190	0,430	-0,833	-0,254	-0,042	-0,138	0,022
Alter	0,499	-0,111	0,034	-0,195	-0,527	0,642	-0,097
Ansässig	0,496	-0,241	0,027	0,121	-0,064	-0,538	-0,623
Anstellung	0,490	-0,277	0,046	-0,018	0,010	-0,312	0,764
Ersparnisse	0,439	0,257	0,005	0,439	0,646	0,360	-0,041
Schulden	-0,038	-0,596	-0,213	-0,531	0,517	0,182	-0,123
Kreditkarten	-0,187	-0,498	-0,507	0,638	-0,179	0,139	0,044

Figure 197: Ergebnistabellen der Hauptkomponentenanalyse

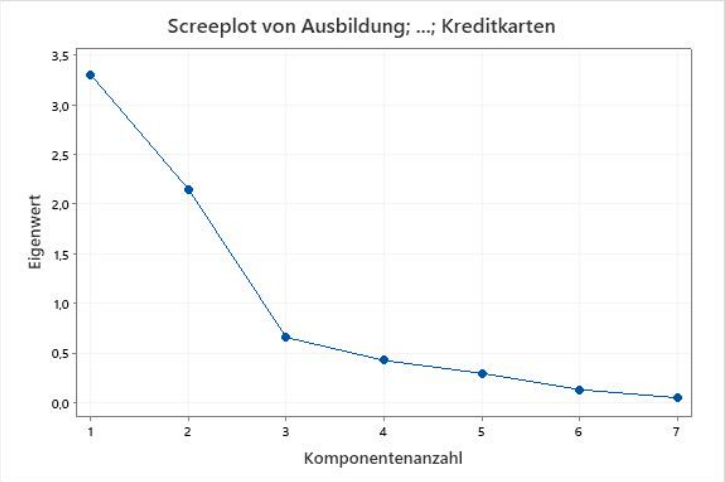


Figure 198: Screeplot zur Hauptkomponentenanalyse

Anhand des Kaiser-Kriteriums würde man sich lediglich nur für 2 Komponenten entscheiden. Jedoch werden mit einer Komponente mehr als 80% der streuung erklärt. Somit entscheidet man sich in diesem Fall für 3 Komponenten.

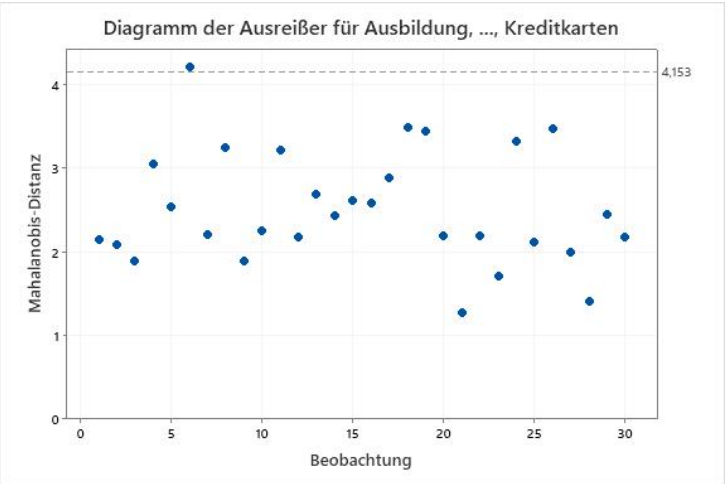


Figure 199: Ausreißer-Diagramm zur Hauptkomponentenanalyse

Lediglich ein Datenwert wird als Ausreißer identifiziert, weshalb die Datenbasis als genügend für die HKA gilt.

Das sog. Ladungsdiagramm zeigt nun auf, welche Variablen sehr ähnlich sind (hier: Ausbildung, Alter, Anstellung):

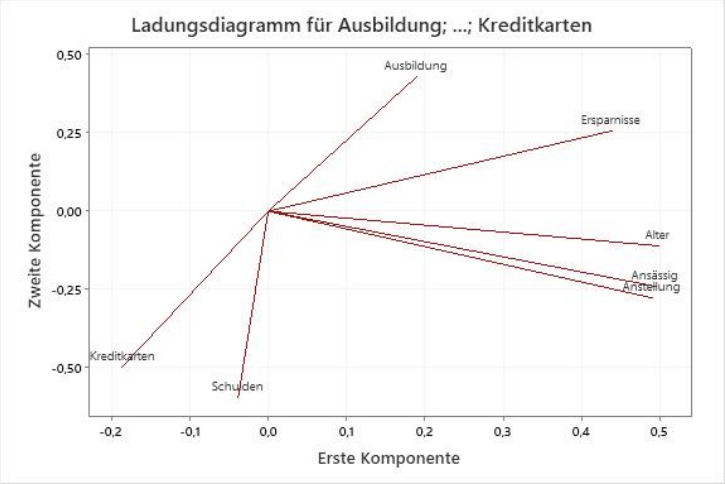


Figure 200: Ladungs-Diagramm zur Hauptkomponentenanalyse

Die Ähnlichkeit der Variablen ist auch gut in der Tabelle der Eigenvektoren zu erkennen (ähnliche Komponentenwerte).

Für die Situation hier werden nun 3 Hauptkomponenten berechnet, diese werden dann in Minitab im Arbeitsblatt abgespeichert:

C11	C12	C13
-1,74850	1,68560	0,62598
1,11841	-0,25555	-1,16531
0,95212	1,50208	-0,40376
-0,75594	1,23308	-1,48616
2,71611	1,35283	-0,42585
2,56989	-4,42063	-0,57402
-1,38690	-1,64319	-0,58010
0,06138	-0,42683	2,03816
1,44380	0,52140	-0,32162
-0,20564	1,80580	0,05406
-1,31435	0,59292	-0,32584

Figure 201: Werte der 3 Hauptkomponenten

Die Matrix der Koeffizienten, die zur Berechnung der Hauptkomponenten verwendet wird, ist:

C15	C16	C17
0,189679	0,429590	-0,832869
0,498624	-0,110622	0,033711
0,495844	-0,240546	0,027458
0,490220	-0,276862	0,045701
0,439173	0,256937	0,005311
-0,038396	-0,595724	-0,213497
-0,186740	-0,497791	-0,506697

Figure 202: Koeffizienten zu den 3 Hauptkomponenten

Anm.: um die Werte der Hauptkomponenten zu berechnen werde die Ausgangswerte der Variablen zunächst standardisiert.

Weitere Analysen können nun anhand der Matrix der Werte der Hauptkomponenten durchgeführt werden. So kann zunächst gezeigt werden, dass die hier errechneten Hauptkomponenten unabhängig sind:

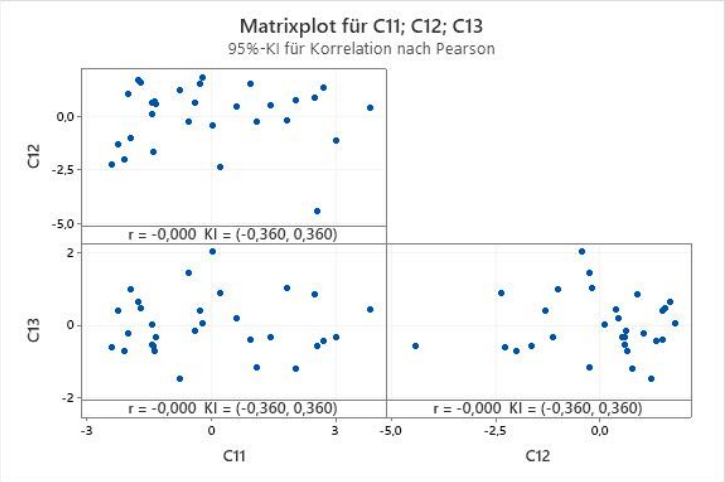


Figure 203: Unabhängigkeit der Hauptkomponenten

Ist eine Zielgrößenoptimierung gewünscht, dann muss das Ergebnis der Optimierung mittels der Inversen der Koeffizientenmatrix und De-Standardisierung in die reale Welt zurück geführt werden.

16.2 Clusteranalyse

16.2.1 Clusterbeobachtungen

Die Analyse der Clusterbeobachtungen gruppiert Objekte anhand gleicher Merkmale, sofern die Merkmale insg. alle ähnlich sind. Die Ähnlichkeit wird gemessen und für die Clusterzuordnung verwendet. Diese Analyse ist geeignet, wenn anfänglich keinerlei Informationen dazu vorliegen, wie die Gruppen gebildet werden sollen.

Bei Clusterbeobachtungen werden die Gruppen anhand eines hierarchischen Verfahrens gebildet. Bei jedem Schritt werden zwei Gruppen (Cluster) zusammengefasst, bis im letzten Schritt nur noch eine Gruppe alle Beobachtungen enthält. Minitab berechnet bei jedem Schritt des Clusterprozesses die Ähnlichkeits- und Distanzwerte für die Gruppen. Ein Dendrogramm (Baumdiagramm) stellt anschaulich dar, wie ähnlich sich die einzelnen Objekte sind.

Da anfangs noch keine sinnvolle Zahl an Clustern bekannt ist, wird analog der Hauptkomponentenanalyse in zwei Schritten vorgegangen. Zunächst wird ein einziges Dendrogramm über alle Elemente angezeigt, das dann anhand der Ähnlichkeit aufzeigt, wie viele Cluster sinnvollerweise zu bilden sind. Als Faustregel zieht man ungefähr mittig im Dendrogramm eine horizontale Linie und zählt die Anzahl der Kreuzungspunkte des Ähnlichkeitsbaumes mit dieser Linie. Ein weiterers Kriterium für eine Clusteranzahlfindung ist, wenn das Dendrogramm große Sprünge macht. Das deutet an, dass dann stärker unterschiedliche Gruppen in den Daten vorhanden sind.

Für die Clusterbildung, also insb. die Ähnlichkeitsberechnung, sind insg. drei verschiedene Distanzmessungen notwendig:

- die Ähnlichkeit der Cluster, gegeben durch die Formel

$$S_{ij} = 100 * \frac{1 - d_{ij}}{d_{max}} \tag{46}$$

  wobei S_{ij} die Ähnlichkeit zwischen den Clustern i und j, d_{ij} die Distanz zwischen diesne Clustern und d_{max} der maximale Wert in der Distanzmatrix ist.

- Ein Distanzmaße für die Clusterbeobachtungen ist z.B. der Euklidische Abstand,

$$d_{ik} = \sqrt{\sum_j (x_{ij} - x_{kj})^2} \tag{47}$$

- das Aggleromative Verfahren. Bei z.B. der Methode 'vollständig' (auch als 'entferntester Nachbar' bezeichnet) ist die Distanz zwischen zwei Clustern die maximale Distanz zwischen einer Beobachtung in einem Cluster und einer Beobachtung in dem anderen Cluster. Die vollständige Distanz wird anhand der folgenden Distanzmatrix berechnet:

$$d_{mj} = \max(d_{kj}, d_{lj}) \tag{48}$$

Anm.: das Aggleromative Verfahren nach Ward liefert i.d.R. die am gleichmässig besetzten Cluster. Jedoch kann bei diesem Verfahren die Distanz zwischen zwei

Clustern größer sein als der ursprüngliche Wert in der Distanzmatrix, was zu negativen Ähnlichkeiten führt.

Für die Daten der Hauptkomponentenanalyse ist das erste Dendrogramm mit der Euklidischen Distanz als Distanzmaß und der Methode von Ward als Aggleromatives Verfahren, zudem werden die Variablen standardisiert, wie folgt:

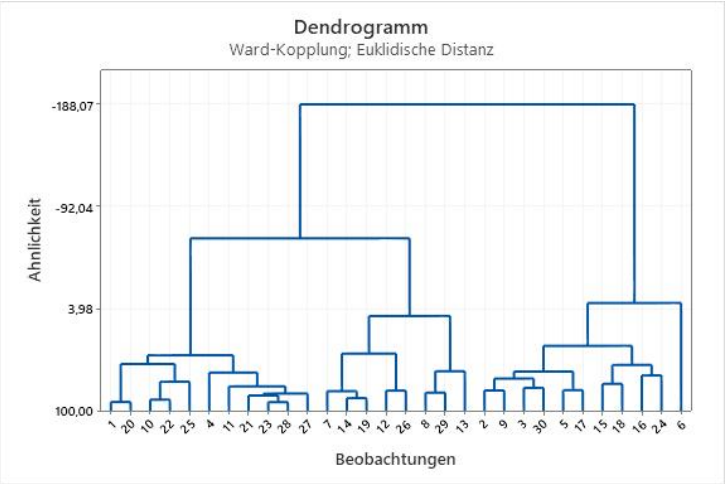


Figure 204: Dendrogramm für einen Cluster

Nach obigem Kriterium wären drei Cluster eine sinnvolle Wahl, um die Datensätze zu gruppieren.

Die Zuordnung der Datensätze zu den jeweiligen Clustern kann dann entsprechend im Arbeitsblatt abgespeichert werden (wobei sich zudem die Grafik ändert, indem die Zuordnung auch farblich gekennzeichnet wird):

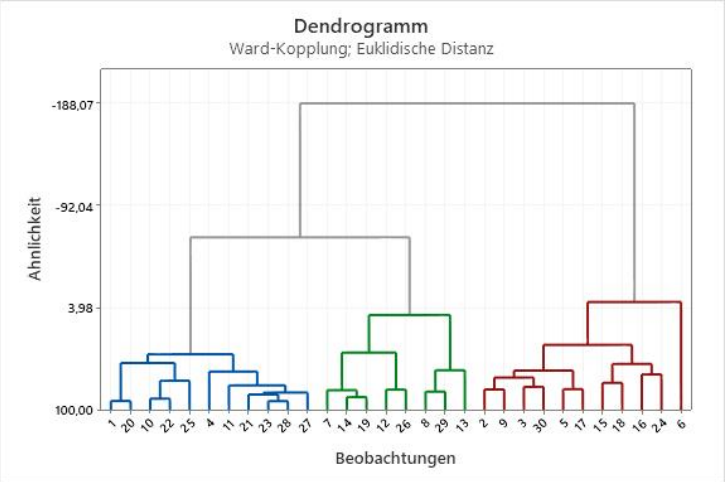


Figure 205: Dendrogramm für die finale Clusterzuordnung

C50
1
2
2
1
2
2

16.2.2 Clusterzentrenanalyse (k-means)

16.3 Diskriminanzanalyse

16.3.1 Ein Diskriminanzmodell für Clustervariablen

Liegen für Beobachtungen Gruppenzuordnungen vor (z.B. aus der Clusteranalyse), dann kann die Diskriminanzanalyse ein Modell erstellen, das diese Gruppenzugehörigkeit anhand der Beobachtungswerte erstellt.

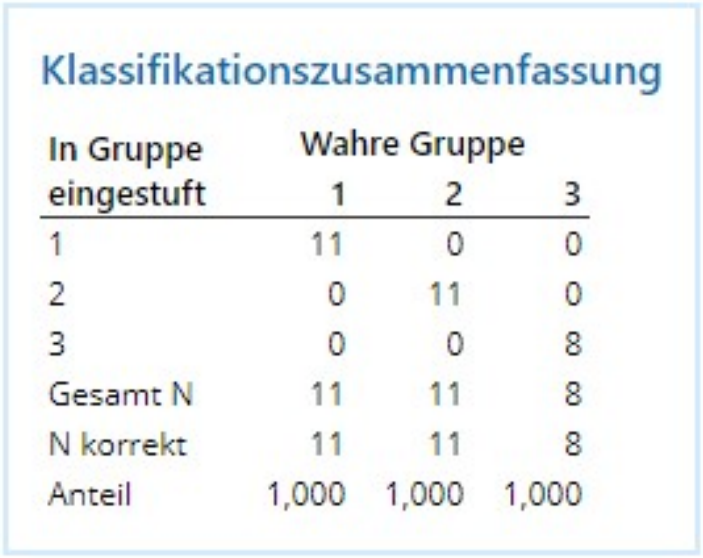
Speziell in Bezug zur Clusteranalyse kann man so mit Hilfe der Diskriminanzanalyse ein Vorhersagemodell neu hinzugekommener Objekte zu den - bekannten Clustern erstellen.

Diese Analyse bietet also folgende Möglichkeiten:

- Feststellen, wie genau die Beobachtungen in den bekannten Gruppen klassifiziert sind
- Auswerten, wie die Prädiktorvariablen die Gruppen differenzieren
- Prognostizieren der Gruppen für Beobachtungen mit bislang unbekannter Gruppenzuordnung

Als Bewertung der Analyse dient insb. die Messung der Güte der Zuordnung, die über die Anteile der korrekt und der nicht korrekt zugeordneten Beobachtungen berrechnet wird (als Anteil, in Prozent).

Für die Daten der Clusteranalyse liefert die Diskriminanzanalyse folegndes Ergebnis:



Klassifikationszusammenfassung

In Gruppe eingestuft	Wahre Gruppe 1	2	3
1	11	0	0
2	0	11	0
3	0	0	8
Gesamt N	11	11	8
N korrekt	11	11	8
Anteil	1,000	1,000	1,000

Figure 207: Ergebnis der Diskriminanzanalyse - Güte

Wie heir zu sehen ist die Zuordnung der Cluster für die Beobachtungen mit Hilfe der Diskriminanz perfekt.

Als Test für die Stabilität dieses Ergebnisses kann eine Kreuzvalidierung verwendet werden. Dadurch ist die Zurodnung nicht mehr perfekt, sondern hat eine Güte von 86%.

Klassifikationszusammenfassung mit Kreuzvalidierung

In Gruppe eingestuft	Wahre Gruppe 1	2	3
1	9	1	0
2	1	9	0
3	1	1	8
Gesamt N	11	11	8
N korrekt	9	9	8
Anteil	0,818	0,818	1,000

Richtige Klassifikationen mit Kreuzvalidierung

N	Richtig	Anteil
30	26	0,867

Figure 208: Ergebnis der Diskriminanzanalyse - Güte nach Kreuzvalidierung

17 Data Mining-Methoden

17.1 Entscheidungsbäume: CART

Die CART-Klassifizierung erstellt einen sog. Entscheidungsbaum für eine binomiale oder multinomiale kategoriale Antwortvariable in Abhängigkeit von ggf. vielen kategorialen oder stetigen Prädiktoren. Die CART-Klassifizierung veranschaulicht dabei Muster und Beziehungen zwischen einer kategorialen Antwortvariablen und wichtigen Prädiktoren innerhalb hochkomplizierter Daten, ohne parametrische Methoden zu verwenden.

Ein Entscheidungsbaum teilt die Daten anhand Vergleichskriterien auf, um so eine bestmögliche Teilgruppierung mit möglichst nur reinen Zuordnungen zu erzeugen, siehe das nachfolgende Bild:

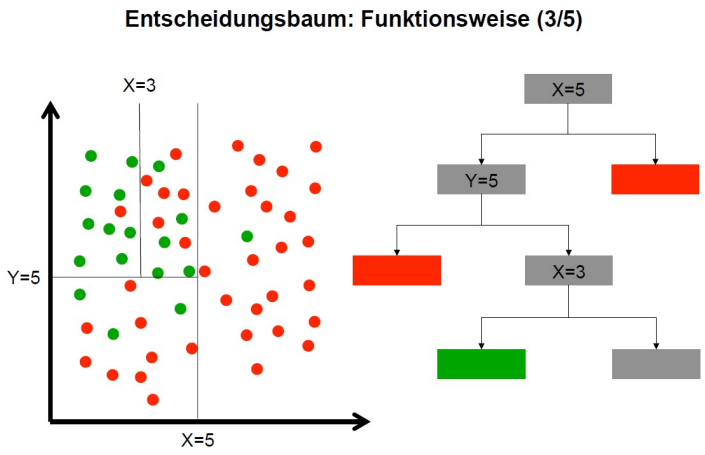


Figure 209: CART - Entscheidungsbaum Beispiel

Um die sog. Split-Punkte zu finden gibt es mehrere Möglichkeiten:

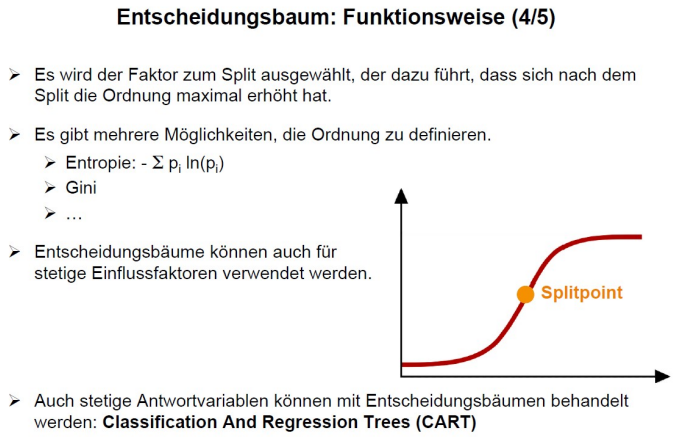


Figure 210: CART - Split-Punktfindung

Für die Daten hier im Fallbeispiel mit Zielvariable 'Ausfall' folgt für die CART-Auswertung ohne Kreuzvalidierung und nach der Split-Methode 'Entropie' das folgende Ergebnis:

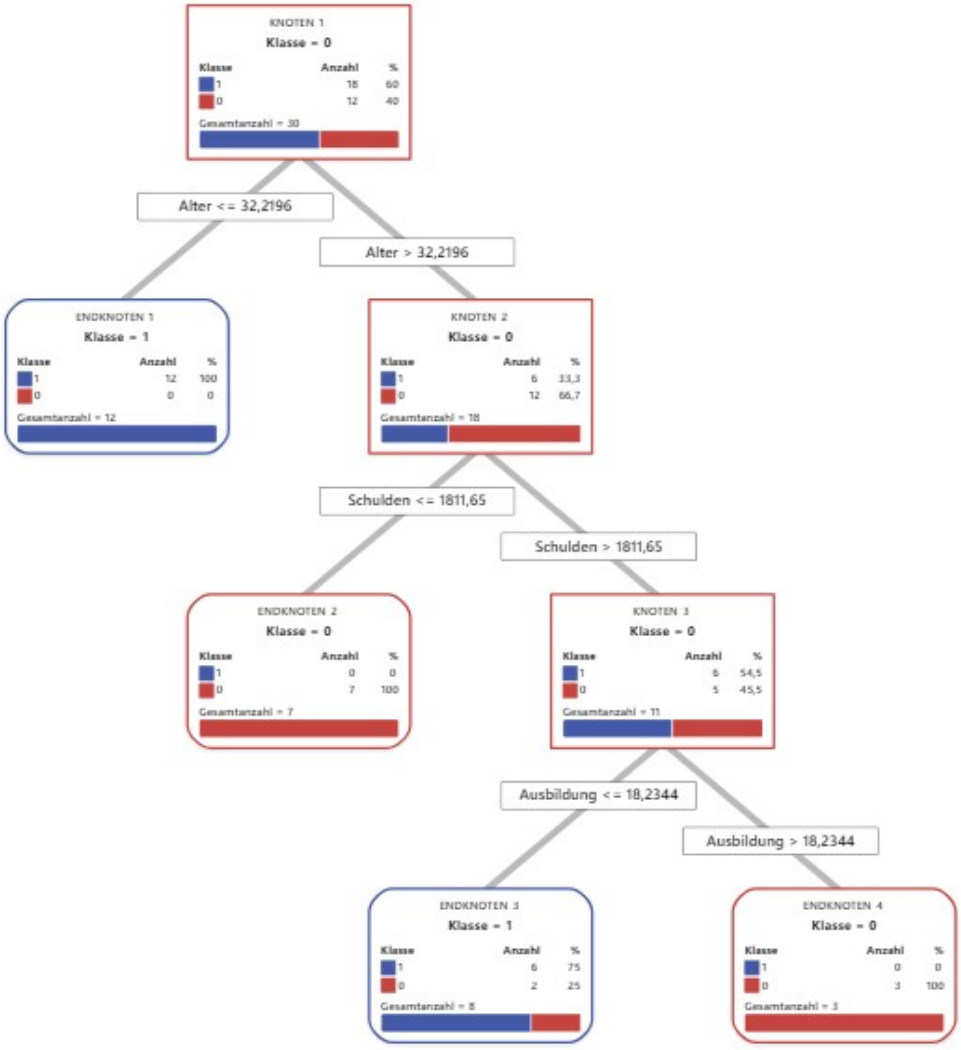


Figure 211: CART - Entscheidungsbaum Ergebnis

Die Korrektheit der Zuordnung in den einzelnen Endknoten des Baumes zeigt an, wie gut die Ordnungsbildung in den Daten durch die Klassifikation geglückt ist.

17.2 Bayes'scher Schätzer

17.3 Treenets

17.4 MARS

17.5 GPS/Generalized Lasso

17.6 Random Forrests

18 Techniken von Industrie 4.0

18.1 Big Data-Datenbanken und -abfragen

18.1.1 Hadoop

18.1.2 Spark

18.2 Cloudservices und -analytics

19 Anhang

19.1 Hypothesentest

Ein Hypothesentest soll eine (verbale) Aussage auf 'Richtigkeit' überprüfen. 'Richtig' ist im statistischen Sinn als 'das Gegenteil kann nicht bewiesen werden' zu interpretieren. So ist z.B. bei stetigen Messwerten nicht zu erwarten, dass das zugehörige Merkmal exakt genau einen bestimmten Wert annimmt. Akzeptiert wird, wenn die Abweichungen nicht zu groß sind ...

Die hier betrachtete Problematik wäre die eines (1-Stichproben-) t-Tests: man misst ein stetiges Merkmal und hinterfragt, ob die Messwerte 'gleich' einem (vorgegebenen oder auch kritischen) Wert sind.

Der Hypothesentest ist prinzipiell wie folgt aufgebaut:

1. Die Aussagenlogik mit Hilfe der Null-Hypothese H_0 und der Alternativ-Hypothese H_1 wird angegeben. Dabei ist die Nullhypothese die Eigenschaft die der Übereinstimmung, also H_0: die Werte sind 'gleich'.

2. Es wird die Prüfgröße angegeben und berechnet. Im Fall des t-Tests hier ist das der folgende Quotient:

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}$$

3. Für diese Prüfgröße wird nun das Quantil einer Test-Verteilung, in diesem Fall der t-Verteilung, abgelesen. Dieser Wert gibt dann die Irrtumswahrscheinlichkeit α dafür an, dass wenn in der Realität H_0 zutrifft, man aber H_1 annehmen würde.

4. man trifft nun seine Entscheidung mit Hilfe der akzeptierten Irrtumswahrscheinlichkeit, i.d.R. 5%. Ist die Irrtumswahrscheinlichkeit, also α, kleiner als z.B. 5%, dann lehnt man H_1 ab, nimmt also H_0 an.

Im Prinzip ist jeder Hypothesentest auf diese Art aufgebaut. Will man also z.B. selbst einen Hypothesentest konstruieren, dann muss man neben der Aussagenlogik des Tests insb. die Testverteilung angeben (definieren oder ermitteln).

Aus der Aussagenlogik 'bei genügend kleinem Irrtumsfehler α lehne ich H_1 ab' wird durch logische Umkehrung dann auch 'die Richtigkeit von H_0, $1 - \alpha$, ist hoch, darum nehme ich H_0 an'. α kann also etwas ungenau, aber logisch durchaus akzeptabel, auch als 'Fehler des Ablehnens von H_0' interpretiert werden - was in mancher Literaturstelle so verwendet wird.

Für die Fehler, die man bei einem Hypothesentest machen kann, gibt es die folgende Darstellung:

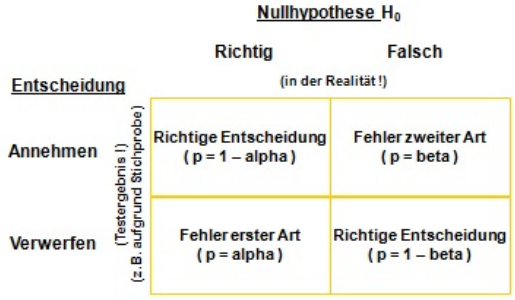


Figure 212: Hypothesentest: Richtigkeit und Fehler

Für die Zeile 'Verwerfen' sind die Begriffe der Signifikanz (Fehler 1. Art) und der Trennschärfe (richtige Entscheidung) definiert.

Die Trennschärfe ist die Umkehrung der obigen Überlegung zur richtigen Ablehnung von H_1 (also der korrekten Annahme von H_0). Die Trennschärfe bzgl. des t-Test-Beispiels kann so interpretiert werden: man stellt mit der Wahrscheinlichkeit der Trennschärfe $(1-\beta)$ einen Unterschied zum vorgegebenen Wert auch tatsächlich fest, wenn dieser in der Realität tatsächlich auch gegeben ist.

Hier noch ein paar Anmerkungen zu Hypothesentests:

Für das richtige Aufstellen der Hypothesen gibt es folgende paar Regeln:

- was ich zeigen will, gehört in die Alternativhypothese.
- Hypothesenaufstellung der Art, dass die irrtümliche Ablehnung der Nullhypothese die schlimmeren Auswirkungen hätte.
- das Gleichheitszeichen gehört in die Nullhypothese.
- für Verteilungstests: dass die Verteilung der Stichprobenverteilung einer bestimmten Verteilung folgt, lässt sich nicht beweisen. Wohl aber lässt sich (statistisch) beweisen, wenn die Stichprobenverteilung einer theoretischen Verteilung nicht folgt, also die Abweichungen von den theoretisch zu erwartenden Werten zu groß sind.

Beim Testen wird nur die Wahrscheinlichkeit α (die Wahrscheinlichkeit der irrtümlichen Ablehnung der Nullhypothese) kontrolliert. Wenn H_0 tatsächlich wahr ist, wird

nur in $\alpha \cdot 100\%$ der Fälle für H1 entschieden. Die Entscheidung für H_1 ist in diesem Sinn statistisch abgesichert.

Bei einer Entscheidung für die Alternativ- und gegen die Nullhypothese spricht man dann von einem signifikanten Ergebnis.

Die Wahrscheinlichkeit β für den Fehler zweiter Art wird bei diesem Vorgehen nicht kontrolliert, daher ist eine Entscheidung für H_0 nicht statistisch abgesichert.

Wenn die Nullhypothese H_0 nicht verworfen wird, bedeutet das nicht, sich aktiv für H_0 zu entscheiden. Es spricht nur nichts gegen die Nullhypothese.

Beispiel: Ein Reifenfabrikant fertigt Fahrradreifen und hat ein neues Modell entwickelt. Die Frage ist nun, ob es besser als das bisherige Modell ist, das eine mittlere Haltbarkeit von 8000 Km aufwies.

Die Laufleistung eines Reifens sei normalverteilt, die Standardabweichung sei bekannt mit 1800 Km.

Nach Angaben der Forschungsabteilung bringt das neue Modell eine Leistung von 10000 Km. Bevor aber mit der teuren Werbekampagne begonnen wird, soll intern mit einem statistischen Test mit der Stichprobengröße $n = 36$ und dem Signifikanzniveau $\alpha = 5\%$ geprüft werden, ob der neue Reifen besser ist als der alte.

Welche Fehler kann die Firma in unserem Fall machen? Bei dem Test könnten die in Wahrheit besseren neuen Reifen als nicht besser als die jetzigen ausgezeichnet werden. Falls die neuen Reifen hingegen nicht besser sind, aber als besser klassifiziert werden, müsste die Firma einen Imageverlust hinnehmen.

Wenn man diese beiden Fälle miteinander vergleicht, wäre der Schaden im zweiten Fall vermutlich größer. Als Konsequenz möchte die Firma die Wahrscheinlichkeit α der irrtümlichen Behauptung 'die neuen Reifen sind besser' nach oben begrenzen.

Folgende Darstellung stellt die Überlegung zur Konstruktion von Hypothesentests dar:

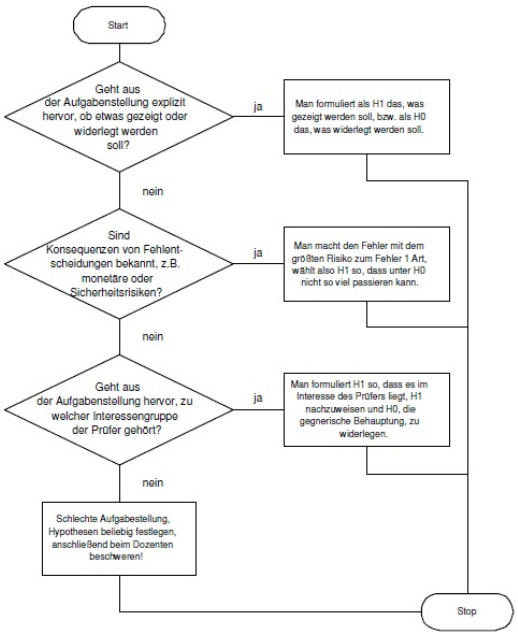


Figure 213: Hypothesentest: Wahl der Fehlerwahrscheinlichkeiten

19.2 Die Operationscharakteristik

Die Operationscharakteristik der Binomialverteilung für die Überdeckung $[n_u, n_o]$ ist:

$$OC(p) = \sum_{k=n_u}^{n_o} \binom{n}{k} p^k (1-p)^{n-k}$$

Für den Fall 'lehne das Los ab, wenn die Fehleranzahl größer ist als c' ist die Operationscharakteristik folgendermaßen:

$$OC(p) = \sum_{k=0}^{c} \binom{n}{k} p^k (1-p)^{n-k}$$

Die Funktion $g(p) := 1 - OC(p)$ heißt Gütefunktion.

19.3 Qualität der Produktion von Einzelstücken oder im Prozess

Bei der Produktion von Einzelstücken, also z.B. bei Häusern und auch bei Software (da soll Quellcode ja nicht wieder und wieder, sondern nur ein Mal geschrieben werden; Reuse of Code ist ein Ziel der sog. Objektorientierung), ist Qualität nur bedingt wiederholt messbar. Ziel ist es zunächst, das Gemeinsame aller Einzelstücke zu identifizieren und dieses dann ggf. der Statistik zuzuführen. So sind z.B. beim Häuserbau, der im Wesentlichen immer gleich verläuft (Aushub, Fundamentierung, Außenwänden und Zwischendecken, usf.), die Projektplanung und (sehr wichtig) die Termintreue gewisse Merkmale, die für einen Baukonzern auch statistisch analysiert werden können.

Für die restlichen individuellen Objekte sind dann Einzelprüfungen durchzuführen.

Bei sehr komplexen Objekten wie Softwarepaketen liegt die Problematik dann eher auf der Strategie des Testens. Die Funktionalitäten der Software sind kombinatorisch und führen zu einer Aufwandsexplosion. Für die Tests von Software sind deshalb verschiedenste Ansätze von Vorgehensweisen definiert worden:

- White Box-Test (Über- oder Abdeckungstest)
- Black Box-Test (Äquivalenzklassenbildung, Randwertanalyse)
- Gray Box-Test (Mischung aus Black- und White Box Test)
- Regressionstest (Überprüfung nach einer Änderung)

Für komplexe zu testende Systeme, die man aufgrund der (kombinatorischen) Komplexität nicht annähernd vollständig testen kann (Auto, Softwarepakete), gibt es eine Methode, die Anzahl der Fehler im System zu schätzen, die Capture-Recapture-Methode (Rückfangmethode):

Das Verfahren ist in der Biologie entwickelt worden (Petersen, dänischer Biostatistiker, ca. 1896) mit dem Ziel, z.B. in einem Teich oder einem weitläufigen Areal die Überlebenswahrscheinlichkeit oder andere Parameter innerhalb dieses Verbreitungsgebietes der Spezies zu schätzen.

Die klassischen Verfahren setzen voraus, dass der Umfang der Grundgesamtheit bekannt ist. Das ist hier nicht der Fall.

Anmerkung:

- bei z.B. Software oder einem Auto ist es sehr schwer, die überhaupt sinnvoll angeben zu wollen. Quellcode-Zeilen, die Anzahl der Komponenten oder die Anzahl der Funktionalitäten sind ungenügende Masse, die völlig unterschiedliche zu gewichtende (aufgrund ihrer Bedeutung) Objekte miteinander vermischt.

Die Rückfangmethode sei nun an dem Biologie-Beispiel erläutert.

1. es werden n_1 Tiere gefangen, markiert, und wieder ausgesetzt
2. nach einem gewissen zeitlichen Abstand (dieser ist plausibel zu wählen, Mischungen der markierten und der unmarkierten Tiere müssen stattfinden können, usw.) wird dann eine zweite Stichprobe vom Umfang n_2 gezogen. Die Anzahl der Tiere mit Markierung sei nun m_2
3. der klassische Schätzer für den Umfang der Grundgesamtheit ist dann:

$$\hat{N} = \frac{n_1 \cdot n_2}{m_2}$$

Die Schätzung basiert auf einem Ziehen ohne Zurücklegen, also der Hypergeometrischen Verteilung $H(N, n_1, n_2)$. Die Wahrscheinlichkeit unter den n_2 gefangenen Tieren gerade m_2 markierte zu fangen, hängt von N ab. Die Maximierung dieser Wahrscheinlichkeit im Sinne des Vorgehens einer Maximum-Likelihood-Methode maximiert diese Wahrscheinlichkeit.

19.4 Schätzung der Gesamt-Ausschussquote bei Kentnis der Teil-Ausschussquoten

Im Abschnitt über die Qualitätsregelkarten (Abschnitt 7) oder der Qualität von Prozessen (Abschnitt 6) sind bei der Betrachtung einzelner metrischer Merkmale

sehr hohe Anforderungen an die Prozesse gestellt worden. Das hat sich insb. durch die Anforderungen an die Streuung im Vergleich mit dem definierten Toleranzintervall ergeben. Ein Prozess ist in etwa fähig, wenn die Ausschussquote bei 0,1% liegt.

Warum das so ist, zeigt folgende Überlegung:

Ein Produkt hat sehr viele (typischerweise mehr als 50) Merkmale; alle diese Merkmale können fehlerhaft sein; ist das der Fall, dann ist das gesamte Produkt defekt.

Setzt man also für jedes Merkmal i eine Fehlerwahrscheinlichkeit von $p_{i_D} = 0.1\%$ an, dann ist die Wahrscheinlichkeit, dass das Merkmal OK ist $= p_{i_{ND}} = 99.99\%$.

Sind die Ereignisse eines Fehlers in einem Merkmal des Produktes unabhängig, dann ist die Wahrscheinlichkeit dafür, dass das gesamte Produkt nicht defekt (P_{ND}) ist:

$$P_{ND} = \prod_{i=1}^{50} p_{i_{ND}} = (0.9999)^{50} = 0,995012$$

Anmerkungen:

- in der Automobilindustrie wird typischerweise eine Quote von max. 300 Defekten auf 1 Mio. Teile, also $P_D = 0.0003$, gefordert; dann ist $P_{ND} = 0.9997 = 99.97\%$. Dieses Maß wird durch den Ansatz der Fehlerquote für jedes (unabhängige) Merkmal von 0.1% fast erreicht.
- die Annahme der Unabhängigkeit wirkt hier stark vereinfachend. In der Praxis muss dieser Aspekt gesondert und präziser untersucht werden (Fehlerballungen über Merkmalsgruppen oder Fehlerhäufungen eines Merkmals abhängig von der Ausprägung eines anderen sind eben in der Praxis oft festzustellen)

19.5 χ^2-Test

Um den χ^2-Test zu demonstrieren, betrachte man die Abschlusstabelle der ersten Fußballbundesliga der Saison 2009/2010:

Platz	Verein	Anzahl Spiele	S	U	N	Tore	Tordiff	Punkte	Punkt Plus	Punkte Minus
1	Bayern München	34	20	10	4	72:31	41	70	90	18
2	FC Schalke 04	34	19	8	7	53:31	22	65	84	22
3	Werder Bremen	34	17	10	7	71:40	31	61	78	24
4	Bayer Leverkusen	34	15	14	5	65:38	27	59	74	24
5	Borussia Dortmund	34	16	9	9	54:42	12	57	73	27
6	VfB Stuttgart	34	15	10	9	51:41	10	55	70	28
7	Hamburger SV	34	13	13	8	56:41	15	52	65	29
8	VfL Wolfsburg	34	14	8	12	64:58	6	50	64	32
9	1. FSV Mainz 05	34	12	11	11	36:42	-6	47	59	33
10	Eintracht Frankfurt	34	12	10	12	47:54	-7	46	58	34
11	TSG Hoffenheim	34	11	9	14	44:42	2	42	53	37
12	Borussia M'gladbach	34	10	9	15	43:60	-17	39	49	39
13	1. FC Köln	34	9	11	14	33:42	-9	38	47	39
14	SC Freiburg	34	9	8	17	35:59	-24	35	44	42
15	Hannover 96	34	9	6	19	43:67	-24	33	42	44
16	1. FC Nürnberg	34	8	7	19	32:58	-26	31	39	45
17	VfL Bochum	34	6	10	18	33:64	-31	28	34	46
18	Hertha BSC	34	5	9	20	34:56	-22	24	29	49

Figure 214: Bundesliga-Abschlusstabelle Saison 2009/2010

Es sind zwei zusätzliche Spalten definiert worden, 'Punkte Plus' und 'Punkte Minus'. Diese berechnen sich wie folgt:

- 'Punkte Plus' ist 2 Mal (Anzahl der Siege + Anzahl der Unentschieden / 2)
- 'Punkte Minus' ist 2 Mal (Anzahl der Niederlagen + Anzahl der Unentschieden / 2)

Das liefert dann eine glatte (ganze) Punktzahl, die als Gewichtung für die Siege bzw. Niederlagen dient.

Anmerkungen:

- diese Aufteilung in zwei Gruppen heißt Dichotomisierung und bedeutet eigentlich Trennung einer Menge in zwei Teilmengen ohne Schnittmenge.
- das Ergebnis hängt von der Art der Dichotomisierung ab ...

Möchte man nun wissen ob zwei Vereine signifikant unterschiedlich leistungsstark sind, dann betrachtet man dazu die Kontingenztabelle bzgl. 'Punkte Plus' und 'Punkte Minus' der beiden Vereine. Für die beiden Vereine 'Bayern München' und 'FC Schalke 04' ist diese dann:

$$\begin{matrix} 90 & 18 \\ 84 & 22 \end{matrix}$$

Mit dem χ^2-Test kann nun die Frage nach der unterschiedlichen Leistungsstärke beantwortet werden; das Ergebnis des Tests ist:

```
Chi-Quadrat-Test: C15; C16

Die erwarteten Anzahlen werden unter den beobachteten Anzahlen ausgegeben.
Die Beiträge zu Chi-Quadrat werden unterhalb der erwarteten Anzahlen
ausgegeben.

C15    C16  Gesamt
1     90     18     108
87,81  20,19
0,054  0,237

2     84     22     106
86,19  19,81
0,055  0,241

Gesamt    174     40     214

Chi-Qd = 0,588; DF = 1; p-Wert = 0,443
```

Diese beiden Vereine sind nach der hier vorgeschlagenen Bewertung (mit Hilfe der 'Punkte' für Siege und Niederlagen) nicht signifikant unterschiedlich in ihrer Leistungsstärke, der p-Wert ist mit 0.443 zu hoch.

Der erste Verein, der - von oben nach unten gelesen - eine signifikant andere Leistungsfähigkeit hat als 'Bayern München', ist auf Platz 7 'VfB Stuttgart'; für diese Paarung wird der p-Wert das erste Mal kleiner als 5%. Der χ^2-Test liefert konkret:

```
Chi-Quadrat-Test: C15; C16
```

```
Die erwarteten Anzahlen werden unter den beobachteten Anzahlen ausgegeben.
Die Beiträge zu Chi-Quadrat werden unterhalb der erwarteten Anzahlen
ausgegeben.

C15    C16  Gesamt
1     90     18     108
83,88  24,12
0,446  1,551

2     70     28      98
76,12  21,88
0,492  1,710

Gesamt    160     46     206

Chi-Qd = 4,198; DF = 1; p-Wert = 0,040
```

Index: