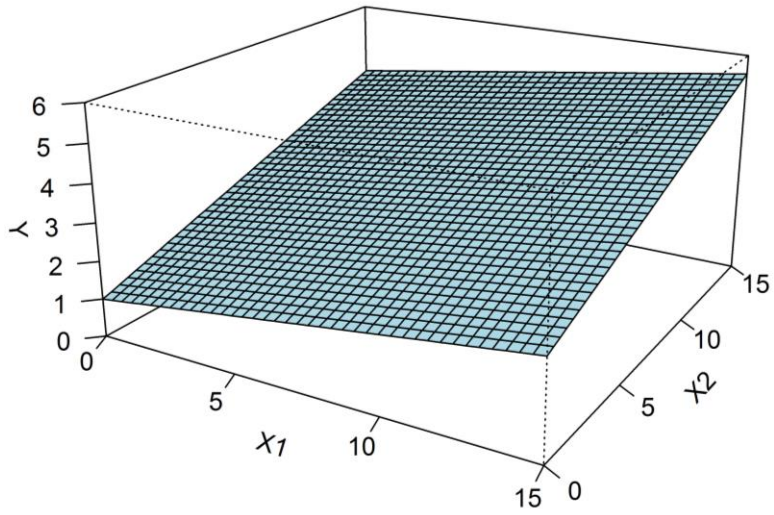
UNIVERSITÄT TRIER

ZIMK - Zentrum für
Informations-, Medien- und
Kommunikationstechnologie

Bernhard Baltes-Götz

Lineare Regressionsanalyse mit SPSS



2022 (Rev. 220610)

Herausgeber: Zentrum für Informations-, Medien- und Kommunikationstechnologie (ZIMK)
an der Universität Trier
Universitätsring 15
D-54286 Trier
WWW: zimk.uni-trier.de
E-Mail: zimk@uni-trier.de
Autor: Bernhard Baltes-Götz
WWW: https://www.uni-trier.de/?id=54994
E-Mail: baltes@uni-trier.de
Copyright: © ZIMK 2022

Vorwort

In diesem Kurs werden elementare Begriffe und Verfahren der linearen Regressionsanalyse in Theorie und Praxis behandelt, wobei den Methoden zur Modelldiagnose besondere Aufmerksamkeit zukommt. Auf der Basis einer statistischen Grundausbildung (zu Begriffen wir *Parameter*, *Statistik*, *Signifikanztest*, etc.) und einer gewissen Erfahrung mit der Regressionsanalyse sollten die Erläuterungen zur Begründung von Analyseschritten und zur Interpretation der Ergebnisse nachvollziehbar sein.

Als Software kommt die Version 27 von IBM SPSS Statistics für Windows (im Manuskript meist kurz als *SPSS* bezeichnet) zum Einsatz, jedoch können die meisten vorgestellten Verfahren auch mit anderen SPSS-Versionen ab 16 unter Windows, macOS oder Linux realisiert werden.

Die aktuelle Version des Manuskripts ist als PDF-Dokument zusammen mit den im Kurs benutzen Daten- und Syntaxdateien auf dem Webserver der Universität Trier zu finden:

http://www.uni-trier.de/index.php?id=22489

Kritik und Verbesserungsvorschläge zum Manuskript werden dankbar entgegen genommen (z. B. unter der Mail-Adresse baltes@uni-trier.de).

Trier, im April 2022 Bernhard Baltes-Götz

Inhaltsübersicht

VORWORT **3**

1 BIVARIATE LINEARE REGRESSION **8**

1.1 Beispiel: Regression von Gewicht auf Größe **8**

1.2 Modell **10**
1.2.1 Fixierter Regressor 10
1.2.2 Annahmen 12
1.2.2.1 Skalenqualität 12
1.2.2.2 Stetigkeit der Kriteriumsverteilung 12
1.2.2.3 Linearität 13
1.2.2.4 Normalität der Residuen 14
1.2.2.5 Varianzhomogenität der Residuen 14
1.2.2.6 Unkorreliertheit der Residuen 14
1.2.2.7 Fehlerfrei ermittelte Regressorwerte 15
1.2.2.8 Technische Voraussetzungen 15
1.2.2.9 Häufige Missverständnisse 16
1.2.3 Stochastische Regressoren 16
1.2.4 Mathematisches Modell und Realität 16

1.3 Schätzung der Modellparameter **17**
1.3.1 Anforderung bei SPSS 17
1.3.2 Herleitung und Eigenschaften der Schätzer 18
1.3.3 Standardisierte Regressionskoeffizienten 19
1.3.4 Empirische Residuen und geschätzte Residualvarianz 20

1.4 Signifikanztests und Konfidenzintervalle zu den Regressionskoeffizienten **21**
1.4.1 Inferenzstatistische Beurteilung von b_1 21
1.4.1.1 Quadratsummenzerlegung und F-Test 21
1.4.1.2 Zwei- und einseitiger t-Test 22
1.4.1.3 Standardfehler 24
1.4.1.4 Konfidenzintervalle 24
1.4.2 Inferenzstatistische Beurteilung von b_0 25

1.5 Determinationskoeffizient **26**
1.5.1 Korrigierter Determinationskoeffizient 26
1.5.2 Relevanz für die Modellbeurteilung 27
1.5.3 Konfidenzintervall 27

1.6 Besonderheiten bei der homogenen Regression **28**

1.7 Modelldiagnose und -modifikation **30**
1.7.1 Linearität 30
1.7.1.1 Diagnose 30
1.7.1.1.1 Residuen-Plots 30
1.7.1.1.2 Linearitätstest 34
1.7.1.2 Linearitätsdefekte beheben 35
1.7.2 Auswahl der zu prüfenden Residuen 38
1.7.2.1 Zentrierte Hebelwerte 38
1.7.2.2 Studentisierte Residuen 39
1.7.2.3 Ausgelassen-studentisierte Residuen 39
1.7.2.4 Standardisierte Residuen 40

1.7.3 Varianzhomogenität der Residuen 40
1.7.3.1 Konsequenzen bei verletzter Varianzhomogenität 41
1.7.3.2 Diagnosemethoden 43
1.7.3.2.1 Residuen-Plot 43
1.7.3.2.2 Spread & Level - Plot 44
1.7.3.2.3 Maximalquotientenkriterium 48
1.7.3.2.4 Breusch-Pagan - Heteroskedastizitätstest 49
1.7.3.3 Transformation des Kriteriums zur Homogenisierung der Fehlervarianzen 52
1.7.3.4 Robuste Inferenzstatistik trotz Heteroskedastizität 56
1.7.3.4.1 Heteroskedastizitäts-konsistente Standardfehler 57
1.7.3.4.2 Bootstrapping 59
1.7.3.5 WLS-Regression 61
1.7.4 Normalverteilung der Residuen 66
1.7.4.1 Relevanz der Normalverteilungsvoraussetzung und zentraler Grenzwertsatz 66
1.7.4.2 Gemeinsame Prüfung für alle Ausprägungen des Regressors 66
1.7.4.3 Auswahl der zu prüfenden Residuen 66
1.7.4.4 Grafische Diagnosemethoden 67
1.7.4.5 Statistische Diagnosemethoden 68
1.7.4.6 Maßnahmen bei einer inakzeptablen Verletzung der Normalität 71
1.7.4.6.1 Elimination von Ausreißern 71
1.7.4.6.2 Box-Cox - Transformation zur Normalisierung der Residualverteilung 76

1.8 Stichprobenumfangsplanung 78
1.8.1 Analysen mit G*Power 79
1.8.1.1 Modell mit einem fixierten Regressor 79
1.8.1.2 Modell mit einem stochastischen Regressor 81
1.8.2 Analysen mit SPSS 83
1.8.3 Simulationsstudie zur Verlässlichkeit von Stichprobenumfangsempfehlungen 85

2 MULTIPLE LINEARE REGRESSION 87

2.1 Beispiel und Anforderung einer multiplen linearen Regression in SPSS 87
2.1.1 Einflussfaktoren auf die Mortalität in amerikanischen Städten 87
2.1.2 Anforderung einer multiplen linearen Regression 88

2.2 Modell und Annahmen 90
2.2.1 Linearität 91
2.2.2 Normalverteilung und Varianzhomogenität der Residuen 92
2.2.3 Unkorreliertheit der Residuen 92
2.2.4 Technische Voraussetzungen 92

2.3 Parameterschätzung 92

2.4 Standardisierte Regressionskoeffizienten 94

2.5 Signifikanztests und Konfidenzintervalle 96
2.5.1 Quadratsummenzerlegung und F-Test zur globalen Nullhypothese 96
2.5.2 Signifikanztests zu den einzelnen Regressionskoeffizienten 96
2.5.3 Konfidenzintervalle zu den Regressionskoeffizienten 98
2.5.4 Multiples Testen 99
2.5.4.1 Scheinbar geschützte t-Tests zu den Regressionskoeffizienten 99
2.5.4.2 Multiple Tests mit α-Fehler - Kumulierungskontrolle nach Bonferroni-Holm 101

2.6 Determinationskoeffizient 102

2.7 Phänomene der multiplen Regression 104
2.7.1 Einfacher und partieller Regressionskoeffizient 105
2.7.2 Partielle Redundanz 105
2.7.3 Vulnerabilitäts-Kompensation 106
2.7.4 Suppressoreffekt 107

2.8 Individuelle Erklärungsbeiträge der Regressoren 109
2.8.1 Quadrierte semipartielle Kriteriumskorrelationen 109
2.8.2 R^2 - Anstiege aus einer Serie geschachtelter Modelle 110
2.8.3 Quadrierte partielle Kriteriumskorrelationen (partielles Eta-Quadrat) 110

2.9 Überprüfung der Modellannahmen 112
2.9.1 Linearität 112
2.9.2 Homoskedastizität 116
2.9.3 Normalverteilung 118

2.10 Power-Analyse 120
2.10.1 Modelle mit fixierten Regressoren 120
2.10.1.1 A priori - Poweranalyse 120
2.10.1.1.1 Globaler F-Test 120
2.10.1.1.2 T-Test zu einem einzelnen Regressor 121
2.10.1.1.3 Mehrere t-Tests zu einzelnen Regressoren 122
2.10.1.2 Post hoc - Poweranalyse 123
2.10.2 Modelle mit stochastischen Regressoren 124

3 GEFAHREN FÜR EINE ERFOLGREICHE MODELLIERUNG 126

3.1 Fälle mit großem Einfluss auf die Schätzung oder Beurteilung des Modells 126
3.1.1 Schlecht prognostizierbare Fälle 126
3.1.2 Fälle mit starker Hebelwirkung 129
3.1.3 Gesamtbeurteilung der Einflussstärke 131

3.2 Multikollinearität 139

4 POLYNOMISCHE REGRESSIONSMODELLE 143

4.1 Ein quadratisches Modell für das volkswirtschaftliche Beispiel 143

4.2 Zentrieren und Multikollinearität bei der polynomischen Regression 147

5 ALTERNATIVEN ZUM SIMULTANEN EINSCHLUSS ALLER REGRESSOREN 150

5.1 Blockbildung zur Untersuchung eines kategorialen Regressors 150

5.2 Automatische Modellsuche 152

5.3 Empfehlungen zur Modellbildung 154

6 SPEZIELLE THEMEN 155

6.1 Kausale Interpretation von Regressionskoeffizienten 155
6.1.1 Verzerrte Effektschätzungen durch vergessene relevante Regressoren 155
6.1.2 Auswirkungen überflüssiger Regressoren auf die Effektbeurteilungen 156

6.2 Mediation 156

6.3 Strukturgleichungsmodelle 158

6.4 Vergleich der Determinationskoeffizienten von nicht geschachtelten Modellen 159

7 REGRESSIONSMODELLE FÜR ZEITREIHENDATEN 160

7.1 Beispiel 160

7.2 Autokorrelation der Residuen 162

7.3 Durbin-Watson-Test 163

7.4 Regression mit integriertem AR(1) - Modell für die Residuen 166
7.4.1 Modell 166
7.4.2 Durchführung mit SPSS 167
7.4.3 Der Box-Ljung - Test als Alternative zum Durbin-Watson - Test 169

ANHANG 171

R 3.6.3 und die R-Integration installieren 171

LITERATUR 174

STICHWORTVERZEICHNIS 177

1 Bivariate lineare Regression

In der einfachen (bivariaten) linearen Regressionsanalyse wird die kausale oder prognostische Relevanz einer **unabhängigen Variablen** *X* (Synonyme: *Regressor, exogene Variable*) für eine **abhängige Variable** *Y* (Synonyme: *Kriterium, Regressand, endogene Variable*) untersucht. Während bei der abhängigen Variablen das metrische Skalenniveau vorausgesetzt wird, kann die unabhängige Variable metrisch oder dichotom-kategorial sein.

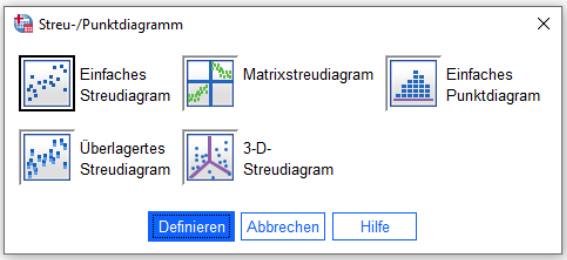
In der Forschungspraxis ist die bivariate lineare Regression kaum einmal das angemessene Forschungsinstrument, weil „mono-kausale" empirische Systeme eher selten anzutreffen sind, und auch bei Interventionsstudien in der Regel die kombinierte Wirkung von *mehreren* Regressoren interessiert. Folglich kommt zur Aufklärung eines metrischen Kriteriums sehr viel häufiger die *multiple* lineare Regression zum Einsatz. Aus didaktischen Gründen scheint es aber sinnvoll, die nicht allzu kleine Zahl von methodischen und technischen Details der regressionsanalytischen Methodologie dosiert vorzustellen. Die im aktuellen Abschnitt 1 über die bivariate Regression behandelten Themen sind weitgehend analog auch für die multiple Regression relevant.

1.1 Beispiel: Regression von Gewicht auf Größe

Zur Erläuterung der einfachen linearen Regression verwenden wir die Datei **ggg.sav**, die für 332 Erwachsene im Alter von 20 bis 30 Jahren (Studierende der Universität Trier) die Körpergröße (SPSS-Variablenname GROESSE), das Körpergewicht (SPSS-Variablenname GEWICHT) und das Geschlecht (SPSS-Variablenname GESCHLECHT) enthält.[1] Im Beispiel übernimmt GEWICHT die Rolle des Kriteriums und GROESSE die Rolle des Regressors. Vor dem Einstieg in die statistische Analyse verschaffen wir uns einen Eindruck von der gemeinsamen Stichprobenverteilung der beiden Variablen. Nach dem Menübefehl

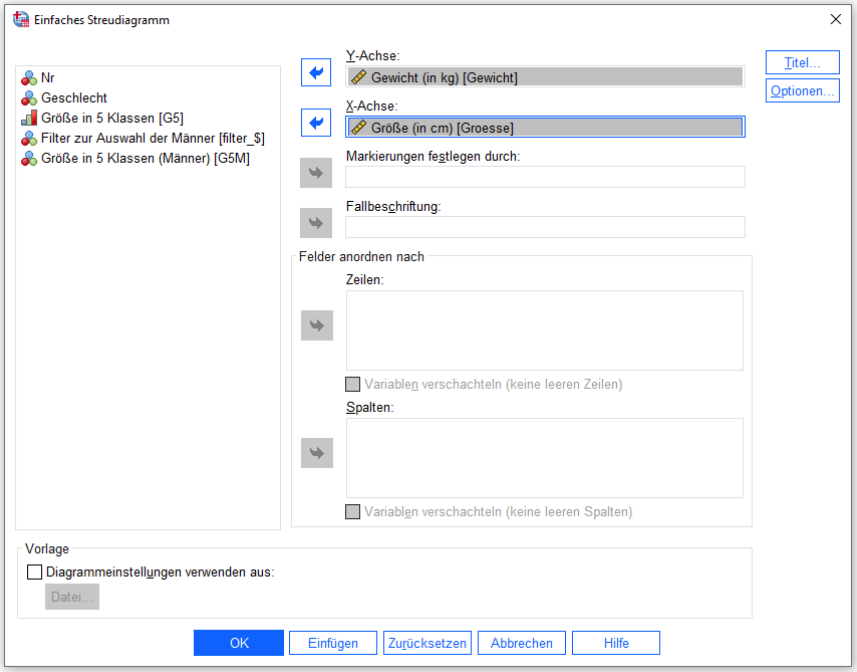
Grafik > Klassische Dialogfelder > Streu-/Punktdiagramm

und der Entscheidung für ein **einfaches Streudiagramm**



wurde mit der Dialogbox

[1] Wie die Datei **ggg.sav** via Internet zu beziehen ist, wird im Vorwort beschrieben. SPSS-Variablennamen werden im Manuskript zur Hervorhebung groß geschrieben.



das folgende bivariate Streudiagramm für GEWICHT und GROESSE erstellt, das per Diagrammeditor noch eine Regressionsgerade erhalten hat (zur Bedienung des Diagrammeditors siehe z. B. Baltes-Götz 2022):

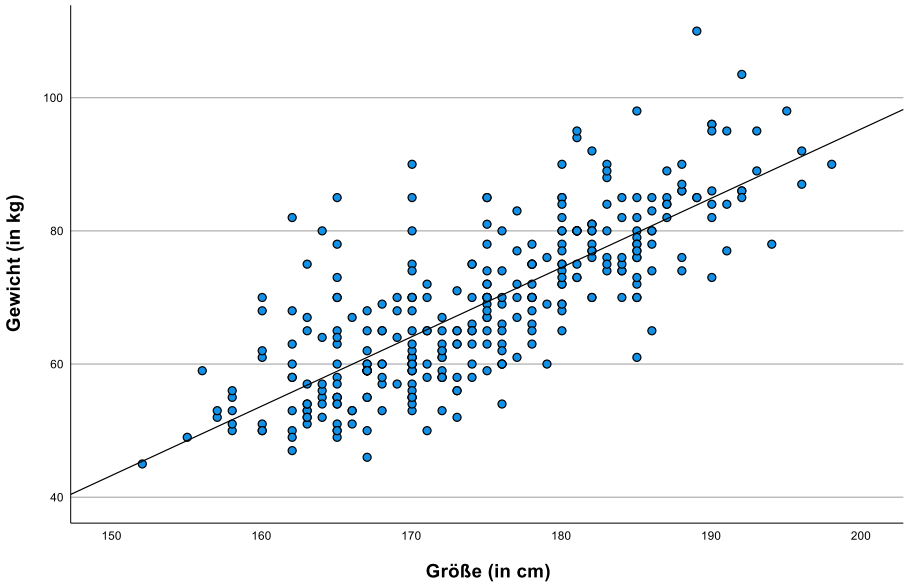


Abbildung 1: Regression von GEWICHT auf GROESSE bei 332 Personen

Die dem Punkteschwarm im Sinne einer minimalen Fehlerquadratsumme (siehe Abschnitt 1.3) angepasste Regressionsgerade kann offenbar den Zusammenhang zwischen den beiden metrischen Merkmalen gut beschreiben. Sie steht für das Modell, dass mit zunehmender Größe das mittlere Gewicht (der größenbedingte Erwartungswert für das Gewicht) linear ansteigt.

1.2 Modell

1.2.1 Fixierter Regressor

Das bivariate Regressionsmodell geht in seiner klassischen, auch in der heutigen Praxis dominierenden Form von einem **fixierten Regressor** aus, d.h.:

- Es wird vorab festgelegt, welche Werte des Regressors mit welchen Häufigkeiten in die Studie einbezogen werden.
- Die Werte des Regressors können fehlerfrei festgestellt werden.
- Die Ergebnisse der Studie können *nicht* über die dort realisierten Regressorwerte hinaus generalisiert werden.

Diese Annahmen scheinen sehr restriktiv und am ehesten auf *experimentelle Studien* mit einem *manipulierten Regressor* anwendbar zu sein (z. B. mit 5 festgelegten Dosierungen einer verabreichten Substanz als Regressorwerten). Es ist allerdings verbreitete Praxis, die aus dem Modell mit einem fixierten Regressor abgeleiteten statistischen Verfahren auch in *Beobachtungsstudien* mit zufällig realisierten Regressorwerten anzuwenden. Im Abschnitt 1.2.3 über *stochastische* Regressoren wird sich herausstellen, dass diese Praxis in der Regel akzeptabel ist.

In einer bivariaten empirischen Regressionsstudie liegen N Wertepaare (x_i, y_i) vor, wobei im klassischen Modell die X-Werte als **fest** vom Versuchsleiter vorgegebene und **fehlerfrei gemessene** Größen, die Y-Werte hingegen als Realisationen von N Zufallsvariablen Y_i aufgefasst werden. Für die N Beobachtungsvariablen Y_i in der Stichprobe wird folgendes Modell behauptet:

$$\begin{aligned} \mathrm{E}(Y_i) &= \beta_0 + \beta_1 x_i \\ \varepsilon_i &:= Y_i - \mathrm{E}(Y_i) \qquad\qquad i = 1,..,N \\ \varepsilon_i &\sim \mathrm{N}(0, \sigma_\varepsilon^2), \mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}_N \end{aligned} \tag{1}$$

Für die zu einem festen Wert x_i des Regressors beobachtete Variable Y_i, liegt modellgemäß der Erwartungswert $\mathrm{E}(Y_i)$ auf einer Geraden, welche durch die Regressionskoeffizienten β_0 und β_1 bestimmt ist:

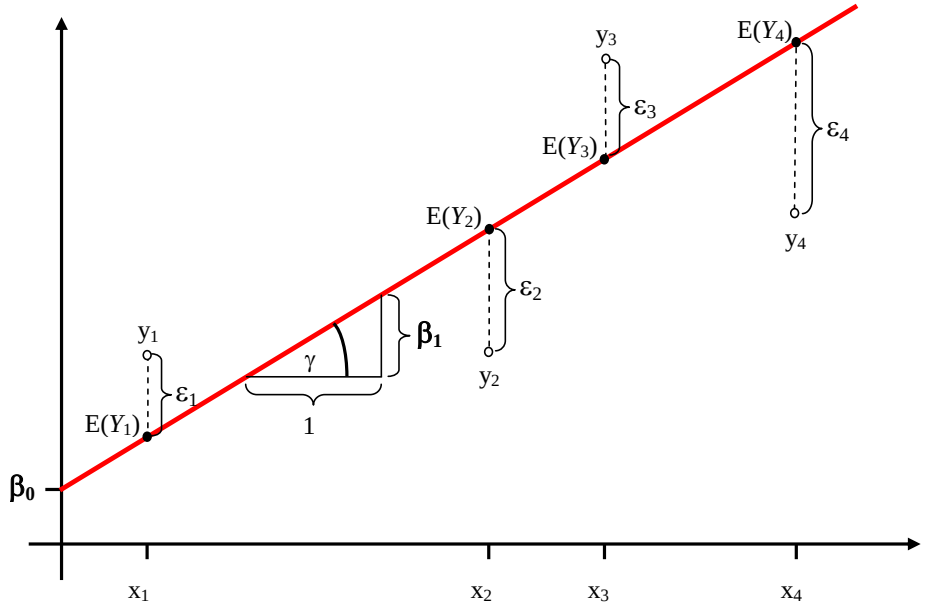


Abbildung 2: Veranschaulichung des Modells der bivariaten linearen Regression

Die Regressionskoeffizienten β_0 und β_1 sind zwei feste, für alle Regressorwerte x_i identische Zahlen, die folgendermaßen zu interpretieren sind:

- Erhöht man *X* um eine Einheit, so steigt modellgemäß der Mittelwert von *Y* um β_1 Einheiten an. Das „Erhöhen“ ist nicht unbedingt im Sinne einer Manipulation bei einem konkreten Fall aus dem Anwendungsbereich der Theorie zu verstehen, sondern bedeutet oft den Übergang zu einem Fall mit einer höheren *X*-Ausprägung. In unserem Beispiel erwarten wir bei Steigerung der Körpergröße um 1 cm ein um β_1 kg erhöhtes mittleres Gewicht.
- β_0 ist das vom Modell für einen Fall mit der Größe 0 prognostizierte Gewicht und offenbar von geringer praktischer Bedeutung. Indem man z.B. den Prädiktor zentriert, d.h. auf den Mittelwert 0 bringt, kann man für eine sinnvolle β_0 - Bedeutung sorgen: prognostiziertes Gewicht bei mittlerer Größe. In der Regel ist der Parameter β_0 von untergeordnetem Interesse, darf aber trotzdem nicht weggelassen werden, weil die Regressionsgerade ansonsten gezwungen wird, durch den Nullpunkt zu verlaufen.

Wenn kein deterministischer Zusammenhang zwischen dem Kriterium und dem Regressor besteht, wird eine beobachtete Y_i - Ausprägung praktisch nie mit ihrem Erwartungswert übereinstimmen. Die Abweichung

$$\varepsilon_i := Y_i - \mathrm{E}(Y_i)$$

zur *i*-ten Beobachtung bezeichnen wir als **Fehler- bzw. Residualvariable** ε_i . Es ist der Anteil von Y_i, der nicht aufgrund der statistischen Beziehung von Regressor und Kriterium vorhergesagt werden kann. Hier artikulieren sich Fehler bei der Messung der Kriteriums und andere nicht im Modell enthaltene Einflüsse auf das Kriterium.

Für jede realisierte Regressorausprägung x_i wird angenommen, dass die zugehörige Residualvariable ε_i einer Normalverteilung mit dem Erwartungswert 0 und der (überall identischen) Varianz σ_ε^2 folgt:

$$\varepsilon_i \sim \mathrm{N}(0, \sigma_\varepsilon^2)$$

Dass die Residuen ε_i modellgemäß (also insbesondere bei erfüllter Linearität) den Erwartungswert 0 besitzen, folgt aus ihrer Definition:

$$\varepsilon_i := Y_i - \mathrm{E}(Y_i) \quad \Rightarrow \quad \mathrm{E}(\varepsilon_i) = \mathrm{E}(Y_i - \mathrm{E}(Y_i)) = \mathrm{E}(Y_i) - \mathrm{E}(Y_i) = 0$$

In der kompakten Aussage

$$\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}_N$$

über die Kovarianzmatrix zum Vektor $\boldsymbol{\varepsilon}$

$$\boldsymbol{\varepsilon} := (\varepsilon_1, ..., \varepsilon_N)$$

mit allen *N* Residuen (siehe Erläuterung im Abschnitt 1.2.2.6) steckt die wichtige Voraussetzung, dass die Residuen zu zwei verschiedenen Beobachtungen unkorreliert sind. Im Abschnitt 1.2.2.6 wird der Ausdruck $\sigma_\varepsilon^2 \mathbf{I}_N$ in ausformulierter Matrixform präsentiert und somit die Annahme der linearen Regression über $\mathrm{Cov}(\boldsymbol{\varepsilon})$ anschaulicher dargestellt.

Anwendungsorientierte Forscher[1] müssen sich beim klassischen linearen Model an die folgende Sichtweise der mathematischen Statistik gewöhnen:

[1] Zur Vermeidung von sprachlichen Umständlichkeiten wird im Manuskript die männliche Form verwendet.

- Wir haben es mit *N* beobachtbaren Zufallsvariablen Y_i und *N* nicht-beobachtbaren Zufallsvariablen ε_i zu tun.
- Die *X*-Ausprägungen werden überhaupt nicht als Realisationen irgendwelcher Zufallsvariablen aufgefasst, sondern als feste Werte, die keinen Messfehler enthalten. Im Abschnitt 1.2.3 über stochastische Regressoren werden wir uns allerdings von dieser klassischen, in den meisten sozialwissenschaftlichen Anwendungen unrealistischen *Fixed-X - Annahme* der linearen Regression lösen.

1.2.2 Annahmen

In diesem Abschnitt werden die in Gleichung (1) kompakt formulierten Annahmen der bivariaten linearen Regression erläutert (vgl. z. B. Darlington & Hayes 2017, S. 88ff; Snedecor & Cochran 1980, S. 153).

Wer Voraussetzungen zur Sicherung der *kausalen* Interpretierbarkeit des Parameters β_1 vermisst, sei auf Abschnitt 6.1 vertröstet. Wir konzentrieren uns zunächst auf die folgenden Fragen:

- Beschreibt das lineare Modell die statistische Beziehung von *X* und *Y* korrekt? Hängt der Erwartungswert $E(Y_i)$ für jeden untersuchten x_i - Wert tatsächlich linear von x_i ab?

$$E(Y_i) = \beta_0 + \beta_1 x_i$$

- Sind neben dieser *primären* Annahme (Linearität der Beziehung) auch die *sekundären* Annahmen des Modells (Normalität, Varianzhomogenität und Unkorreliertheit der Residuen) erfüllt? Die Unterscheidung zwischen der primären und den sekundären Annahmen eines linearen Modells wird von Darlington & Hayes (2017, S. 88) vorgeschlagen.
- Ist der Parameter β_1 von 0 verschieden? In welcher Richtung weicht der Parameter vom Wert 0 ab?

1.2.2.1 Skalenqualität

Bei der abhängigen Variablen wird metrisches Skalenniveau vorausgesetzt. Diese Voraussetzung taucht zwar im mathematischen Modell nicht explizit auf, ergibt sich aber aus der Annahme normalverteilter Residuen, die ohne das metrische Skalenniveau sinnlos wäre.

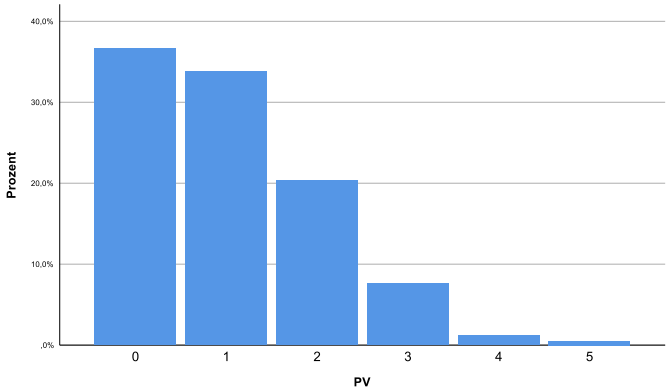
Wenn das Kriterium nur ein geordnet-kategoriales (ordinales) Skalenniveau besitzt, dann bietet sich die ordinale logistische Regressionsanalyse an, die bei den Regressoren dieselben Skalenniveaus unterstützt wie die lineare Regressionsanalyse (siehe Baltes-Götz 2012). Von der logistischen Regressionsanalyse existieren auch Varianten zur Analyse eines ungeordnet-kategorialen (nominalen) Kriteriums mit zwei oder mehr Ausprägungen (binäre bzw. multinomiale logistische Regression).

Die unabhängige Variable kann metrisch oder dichotom-kategorial sein. Auch bei einem dichotomen Regressor ist die lineare Modellierung sinnvoll, wobei die Parameter β_0 und β_1 je nach gewählter Kodierung z. B. für Gruppenmittelwerte oder Mittelwertsunterschiede stehen. Ein kategorialer Regressor mit $w > 2$ Ausprägungen muss im Rahmen einer *multiplen* Regression durch $(w - 1)$ Kodiervariablen repräsentiert werden. Ein geordnet-kategorialer (ordinaler) Regressor muss auf kategoriales Niveau herabgestuft oder als metrisch behandelt werden.

1.2.2.2 Stetigkeit der Kriteriumsverteilung

Die Verteilung der abhängigen Variablen sollte vom Ideal der Stetigkeit nicht zu weit entfernt sein. Eine Verteilung ist stetig im Sinne der Wahrscheinlichkeitstheorie, wenn ihre Verteilungsfunktion stetig im Sinn der Analysis ist, also keine Sprünge macht. Bei einer stetigen Verteilung ist die Wahrscheinlichkeit für jede konkrete Ausprägung exakt gleich 0, was für beobachtbare Variablen schon aus Gründen der Messgenauigkeit nicht der Fall ist.

Bei einer Zählvariablen als Kriterium (z. B. Anzahl der im letzten Jahr gelesenen Bücher)



ist oft ein *generalisiertes lineares Modell* mit der Annahme einer diskreten Verteilung (z.B. Poisson-Verteilung oder negative Binomialverteilung) für die Residuen besser geeignet als das lineare Modell, das für die Residuen die stetige Normalverteilung annimmt (siehe z. B. Baltes-Götz 2016a).

Wenn für diskret verteilte Residuen kein konkretes Verteilungsgesetz angenommen werden soll, dann bietet sich die logistische Regressionsanalyse an, die im Abschnitt 1.2.2.1 schon für Kriterien mit geordnet-kategorialem oder nominalem Skalenniveau empfohlen wurde (siehe Baltes-Götz 2012).

In der wirtschafts- und sozialwissenschaftlichen Forschung ist es eine verbreitete Praxis, Antworten von Personen auf einer Skala mit 5 oder mehr Stufen als intervallskaliert zu behandeln, obwohl streng genommen nur ein geordnet-kategoriales Messniveau vorliegt (vgl. Warner 2013, S, 6ff).

1.2.2.3 Linearität

Laut Modell gilt für zwei unbekannte (zu schätzende) reelle Zahlen (Regressionskoeffizienten) β_0 und β_1, dass zu allen in der Studie realisierten Regressorwerten x_i die Erwartungswerte $E(Y_i)$ der zugehörigen Zufallsvariablen Y_i auf der Regressionsgeraden durch die Punktepaare $(x_i, \beta_0 + \beta_1 x_i)$ liegen:

$$\exists \beta_0, \beta_1 \in \mathrm{R}: \quad \mathrm{E}(Y_i) = \beta_0 + \beta_1 x_i \quad \forall i = 1,.., N$$

Dabei ist β_0 der Schnittpunkt der Regressionsgeraden mit der Y-Achse (Ordinatenabschnitt) und β_1 die Steigung der Regressionsgeraden (siehe Abbildung 2).

Für die Residuen ε_i folgt aufgrund ihrer Definition

$$\varepsilon_i := Y_i - \mathrm{E}(Y_i)$$

der Erwartungswert 0, sodass in einer Stichprobe anhand der *geschätzten* Residuen (siehe Abschnitt 1.7.2) die Linearitätsannahme überprüft werden kann. Ist die Linearitätsannahme erfüllt, dann haben auch die geschätzten Residuen für beliebige Werte des Regressors den Erwartungswert 0.

Die Linearitätsannahme ist bei der Modellierung der Regression von *Y* auf *X* weit weniger restriktiv, als ihr Name es vermuten lässt. Der Regressor *X* kann z. B. durch Quadrieren einer anderen Variablen *Z* entstanden sein, sodass im Rahmen unseres Modellansatzes ohne weiteres die (kurvilineare) Regression von *Y* auf Z^2 analysiert werden kann. Allerdings wird neben Z^2 in der Regel auch *Z* als Regressor auftreten, sodass wir das Modell der *bivariaten* Regression verlassen (siehe Abschnitt 4 über polynomische Regressionsmodelle).

In einem weiteren Vorausblick auf die Modellierungsflexibilität der *multiplen* Regression soll noch erwähnt werden, das auch die *Interaktion* zweier Regressoren *X* und *Z* bzgl. des Kriteriums *Y* mit der multiplen linearen Regressionsanalyse untersucht werden kann, indem neben *X* und *Z* auch das Produkt $X \cdot Z$ als Regressor einbezogen wird. Über die Behandlung von Interaktionseffekten informiert ein ZIMK-

Manuskript (Baltes-Götz 2020a), das auf dem Webserver der Universität Trier von der Startseite (http://www.uni-trier.de/) ausgehend folgendermaßen finden ist:

IT-Services > Downloads & Broschüren > Downloads
Mediator- und Moderatoranalyse mit SPSS und PROCESS

Manche nichtlineare Modelle lassen sich durch die Transformation von Kriterium und/oder Regressor linearisieren. Ein bekanntes Beispiel aus der Wahrnehmungspsychologie ist das Stevens'sche Potenzgesetz (gefunden bei Wentura & Pospeschill 2015, S. 57). Für die objektive Reizstärke *S* (z. B. Lichtstärke) und die subjektive Wahrnehmung *W* (z. B. empfundene Helligkeit) wird der folgende Zusammenhang mit Beteiligung der Parameter *a* und *b* behauptet:

$$W = aS^b$$

Wendet man auf beide Seiten der Gleichung die Logarithmusfunktion an, dann resultiert ein bivariates lineares Modell mit dem Kriterium $\log(W)$ und dem Regressor $\log(S)$. Zur Anwendung auf empirische Daten muss im Vergleich zum Stevens'schen Potenzgesetz noch ein Residualterm ergänzt werden:

$$\log(W) = a + b\log(S) + \varepsilon$$

In sehr seltenen Fällen ist in der bivariaten linearen Regression ein Modell *ohne* Ordinatenabschnitt β_0 gewünscht, sodass die Regressionsgerade durch den Ursprung des Koordinatensystems verläuft. Ist der Regressorwert $x_i = 0$ im Untersuchungsplan enthalten, dann behauptet ein solches Modell über die zugehörige Zufallsvariable Y_i:

$$E(Y_i) = \beta_1 x_i = \beta_1 0 = 0$$

Man erhält das Modell der sogenannten *homogenen Regression*, bei dem einige Besonderheiten zu beachten sind, die im Abschnitt 1.6 behandelt werden (siehe auch Kockläuner 1988, S. 44ff). In der Regel verwendet man das *inhomogene* Modell *mit* Ordinatenabschnitt, dessen verdächtig klingender Name nicht zu der falschen Annahme verleiten sollte, es handele sich um ein minderwertiges Modell. Von der (*in-*)*homogenen Regression* zu sprechen, ist in der mathematischen Statistik üblich und auch in der angewandten Statistik verbreitet (siehe z. B. Eckstein 2016).

1.2.2.4 Normalität der Residuen

Für die (nicht beobachtbaren) Fehler- bzw. Residualvariablen ε_i wird angenommen, dass sie **normalverteilt** sind. Sie dürfen sich vorstellen, dass für jeden realisierten Regressorwert x_i eine Normalverteilung potentieller ε_i - Werte existiert, aus der eine zufällige Realisation gezogen wird, die zusammen mit dem konstanten Anteil $\beta_0 + \beta_1 x_i$ die Realisation der abhängigen Variablen Y_i ergibt.

1.2.2.5 Varianzhomogenität der Residuen

Die Normalverteilungen der Residualvariablen ε_i haben **alle dieselbe Varianz** σ_ε^2. Statt von *Varianzhomogenität* spricht man auch von *Homoskedastizität*.

1.2.2.6 Unkorreliertheit der Residuen

Modellgemäß sind die *N* Residualvariablen ε_i unkorreliert. Ihre Kovarianzmatrix ist eine Diagonalmatrix der Ordnung *N* mit dem identischen Eintrag σ_ε^2 auf der Hauptdiagonalen (für die als identisch angenommenen Fehlervarianzen) und Nullen an allen anderen Positionen (für die Kovarianzen). Mit Hilfe der Einheitsmatrix der Ordnung *N* (Symbol: $\mathbf{I}_N$) lässt sich diese Situation kompakt so beschreiben:

$$\text{Cov}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}_N = \begin{bmatrix} \sigma_\varepsilon^2 & 0 & . & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & . & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & . & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & . & 0 & \sigma_\varepsilon^2 \end{bmatrix}$$

Weil die Residuen normalverteilt sind, folgt aus der Unkorreliertheit die stochastische Unabhängigkeit.

Begründete Zweifel an der Unabhängigkeitsannahme bestehen z. B. in folgenden Situationen:

- **Zeitreihendaten** (z. B. Arbeitsmarktdaten aus 40 aufeinander folgenden Jahren)
  Mit Zeitreihendaten werden wir uns im Abschnitt 7 beschäftigen.
- **Cluster-Stichproben** (z. B. 300 Schüler aus insgesamt 20 Schulklassen) oder **Panel-Stichproben** (z. B. Evaluationsstudie mit 100 Teilnehmern, die zu 5 Zeitpunkten beobachtet werden)
  In einem solchen Fall kann man u.a. ....
  - eine sogenannte **Mehrebenenanalyse** durchführen (siehe z. B. Baltes-Götz 2020b; Baltes-Götz 2016b)
  - oder ein **GEE-Modell** (*Generalized Estimating Equation*) anwenden (siehe z. B. Baltes-Götz 2016a).

Bei einer Verletzung der Unabhängigkeitsannahme erhält man zwar unverzerrte Schätzer für die Regressionskoeffizienten, doch resultiert eine fehlerhafte Inferenzstatistik:

- Die geschätzten Vertrauensintervalle sind meist zu klein, sodass die erwartete Überdeckungswahrscheinlichkeit (von meist 95%) nicht eingehalten wird.
- Die Signifikanztests sind zu liberal, sodass die Wahrscheinlichkeit für Fehler erster Art über der akzeptierten Rate (von meist 5%) liegt.

1.2.2.7 Fehlerfrei ermittelte Regressorwerte

Ist die Annahme fehlerfrei ermittelter Regressorwerte *nicht* gerechtfertigt, resultiert ein betragsmäßig geminderter Schätzer für den Regressionskoeffizienten β_1. Zur Korrektur ist ein Modell mit einer **latenten Variablen** als Regressor anzuwenden, was bei Verfügbarkeit von mehreren Indikatoren für den Regressor in einem Strukturgleichungsanalyseprogramm wie z.B. IBM SPSS Amos (siehe z. B. Baltes-Götz 2015) oder Mplus (Muthén & Muthén 2017) möglich ist.

1.2.2.8 Technische Voraussetzungen

Damit der regressionsanalytische Algorithmus (siehe unten) ablaufen kann, sind noch zwei unproblematische technische Voraussetzungen zu erfüllen:

- Die N in der Stichprobe realisierten X-Werte dürfen nicht alle gleich sein. Man wird sich in der Regel um eine große Variabilität bemühen. Extreme (untypische) und daher sehr einflussreiche Fälle (siehe Abschnitt 3.1.3) sind aber zu vermeiden.
- Der Stichprobenumfang N muss größer sein als 2 (= Anzahl der Parameter).

Wie man sich leicht vorstellen kann, garantiert eine Beachtung des minimalen Stichprobenumfangs *keine* geeignete Power für die üblichen Hypothesentests zu den Regressionskoeffizienten (siehe Abschnitt 1.4). Wir werden uns im Abschnitt 1.8 mit der Power-orientierten Stichprobenumfangsplanung beschäftigen.

1.2.2.9 Häufige Missverständnisse

Um häufig anzutreffenden Missverständnissen entgegenzuwirken, soll für das lineare Regressionsmodell mit einem fixierten Regressor noch einmal ausdrücklich betont werden:

- Es wird keine Annahme über die Verteilung des Regressors benötigt.
- Es wird keine Annahme über die *Randverteilung* des Kriteriums benötigt. Die Normalverteilungsvoraussetzung besteht für die *Residuen des Modells*.
- Es ist keine bivariate Normalverteilung von Regressor und Kriterium erforderlich.

Im gleich folgenden Abschnitt 1.2.3 wird sich zeigen, dass im Rahmen des klassischen Modells und mit der darauf basierenden Software auch *stochastische* Regressoren analysiert werden dürfen.

1.2.3 Stochastische Regressoren

Das bislang dargestellte regressionsanalytische Modell scheint wegen der Voraussetzung fester *X*-Werte z. B. auf die im Abschnitt 1.1 präsentierten Daten zur Regression von Gewicht auf Größe *nicht* anwendbar zu sein. Statt Personen mit vorher genau festgelegten Körpergrößen auszuwählen, wurden die Probanden nämlich zufällig aus einer Grundgesamtheit gezogen. Es lässt sich allerdings zeigen (vgl. Fahrmeir et al. 2007, S. 61f; Snedecor & Cochran 1980, S. 160), dass alle Aussagen, die unter den im Abschnitt 1.2.2 beschriebenen Voraussetzungen gelten, auch auf Studien mit zufällig gewählten *X*-Ausprägungen übertragbar sind, sofern für die **bedingte Verteilung** $\mathrm{L}(Y \mid X)$ der Zufallsvariablen *Y* für beliebige Werte der Zufallsvariablen *X*, gilt (*L* steht für *Law*, und *N* steht für die Normalverteilung):

$$\mathrm{L}(Y \mid X) = \mathrm{N}(\beta_0 + \beta_1 X, \sigma_\varepsilon^2)$$

In diesem Verteilungsgesetz sind Linearität, Normalität und Homoskedastizität enthalten. Außerdem benötigt man die Unabhängigkeit der Residuen und die technischen Voraussetzungen (siehe Abschnitt 1.2.2.8).

Wird im Modell der bivariaten Regression mit einem stochastischen Regressor die **bivariate Normalverteilung** von Regressor und Kriterium zurecht angenommen, dann sind die Linearität sowie die Normalität und Varianzhomogenität der Residuen mit mathematischer Notwendigkeit erfüllt, und bei der Stichprobenumfangsplanung kann die stochastische Natur des Regressors korrekt berücksichtigt werden (siehe Abschnitt 1.8.1.2).

1.2.4 Mathematisches Modell und Realität

Durch die oben aufgelisteten Voraussetzungen der bivariaten linearen Regression werden hochgradig spezielle Verhältnisse angenommen, z. B. die Identität der Varianzen für alle Variablen ε_i. Man müsste von einem Wunder reden, wenn in einer empirischen Studie eine solche Voraussetzung tatsächlich perfekt erfüllt wäre. Für inferenzstatistische Verfahren, die eine Voraussetzung (z. B. die Homoskedastizität oder die Normalität der Residuen) als Nullhypothese prüfen sollen, gilt in der Regel:

- Die Voraussetzung ist mit großer Wahrscheinlichkeit *falsch*.
- Ob die Verletzung der Voraussetzung tatsächlich „entdeckt“ wird, hängt von ihrem Ausmaß und von der Power des Prüfverfahrens, also im Wesentlichen von der Stichprobengröße, ab.

Die eigentlich relevante Frage lautet aber: Ist die Verletzung einer Voraussetzung tolerierbar oder nicht. Leider gibt es zur Klärung dieser Frage oft kein objektives Verfahren (ohne Ermessensentscheidungen).

Ein Forschungsprojekt mit dem Ziel eines allgemeingültigen Modells liefert in der Regel eine *vereinfachte* Sicht der Realität zu liefern, was beim Statistiker George E. P. Box zur desillusionierten, aber nicht resignierten Einsicht geführt hat (Box 1979, S. 202):

> All models are wrong but some are useful.

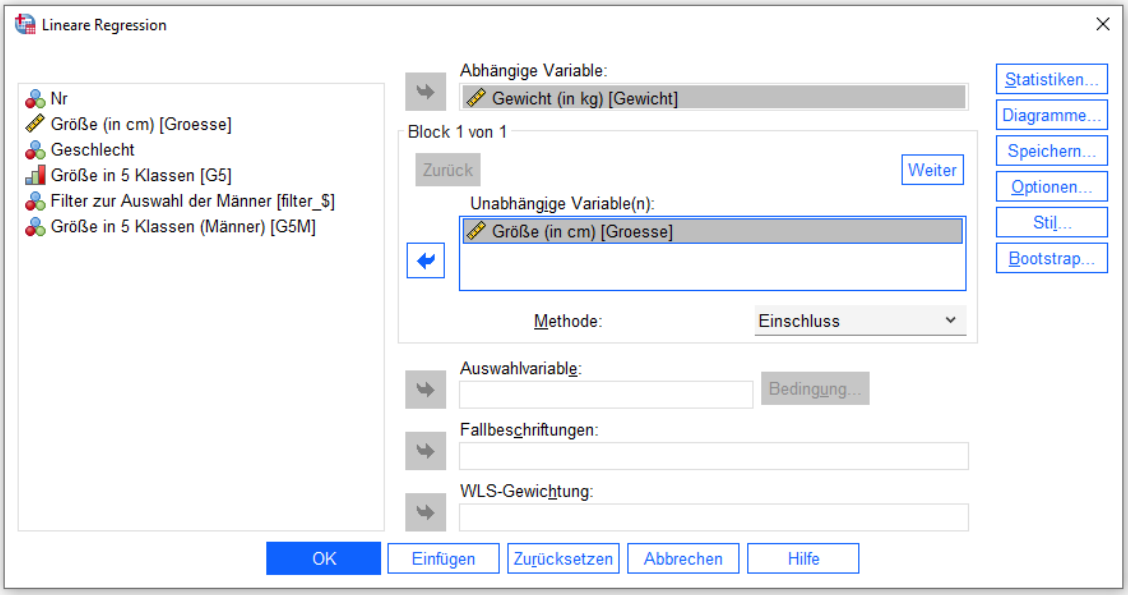
1.3 Schätzung der Modellparameter

1.3.1 Anforderung bei SPSS

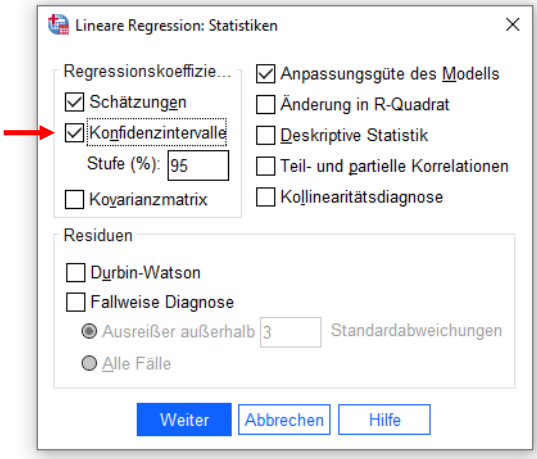
In diesem Abschnitt beschäftigen wir uns mit der Parameterschätzung in der bivariaten linearen Regression und verwenden zur Illustration die SPSS-Ausgaben zur (inhomogenen) Regression von GEWICHT auf GROESSE, die Sie nach dem Menübefehl

Analysieren > Regression > Linear

in der folgenden Dialogbox anfordern können:



Verlangen Sie in der **Statistiken** - Subdialogbox über die Voreinstellung hinausgehend noch die Berechnung von Konfidenzintervallen zu den Regressionskoeffizienten β_0 und β_1:



Vor einer Prüfung der Regressionsvoraussetzungen dürfen die Ergebnisse noch nicht interpretiert werden:[1]

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,780 [a]	,609	,608	7,658

a. Einflußvariablen : (Konstante), Größe (in cm)

[1] Je nach SPSS-Version kann die Beschriftung der Regressionsergebnistabellen leicht abweichen.

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	30113,460	1	30113,460	513,492	,000 [b]
	Nicht standardisierte Residuen	19352,663	330	58,644		
	Gesamt	49466,123	331			

a. Abhängige Variable: Gewicht (in kg)

b. Einflußvariablen : (Konstante), Größe (in cm)

Koeffizienten [a]

		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
Modell		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	-112,767	8,033		-14,038	,000	-128,569	-96,964
	Größe (in cm)	1,041	,046	,780	22,660	,000	,950	1,131

a. Abhängige Variable: Gewicht (in kg)

Es ist aber der richtige Moment gekommen, das Schätzverfahren der bivariaten linearen Regression zu beschreiben. Zur Prüfung der Voraussetzungen benötigen wir nämlich die geschätzten Residuen, die wiederum nur mit Hilfe der geschätzten Regressionskoeffizienten zu bestimmen sind.

1.3.2 Herleitung und Eigenschaften der Schätzer

Im Folgenden werden durch x_i bzw. y_i, $i = 1,..,N$, die Werte der unabhängigen bzw. abhängigen Variablen in der Stichprobe sowie durch $\bar{x}$ bzw. $\bar{y}$ die zugehörigen Stichprobenmittelwerte bezeichnet. Aus den empirischen Daten werden Schätzer b_0 bzw. b_1 für β_0 bzw. β_1 so bestimmt, dass die Summe der quadrierten Abweichungen der geschätzten Werte:

$$\hat{y}_i := b_0 + b_1 x_i, i = 1,.., N \tag{2}$$

von den beobachteten Werten y_i minimal wird:

$$\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \overset{!}{=} \min$$

Für diese **Methode der kleinsten Quadrate** wird gelegentlich auch die aus dem Englischen übernommene Bezeichnung **OLS-Schätzung** (*Ordinary Least Squares*)) verwendet. Wir bezeichnen anschließend b_0 und b_1 gelegentlich als *KQ-Schätzer* (*Kleinst-Quadrat - Schätzer*).

Mit den Methoden der Differentialrechnung (Nullsetzen der partiellen Ableitungen nach b_0 und b_1 sowie Auflösen der entstehenden *Normalgleichungen*) erhält man die folgenden Schätzer:

$$b_1 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{3}$$

$$b_0 = \bar{y} - b_1 \bar{x} \tag{4}$$

Ist die Linearitätsvoraussetzung erfüllt ($E(Y_i) = \beta_0 + \beta_1 x_i, i = 1,.., N$), dann sind b_0 und b_1 **erwartungstreue Schätzer** für β_0 und β_1.

Sind außerdem die Varianzhomogenität und die Unabhängigkeit der Residuen gegeben, dann sind b_0 und b_1 nach dem Satz von Gauß-Markov sogar **BLUE-Schätzer** (*Best Linear Unbiased Estimators*), d.h. es sind die besten erwartungstreuen Schätzer, die als lineare Funktion der Y_i - Werte dargestellt werden können. Zu den besten Schätzern macht sie die Eigenschaft, unter allen Konkurrenten die geringste Varianz (also die höchste Präzision) zu besitzen.

Ist außerdem die Normalverteilung der Residuen gegeben, dann können zu den Regressionskoeffizienten auch Konfidenzintervalle berechnet und Signifikanztests durchgeführt werden (siehe unten).

In dieser Situation sind die OLS-Schätzer übrigens identisch mit den Schätzern nach dem Maximum-Likelihood-Prinzip.

In der SPSS-Ausgabe finden Sie die Schätzer in der mit **Regressionskoeffizient B** bezeichneten Spalte der **Koeffizienten**-Tabelle, wobei b_0 in der **(Konstante)** beschrifteten Zeile zu finden ist. Für unser Beispiel erhalten wir:

$$b_0 = -112{,}767$$
$$b_1 = 1{,}041$$

Das Gewicht steigt also in unserer Stichprobe pro Zentimeter Größe um ca. 1 kg an.

1.3.3 Standardisierte Regressionskoeffizienten

Bei der Regression des *standardisierten*, d.h. auf den Mittelwert 0 und die Varianz 1 gebrachten Kriteriums auf den standardisierten Regressor resultiert der **standardisierte Regressionskoeffizient**. Er ist auch als **Beta-Koeffizient** bekannt und im Fall der bivariaten Regression mit der Produktmomentkorrelation identisch, was ihn zu einem Maß für die **Effektstärke** macht (vgl. Abschnitte 1.5 und 1.8). Der standardisierte Regressionskoeffizient wird von SPSS automatisch berechnet und in der **Beta**-Spalte der **Koeffizienten**-Tabelle präsentiert. Einen standardisierten Ordinatenabschnitt protokolliert SPSS nicht, weil im Modell für standardisierte Variablen notwendigerweise der Ordinatenabschnitt 0 resultiert (siehe Gleichung (4)).

Aus der Bezeichnung *Beta-Koeffizient* resultiert ein Bezeichnungsproblem, weil wir den griechischen Buchstaben Beta bereits für die Populationsparameter zu den unstandardisierten Variablen verwenden. Wir behelfen uns mit dem folgenden Symbol für die Stichprobenschätzung des Beta-Koeffizienten zum Regressor X:

$$\tilde{b}_1$$

Er steht in folgender Beziehung zum korrespondierenden Regressionskoeffizienten b_1 für die unstandardisierten Variablen:

$$\tilde{b}_1 = b_1 \frac{s_X}{s_Y}$$

Dabei stehen:

- s_X für die geschätzte Standardabweichung von X
- s_Y für die geschätzte Standardabweichung von Y

Lässt man für die Beispieldaten über den Subdialog **Statistiken** der Regressionsprozedur die deskriptiven Statistiken

Deskriptive Statistiken

	Mittelwert	Std.-Abweichung	N
Gewicht (in kg)	69,02	12,225	332
Größe (in cm)	174,70	9,167	332

ausgeben, dann kann man die angegebene Formel für den standardisierten Regressionskoeffizienten leicht nachrechnen:

$$0{,}7801 \approx 1{,}041 \frac{9{,}167}{12{,}225}$$

Sind reale Maßeinheiten (wie z. B. cm und kg) vorhanden, dann vermittelt das unstandardisierte Regressionsgewicht einen besseren Eindruck vom Effekt eines Regressors als die standardisierte Variante. Beim Standardisieren werden nämlich die Maßeinheiten eliminiert. Allerdings bieten der Regressionskoeffizient und der Korrelationskoeffizient keine konkurrierenden, sondern sich ergänzende Informationen:

- Der Regressionskoeffizient informiert über die Veränderung der mittleren Kriteriumsausprägung bei einer Steigerung des Regressors um eine Einheit und kann z.B. zur Planung von Interventionen benutzt werden.
- Der Korrelationskoeffizient informiert über den Anteil der Kriteriumsvarianz, die vom Regressor erklärt werden kann (siehe Abschnitt 1.5). Er hängt vom Regressionskoeffizienten, von der in einer Studie realisierten Varianz des Regressors und von der Residualvarianz ab.

Bei Variablen mit „einheitsfreier" Messung (z. B. soziales Ansehen, Kreativität) ist der *Betrag* des unstandardisierten Regressionskoeffizienten eventuell nur schwer zu interpretieren, sodass man sich vor allem dafür interessiert, welche mittlere Änderung des Kriteriums in Standardabweichungseinheiten resultiert, wenn der Regressor um eine Standardabweichung steigt. Es darf aber nicht vergessen werden, dass der standardisierte Regressionskoeffizient nicht nur von der Wirkung des Regressors abhängt, sondern auch von der in einer Studie realisierten Varianz des Regressors und von der Residualvarianz. In zwei Populationen bzw. Studien können für einen Regressor trotz identischer Wirkung stark abweichende standardisierte Regressionskoeffizienten resultieren.

1.3.4 Empirische Residuen und geschätzte Residualvarianz

Die empirischen Fehler $(y_i - \hat{y}_i)$ stellen Schätzer der (nicht beobachtbaren) Modellresiduen ε_i dar. Weil sie in den folgenden Ausführungen häufig auftreten, wollen wir eine bequeme Abkürzung vereinbaren:

$$e_i := y_i - \hat{y}_i$$

In der inhomogenen Regression (*mit* β_0 in der Regressionsgleichung) haben die geschätzten Werte $\hat{y}_i = b_0 + b_1 x_i$ (auf Basis der KQ-Schätzer b_0 und b_1) denselben Mittelwert wie die beobachteten *Y*-Werte:

$$\bar{\hat{y}} = \bar{y} \tag{5}$$

Folglich haben die *N* in der Stichprobe beobachteten empirischen Residuen e_i stets den Mittelwert 0:

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) = \underbrace{\frac{1}{N}\sum_{i=1}^{N} y_i}_{=\bar{y}} - \underbrace{\frac{1}{N}\sum_{i=1}^{N} \hat{y}_i}_{=\bar{\hat{y}}=\bar{y}} = 0$$

Für die Residualvarianz σ_ε^2, also für den dritten Modellparameter neben β_0 und β_1, wird im Fall der inhomogenen Regression der folgende Schätzer berechnet:

$$\hat{\sigma}_\varepsilon^2 := \frac{1}{N-2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 = \frac{1}{N-2}\sum_{i=1}^{N} e_i^2 \tag{6}$$

In der SPSS-Ausgabe finden Sie $\hat{\sigma}_\varepsilon^2$ als mittlere Quadratsumme der Residuen in der **ANOVA**-Tabelle. Wir erhalten den Wert 58,644. Die Wurzel aus der geschätzten Fehlervarianz (= 7,658) findet sich unter der Bezeichnung **Standardfehler des Schätzers** in der Tabelle **Modellzusammenfassung**.

1.4 Signifikanztests und Konfidenzintervalle zu den Regressionskoeffizienten

Die Schätzer b_0 und b_1 hängen von den *Y*-Variablen ab und sind daher ebenfalls Zufallsvariablen, die bei jeder Durchführung einer Studie andere Werte annehmen. Zum Glück ist ihre Verteilung *unter den Voraussetzungen der Regressionsanalyse* bekannt, sodass aufgrund einer einmaligen Durchführung der Studie wesentliche Aussagen über die eigentlich interessierenden Populationsparameter β_0 und β_1 möglich sind.

Da wir gerade mit den KQ-Schätzern vertraut geworden sind, setzen wir die Behandlung ihrer Eigenschaften bei gültigen Voraussetzungen fort und schieben die Modelldiagnose noch etwas auf.

1.4.1 Inferenzstatistische Beurteilung von b_1

1.4.1.1 Quadratsummenzerlegung und F-Test

Die totale Quadratsumme QST (mit N - 1 Freiheitsgraden):

$$\mathrm{QST} := \sum_{i=1}^{N} (y_i - \bar{y})^2$$

lässt sich im Modell der bivariaten inhomogenen Regression (*mit* Achsenabschnitt β_0) additiv zerlegen in einen durch die Regression erklärten Anteil QSR (mit einem Freiheitsgrad):

$$\mathrm{QSR} := \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2$$

und die unerklärte Fehlerquadratsumme QSF (mit N - 2 Freiheitsgraden):

$$\mathrm{QSF} := \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} e_i^2$$

Es gilt also:

$$\sum_{i=1}^{N} (y_i - \bar{y})^2 = \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \qquad (7)$$

SPSS liefert die Quadratsummenzerlegung in der **ANOVA**-Tabelle:

$$\begin{array}{r} 30113{,}460 \\ +\ 19352{,}663 \\ \hline = 49466{,}123 \end{array}$$

Unter der (ungerichteten, zweiseitigen) Nullhypothese

$$H_0: \beta_1 = 0$$

mit der zugehörigen Alternativhypothese

$$H_1: \beta_1 \neq 0$$

ist der Quotient

$$\mathrm{F} := \frac{\frac{\mathrm{QSR}}{1}}{\frac{\mathrm{QSF}}{N-2}}$$

bei Gültigkeit aller Voraussetzungen F-verteilt mit *einem* Zählerfreiheitsgrad sowie (N - 2) Nennerfreiheitsgraden und erlaubt damit eine Beurteilung der Nullhypothese. SPSS berichtet die Stichprobenausprägung dieser Prüfgröße zusammen mit der Quadratsummenzerlegung in der **ANOVA**-Tabelle. Wir erhalten den Wert 513,492, dessen Überschreitungswahrscheinlichkeit unter der Nullhypothese kleiner als 0,001 ist, sodass die Nullhypothese abzulehnen ist, sofern die noch ausstehende Prüfung der Voraussetzungen ohne Beanstandungen verläuft

1.4.1.2 Zwei- und einseitiger t-Test

Für die ungerichtete (zweiseitige) Nullhypothese zu β_1 kann zum eben beschriebenen F-Test ein äquivalenter t-Test hergeleitet werden. Die folgende Prüfgröße

$$t := \frac{b_1}{\hat{\sigma}_\varepsilon}\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2} = \frac{b_1}{\hat{\sigma}_{b_1}} \quad \text{mit } \hat{\sigma}_{b_1} := \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}}$$

ist bei Gültigkeit aller Voraussetzungen unter der H_0 t-verteilt mit N - 2 Freiheitsgraden, wobei $\hat{\sigma}_\varepsilon$ die Wurzel aus dem Fehlervarianzschätzer $\hat{\sigma}_\varepsilon^2$ und $\hat{\sigma}_{b_1}$ der gleich vorzustellende geschätzte Standardfehler von b_1 ist. Der t-Wert steht in folgender Beziehung zum oben definierten F-Wert:

$$|t| = \sqrt{F}$$

In der SPSS-Ausgabe erscheint der t-Test in der **Koeffizienten**-Tabelle neben dem zugehörigen Schätzer b_1. Wir erhalten den t-Wert 22,66 $\approx \sqrt{513{,}492}$, der dieselbe Überschreitungswahrscheinlichkeit besitzt wie der zugehörige F-Wert.

Man kann auch einen **einseitigen t-Test** zu β_1 durchführen, z. B. mit dem Hypothesenpaar:

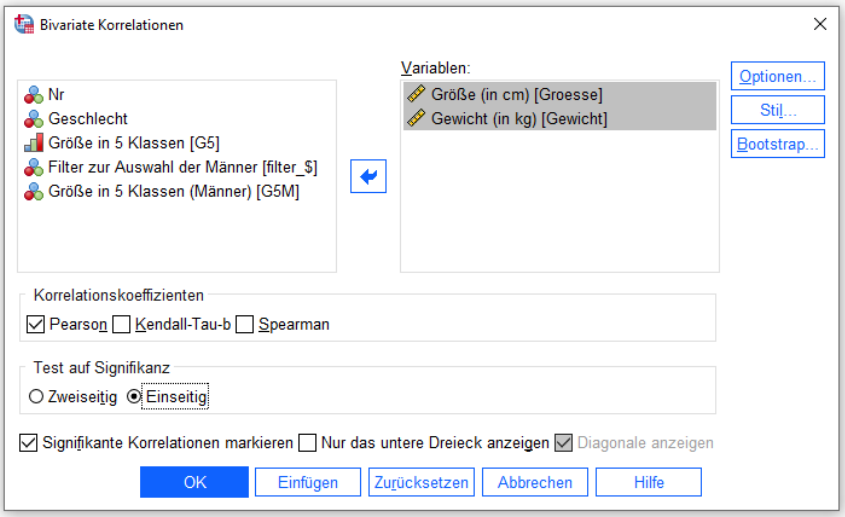
$$H_0: \beta_1 \le 0 \text{ versus } H_1: \beta_1 > 0$$

In diesem Fall muss die von SPSS ausgegebene empirische Überschreitungswahrscheinlichkeit *halbiert* werden, bevor sie mit der akzeptierten α-Fehlerwahrscheinlichkeit (meist 0,05) verglichen wird. Außerdem darf die einseitige Nullhypothese natürlich nur dann verworfen werden, wenn der geschätzte Regressionskoeffizient das (vor der Datenerhebung) vorhergesagte Vorzeichen hat.

Äquivalent zum Test für den Regressionskoeffizienten β_1 ist übrigens der Test zur **Pearson-Korrelation** zwischen *Y* und *X*. In SPSS können Sie die Pearson-Korrelation samt Test über den Menübefehl

Analysieren > Korrelation > Bivariat

anfordern, wobei Sie diesmal zwischen ein- und zweiseitiger Testung wählen dürfen, z. B.:



Die von SPSS beim einseitigen Test ausgegebene empirische Überschreitungswahrscheinlichkeit

Korrelationen

		Größe (in cm)	Gewicht (in kg)
Größe (in cm)	Korrelation nach Pearson	1	,780 **
	Signifikanz (1-seitig)		,000
	N	332	332
Gewicht (in kg)	Korrelation nach Pearson	,780 **	1
	Signifikanz (1-seitig)	,000	
	N	332	332

**. Die Korrelation ist auf dem Niveau von 0,01 (1-seitig) signifikant.

ist genau halb so groß wie die im Rahmen der Regressionsanalyse berechnete zweiseitige Variante:

$$1{,}5837\text{E-}69 \approx \frac{3{,}1675\text{E-}69}{2}$$

Im Beispiel wird sowohl die einseitige als auch die zweiseitige Nullhypothese verworfen. Bei einem schwachen Effekt und/oder einer kleinen Stichprobe kann aber die bessere Power des einseitigen Tests durchaus für unterschiedliche Entscheidungen sorgen.

Mit Hilfe der **verallgemeinerten t-Test** - Prüfgröße (ebenfalls mit N - 2 Freiheitsgraden)

$$t := \frac{b_1 - \beta_1^{(H_0)}}{\hat{\sigma}_{b_1}}$$

lassen sich (zwei- oder einseitige) Tests zu beliebigen Nullhypothesenbehauptungen $\beta_1^{(H_0)}$ durchführen. Im Anwendungsbeispiel kann zum Testproblem

$$H_0: \beta_1 \leq 1 \quad \text{versus} \quad H_1: \beta_1 > 1$$

aus den oben beschriebenen SPSS-Ausgaben die Prüfgröße

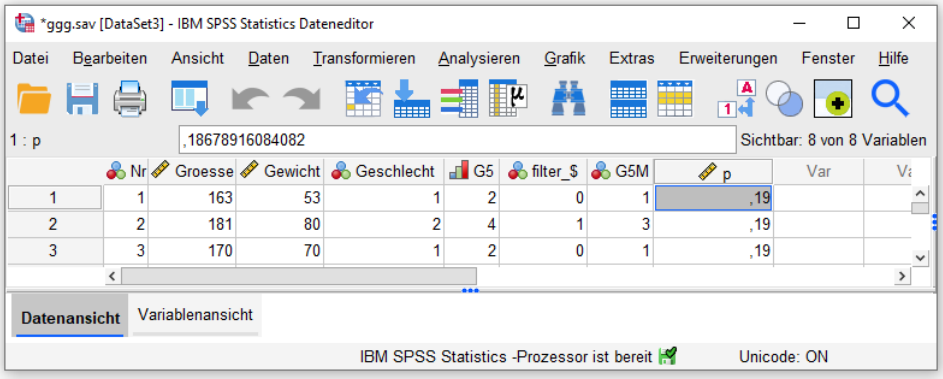
$$\frac{1{,}041 - 1{,}0}{0{,}046} = 0{,}891$$

berechnet werden. Bei einseitiger Testung und 330 Freiheitsgraden (N - 2) ergibt sich eine Überschreitungswahrscheinlichkeit von 0,187, sodass die Nullhypothese das Rennen macht.

Zur Berechnung der Überschreitungswahrscheinlichkeit taugt in SPSS das folgende COMPUTE-Kommando unter Verwendung der Funktion CDF.T (*cumulative distribution function t*):

```
compute  p = 1 - cdf.t(0.891, 330).
execute.
```

Man erhält in der Arbeitsdatei eine neue Spalte mit dem gesuchten p-Wert.



1.4.1.3 Standardfehler

In der Definitionsgleichung zur Prüfgröße des t-Tests ist uns schon der geschätzte **Standardfehler** $\hat{\sigma}_{b_1}$ zum Regressionskoeffizienten b_1 begegnet:

$$\hat{\sigma}_{b_1} := \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}}$$

Er schätzt die Standardabweichung der Verteilung von b_1 um den Erwartungswert β_1 und spielt eine wichtige Rolle in der Inferenzstatistik. Je kleiner der Standardfehler $\hat{\sigma}_{b_1}$ ist, desto ...

- größer ist die Power der Signifikanztests über b_1 (siehe Abschnitt 1.4.1.2)
- kleiner ist das Vertrauensintervall für β_1 (siehe Abschnitt 1.4.1.4)

Mit der Definition

$$s_X := \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

lässt sich der geschätzte Standardfehler $\hat{\sigma}_{b_1}$ zum Regressionskoeffizienten b_1 auch so schreiben:

$$\hat{\sigma}_{b_1} := \frac{\hat{\sigma}_\varepsilon}{\sqrt{N}\, s_X}$$

Nun sieht man besonders deutlich, welche Faktoren auf $\hat{\sigma}_{b_1}$ einwirken:

- Offenbar wächst der Standardfehler mit der Residualvarianz des Modells,
- während sich die Standardabweichung der *X*-Werte
- und der Stichprobenumfang dämpfend auswirken.

Werden bei einer Studie die Regressorwerte tatsächlich geplant (vgl. Abschnitt 1.2.3), sollte man möglichst unterschiedlich *X*-Werte wählen. Untypische und daher sehr einflussreiche Fälle (siehe Abschnitt 3.1.3) sind aber zu vermeiden. Außerdem wirkt sich natürlich der mit dem Stichprobenumfang steigende Informationsgehalt positiv aus.

In der SPSS-Ausgabe erscheint der Standardfehler neben dem zugehörigen Schätzer b_1. In unserem Beispiel resultiert zum Schätzer 1,041 ein Standardfehler von 0,046.

1.4.1.4 Konfidenzintervalle

Mit Hilfe des im Abschnitt 1.4.1.3 vorgestellten Standardfehlers $\hat{\sigma}_{b_1}$ zum Steigungskoeffizienten b_1 lässt sich ein **Konfidenzintervall** für β_1 berechnen. Das folgende Intervall

$$[b_1 - t_{1-\alpha/2;N-2}\,\hat{\sigma}_{b_1};\, b_1 + t_{1-\alpha/2;N-2}\,\hat{\sigma}_{b_1}]$$

enthält bei Gültigkeit aller Voraussetzungen mit einer Wahrscheinlichkeit von $(1 - \alpha)$ den wahren Steigungskoeffizienten β_1. Dabei ist $t_{1-\alpha/2;N-2}$ das $(1 - \alpha/2)$ - Quantil der t-Verteilung mit $(N - 2)$ Freiheitsgraden (= kritischer Wert beim zweiseitigen Signifikanztest zum Niveau α). In unserem Beispiel ($N = 332$, $\alpha = 0{,}05$) erhalten wir als relevantes Quantil der t-Verteilung

$$t_{0{,}975;330} \approx 1{,}967$$

und als Vertrauensintervall für β_1:

$$[1{,}041 - 1{,}967 \cdot 0{,}046;\ 1{,}041 + 1{,}967 \cdot 0{,}046] \approx [0{,}951;\ 1{,}131]$$

SPSS liefert die 95% - Vertrauensintervalle zu den Regressionskoeffizienten (aufgrund unserer Zusatzanforderung in der **Statistiken**-Subdialogbox, siehe Abschnitt 1.3) in der **Koeffizienten**-Tabelle und findet dabei für b_1 aufgrund der höheren Rechengenauigkeit ein leicht abweichendes Ergebnis.

Der *zweiseitige* Signifikanztest zum Regressionskoeffizienten b_1 (siehe Abschnitt 1.4.1.2) verwirft seine Nullhypothese übrigens genau dann, wenn das 95% - Konfidenzintervall zu b_1 den Wert 0 *nicht* enthält.

Es wird mittlerweile in der methodologischen Literatur zu Recht betont, dass ein Vertrauensintervall bedeutend mehr Information über den Populationsparameter β_1 liefert als ein zweiseitiger Signifikanztest (z. B. Cohen et al. 2003). In Publikationsrichtlinien wird nachdrücklich verlangt, Vertrauensintervalle zu berichten (z. B. APA 2010, S. 34).

Passend zum *einseitigen* Signifikanztest lässt sich ein *einseitiges* Konfidenzintervall konstruieren. Bei

$$H_1: \beta_1 > 0$$

resultiert zu b_1 z. B. das folgende rechtseitig offene, einseitige Vertrauensintervall:

$$[b_1 - t_{1-\alpha;N-2}\,\hat{\sigma}_{b_1}; \infty)$$

In unserem Beispiel ($N = 332$, $\alpha = 0{,}05$) erhalten wir als relevantes Quantil der t-Verteilung

$$t_{0,95;330} \approx 1{,}649$$

und als rechtseitig offenes, einseitiges Vertrauensintervall zu b_1:

$$[1{,}041 - 1{,}649 \cdot 0{,}046; \infty) \approx [0{,}965; \infty)$$

Durch den Verzicht auf eine rechte Schranke erhalten wir eine präzisere linke Schranke.

Der *einseitige* Signifikanztest zum Regressionskoeffizienten b_1 (siehe Abschnitt 1.4.1.2) verwirft seine Nullhypothese übrigens genau dann, wenn das zugehörige einseitige Konfidenzintervall zu b_1 den Wert 0 *nicht* enthält.

1.4.2 Inferenzstatistische Beurteilung von b_0

Bei der inhomogenen Regression lassen sich auch für den Achsenabschnitt (engl.: *intercept*) analog zum Vorgehen beim Steigungskoeffizienten t-Tests durchführen (z. B. zur Hypothese $H_0: \beta_0 = 0$) und Konfidenzintervalle (z. B. zum Risiko $\alpha = 0{,}05$) bestimmen.

Weil β_0 gerade die Modellprognose für $X = 0$ ist, kommt dem Signifikanztest zu b_0 genau dann eine Bedeutung zu, wenn die X-Ausprägung 0 von Interesse ist. Diese Lage lässt sich durch eine Transformation (z. B. Zentrierung) von X herstellen.

Wenn der zweiseitige t-Test trotz einer guten Power (vgl. Abschnitt 1.8) seine Nullhypothese beibehält, könnte das sparsamere homogene Modell gegenüber dem inhomogenen Modell bevorzugt werden. In der Regel wird aber der Ordinatenabschnitt im Modell belassen, sofern nicht theoretische Gründe das homogene Modell nahelegen. Ein praktisches Argument gegen das homogene Modell besteht darin, dass in der SPSS-Ergebnisausgabe zu diesem Modell mit dem Determinationskoeffizienten ein wichtiger Bestandteil schwer zu interpretieren ist siehe Abschnitt 1.6).

In unserem Beispiel signalisiert eine empirische Überschreitungswahrscheinlichkeit unter 0,001, dass der Achsenabschnitt auf jeden Fall im Modell verbleiben muss. Allerdings steht dieses Ergebnis wie alle anderen noch unter dem Vorbehalt der ausstehenden Voraussetzungsprüfung.

1.5 Determinationskoeffizient

Als Maß für die Erklärungsleistung (Effektstärke) eines Modells wird der Determinationskoeffizient R^2 verwendet, der bei einer inhomogenen Regression (*inklusive* Achsenabschnitt β_0) als Quotient aus der erklärten Quadratsumme QSR und der totalen Quadratsumme QST definiert ist:

$$R^2 := \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = \frac{\text{QSR}}{\text{QST}} = \frac{\text{QST} - \text{QSF}}{\text{QST}} = 1 - \frac{\text{QSF}}{\text{QST}} \tag{8}$$

Im Beispiel erhalten wir:

$$R^2 = \frac{30113{,}460}{49466{,}123} = 1 - \frac{19352{,}663}{49466{,}123} \approx 0{,}609$$

Bei der inhomogenen *bivariaten* Regression ist R^2 übrigens identisch mit dem Quadrat der Pearsonschen Stichprobenkorrelation zwischen dem Kriterium und dem Regressor. Diese Korrelation findet sich (unquadriert) in der **Beta**-Spalte der **Koeffizienten**-Tabelle.

1.5.1 Korrigierter Determinationskoeffizient

Der Stichproben-Determinationskoeffizient R^2 überschätzt den determinierten Varianzanteil in der Population, weil die R^2 - Definition die in den Quadratsummen enthaltenen Freiheitsgrade nicht berücksichtigt. Um zu einem nahezu unverzerrten Schätzwert zu gelangen, werden im **korrigierten** bzw. **adjustierten** R_a^2 alle Quadratsummen durch die Anzahl der enthaltenen Freiheitsgrade dividiert, sodass sich im Fall der bivariaten Regression ergibt:

$$R_a^2 := 1 - \frac{\frac{\sum_{i=1}^{N} e_i^2}{N-2}}{\frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}}$$

Für die Beziehung zwischen R_a^2 und R^2 gilt:

$$R_a^2 = 1 - (1 - R^2)\frac{N-1}{N-2} = R^2 - (1 - R^2)\frac{1}{N-2}$$

Wegen $N > 2$ (technische Voraussetzung, siehe Abschnitt 1.2.2.8) und $R^2 \in [0; 1]$ gilt:

- $R_a^2 \leq R^2$
- $R_a^2 < R^2$, falls $R^2 < 1$

Der minimale Wert von R_a^2 ist *negativ* ($-\frac{1}{N-2}$) und wird erreicht, wenn $R^2 = 0$ ist. Negative R_a^2 - Werte sollten auf 0 gesetzt werden.

Dem korrigierten R_a^2 sollte zur Beurteilung der Erklärungsleistung eines Modells regelmäßig der Vorzug gegeben werden. SPSS liefert in der Tabelle **Modellzusammenfassung** für unser Beispiel ($N = 332$) einen korrigierten R_a^2 - Wert von 0,608, den wir der Übung halber nachrechnen:

$$R_a^2 = 1 - \frac{\frac{19352{,}663}{330}}{\frac{49466{,}123}{331}} = 0{,}608$$

Weil die Stichprobe relativ groß und nur ein Prädiktor im Modell ist, findet im Beispiel fast keine Korrektur statt.

In der bivariaten Regression ist der korrigierte Determinationskoeffizient nur bei sehr kleinen Stichproben relevant. In der multiplen Regression ist er von größerem Nutzen aufgrund seiner Tendenz, die Verwendung von zahlreichen Regressoren ohne theoretische Begründung zu bestrafen (siehe Abschnitt 2.6).

1.5.2 Relevanz für die Modellbeurteilung

Die gelegentlich zu lesende Forderung (siehe z. B. Kockläuner 1988, S. 26), bei einem sinnvollen Modell solle das adjustierte R_a^2 mindestens 0,50 betragen, ist *nicht* gerechtfertigt. Ein solches R_a^2 ist keine notwendige Voraussetzung für eine sinnvolle regressionsanalytische Modellierung. Goldberger (1991, S. 177) bemerkt zu Recht:

> Nothing in the CR (Classical Regression) model requires that R^2 be high. Hence a high R^2 is not evidence in favor of the model, and a low R^2 is not evidence against it.

Auch Urban & Mayerl (2018, S. 101ff) argumentieren gegen die Jagd nach hohen Determinationskoeffizienten.

Zur Beschreibung (nicht zur Bewertung) von Populations-Effektstärken in *bivariaten* Regressionsmodellen nennt Cohen (1988, S. 79ff) folgende Orientierungsgrößen für die Verhaltens- und Sozialwissenschaften:

Effektstärke in der Population	erklärter Varianzanteil
klein	0,01
mittel	0,09
groß	0,25

1.5.3 Konfidenzintervall

Wer ein Konfidenzintervall zum Determinationskoeffizienten wünscht bzw. benötigt, wird von SPSS enttäuscht. Von den frei verfügbaren Ersatzlösungen bietet sich das R-Paket **MBESS** (Kelley 2018) an:

- MBESS ist relativ schnell einsetzbar, weil eine R-Installation leicht zu bewerkstelligen und bei vielen Statistik-Anwendern bereits vorhanden ist. Zur Bedienung von R siehe z.B. Baltes-Götz, B. (2017).
- MBESS berechnet das Konfidenzintervall sowohl für das Modell mit einem fixierten Regressor als auch für das Modell mit einem stochastischen Regressor.

Das folgende kommentierte R-Skript enthält die erforderlichen Kommandos:

```
# R-Paket MBESS installieren (nur einmal erforderlich)
install.packages("MBESS")
# R-Paket MBESS laden
library(MBESS)
# Konfidenzintervall für das Modell mit einem fixierten Regressor berechnen
ci.R2(F.value=513.492212, df.1 = 1, df.2 = 330, conf.level = .95, Random.Predictors=FALSE)
# Konfidenzintervall für das Modell mit einem stochastischen Regressor berechnen
ci.R2(F.value=513.492212, df.1 = 1, df.2 = 330, conf.level = .95, Random.Predictors=TRUE)
```

Zur Berechnung der Konfidenzintervalle wird die Funktion **ci.R2** aus dem Paket **MBESS** aufgerufen. Alle dabei erforderlichen Argumente liefert die SPSS-Regressionsprozedur in der **ANOVA**-Tabelle:

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	30113,460	1	30113,460	513,49	,000 [b]
	Nicht standardisierte Residuen	19352,663	330	58,644		
	Gesamt	49466,123	331			

a. Abhängige Variable: Gewicht (in kg)

b. Einflußvariablen : (Konstante), Größe (in cm)

Die R-Kommandos können in einer beliebigen R-Bedienoberfläche (z.B. RGui oder RStudio) oder nach der im Anhang beschriebenen Kopplung von R und SPSS in einem SPSS-Syntaxfenster abgeschickt werden.

Im Beispiel mit der Regression von Gewicht auf Größe erhalten wir für das (streng genommen unangemessene) Modell mit einem fixierten Regressor das 95% - Vertrauensintervall [0,547; 0,658]:

```
$Lower.Conf.Limit.R2
[1] 0.5473722

$Upper.Conf.Limit.R2
[1] 0.6578152
```

Für das Modell mit einem stochastischen Regressor erhalten wir das erwartungsgemäß größere 95% - Vertrauensintervall [0,538; 0,670]:

```
$Lower.Conf.Limit.R2
[1] 0.5384063

$Upper.Conf.Limit.R2
[1] 0.6702115
```

1.6 Besonderheiten bei der homogenen Regression

Bei der nur selten verwendeten homogenen Regression *(ohne* Ordinatenabschnitt) haben die geschätzten Werte $\hat{y}_i = b_1 x_i$ i.a. *nicht* denselben Mittelwert wie die beobachteten *Y*-Werte:

$$\overline{\hat{y}} \neq \overline{y}$$

und infolge dessen gilt i.a. auch *keine* Quadratsummenzerlegung analog zu Gleichung (7):

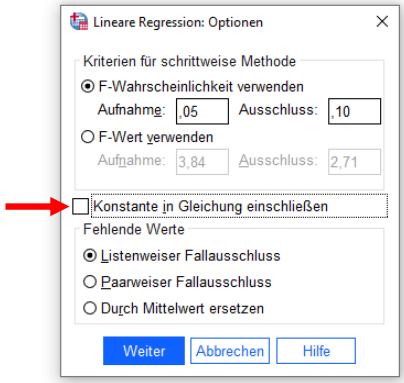
$$\sum_{i=1}^{N}(y_i - \overline{y})^2 \neq \sum_{i=1}^{N}(\hat{y}_i - \overline{\hat{y}})^2 + \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Dies hat wiederum Konsequenzen für den in Gleichung (8) definierten Determinationskoeffizienten, der Werte außerhalb des Intervalls [0, 1] annehmen könnte, sodass man für homogene Modelle ein alternatives Bestimmtheitsmaß bevorzugt, das auch von SPSS verwendet wird:

$$\frac{\sum_{i=1}^{N}\hat{y}_i^{\,2}}{\sum_{i=1}^{N}y_i^{\,2}}$$

Es gibt an, welcher Anteil der Variabilität der *Y*-Werte um den Nullpunkt durch das Regressionsmodell erklärt werden kann.

Bei der Anforderung einer homogenen Regression in SPSS ist in der Subdialogbox **Optionen** die Markierung beim Kontrollkästchen **Konstante in Gleichung einschließen** zu *entfernen*:



Mit unseren Beispieldaten erhalten wir für das hier *sicher untaugliche* homogene Modell folgende Ergebnisse:

Modellzusammenfassung

Modell	R	R-Quadrat [b]	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,990 [a]	,981	,981	9,663

a. Einflußvariablen : Größe (in cm)

b. Bei der Regression durch den Ursprung (Modell ohne konstanten Term) mißt das R-Quadrat den Anteil der Variabilität in der abhängigen Variable durch den Ursprung, der durch Regression erklärt werden kann. Dieses Verfahren KANN NICHT mit dem R-Quadrat bei Modellen verglichen werden, die einen konstanten Term enthalten.

ANOVA[a,b]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	1600106,538	1	1600106,538	17135,469	,000 [c]
	Nicht standardisierte Residuen	30908,712	331	93,380		
	Gesamt	1631015,250 [d]	332			

a. Abhängige Variable: Gewicht (in kg)

b. Lineare Regression durch den Ursprung

c. Einflußvariablen : Größe (in cm)

d. Die Gesamtsumme der Quadrate wurde nicht für die Konstante korrigiert, weil die Konstante für die Regression durch den Ursprung gleich null ist.

Koeffizienten [a,b]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressionskoeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	Größe (in cm)	,397	,003	,990	130,903	,000	,391	,403

a. Abhängige Variable: Gewicht (in kg)

b. Lineare Regression durch den Ursprung

SPSS weist per Fußnote auf die Konsequenzen des homogenen Modells für die Berechnung der Determinationskoeffizienten und der Quadratsummen hin. Es wäre ein eklatanter Fehler, aus dem korrigierten R^2 - Wert von 0,981 für das homogene Modell auf dessen Überlegenheit gegenüber dem inhomogenen Modell zu schließen, das einen korrigierten R^2-Wert von 0,608 aufweist.

Kockläuner (1988, S. 47) kritisiert zu Recht, dass SPSS in der Ausgabe zum homogenen Fall das Bestimmtheitsmaß mit **R-Quadrat** und die Wurzel daraus mit **R** überschreibt. Damit wird der falsche Eindruck geweckt, es handle sich um die (quadrierte) Korrelation zwischen Kriterium und Regressor, was aber im homogenen Fall *nicht* zutrifft.

1.7 Modelldiagnose und -modifikation

In diesem Abschnitt werden von den im Abschnitt 1.2.2 beschriebenen Modellvoraussetzungen behandelt:

- Linearität
- Varianzhomogenität der Residuen
- Normalität der Residuen

In unserem Beispiel zur bivariaten Regression ist eine autoregressive Abhängigkeit der Residuen aufgrund des querschnittlichen Untersuchungsplans (ohne serielle Anordnung der Fälle) praktisch ausgeschlossen. Weil auch keine Stichprobe mit Cluster-Struktur vorliegt, kann die Annahme unkorrelierter Residuen ohne Prüfung akzeptiert werden.

Begründete Zweifel an der Unabhängigkeit der Residuen bestehen bei Zeitreihenstudien sowie bei Stichproben mit Cluster- oder Panelstruktur (siehe Abschnitt 1.2.2.6)

1.7.1 Linearität

Es ist zu prüfen, ob für den Erwartungswert von Y_i zu jedem in der Studie realisierten x_i-Wert tatsächlich gilt:

$$E(Y_i) = \beta_0 + \beta_1 x_i$$

In der Terminologie von Darlington & Hayes (2017, S. 88) ist dies die *primäre* Annahme des Modells.

1.7.1.1 Diagnose

1.7.1.1.1 Residuen-Plots

Für die wahren Residuen ε_i gilt aufgrund ihrer Definition:

$$E(\varepsilon_i) = 0$$

Die wahren Residuen ε_i werden durch die empirischen Residuen

$$e_i := y_i - b_0 - b_1 x_i$$

geschätzt, und bei gültiger Linearitätsannahme gilt auch für die geschätzten Residuen:

$$E(e_i) = 0$$

Zur Überprüfung der Linearität werden wir untersuchen, ob alle geschätzten Residuen e_i zufällig um ihren gemeinsamen Erwartungswert 0 variieren. Der *globale* Mittelwert aus allen geschätzten Residuen ist bei der üblichen inhomogenen Regression in jedem Fall gleich 0 (siehe Abschnitt 1.3.4). Wir müssen daher anhand der Stichprobendaten überprüfen, ob *für alle realisierten x_i-Werte* die zugehörigen e_i zufällig um 0 variieren. Daher lassen wir von SPSS das Streuungsdiagramm mit den (e_i, x_i) - Paaren erzeugen. Hier sollten sich alle Residuen in einem Streifen um die Null-Lage aufhalten, wenn neben der Linearitäts- auch die Varianzhomogenitätsannahme erfüllt ist:

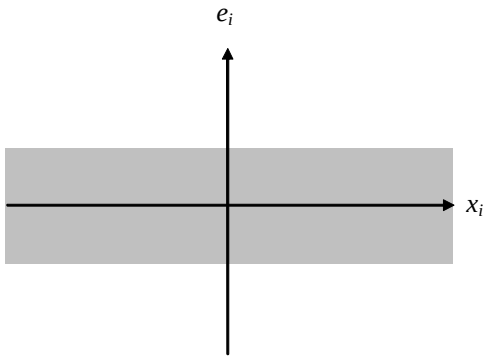
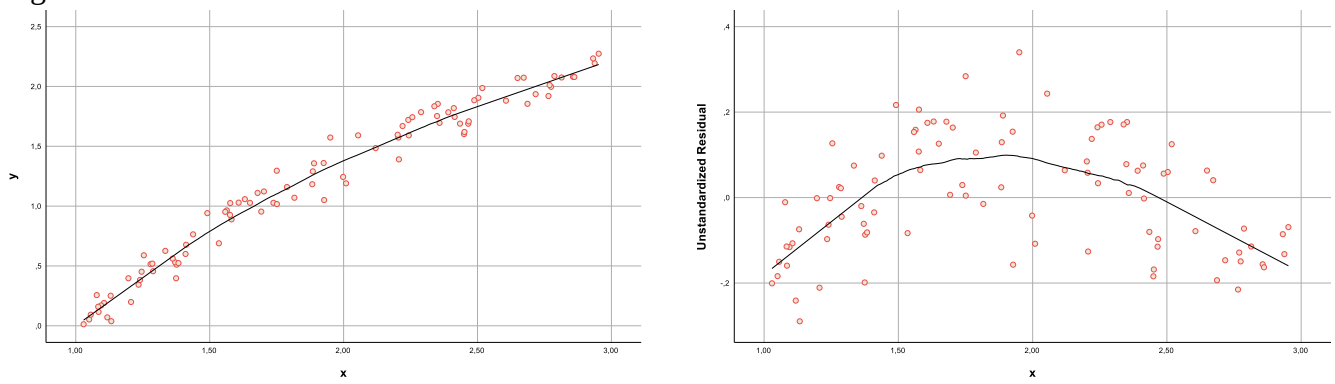


Abbildung 3: Modellkonformer Erwartungsbereich für die e_i - Werte im Plot gegen die x_i - Werte

Grundsätzlich lässt sich die Linearitätsannahme bereits mit dem Plot der abhängigen gegen die unabhängige Variable überprüfen, doch sind Verstöße im Residuenplot oft besser zu erkennen, wie das folgende Beispiel zeigt:

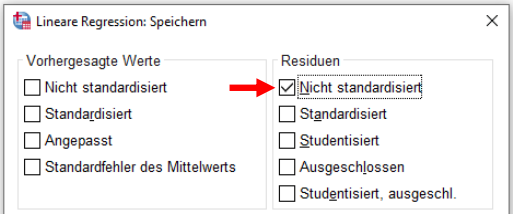


Der linke Plot zeigt ein (X, Y) - Stichprobenstreudiagramm aus einer künstlichen Population mit dem folgenden wahren Modell:

$$Y = 2\ln(X) + \varepsilon,\ \varepsilon \sim N(0, \sigma_\varepsilon^2)$$

Im Diagramm rechts daneben werden die Residuen e_i aus der fehlspezifizierten linearen Regression von Y auf X gegen den Regressor geplottet, wobei die Verletzung der Linearitätsannahme besser zu erkennen ist. Wie man derartige Diagramme erstellt, ist gleich zu erfahren.

Aus didaktischen Gründen beschränken wir uns vorläufig auf eine Darstellung der **unstandardisierten Residuen** $(y_i - \hat{y}_i)$, die wir im Manuskript mit e_i bezeichnen, gegen den Regressor.[1] Dazu fordern wir in der Subdialogbox **Speichern** der Regressionsprozedur die Erzeugung einer neuen SPSS-Variablen mit den unstandardisierten Residuen an:



Im Ausgabefenster von SPSS findet man den Namen der neu in die Arbeitsdatei aufgenommenen Variablen im Abschnitt **Hinweise** der Regressionsausgabe, der per Voreinstellung ausgeblendet ist:

Hinweise

Erstellte oder modifizierte Variablen	RES_1	Unstandardized Residual

Unter Verwendung dieser Variablen erstellt man z. B. über

Grafik > Klassische Dialogfelder > Streu-/Punktdiagramm > Einfaches Streudiagramm

das gewünschte Diagramm. Für unsere Beispieldaten zur Regression von Gewicht auf Größe resultiert:

[1] Wir werden bald ein für unsere Diagnosezwecke sehr gut geeignetes, bequem von der SPSS-Regressionsprozedur erhältliches Streudiagramm unter Verwendung von aufwändiger geschätzten Residuen kennenlernen (siehe Abschnitt 1.7.3.2).

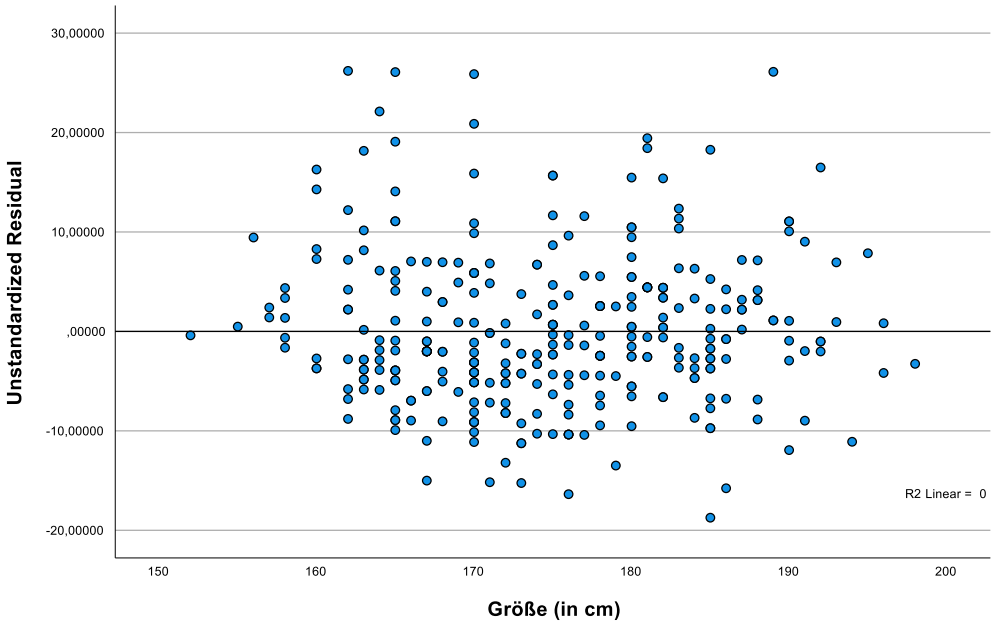


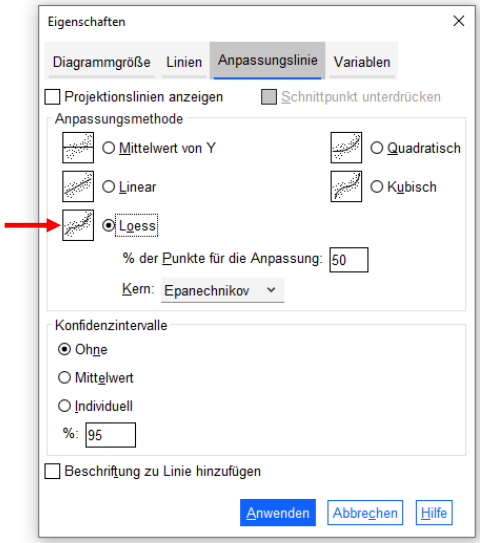
Abbildung 4: Unstandardisierte Residuen versus Prädiktor für die Regression von GEWICHT auf GROESSE

Die Regressionsgerade lässt sich im Diagrammeditor über den Menübefehl

Elemente > Anpassungslinie bei Gesamtsumme

oder den Symbolschalter einfügen. Sie zeigt beim Ordinatenabschnitt und beim Steigungskoeffizienten jeweils eine modellgemäße 0.

Auf den ersten Blick scheinen die Residuen zu allen *X*-Werten um den korrekten bedingten Erwartungswert 0 zu schwanken. Vor einem endgültigen Urteil über die Linearitätsannahme wählen wir zur **Anpassungslinie** per **Eigenschaften**-Fenster die nonparametrische **Anpassungsmethode Loess** (***lo****cal regr****ess****ion*):



Man erhält eine Regressionslinie mit lokalen Anpassungen an den Punkteschwarm.

In unserem Beispiel zeigen sich nun doch Hinweise auf mögliche Schwächen des Modells:

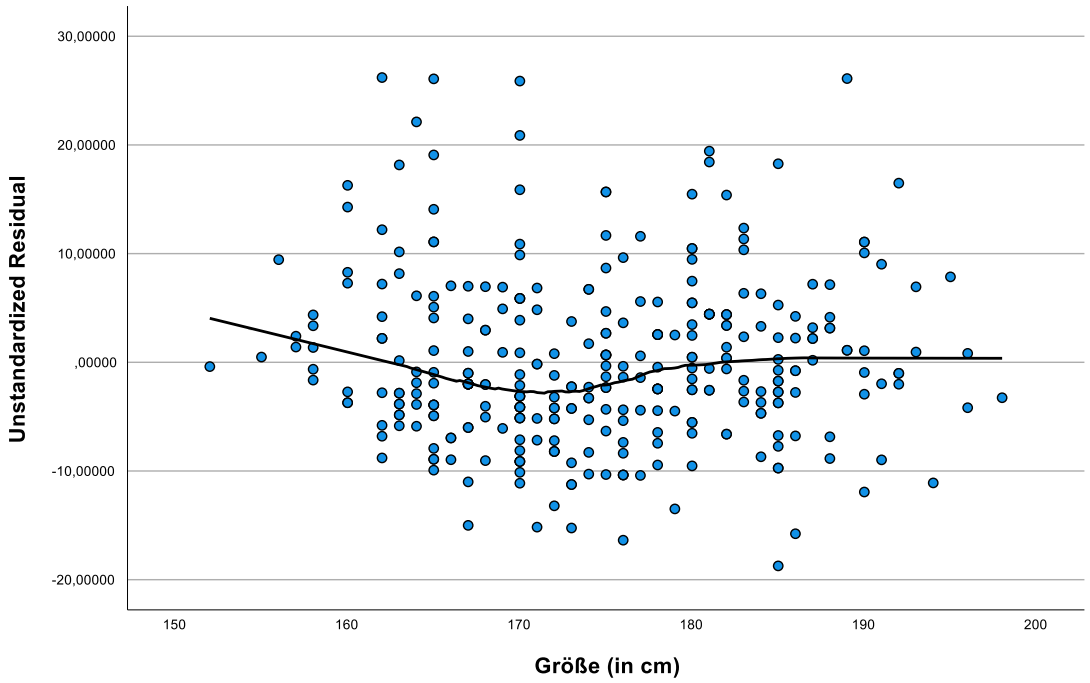


Abbildung 5: Unstandardisierte Residuen versus Prädiktor für die Regression von GEWICHT auf GROESSE mit lokal optimierter Anpassungslinie

Um die Abweichungen von der modellgemäßen Ideallinie einer für alle x_i - Werte konstanten Erwartung von 0 besser zu verstehen, betrachten wir erneut das Streuungsdiagramm für die Originaldaten, diesmal mit eingezeichneter LOESS-Linie und unterschiedlichen Markierungen für Frauen und Männer:

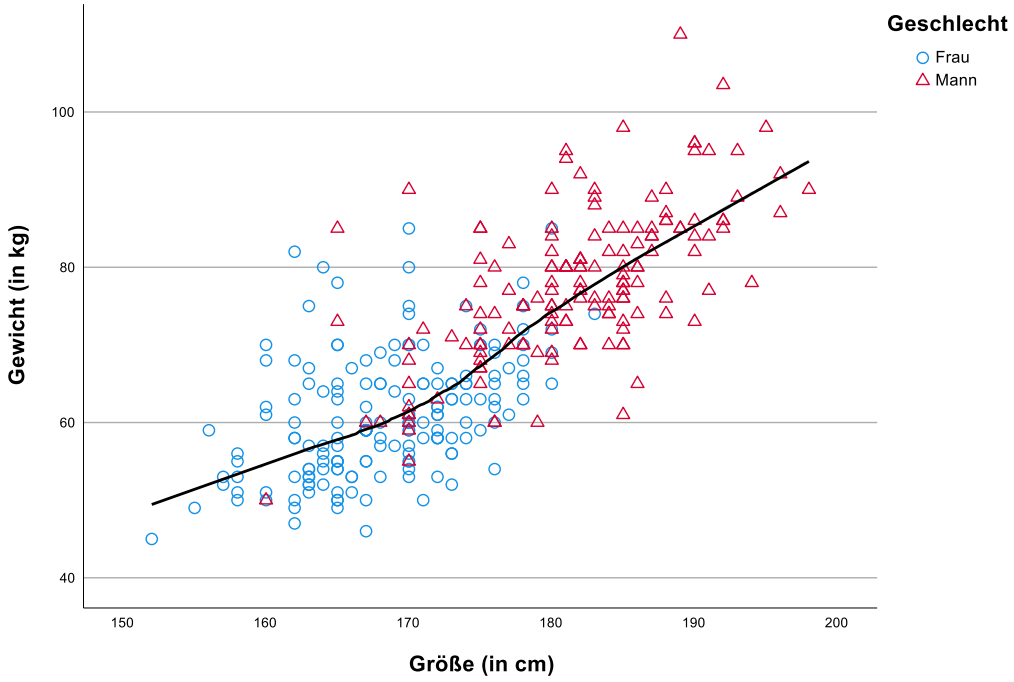


Abbildung 6: Regression von GEWICHT auf GROESSE mit geschlechtsspezifischen Symbolen und lokal optimierter Anpassungslinie

Es drängt sich die Vermutung auf, dass die Regression von Gewicht auf Größe bei Frauen (oder bei kleineren Personen) flacher verläuft als bei den Männern (oder bei größeren Personen). Wenn sich diese Hypothese erhärten lässt, dann liegt bei unserem Modell ein **Spezifikationsfehler** vor.

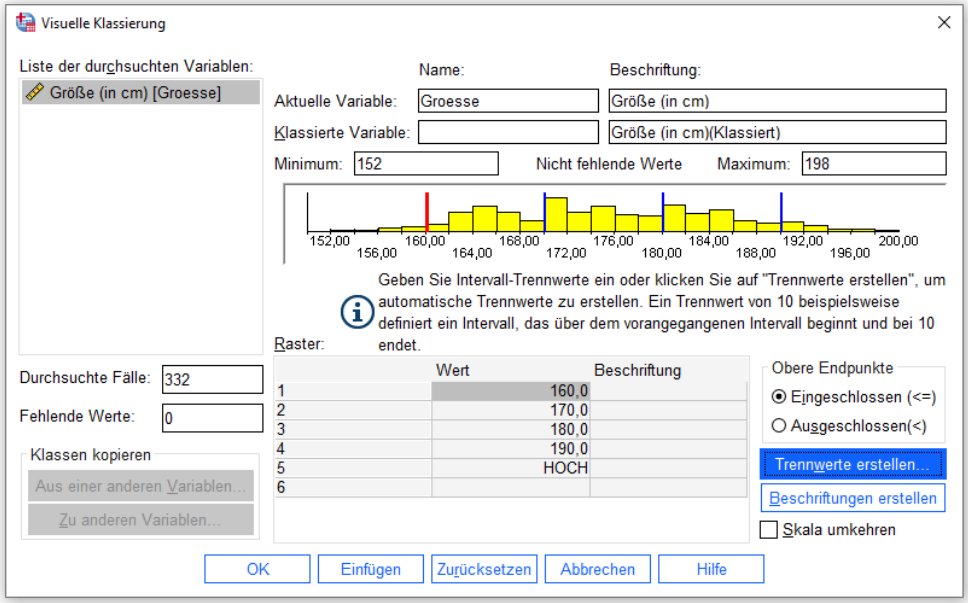
Es ist generell anzumerken, dass wir bei der Suche nach Ursachen für eine mutmaßlich verletzte Linearitätsannahme *explorativ* arbeiten, so dass die generierten Hypothesen nur an einer unabhängigen Stichprobe überzeugend geprüft werden können.

1.7.1.1.2 Linearitätstest

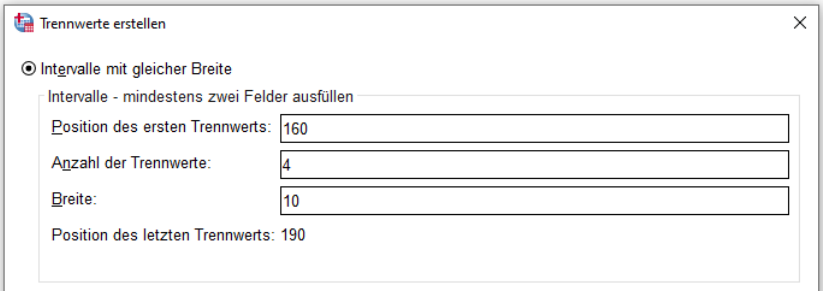
Für die bivariate Regression mit einem metrischen Regressor bietet SPSS in der Prozedur MEANS einen Linearitätstest an, wobei allerdings gruppierte Daten mit hinreichend stark besetzten Regressorkategorien benötigt werden. Um den Test auf die Regression von Gewicht auf Größe anwenden zu können, wurde über den Menübefehl

Transformieren > Visuelle Klassierung

eine Variante des Regressors mit 5 Ausprägungen basierend auf 5 gleich breiten Intervallen erstellt (SPSS-Variablenname: G5):



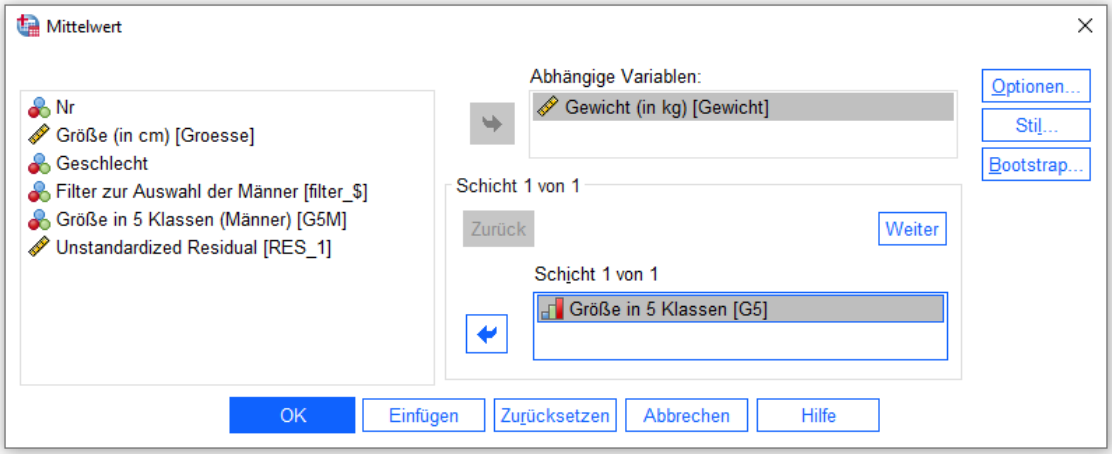
Die Trennwerte lassen sich nach einem Mausklick auf den Schalter **Trennwerte erstellen** hier definieren:



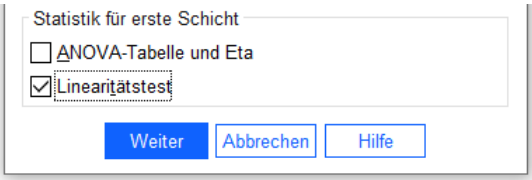
Nach dem Menübefehl

Analysieren > Mittelwerte vergleichen > Mittelwerte

geben wir im folgenden Dialog das Gewicht als **abhängige** und die vergröberte Größenvariante als **unabhängige Variable** an:



Im **Optionen**-Subdialog fordern wir schließlich den **Linearitätstest** an:



Das Testergebnis spricht für eine **Abweichung von der Linearität** (p = 0,023):

ANOVA-Tabelle

			Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Gewicht (in kg) * Größe in 5 Klassen	Zwischen den Gruppen	(Kombiniert)	27715,065	4	6928,766	104,165	,000
		Linearität	27071,550	1	27071,550	406,987	,000
		Abweichung von der Linearität	643,515	3	214,505	3,225	,023
	Innerhalb der Gruppen		21751,057	327	66,517		
	Insgesamt		49466,123	331			

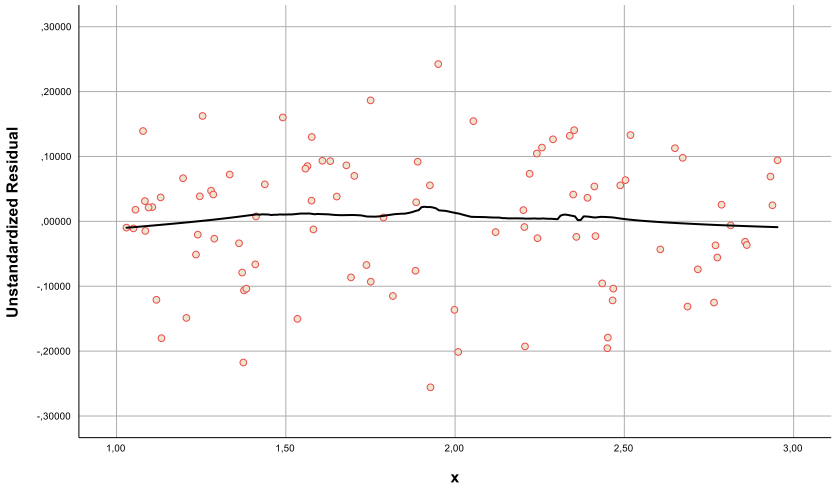
Dieser Befund stützt die Vermutung, die wir aus der Analyse von Streudiagrammen gewonnen haben.

1.7.1.2 Linearitätsdefekte beheben

Die generelle Empfehlung für den Fall einer verletzten Linearitätsannahme kann nur lauten, den Fehler in der Modellspezifikation zu beheben:

- Eventuell ist für die Abhängigkeit des Kriteriums vom Regressor eine **alternative funktionale Form** zu wählen. Viele nichtlineare Zusammenhänge lassen sich durch geeignete Transformationen von Regressor und/oder Kriterium linearisieren. Bei einer nichtlinearen Transformation des Kriteriums ist mit Auswirkungen auf die Varianzhomogenität und die Verteilungsform der Residuen zu rechnen.

  Ersetzt man in dem auf Seite 31 vorgestellten Modell für simulierte Daten den Regressor X durch $\ln(X)$, dann resultiert ein Residuenplot (mit LOESS-Anpassungslinie) ohne Hinweise auf Linearitätsprobleme:



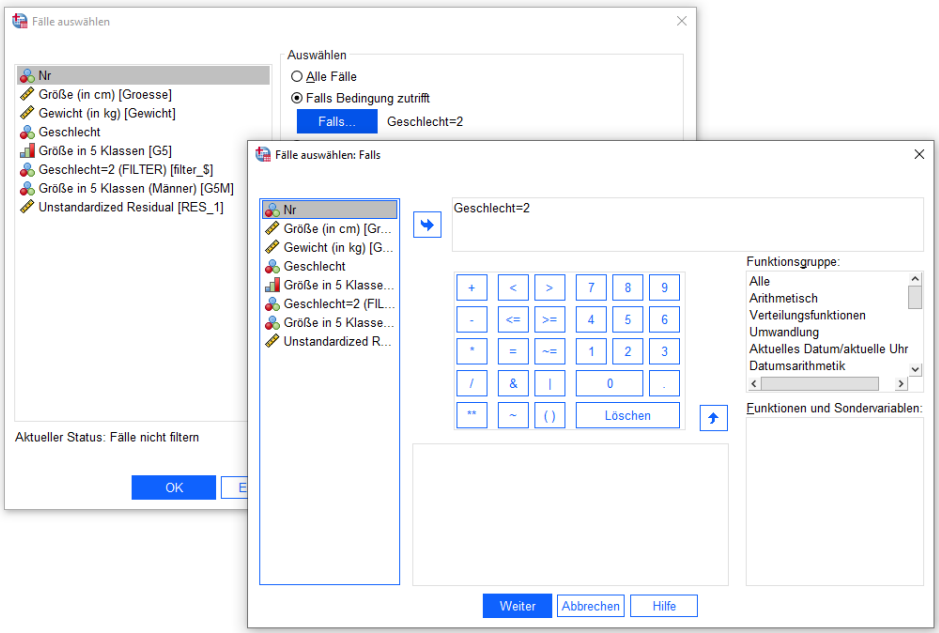
- Möglicherweise muss das Modell um **zusätzliche Regressoren** erweitert werden, womit wir den Bereich der bivariaten Regression verlassen. Das Streuungsdiagramm mit Gewicht, Größe und Geschlecht erlaubt die Hypothese, dass das Geschlecht sowie die Wechselwirkung von Größe und Geschlecht (vertreten durch das Produkt der beiden Variablen) in das Modell aufgenommen werden sollten (siehe Baltes-Götz 2020a zur Moderatoranalyse). Alternativ kann das Streuungsdiagramm aber auch so gedeutet werden, dass die Regression von Gewicht auf Größe mit zunehmender Größe steiler verläuft. Diese Annahme eines positiv beschleunigten Effekts von Größe auf Gewicht lässt sich durch die Aufnahme der quadrierten Größe in die Gleichung modellieren. Es resultiert der wohl wichtigste Vertreter aus der Klasse der polynomischen Regressionsmodelle (siehe Abschnitt 4).

Beachten Sie den Unterschied zwischen der Prüfung eines *vor* der Datenerhebung formulierten Modells (konfirmatorische Forschung) und der Überarbeitung eines Modells während der explorativen Datenanalyse. Bei der explorativen Forschung entstehen Modelle, die nur in einer neuen Stichprobe überzeugend überprüft werden können.

Um die Gewicht-Größe - Daten weiterhin als Beispiel für die *bivariate* Regression mit überzeugender Linearitätsannahme verwenden zu können, schränken wir unsere Stichprobe über den Menübefehl

Daten > Fälle auswählen

mit dem folgenden Dialog



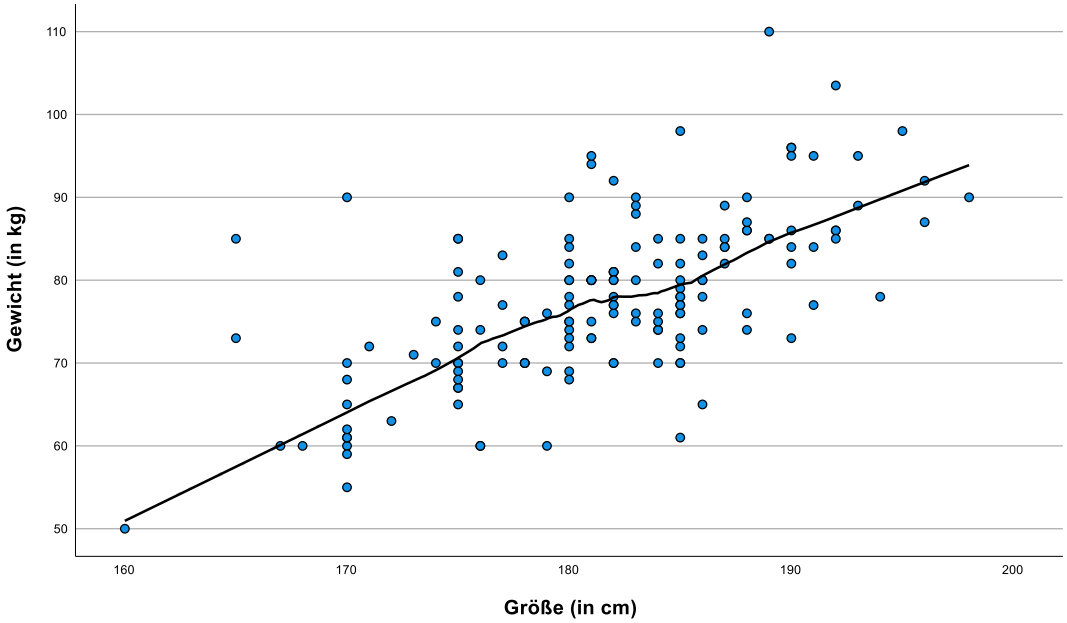
auf die Männer ein (N = 159).

In der Datei **ggg.sav** befindet sich die Variable G5M mit 5 aus der männlichen Größenverteilung abgeleiteten Gruppen. Damit akzeptiert der im Abschnitt 1.7.1.1.2 beschriebene Linearitätstest seine Nullhypothese (p = 0,541):[1]

ANOVA-Tabelle

			Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Gewicht (in kg) * Größe in 5 Klassen (Männer)	Zwischen den Gruppen	(Kombiniert)	6644,030	4	1661,008	29,693	,000
		Linearität	6522,962	1	6522,962	116,607	,000
		Abweichung von der Linearität	121,068	3	40,356	,721	,541
	Innerhalb der Gruppen		8614,734	154	55,940		
	Insgesamt		15258,764	158			

Auch das Streuungsdiagramm mit den Variablen GEWICHT und GROESSE und einer eingezeichneten LOESS-Linie bestätigt für die männliche Teilstichprobe das lineare Modell:



Wir erhalten die folgenden Schätz- und Testergebnisse für die Teilstichprobe der Männer:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,643 [a]	,413	,409	7,552

a. Einflußvariablen : (Konstante), Größe (in cm)

[1] Die Variable G5M wurde durch das folgende RECODE-Kommando erstellt:

```
recode GROESSE (lo thru 170 = 1)  (170 thru 178 = 2)  (178 thru 186 = 3)
               (186 thru 192 = 4)  (192 thru hi =5)  into G5M.
execute.
```

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	6304,375	1	6304,375	110,536	,000 [b]
	Nicht standardisierte Residuen	8954,390	157	57,034		
	Gesamt	15258,764	158			

a. Abhängige Variable: Gewicht (in kg)

b. Einflußvariablen : (Konstante), Größe (in cm)

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.	95,0% Konfidenzintervalle für B	
		Regressionskoeffizient B	Std.-Fehler	Beta			Untergrenze	Obergrenze
1	(Konstante)	-89,005	15,875		-5,607	,000	-120,362	-57,648
	Größe (in cm)	,919	,087	,643	10,514	,000	,746	1,091

a. Abhängige Variable: Gewicht (in kg)

Obwohl nun die funktionale Form der Regression Gewicht auf Größe geklärt ist, werden im Abschnitt 1.7.4 im Zusammenhang mit der Normalverteilungsannahme noch einmal Zweifel aufkommen, ob wirklich für *alle* beobachteten Zufallsvariablen Y_i der Erwartungswert auf der Regressionsgerade liegt.

Die Ergebnisse demonstrieren u.a., dass man ein geschätztes Regressionsmodell nicht über den realisierten X-Wertebereich hinaus generalisieren darf. Ansonsten gelangt man im Beispiel zur „Erkenntnis“, dass bei sehr kleinen Personen ein negatives Körpergewicht zu erwarten sei (siehe Schätzwert für die **Konstante**).

1.7.2 Auswahl der zu prüfenden Residuen

Bislang haben wir uns aus didaktischen Gründen auf die einfachen Residuen e_i $(= y_i - \hat{y}_i)$ beschränkt. Nun werden deren Nachteile diskutiert und Alternativen vorgestellt, die zur anstehenden Prüfung der Homoskedastizität und der Normalität der Residuen etwas besser geeignet sind.

1.7.2.1 Zentrierte Hebelwerte

Für die anschließenden Definitionen wird mit dem **zentrierten Hebelwert** (engl.: *leverage*- oder *hat-value*) ein Begriff benötigt, für den wir uns später im Zusammenhang mit der Diagnose einflussreicher Fälle nochmals interessieren werden. In der bivariaten Regression (mit Ordinatenabschnitt) wird der zentrierte Hebelwert h_i von Fall *i* folgendermaßen definiert:

$$h_i := \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Er wird als relative quadrierte Entfernung des Regressorwerts x_i vom Mittelwert $\bar{x}$ berechnet und quantifiziert die Hebelwirkung eines Falles auf die Schätzungen der Regressionskoeffizienten. Seine Werte liegen zwischen 0 und $1 - \frac{1}{N}$ (Urban & Mayerl 2018, S. 180ff).

Manche Autoren (z. B. Darlington & Hayes 2017, S. 484ff; Fox & Weisberg 2011) verwenden eine abweichende Hebeldefinition, die sich vom zentrierten Hebel sensu SPSS durch die additive Konstante 1/*N* unterscheidet:

$$\tilde{h}_i := h_i + \frac{1}{N}$$

1.7.2.2 Studentisierte Residuen

Die bisher betrachteten unstandardisierten Residuen e_i sind zur Prüfung der Homoskedastizitätsannahme nicht optimal geeignet. Man kann zeigen (siehe z. B. Fahrmeir et al. 2007, S. 107; Snedecor & Cochran, 1980, S. 144 und 151), dass auch *bei Gültigkeit aller Regressionsannahmen* die Varianzen der geschätzten Residuen e_i von den zugehörigen x_i-Werten abhängen. Beachten Sie, dass diese Aussage für die *geschätzten* Residuen e_i gilt, nicht jedoch für die wahren Residuen ε_i. Nach Fox & Weisberg (2011, S. 286) gilt für die Varianz von e_i:[1]

$$\text{Var}(e_i) = \sigma_\varepsilon^2(1 - \frac{1}{N} - h_i)$$

Sie hängt u.a. von dem im Abschnitt 1.7.2.1 definierten zentrierten Hebelwert h_i ab. h_i variiert von 0 bis $1 - \frac{1}{N}$, wobei Fälle mit einem extremen (untypischen) x_i - Wert einen großen Hebelwert und dementsprechend eine kleine Residualvarianz besitzen.

Dividiert man die einfachen Residuen e_i durch ihre jeweils individuell geschätzte Standardabweichung,

$$\tilde{e}_i := \frac{e_i}{\hat{\sigma}_\varepsilon \sqrt{1 - \frac{1}{N} - h_i}}$$

so erhält man die von SPSS als **studentisiert** bezeichneten Residuen, die bei Gültigkeit des Modells die konstante Varianz 1 besitzen und somit z. B. zur Beurteilung der Homoskedastizitätsfrage besser geeignet sind als die einfachen Residuen (siehe Draper & Smith, 1981, Abschnitt 3.7; Kockläuner, 1988, Abschnitt 3.2.1).

Leider werden die Begriffe für Regressionsresiduen uneinheitlich verwendet. Die studentisierten Residuen sensu SPSS werden z. B. von Fahrmeir et al. (2007, S. 108) sowie von Fox & Weisberg (2011, S. 286) als *standardisierte* Residuen bezeichnet.

1.7.2.3 Ausgelassen-studentisierte Residuen

Sollen Residuen *individuell* beurteilt werden (z. B. bei einer Ausreißeranalyse, siehe Abschnitt 3.1.1), dann muss ihre Verteilung bei gültigem Modell bekannt sein. Für die von SPSS als *studentisiert* bezeichneten Residuen ist eine solche Verteilungssausage *nicht* möglich, weil der standardisierende Nenner auch vom Residuum im Zähler abhängt, sodass kein Quotient aus zwei unabhängigen Zufallsgrößen vorliegt. Genau diese Voraussetzung wird aber benötigt, um die Verteilung herleiten zu können. Bei den sogenannten **ausgelassen-studentisierten Residuen**, die in diesem Manuskript mit r_i bezeichnet und in der Regel für Diagnosezwecke bevorzugt werden, ist das Problem auf simple Weise eliminiert. Im Nenner wird eine Schätzung der Residualstandardabweichung verwendet, die ohne Beteiligung der *i*-ten Beobachtung zustande gekommen ist:

---

[1] Bei Fox & Weisberg (2011, S. 286) sieht die Formel für Var(e_i) etwas anders aus, weil dort ein leicht abweichender Hebelbegriff verwendet wird, für den wir im Abschnitt 1.7.2.1 die Bezeichnung $\tilde{h}_i$ eingeführt haben:

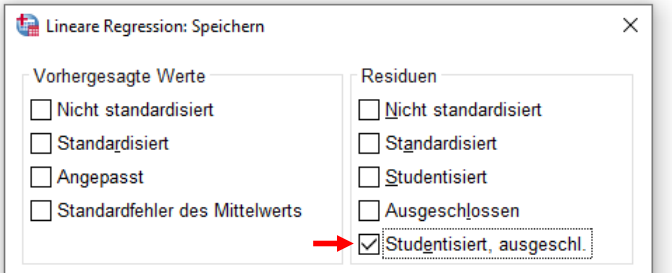
$$\text{Var}(e_i) = \sigma_\varepsilon^2(1 - \tilde{h}_i)$$

$$r_i := \frac{e_i}{\hat{\sigma}_{\varepsilon.i}\sqrt{(1-\frac{1}{N}-h_i)}} \tag{9}$$

Bei einem gültigen Modell mit k Regressoren folgen die r_i - Werte einer t-Verteilung mit N - k - 2 Freiheitsgraden (Fox & Weisberg 2011, S. 287).

In diesem Manuskript wird bewusst von *ausgelassen-studentisierten Residuen* gesprochen in Abgrenzung von der verbreiteten Bezeichnung *studentisierte ausgeschlossen Residuen*. Mit der zuletzt genannten Bezeichnung ist meist eine alternative Definition verbunden (siehe z. B. Eid et al. 2013, S. 683; Fahrmeir et al. 2007, S. 109), wobei im Zähler nicht das einfache, sondern das sogenannte *ausgeschlossene Residuum* steht. Man erhält es als Differenz aus dem Beobachtungswert von Fall *i* und der Prognose unter Verwendung eines Modells, das ohne Beteiligung von Fall *i* geschätzt worden ist.

SPSS berechnet r_i - Werte nach Formel (9) und schreibt diese in eine neue Variable der Arbeitsdatei, wenn Sie in der **Speichern**-Subdialogbox der Regressionsprozedur das Kontrollkästchen **Studentisiert, ausgeschl.** markieren:



Die neue Variable erhält (falls noch nicht vergeben) den Namen SDR_1 und die Variablenbeschriftung *Studentized Deleted Residual*. SPSS verwendet also die übliche Bezeichnung, weicht aber durch eine Berechnung nach Formel (9) leicht von der üblichen Definition ab.

Darlington & Hayes (2017, S. 488) bezeichnen die durch Formel (9) definierten Residuen als *t-residuals*. Fox & Weisberg (2011, S. 287) sprechen von *studentized residuals*.

Wir werden ab jetzt zur Modellprüfung die ausgelassen-studentisierten Residuen bevorzugen. Im Abschnitt 1.7.1 haben wir uns aus didaktischen Gründen auf die einfachen (unstandardisierten) Residuen beschränkt.

1.7.2.4 Standardisierte Residuen

Die von SPSS als *standardisiert* bezeichneten Residuen sind schlicht durch den Quotienten aus dem einfachen Residuum und dem Standardfehler der Schätzung (siehe Seite 20) definiert:

$$\frac{e_i}{\hat{\sigma}_\varepsilon} \quad \text{mit } \hat{\sigma}_\varepsilon = \sqrt{\frac{1}{N\text{ - }2}\sum_{i=1}^{N} e_i^2}$$

Sie besitzen daher wie die einfachen Residuen auch bei gültigen Modellvoraussetzungen heterogene Varianzen.

1.7.3 Varianzhomogenität der Residuen

Anschließend werden im Abschnitt 1.7.3.1 die möglichen Konsequenzen bei verletzter Homoskedastizitätsannahme behandelt. Danach werden im Abschnitt 1.7.3.2 grafische und statistische Verfahren zur Beurteilung der Homoskedastizität beschrieben.

Wird eine relevante Verletzung der Varianzhomogenität festgestellt, sollte zunächst geprüft werden, ob einzelne *Ausreißer* dafür verantwortlich sind. Eventuell müssen bzw. dürfen diese Fälle von der Analyse ausgeschlossen werden, weil sie nicht zu der interessierenden Population gehören (siehe Abschnitt 3.1). Besteht weiterhin eine ernsthafte Varianzheterogenität, kommen folgende Maßnahmen in Betracht:

- Nichtlineare Transformation des Kriteriums zur Homogenisierung der Fehlervarianzen, wobei auch die Linearität des Modells tangiert wird (siehe Abschnitt 1.7.3.3)
- Verwendung von inferenzstatistischen Verfahren mit Robustheit gegenüber Varianzheterogenität (siehe Abschnitt 1.7.3.4)
- Gewichtete Kleinst-Quadrat - Schätzung (WLS-Regression, siehe Abschnitt 1.7.3.5)

1.7.3.1 Konsequenzen bei verletzter Varianzhomogenität

Auch bei verletzter Varianzhomogenität sind die gewöhnlichen Kleinst-Quadrat-Schätzer immer noch erwartungstreu und konsistent, also durch Steigerung des Stichprobenumfangs auf jede beliebige Präzision zu bringen. Sie haben jedoch nicht mehr die im Abschnitt 1.3 beschriebenen BLUE-Optimalitätseigenschaften, sondern besitzen größere Standardfehler als die bestmöglichen Schätzer (Ryan 1997, S. 61). Dieser Qualitätsverlust tritt auf, weil alle Beobachtungen mit gleichem Gewicht eingehen, obwohl sie mit unterschiedlichen Fehlerstreuungen belastet sind.

Besonders gravierend sind die Auswirkungen auf die Schätzungen der *Standardfehler* zu den Regressionskoeffizienten. Diese Schätzungen sind:

- verzerrt, also *nicht* erwartungstreu
- inkonsistent
  Die Verzerrung wird durch eine Steigerung des Stichprobenumfangs nicht kleiner, sondern eventuell sogar größer (Long & Ervin 2000).

Im Ergebnis erhält man fehlerhafte Signifikanztests und Vertrauensintervalle zu den Regressionskoeffizienten, wobei die Tests je nach dem Muster der Heteroskedastizität zu liberal oder zu konservativ ausfallen können. Bei einer multiplen Regression kann auch der globale F-Test zum gesamten Modell analog betroffen sein (vgl. Abschnitt 2.5.1).

Einige Befunde zu den Effekten verschiedener Heteroskedastizitätsmuster (siehe Darlington & Hayes 2017, S. 501f; Hayes & Cai 2007, S. 710):

- Eine relativ milde Heteroskedastizität hat in der Regel keine gravierenden Auswirkungen auf die Ergebnisse einer Regressionsanalyse. In den Abschnitten 1.7.3.2.1 (Residuen-Plot), 1.7.3.2.2 (Spread & Level - Plot), 1.7.3.2.3 (Maximalquotientenkriterium) und 1.7.2.4 (Breusch-Pagan - Test) folgen konkrete Vorschläge zur Diagnose einer inakzeptablen Heteroskedastizität.
- Fällt die Fehlervarianz bei extremen (sehr kleinen oder sehr großen) Werten des Regressors niedriger aus im Vergleich zu mittleren Ausprägungen, dann resultieren überschätzte Standardfehler, also zu konservative Tests und überbreite Konfidenzintervalle.
- Wächst die Fehlervarianz mit den Werten des Regressors, dann resultieren unterschätzte Standardfehler, also zu liberale Tests und zu schmale Konfidenzintervalle. Mit diesem Fall ist in der Praxis eher zu rechnen.

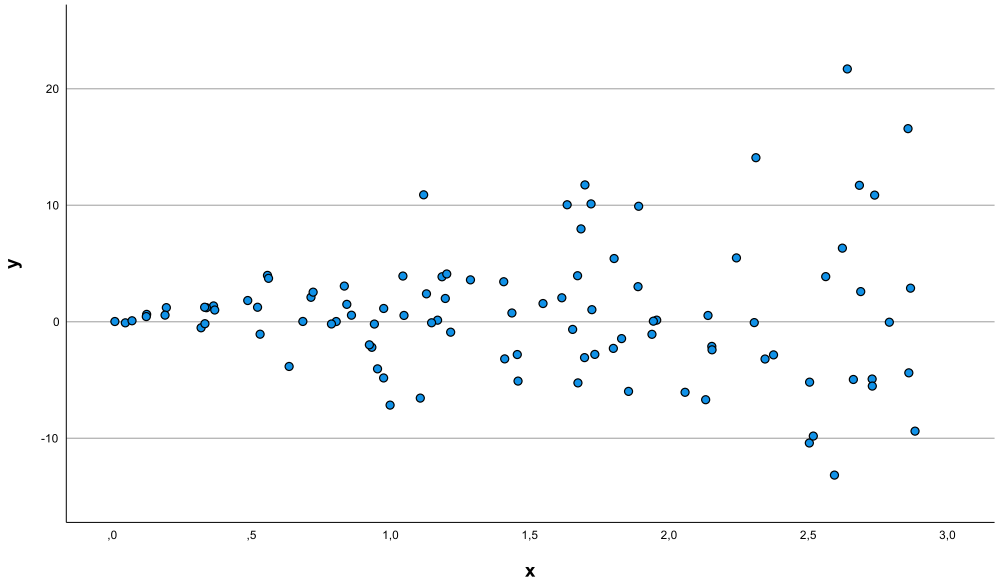
Als Ursachen für Heteroskedastizität kommen z. B. in Frage:

- Mit dem Kriteriumswert wachsender Messfehler
  Ist β_1 von 0 verschieden, ändert sich mit dem Regressor der erwartete Kriteriumswert. Wenn das Instrument zur Messung der Kriteriumswerte mit zunehmender Merkmalsausprägung tendenziell größere Fehler macht, dann resultiert Heteroskedastizität.
- Modellspezifikationsfehler (z. B. vergessene Interaktion zwischen zwei Regressoren)
  Eine erkannte Heteroskedastizität stellt immer einen Ansporn zur Verbesserung des Modells dar.

Um einen Eindruck von der Verzerrung des Standardfehlers zu einem OLS-Regressionskoeffizienten durch Heteroskedastizität zu vermitteln, wurde eine Simulationsstudie mit dem folgenden wahren Modell

$$Y = 0 \cdot X + \varepsilon,\ \varepsilon \sim N(0,\ 9 \cdot X^2),\ X \text{ gleichverteilt auf dem Intervall } [0,\ 3]$$

500-mal mit jeweils 100 Fällen durchgeführt. Es besteht *kein* Effekt von *X* auf *Y*, aber eine ausgeprägte Heteroskedastizität der Residuen, die bei diesem Null-Effekt - Modell mit den Kriteriumswerten übereinstimmen:



Im Abschnitt 1.7.3.2.2 werden wir reale Daten mit einem vergleichbaren Heterogenitätsproblem analysieren.

In der folgenden Tabelle ist die Standardabweichung der Steigungskoeffizienten aus den 500 Stichproben sowie der Mittelwert der 500 per OLS-Technik geschätzten Standardfehler zum Steigungskoeffizienten zu sehen:

Statistiken

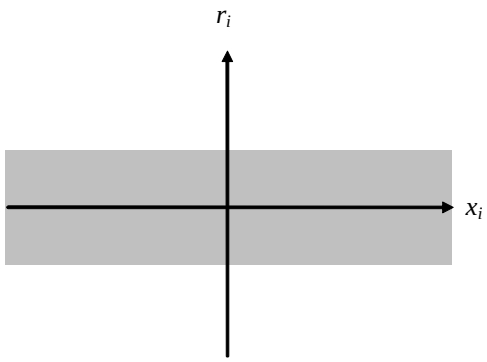
		Regressionskoeffizient	Standardfehler
N	Gültig	500	500
	Fehlend	0	0
Mittelwert		,00814	,59991
Standardabweichung		,655421	,062368

Es zeigt sich, dass der Standardfehler durchschnittlich um den Faktor $\frac{0{,}600}{0{,}655} = 0{,}916$, also um ca. 8% unterschätzt wird. Als Folge stellt sich eine erhöhte α-Fehlerrate von 7,8 % (statt der korrekten 5%) bzw. ein entsprechend reduzierter Anteil von 95% - Vertrauensintervallen, die den wahren Wert 0 enthalten (92,2% statt 95%). Um diese Störungen der Inferenzstatistik in relevanter Größenordnung zu provozieren, musste allerdings eine sehr gravierende Heteroskedastizität realisiert werden.

1.7.3.2 Diagnosemethoden

1.7.3.2.1 Residuen-Plot

Aufgrund der Erläuterungen im Abschnitt 1.7.2 bietet es sich an, zur optischen Beurteilung der Homoskedastizität das (r_i, x_i) - Streudiagramm der ausgelassen-studentisierten Residuen gegen den Regressor zu betrachten. Die Residuen sollten in einem Streifen um die Null-Lage aufhalten, wenn neben der Linearitäts- auch die Varianzhomogenitätsannahme erfüllt ist:



Für die Regression von Gewicht auf Größe (mittlerweile nur noch bei der männlichen Teilstichprobe) ergibt sich das folgende Bild mit wenig Anlass zur Sorge um die Homoskedastizität:

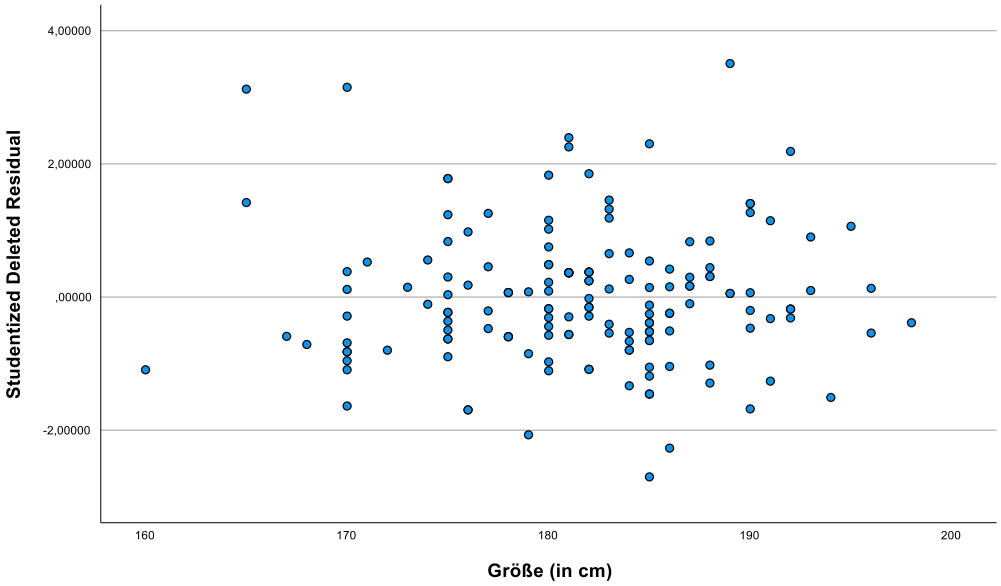
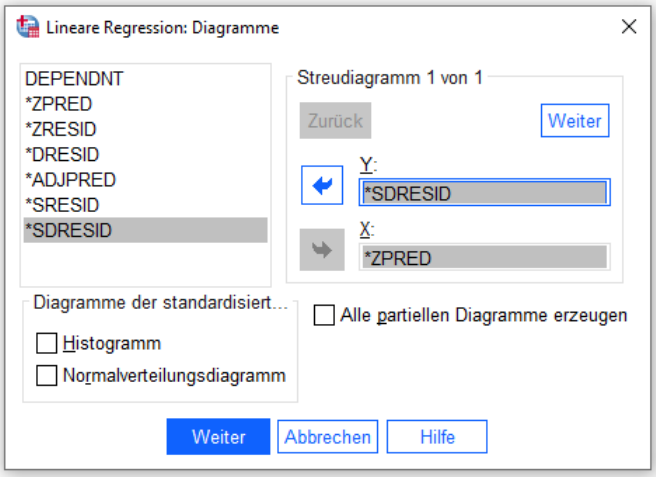


Abbildung 7: Ausgelassen-studentisierte Residuen versus Prädiktor für die Regression von GEWICHT auf GROESSE (bei N = 159 Männern)

Wir haben bislang in naheliegender Vorgehensweise Residuen gegen die *X-Variable* geplottet und mussten dabei ein relativ umständliches Vorgehen wählen. SPSS bietet *innerhalb* der Regressionsprozedur bequem zugängliche Plots an, die zu unseren obigen Abbildungen perfekt äquivalent sind. Darin wird als X-Achsen - Variable statt des Regressors die standardisierte Modellprognose verwendet, die bei einer bivariaten Regression stets eine perfekte lineare Funktion des Regressors ist. Ein Streudiagramm mit den ausgelassen-studentisierten Residuen (SPSS-Bezeichnung: SDRESID) und den standardisierten Modellprognosen (SPSS-Bezeichnung: ZPRED) wird in der Subdialogbox **Diagramme** folgendermaßen angefordert:



Das per Regressionsprozedur erstellte Streudiagramm zeigt erwartungsgemäß exakt dasselbe Bild wie Abbildung 7:

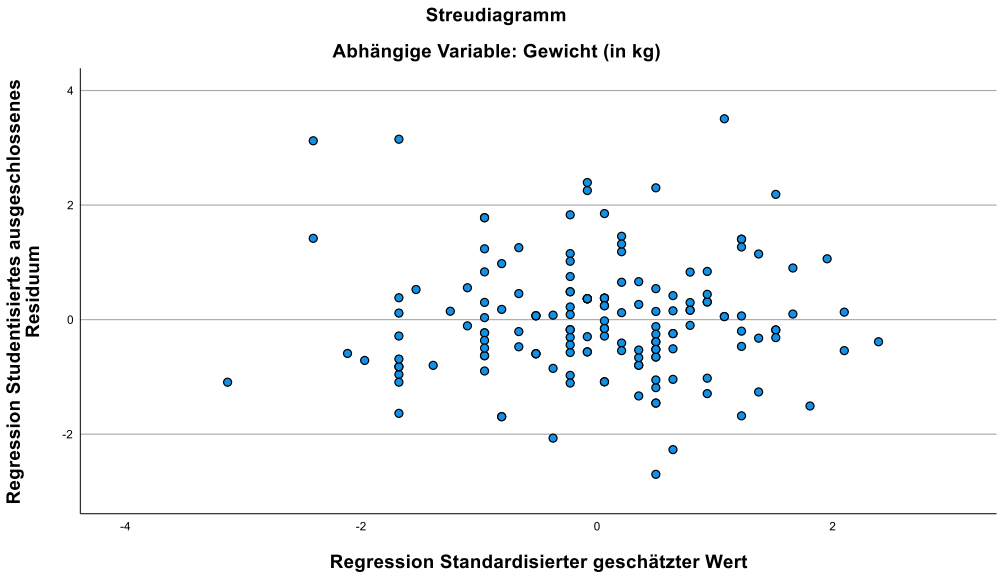
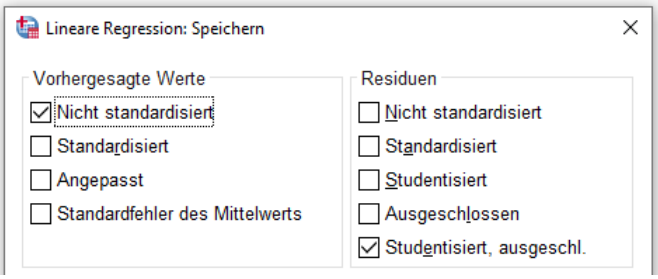


Abbildung 8: Ausgelassen-studentisierte Residuen vs. standardisierte geschätzte Werte für die Regression von GEWICHT auf GROESSE (bei N = 159 Männern)

1.7.3.2.2 Spread & Level - Plot

Im sogenannten *Spread & Level - Plot* (siehe z. B. Fox & Weisberg 2011, S. 315) werden die logarithmierten Beträge der ausgelassen-studentisierten Residuen gegen die logarithmierten Modellprognosen geplottet. Fälle mit einem Prognosewert kleiner oder gleich 0 müssen ausgeschlossen werden, weil der Logarithmus für solche Werte nicht definiert ist. Um den Plot mit SPSS zu erstellen, lässt man von der Regressionsprozedur die ausgelassen-studentisierten Residuen sowie die vorhergesagten Werte des Modells als neue Variablen in der Arbeitsdatei speichern:



Daraus werden die für den Spread & Level - Plot benötigten Variablen berechnet, z. B.:

```
compute LnAbsASR = ln(abs(SDR_1)).
compute LnProg = ln(PRE_1).
execute.
```

Schließlich kann man das gewünschte Streudiagramm z. B. über den Menübefehl

Grafik > Klassische Dialogfelder > Streu-/Punktdiagramm > Einfaches Streudiagramm

erstellen lassen. Wächst die Residualvarianz mit dem prognostizierten Wert (ein typisches Bild bei verletzter Homoskedastizität), dann ergibt sich im Spread & Level - Plot eine ansteigende Regressionsgerade, was bei der Regression von GEWICHT auf GROESSE (bei N = 159 Männern) *nicht* der Fall ist:

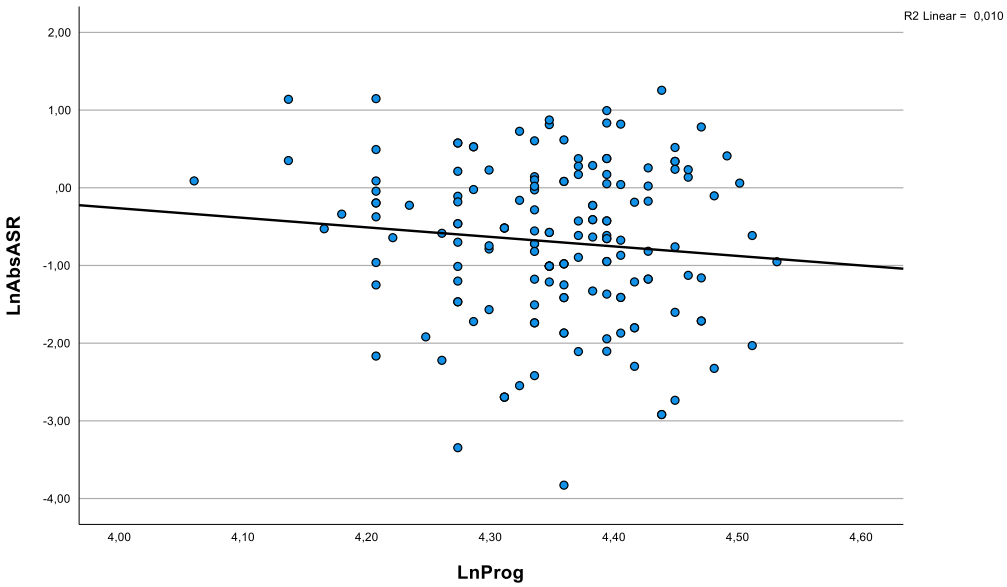


Abbildung 9: Spread & Level - Plot für die Regression von GEWICHT auf GROESSE (bei N = 159 Männern)

Aus einem signifikanten Steigungskoeffizienten b_1 im Spread & Level - Plot lässt sich eine Empfehlung für eine Fehlervarianz-stabilisierende Transformation des Kriteriums ableiten (siehe Fox 1997, S. 302):

$$Y \rightarrow Y^{1-b_1}$$

Im Beispiel wird die Nullhypothese zum Steigungskoeffizienten in der Regression von LNABSASR auf LNPROG beibehalten (p = 0,201), sodass keine Transformation des Kriteriums erforderlich ist:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	4,640	4,159		1,116	,266	-3,575	12,855
	LnProg	-1,226	,956	-,102	-1,283	,201	-3,114	,662

a. Abhängige Variable: LnAbsASR

Damit nicht der Eindruck entsteht, dass wir im Manuskript mit bequemen Daten den Schwierigkeiten des statistischen Alltags aus dem Weg gehen, wird noch ein empirisches Beispiel mit deutlich verletzter Varianzhomogenität präsentiert. In einer bei Kockläuner (1988) vorgestellten volkswirtschaftlichen Studie mit 102 Nationen als Beobachtungseinheiten geht es um ein Modell zur Erklärung des Pro-Kopf - Bruttosozialprodukts (SPSS-Variablenname BSP) durch verschiedene Entwicklungsindikatoren (Daten aus dem Jahr 1974).[1] In einem ersten Modell, dessen Ergebnisse anschließend diskutiert werden, verwendet Kockläuner einen Ernährungsindex (SPSS-Variablenname ERN) als einzigen Regressor. Er wurde aus einer Hauptkomponentenanalyse gewonnen und reflektiert u.a. den Pro-Kopf - Kalorien- bzw. - Proteinverbrauch. In der linearen Regression zeigt sich ein starker Effekt des Regressors, wobei aber die Prüfung der Modellvoraussetzungen noch aussteht:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	1693,941	120,663		14,039	,000	1454,549	1933,334
	Ernährungsindex	1604,020	121,261	,798	13,228	,000	1363,441	1844,599

a. Abhängige Variable: Brutto-Sozialprodukt pro Kopf

Zur Beurteilung der Homoskedastizität (und der Linearität) betrachten wir die gemeinsame Verteilung der Variablen BSP und ERN mit eingezeichneter Regressionsgeraden:

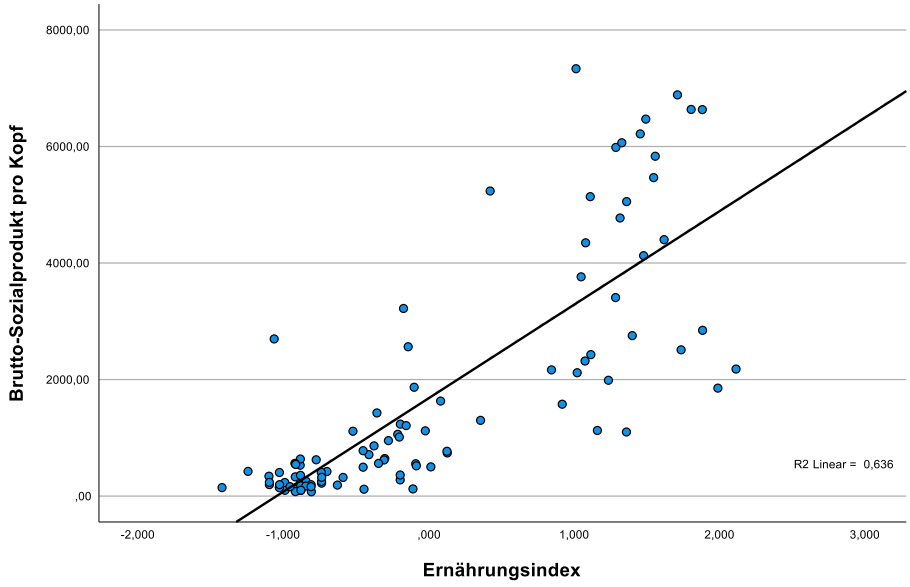


Abbildung 10: Regression von BSP auf ERN im volkswirtschaftlichen Beispiel

Während das Streudiagramm bzgl. der Linearitätsannahme keinen eindeutigen Schluss erlaubt, spricht es deutlich für einen Anstieg der Fehlervarianz mit dem Prädiktor. Derselbe Befund zeigt sich im Plot der ausgelassen-studentisierten Residuen gegen die vorhergesagten Werte, der sich bequem per Regressionsprozedur anfordern lässt (vgl. Abschnitt 1.7.3.2.1):

[1] Die Datei **Kockl.sav** mit den Daten befindet sich an der im Vorwort vereinbarten Stelle.

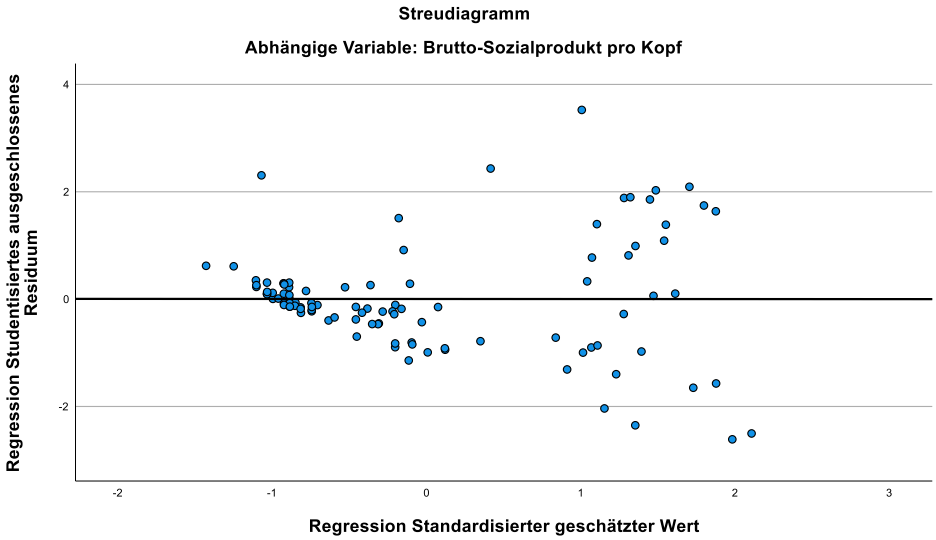


Abbildung 11: Plot der ausgelassen-studentisierte Residuen gegen die standardisierte Schätzwerte für die Regression von BSP auf ERN im volkswirtschaftlichen Beispiel

Wir erhalten ein annähernd trichterförmiges Bild:

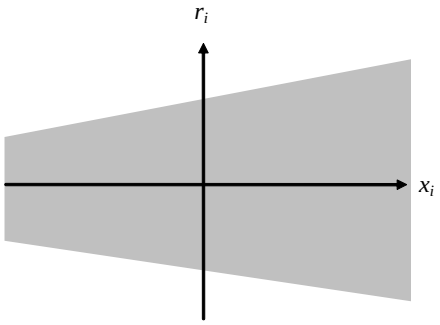


Abbildung 12: Nach diesem trichterförmigen Muster wächst oft die Residualvarianz mit dem Regressor

Im Spread & Level - Plot der logarithmierten absoluten ausgelassen-studentisierten Residuen gegen die logarithmierten Modellprognosen spricht eine deutlich ansteigende Regressionsgerade ($b_1 = 0{,}75$) gegen die Homoskedastizität:

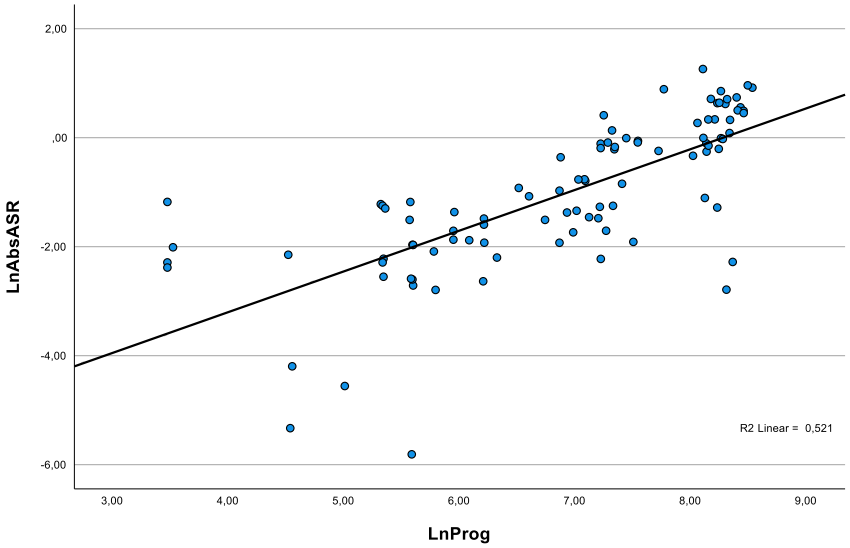


Abbildung 13: Plot der logarithmierten absoluten ausgelassen-studentisierten Residuen gegen die logarithmierten Modellprognosen im volkswirtschaftlichen Beispiel

In diesem Diagramm fehlen 6 Fälle, weil für den Prognosewert kleiner oder gleich 0 der Logarithmus nicht berechnet werden konnte. Das Vertrauensintervall zum Regressionskoeffizienten ist weit vom Wert 0 entfernt:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	-6,191	,519		-11,930	,000	-7,221	-5,160
	LnProg	,748	,074	,722	10,115	,000	,601	,895

a. Abhängige Variable: LnAbsASR

Es resultiert die folgende Empfehlung zur Transformation des Kriteriums mit dem Ziel einer verbesserten Homoskedastizität (siehe Abschnitt 1.7.3.3):

$$Y \rightarrow Y^{1-0,75} = Y^{0,25}$$

1.7.3.2.3 Maximalquotientenkriterium

Im Abschnitt 1.2.4 (mit dem Titel *Mathematisches Modell und Realität*) haben wir uns schon damit abgefunden, dass es ein perfektes Modell (und damit auch eine *perfekte* Homoskedastizität) praktisch nie geben wird. Wenn wir uns bei einer Studie mit den Voraussetzungen der linearen Regression beschäftigen, geht es also kaum um die Frage, ob eine perfekte Homoskedastizität besteht, sondern es ist zu klären, ob die beobachtete Verletzung der Homoskedastizität toleriert werden kann oder nicht. Nach einer bei Ryan (1997, S. 61) mitgeteilten Daumenregel besteht kein Grund zur Sorge, solange der Quotient aus der maximalen und der minimalen Fehler*standardabweichung* den Wert 1,5 nicht übersteigt, während Quotienten ab drei auf jeden Fall inakzeptabel sind. Übertragen auf Fehler*varianzen* sind also Quotienten ab neun als kritisch anzusehen. Ähnlich äußern sich auch Cohen et al. (2003, S. 120) und Fox (1997, S. 306f), die zehn als kritisches Verhältnis zwischen der maximalen und der minimalen Fehlervarianz nennen.

Um die Fehlervarianz für einen bestimmten *X*-Wert schätzen zu können, muss man natürlich *mehrere Y*-Werte unter dieser Bedingung ermitteln. In einer Beobachtungsstudie mit einem metrischen Regressor liegen solche Replikationen in der Regel nicht vor, doch kann man durch Zusammenfassen bzgl. des Prädiktors zu verwertbaren Schätzungen von bereichsspezifischen Fehlervarianzen kommen. Für die folgende Abbildung wurden im Beispiel mit der Regression von Gewicht auf Größe fünf etwa gleich stark besetzte GROESSE-Intervalle gebildet, zu denen jeweils der mittlere GROESSE-Wert und die Standardabweichung der einfachen Residuen ermittelt wurden:[1]

[1] Die SPSS-Syntaxdatei **GgAgr.sps** (zu finden an der verabredeten Stelle) enthält die zum Erstellen der Abbildung und zum Ermitteln der intervallspezifischen Fehlerstandardabweichungen erforderlichen SPSS-Kommandos. Trotz der Argumente aus Abschnitt 1.7.2 werden aktuell die *einfachen* Residuen betrachtet, weil sich die zitierten Empfehlungen zum Maximalquotienten auf einfache Residuen beziehen.

Während die aktuell betrachteten GROESE-Intervalle eine ähnliche *Häufigkeit* besitzen, wurden im Abschnitt 1.7.1.1.2 zur Beurteilung der Linearität Intervalle mit *identischer Breite* benötigt.

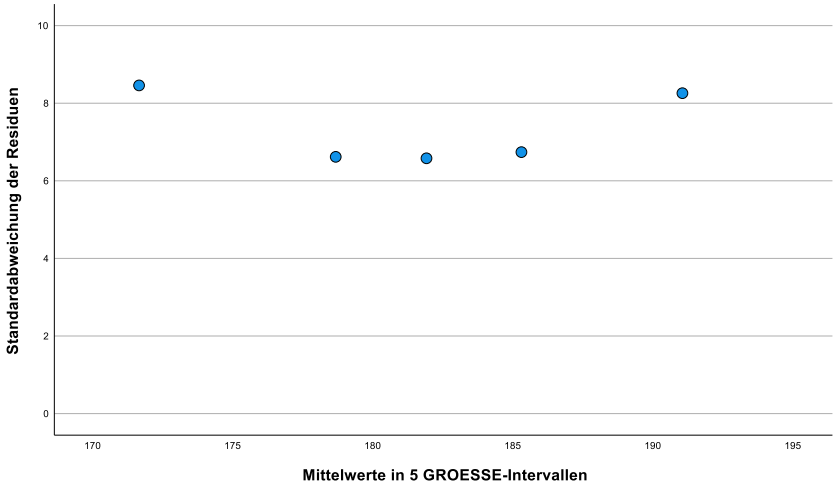


Abbildung 14: Standardabweichungen der Residuen für 5 GROESSE-Intervalle (bei N = 159 Männern)

Nach dem Maximalquotienten-Kriterium liegt eine tolerierbare Abweichung von der Varianzhomogenität vor:

$$\frac{8{,}46}{6{,}58} = 1{,}29$$

Analog zum Vorgehen bei der Regression von GEWICHT auf GROESSE wurden auch für das volkswirtschaftlichen Beispiel geschätzte Standardabweichungen der (einfachen) Residuen für fünf gleich stark besetzte Intervalle bzgl. der Prädiktorvariablen ermittelt:

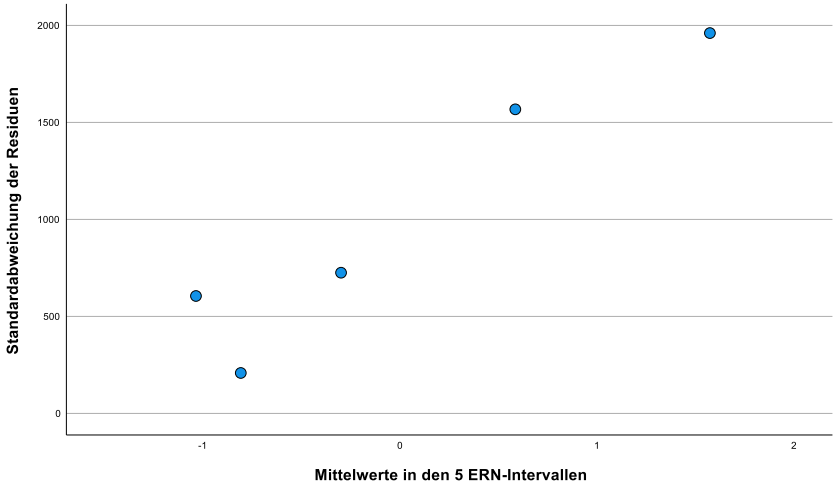


Abbildung 15: Standardabweichungen der einfachen Residuen für 5 ERN-Intervalle im volkswirtschaftlichen Beispiel

Diesmal ist der Quotient aus der größten und der kleinsten Standardabweichung so extrem (deutlich größer als 3), dass nach der oben angegebenen Regel eine inakzeptable Verletzung der Homogenitätsannahme besteht:

$$\frac{1960{,}63}{208{,}53} = 9{,}40$$

1.7.3.2.4 Breusch-Pagan - Heteroskedastizitätstest

Es sind verschiedene Signifikanztests zur Beurteilung der Varianzhomogenität vorgeschlagen worden (siehe z. B. Backhaus et al. 2008, S. 86; Cohen et al. 2003, S. 133; Groß 2003, S. 318; Fox 1997, S. 320). Ein generelles Problem dieser Verfahren ist die im Abschnitt 1.2.4 diskutierte Abhängigkeit von der Stichprobengröße. Trotz der Bedenken soll exemplarisch der häufig verwendete Breusch-Pagan - Hetero-

skedastizitätstest vorgestellt werden, den SPSS seit der Version 25 in der modifizierten und im allgemeinen bevorzugten Variante (siehe Wooldridge 2013, S. 267) in etlichen Prozeduren (z. B. UNIANOVA, GLM) anbietet.

Die Konstruktion des modifizierten Breusch-Pagan - Heteroskedastizitätstests ist plausibel und einfach, sodass man ihn mit fast jeder Regressionsprozedur in Eigeninitiative durchführen kann, wenn er nicht als Fertigprodukt angeboten wird:

- Man lässt die einfachen Residuen (e_i) abspeichern und quadriert sie anschließend.
- Man rechnet eine Regressionsanalyse mit den quadrierten Residuen als Kriterium und dem Regressor aus dem ursprünglichen Modell.
- Das unkorrigierte R^2 aus der neuen Analyse ergibt nach einer Multiplikation mit der Stichprobengröße N die Breusch-Pagan - Prüfgröße. Bei gültiger Nullhypothese ist diese Prüfgröße χ^2-verteilt mit einem Freiheitsgrad.

Für unser Beispiel mit der Regression von Gewicht auf Größe bei 159 Männern resultiert für das Modell mit den quadrierten Residuen als Kriterium ein unkorrigierter R^2-Wert von 0,004190:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,06473 [a]	,004190	-,002153	99,283675

a. Einflußvariablen : (Konstante), Größe (in cm)

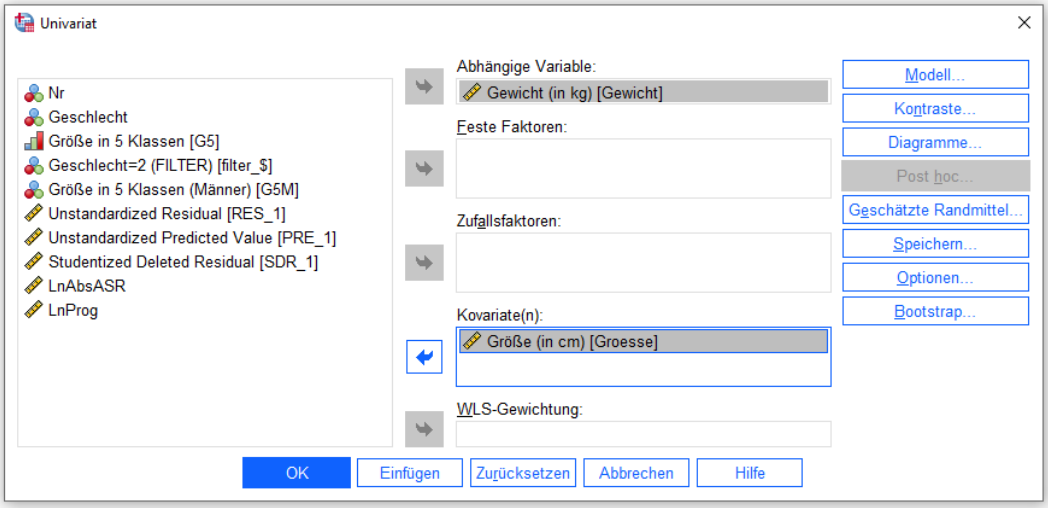
Für die Breusch-Pagan - Prüfgröße erhält man also den Wert:

$$159 \cdot 0{,}004190 = 0{,}66621$$

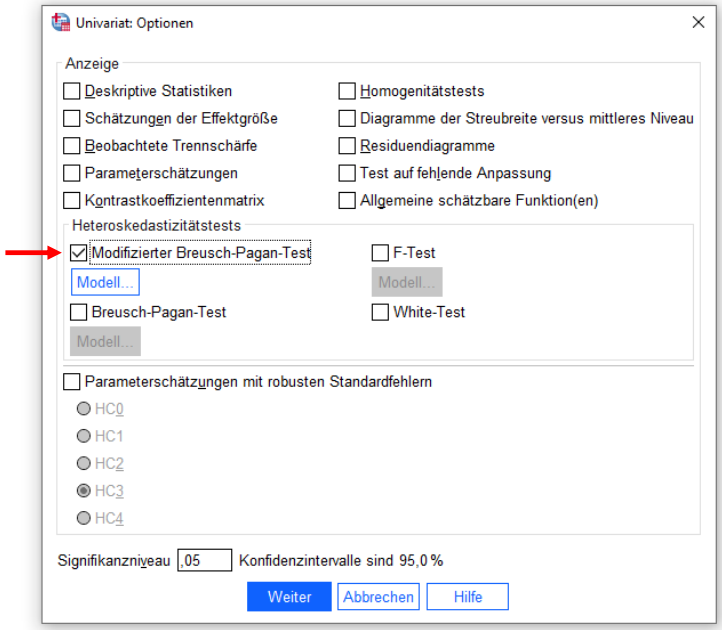
Eben wurde die Breusch-Pagan - Prüfgröße mit Hilfe der SPSS-Prozedur REGRESSION ermittelt, um ihre Konstruktion zu illustrieren. Mit der SPSS-Prozedur UNIANOVA kommt man bequemer ans Ziel und erhält auch noch ein Signifikanztestergebnis. Wir deklarieren nach dem Menübefehl

Analysieren > Allgemeines lineares Modell > Univariat

im folgenden Dialog die Regression von Gewicht auf Größe:



Nach einem Mausklick auf den Schalter **Optionen** fordern wir den **modifizierten Breusch-Pagan-Test** an:



Wir erhalten eine Bestätigung für das eben semi-automatisch ermittelte Ergebnis und den zugehörigen *p*-Wert (0,414):

Modifizierter Breusch-Pagan-Test auf Heteroskedastizität[a,b,c]

Chi-Quadrat	df	Sig.
,666	1	,414

a. Abhängige Variable: Gewicht (in kg)

b. Testet die Nullhypothese, dass die Varianz der Fehler nicht von den Werte der unabhängigen Variablen abhängt.

c. Vorhergesagte Werte aus Design: Konstanter Term + Groesse

Somit kann die Nullhypothese homogener Fehlervarianzen beibehalten werden.

Bei den im Abschnitt 1.7.3.2.2 beschriebenen volkswirtschaftlichen Daten (aus Kockläuner 1988) zwingt hingegen ein hochsignifikantes Testergebnis dazu, die Homogenitäts-Nullhypothese abzulehnen:

Modifizierter Breusch-Pagan-Test auf Heteroskedastizität[a,b,c]

Chi-Quadrat	df	Sig.
33,993	1	,000

a. Abhängige Variable: Brutto-Sozialprodukt pro Kopf

b. Testet die Nullhypothese, dass die Varianz der Fehler nicht von den Werte der unabhängigen Variablen abhängt.

c. Vorhergesagte Werte aus Design: Konstanter Term + ern

1.7.3.3 Transformation des Kriteriums zur Homogenisierung der Fehlervarianzen

Eine *lineare* Transformation kann keine Stabilisierung heterogener Fehlervarianzen bewirken. Bei der Suche nach einer geeigneten *nichtlinearen* Transformation des Kriteriums ist zu beachten:

- Von der Transformation ist auch die Linearitätsannahme des Modells betroffen: War die Linearitätsannahme *vor* der Transformation erfüllt (d.h. $E(Y_i) = \beta_0 + \beta_1 x_i, i = 1,.., N$), kann sie es anschließend nicht mehr sein.
- Ebenso ist die Gestalt der Residualverteilung betroffen. Mit etwas Glück findet man eine Transformation, die sowohl die Homoskedastizität als auch die Normalität der Residuen fördert (vgl. Abschnitt 1.7.4.6.2).

Aus der nichtlinearen Transformation der abhängigen Variablen resultiert eine Bedeutungsveränderung für ihre Ausprägungen. So rücken z. B. bei der häufig eingesetzten logarithmischen Transformation hohe Werte näher zusammen. Wo vorher ein großer, durch eine starke Einwirkung des Regressors zu erklärender Unterschied bestand, differieren die transformierten Werte deutlich weniger, sind also hinsichtlich ihrer Aussage über den Merkmalsträger ähnlicher. Eine solche Bedeutungsveränderung kann durchaus erwünscht sein und die Modellierung erleichtern.

Von den möglichen Transformationen des Kriteriums, die bei unterschiedlichen Heterogenitätsmustern indiziert sind (siehe z. B. Draper & Smith, 1981, 237ff), betrachten wir:

a) Logarithmus
Verhält sich die Standardabweichung der Y_i - Variablen proportional zu ihrem Erwartungswert,

$$\sqrt{\mathrm{Var}(Y_i)} = \omega \, E(Y_i), \; i = 1,.., N \; ,$$

dann empfiehlt sich eine Transformation mit dem natürlichen Logarithmus, wenn fast alle Y_i - Werte positiv sind:

$$Y_i \longrightarrow \ln(Y_i)$$

b) Potenzieren mit Exponentenwahl gemäß Spread & Level - Plot
Aus der Steigung b_1 der Regressionsgeraden zum Spread & Level - Plot (siehe Abschnitt 1.7.3.2.2) lässt sich eine Empfehlung für eine Fehlervarianz-stabilisierende Transformation des Kriteriums ableiten (siehe Fox 1997, S. 302):

$$Y_i \rightarrow Y_i^{1-b_1}$$

Beide Verfahren sollen auf die im Abschnitt 1.7.3.2.2 beschriebenen, mit einem Heteroskedastizitätsproblem belasteten volkswirtschaftlichen Daten angewendet werden. Teilt man die Stichprobe nach der empirischen Verteilung des Kriteriums (BSP) in 5 ungefähr gleich stark besetzte Segmente auf und plottet die BSP-Teilstichproben-Standardabweichungen gegen die BSP-Teilstichproben-Mittelwerte, dann ergibt sich die folgende annähernd lineare Beziehung:[1]

[1] Im Abschnitt 1.7.3.5 hatten wir ebenfalls gruppierte Daten betrachtet, doch wurden dort Intervalle bzgl. des Prädiktors gebildet, während wir nun Intervalle bzgl. des Kriteriums betrachten.

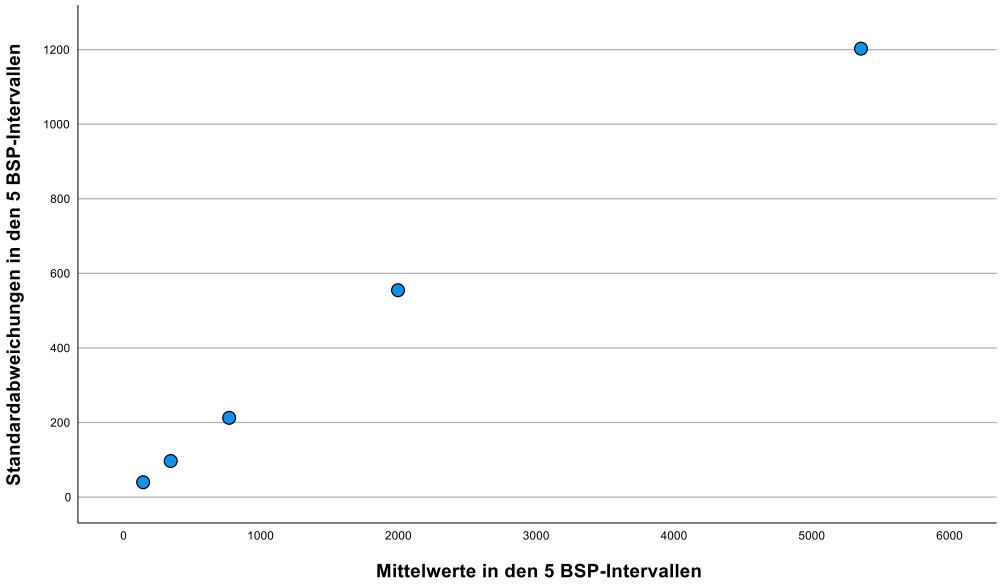


Abbildung 16: BSP-Standardabweichung versus BSP-Mittelwert
für 5 gleich stark besetzte BSP-Intervalle im volkswirtschaftlichen Beispiel

Folglich kommt eine logarithmische Transformation in Betracht.

Beim Spread & Level - Plot (siehe Abschnitt 1.7.3.2.2) zeigt sich ein Steigungskoeffizient von 0,75 (in der Regression der logarithmierten Beträge der ausgelassen-studentisierten Residuen auf die logarithmierten Modellprognosen):

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	-6,191	,519		-11,930	,000	-7,221	-5,160
	LnProg	,748	,074	,722	10,115	,000	,601	,895

a. Abhängige Variable: LnAbsASR

Daraus resultiert die folgende Empfehlung zur Transformation des Kriteriums:

$$Y_i \rightarrow Y_i^{1-0,75} = Y_i^{0,25}$$

Unter Verwendung des Konfidenzintervalls für den Steigungskoeffizienten aus der Spread & Level - Regression liegt der empfohlene Exponent für die Transformation im folgenden Intervall:

$$[0,105;\ 0,399]$$

Die beiden Empfehlungen ($\ln(Y)$, $Y^{0,25}$) sind durchaus miteinander verträglich, weil der Logarithmus als „Grenzwert“ für die Potenztransformation Y^{λ} für ($\lambda \rightarrow 0$) betrachtet werden kann (vgl. Abschnitt 1.7.4.6.2).[1] Im Beispiel resultieren aus den beiden Vorschlägen transformierte Variablen mit einer Korrelation von 0,99. Wir entscheiden uns für die vertrautere, logarithmische Transformation und erstellen die neue Variable ln(BSP) (SPSS-Variablenname: LNBSP).

Die logarithmische Transformation des Kriteriums beeinflusst nicht nur die die Homogenität der Residuen, sondern auch die Linearität des Modells. Zur Beurteilung des modifizierten Modells betrachten wir zunächst das Streudiagramm von LNBSP und ERN:

[1] Im Manuskript wird der natürliche Logarithmus (mit der Eulerschen Zahl als Basis) verwendet und *ln()* als Funktionsname genutzt. Diese Wahl hat keine große Bedeutung, weil sich die verschiedenen Logarithmus-Varianten nur durch einen konstanten Faktor voneinander unterscheiden.

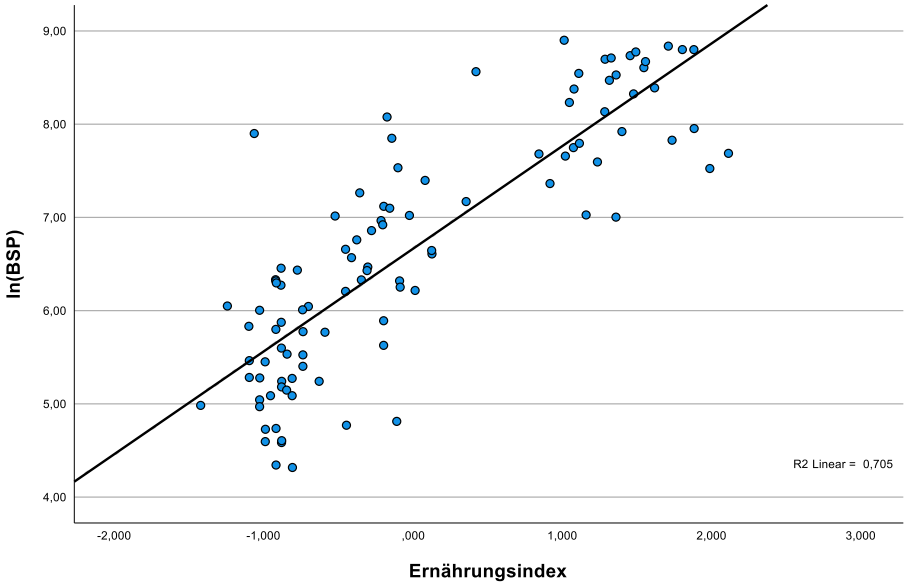


Abbildung 17: Regression von LNBSP auf ERN im volkswirtschaftlichen Beispiel

Im Vergleich zum Plot von BSP gegen ERN ist eine erheblich bessere Anpassung der Regressionsgeraden festzustellen. Dies zeigt sich auch im korrigierten R_a^2 von 0,702 für die Regression von LNBSP auf ERN (im Vergleich zum Wert 0,633 für die Regression von BSP auf ERN):

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,840 [a]	,705	,702	,71533

a. Einflußvariablen : (Konstante), Ernährungsindex

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	6,673	,071		94,208	,000	6,532	6,813
	Ernährungsindex	1,101	,071	,840	15,464	,000	,959	1,242

a. Abhängige Variable: ln(BSP)

Im Breusch-Pagan - Test zur Varianzhomogenität (vgl. Abschnitt 1.7.3.2.4) zeigt sich im Unterschied zum Model für das ursprüngliche Kriterium *keine* Signifikanz, sodass die Homogenitätsannahme für das modifizierte Modell akzeptiert werden kann:

Modifizierter Breusch-Pagan-Test auf Heteroskedastizität [a,b,c]

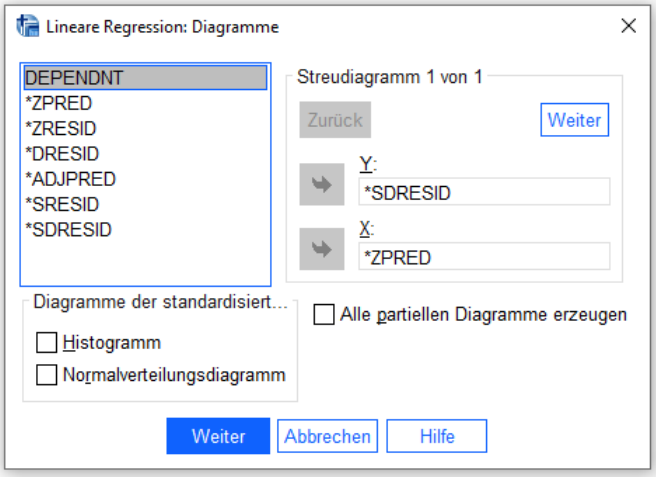
Chi-Quadrat	df	Sig.
,635	1	,426

a. Abhängige Variable: ln(BSP)

b. Testet die Nullhypothese, dass die Varianz der Fehler nicht von den Werte der unabhängigen Variablen abhängt.

c. Vorhergesagte Werte aus Design: Konstanter Term + ern

Allerdings sind diese freundlichen Ergebnisse noch kein Beweis für die Modellgültigkeit, sodass wir zur Beurteilung der Linearität unter Verwendung des **Diagramme**-Subdialogs der Regressionsprozedur



die ausgelassen-studentisierten Residuen gegen die standardisierten Prognosewerte (also letztlich gegen den Regressor) plotten lassen:

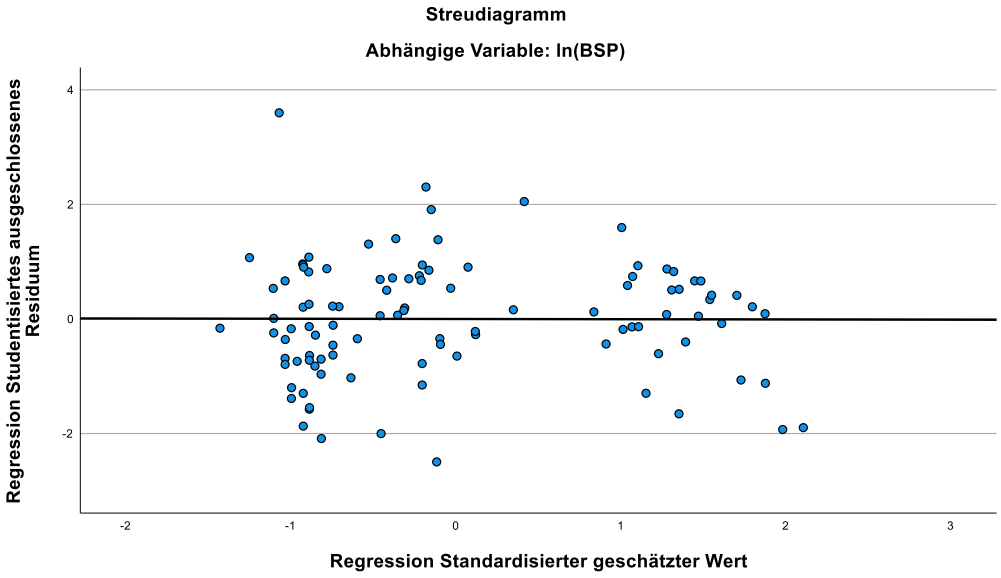


Abbildung 18: Ausgelassen-studentisierte Residuen vs. prognostizierte Werte für die Regression von LNBSP auf ERN im volkswirtschaftlichen Beispiel

Laut Kockläuner (1988, S. 99) liegen hier „keine offenkundigen nichtlinearen Strukturen“ vor, sodass die Linearitätsannahme für das Modell mit dem logarithmierten Kriterium akzeptiert wäre. An dieser Stelle sollte die Behandlung des volkswirtschaftlichen Beispiels eigentlich abgeschlossen sein. Es sollte ja nur dazu dienen, mögliche Maßnahmen zur Korrektur einer Varianzheterogenität zu demonstrieren. Allerdings fällt es schwer, Kockläuners Urteil zu Abbildung 18 zu teilen. Es zeigt sich eine umgekehrt U-förmige Beziehung, sodass die Linearitätsannahme eher abzulehnen ist:

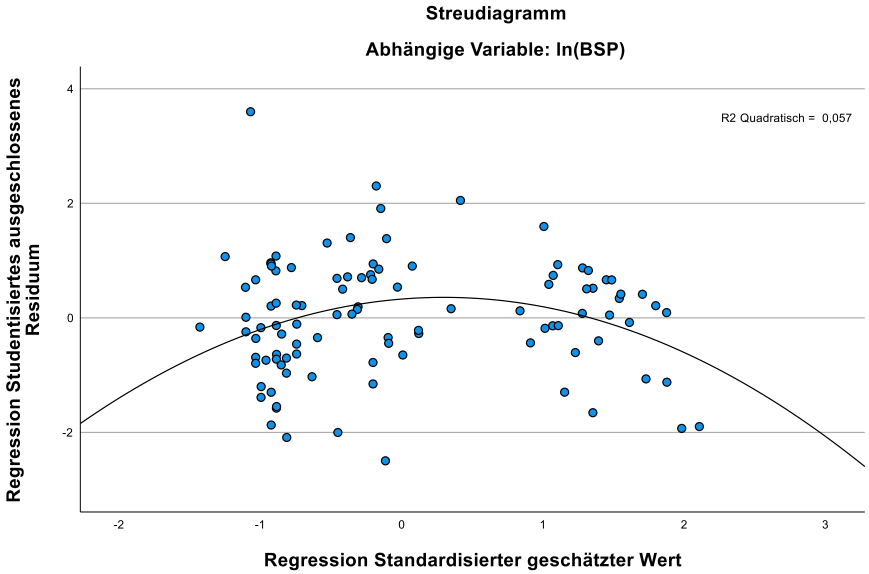


Abbildung 19: Ausgelassen-studentisierte Residuen vs. prognostizierte Werte für die Regression von LNBSP auf ERN im volkswirtschaftlichen Beispiel (mit quadratischer Regressionsfunktion)

Eine nochmalige Inspektion von Abbildung 17 (Regression von LNBSP auf ERN) lässt vermuten, dass die Regression von quadratischer Gestalt ist, sodass in der Regressionsgleichung neben ERN auch die quadrierte Variable

$$ERN2 := ERN^2$$

als Prädiktor benötigt wird. Allerdings haben wir es nun mit einer *multiplen* Regressionsgleichung zu tun, die außerdem auch noch von spezieller (polynomischer) Bauart ist. Da wir die Theorie der multiplen Regression noch nicht kennen, vertagen wir die weitere Behandlung des volkswirtschaftlichen Beispiels auf den Abschnitt 4 über die polynomische Regression. Dort sind auch die (erfreulichen) Ergebnisse zur Varianzhomogenitätsprüfung für das revidierte Modell zu finden.

Im volkswirtschaftlichen Beispiel hat das Logarithmieren zu einer „verbesserten", wissenschaftlich relevanteren Kriteriumsvariablen geführt, die erfolgreicher modelliert werden kann als die Ausgangsversion.

1.7.3.4 Robuste Inferenzstatistik trotz Heteroskedastizität

In diesem Abschnitt geht es um statistische Techniken, welche die Heteroskedastizität nicht beseitigen, aber eine daraus resultierende fehlerhafte Inferenzstatistik verhindern können. Diese Techniken haben im Vergleich zur Transformation des Kriteriums (siehe Abschnitt 1.7.3.3) oder zur WLS-Regression (siehe Abschnitt 1.7.3.5) den Vorteil, dass die Struktur der Heteroskedastizität *nicht* bekannt sein muss. Allerdings sind sie trotzdem nicht in jeder Situation überlegen, denn:

- Bei der Transformation des Kriteriums wird eventuell eine Verbesserung des Modells erreicht, was im vorherigen Abschnitt am volkswirtschaftlichen Beispiel zu sehen war.
- Bei der WLS-Regression werden die Regressionskoeffizienten präziser geschätzt.

Die Verfahren bieten sich an, wenn ...

- die Linearität eines Modells akzeptiert und daher eine nichtlineare Transformation des Kriteriums wenig attraktiv ist,
- für die aufwändige WLS-Regression zu wenig Wissen über die Struktur der Heteroskedastizität vorhanden ist.

1.7.3.4.1 Heteroskedastizitäts-konsistente Standardfehler

Seit geraumer Zeit stehen Verfahren bereit, die eine konsistente Schätzung der Kovarianzmatrix der OLS-Regressionskoeffizienten trotz Heteroskedastizität von beliebiger (unbekannter) Form ermöglichen (siehe Hayes & Cai 2007; Long & Erwin 2000). Bei diesen Verfahren werden die Regressionskoeffizienten wie gewohnt per Kleinst-Quadrat-Kriterium geschätzt. Zur Schätzung der *Kovarianzmatrix* der OLS-Regressionskoeffizienten wird hingegen eine Methode verwendet, die auch bei Heteroskedastizität ein konsistentes Verhalten zeigt, d.h. bei wachsender Stichprobengröße an Präzision gewinnt. Auf diesem heterogenitäts-robusten Schätzergebnis basieren Signifikanztests und Vertrauensintervalle. Long & Erwin (2000) sprechen von der *heteroscedasticity consistent covariance matrix* (HCCM). Hayes & Cai (2007) konzentrieren sich auf das Verteilungsverhalten eines einzelnen Regressionskoeffizienten und verwenden den Begriff *heteroskedasticity-consistent standard error* (HCSE).[1] Weitere, in der Literatur gebräuchliche Bezeichnungen sind:

- *Sandwich-Schätzer* (abgeleitet von der mathematischen Struktur)
- *Huber-White -*, *Eicker-White* - oder *Eicke-Huber-White - Schätzer*, womit die Urheber der Technologie gewürdigt werden (siehe Eicker 1967, Huber 1967, White 1980)

Über die unter dem Namen HC0 bekannte, ursprüngliche Variante des heteroskedastizitäts-konsistenten Schätzers weiß man inzwischen, dass sie zwar konsistent ist, bei kleinen Stichproben ($N \leq 100$) aber unterschätzte Standardfehler und damit zu liberale Signifikanztests liefert (Long & Erwin 2000). Von der Weiterentwicklung HC3 sind auch bei kleinen Stichproben zuverlässige Ergebnisse zu erwarten, und die Variante HC4 wird von Hayes & Cai (2007, S. 712) bei Anwesenheit von Fällen mit großen Hebelwerten empfohlen (vgl. Abschnitte 1.7.2.1 und 3.1.3).

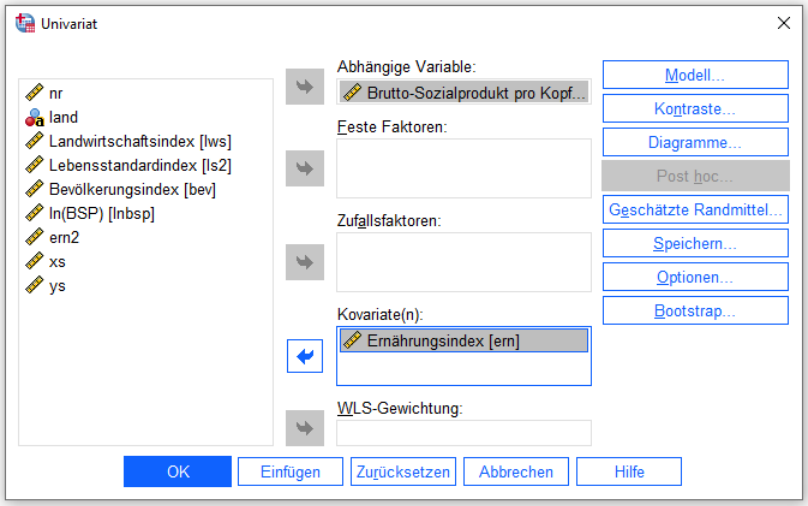
Nach Darlington & Hayes (2017, S. 511f) arbeiten die heteroskedastizitäts-konsistenten Standardfehler auch bei approximativer Homoskedastizität korrekt. Die Autoren halten es daher für vertretbar, die robusten Standardfehler routinemäßig einzusetzen und sich damit von der Homoskedastizitätsannahme zu befreien.

In SPSS wird die heteroskedastizitäts-konsistente Schätzung der Standardfehler zu Regressionskoeffizienten (inklusive HC3 und HC4) seit der Version 25 u.a von der Prozedur UNIANOVA unterstützt, auf die wir im Abschnitt 1.7.3.2.4 schon zur Durchführung des Breusch-Pagan - Heteroskedastizitätstests zurückgegriffen haben. Man erhält ihren Dialog über den Menübefehl:

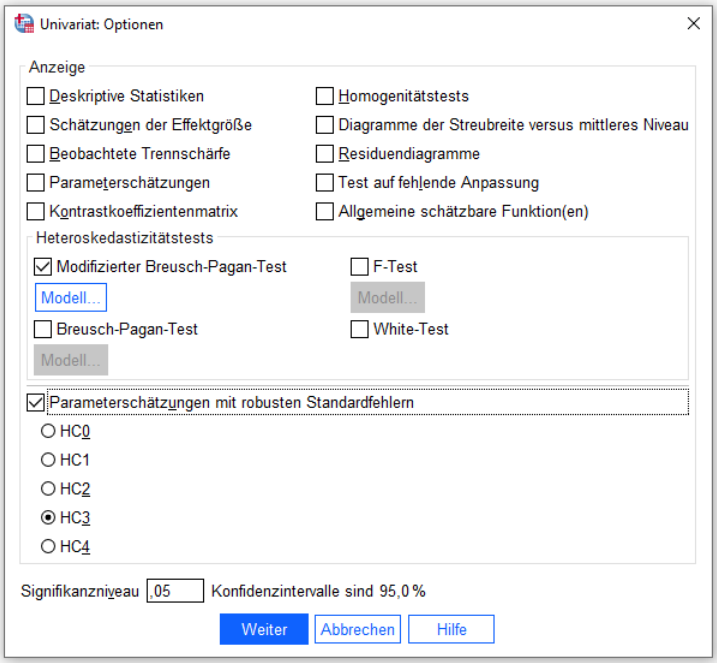
Analysieren > Allgemeines lineares Modell > Univariat

Wir analysieren das volkswirtschaftliche Modell (nach Kockläuner 1988), das in seiner ursprünglichen Form bei allen Verfahren zur Homoskedastizitätsdiagnose negativ aufgefallen ist und wählen BSP als **abhängige Variable** sowie ERN als **Kovariate**:

[1] Für die *heteroskedasticity* sind in der englischen Literatur die Schreibeweisen mit *k* und *c* anzutreffen.



Nach einem Mausklick auf den Schalter **Optionen** fordern wir **Parameterschätzungen mit robusten Standardfehlern** in der voreingestellten Variante **HC3** an:



Im Vergleich zu den Ergebnissen der Regressionsprozedur für das Kriterium BSP (vgl. Abschnitt 1.7.3.2.2)

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	1693,941	120,663		14,039	,000	1454,549	1933,334
	Ernährungsindex	1604,020	121,261	,798	13,228	,000	1363,441	1844,599

a. Abhängige Variable: Brutto-Sozialprodukt pro Kopf

liefert UNIANOVA für den Steigungskoeffizienten denselben Schätzwert (1604,02), aber einen deutlich größeren Standardfehler (167,816 statt 121,261):

Parameterschätzungen mit robusten Standardfehlern

Abhängige Variable: Brutto-Sozialprodukt pro Kopf

Parameter	Regressions-koeffizient B	Robuster Standardfehler[a]	T	Sig.	95%-Konfidenzintervall Untergrenze	95%-Konfidenzintervall Obergrenze
Konstanter Term	1693,941	122,893	13,784	,000	1450,124	1937,758
ern	1604,020	167,816	9,558	,000	1271,077	1936,963

a. HC3-Methode

Um die Überlegenheit der HCSE-Technik im Vergleich zur Standardtechnik bei Vorliegen von Heteroskedastizität zu überprüfen, wurde bei der im Abschnitt 1.7.3.1 beschriebenen Simulationsstudie auch die Berechnung von HC3-Standardfehlern über die Prozedur UNIANOVA einbezogen. In der folgenden Tabelle ist die Standardabweichung der Steigungskoeffizienten aus 500 Stichproben sowie der Mittelwert der 500 per HCSE-Technik geschätzten Standardfehler zum Steigungskoeffizienten zu sehen:

Statistiken

		Regressionskoeffizient	Robuster Standardfehler
N	Gültig	500	500
	Fehlend	0	0
Mittelwert		,00814	,65993
Standardabweichung		,655421	,101209

Der Standardfehler wird im Mittel sehr präzise geschätzt, und die α-Fehlerrate (in der Tabelle nicht enthalten) erreicht exakt den erwarteten Wert von 5%.

1.7.3.4.2 Bootstrapping

In der amerikanischen Variante einer bekannten Münchhausen-Geschichte schafft es der Held, sich an den eigenen Stiefelriemen aus dem Sumpf zu ziehen. Dieser „Trick“ lieferte den Namen für eine durchaus solide statistische Schätz- und Testmethodologie, die erstmals von Efron (1982) ausformuliert worden ist.

Man behandelt die Stichprobe als *Population*, ermittelt durch Ziehen mit Zurücklegen zahlreiche Sekundärstichproben (z. B. 1000) mit derselben Größe wie die Original- bzw. Primärstichprobe, wobei in der Regel etliche Fälle *mehrfach* in einer Sekundärstichprobe vertreten sind. Aus jeder Sekundärstichprobe wird mit den üblichen Methoden (z. B. OLS) ein Schätzer für den interessierenden Parameter gewonnen, sodass man aus der Gesamtheit der Sekundärstichproben eine *empirische* Kennwerteverteilung erhält. Diese ersetzt die *theoretische* Stichprobenkennwerteverteilung, die auf Modellannahmen basiert. Aus der empirischen Stichprobenkennwerteverteilung lassen sich Vertrauensintervalle und Testentscheidungen konstruieren, die auch bei einer Verletzung der Varianzhomogenität und/oder der Normalität der Residuen gültig sind.

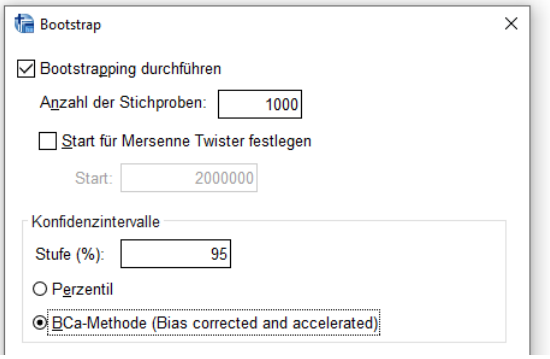
Lange war die benötigte Rechenleistung ein Hindernis für die Anwendung der Bootstrap-Technologie, doch mittlerweile suchen die CPU-Hersteller nach relevanten Anwendungen für ihre Gigahertz- und Multicore-Boliden.

In folgenden Situationen ist die sehr generell einsetzbare Inferenzstatistik mit Bootstrap-Methoden von Interesse:

- Bei grob verletzten Voraussetzungen sind darauf basierende Standardfehler potentiell verzerrt, und es droht eine unzuverlässige Inferenzstatistik. Mit Hilfe der Bootstrap-Technik gelangt man in vielen Fällen zu realistischeren Standardfehlern und inferenzstatistischen Ergebnissen.
- Für manche Statistiken lassen sich aus den Modellannahmen keine Standardfehler herleiten.

SPSS kann bei vielen Prozeduren das Erstellen von Bootstrap-Sekundärstichproben und die Zusammenfassung der Ergebnisse automatisieren. Wenn eine Prozedur das Bootstrapping unterstützt, ist ein entspre-

chender Schalter in ihrer Dialogbox vorhanden, z. B. bei der linearen Regression. Nach einem Klick auf den Schalter **Bootstrap** erscheint der folgenden Dialog:



Hier aktiviert man das Bootstrapping und wählt eine Anzahl von Sekundärstichproben (z. B. 1000). Ein Startwert für den Pseudozufallszahlengenerator (**Mersenne Twister**) macht das Bootstrap-Ergebnis reproduzierbar. Für die **Vertrauensintervalle** legt man des Konfidenzniveau fest und wählt in der Regel die **BCa-Methode** (*Bias corrected and accelerated*).

Mit den volkswirtschaftlichen Daten nach Kockläuner (1988) erhalten wir folgende Ergebnisse für die Regression des Pro-Kopf-Bruttosozialprodukts (BSP) auf den Ernährungsindex (ERN):

Bootstrap für Koeffizienten

Modell		Regressions-koeffizient B	Bootstrap[a] Verzerrung	Std.-Fehler	Sig. (2-seitig)	BCa 95% Konfidenzintervall Unterer Wert	Oberer Wert
1	(Konstante)	1693,941	-3,178	120,517	,001	1462,881	1939,975
	Ernährungsindex	1604,020	-2,206	161,510	,001	1294,750	1926,937

a. Sofern nicht anders angegeben, beruhen die Bootstrap-Ergebnisse auf 1000 Bootstrap-Stichproben

Im Vergleich zu den Ergebnissen der Regressionsprozedur liefert das Bootstrap-Verfahren für den Steigungskoeffizienten denselben Schätzwert, aber einen deutlich größeren Standardfehler (161,510 statt 121,261).

Um die Überlegenheit der Bootstrap-Schätzung im Vergleich zu Standardtechnik bei Vorliegen von Heteroskedastizität zu überprüfen, wurde bei der im Abschnitt 1.7.3.1 beschriebenen Simulationsstudie auch die Berechnung von Bootstrap-Standardfehlern einbezogen. In der folgenden Tabelle ist die Standardabweichung der Steigungskoeffizienten aus 500 Stichproben sowie der Mittelwert der 500 per Bootstrap-Technik geschätzten Standardfehler zum Steigungskoeffizienten zu sehen:

Statistiken

		Regressionskoeffizient	Bootstrap-Standardfehler
N	Gültig	500	500
	Fehlend	0	0
Mittelwert		,00814	,64370
Standardabweichung		,655421	,101722

Es zeigt sich, dass der Standardfehler im Mittel recht präzise geschätzt wird. Dementsprechend liegt die α-Fehlerrate mit 5,2 % sehr nahe am erwarteten Wert (5%).

Im Vergleich zur Technik der heteroskedastizitäts-konsistenten Standardfehler (siehe Abschnitt 1.7.3.4.1) bietet die Bootstrap-Technologie den Vorteil, auch die Normalverteilungsannahme „abzuschütteln“.

In Urban & Mayerl (2018, S. 364ff) wird die Durchführung einer Regressionsanalyse per Bootstrapping mit SPSS ausführlich beschrieben.

1.7.3.5 WLS-Regression

Bei der WLS-Regression[1] (*Weighted Least Squares*) versucht man zu einem neuen Modell mit homogenen Fehlervarianzen zu gelangen, indem man bei jeder Beobachtung die gesamte Regressionsgleichung (Kriterium, Regressor und Ordinatenabschnitt) mit dem Kehrwert

$$\frac{1}{\sqrt{\mathrm{Var}(\varepsilon_i)}}$$

der Fehlerstreuung aus dem ursprünglichen Modell multipliziert. Die resultierenden Parameterschätzungen sind erwartungstreu und effizient. Außerdem werden die Standardfehler zu den Regressionskoeffizienten erwartungstreu geschätzt (Long & Erwin 2000).

Während die im Abschnitt 1.7.3.4 beschriebenen robusten Verfahren die gewöhnlichen OLS-Schätzer zu den Regressionskoeffizienten berechnen und dazu einen trotz Heteroskedastizität korrekten Standardfehler liefern, leistet die WLS-Regression mehr: Man erhält eine gültige Inferenzstatistik *und* präzisere (effiziente) Schätzer zu den Regressionskoeffizienten.

Allerdings ist beim WLS-Verfahren der Aufwand höher, weil man Gewichtungsfaktoren $\frac{1}{\sqrt{\mathrm{Var}(\varepsilon_i)}}$ benötigt und folglich die Struktur der Heteroskedastizität kennen muss. In günstigen Fällen ist für eine Gewichtungsvariable W_i bekannt, dass sich die Residualvarianzen proportional zu einer Potenz von W_i (also proportional zu W_i^h) verhalten. Dann lassen sich die Fehlervarianzen homogenisieren, indem man beide Seiten der Regressionsgleichung durch $\sqrt{W_i^h}$ dividiert. Wenn z. B. für die Regressionsgleichung

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,\ i = 1,..,\ N$$

gilt:

$$\mathrm{Var}(\varepsilon_i) = \omega X_i^2,\ i = 1,..,N \tag{10}$$

dann sind zur Varianzhomogenisierung beide Seiten durch X_i zu dividieren.[2] Als Gewichtungsvariable fungiert hier X_i, und der Exponent *h* ist 2, sodass er beim Radizieren gerade verschwindet. Beim Dividieren durch X_i resultiert folgendes Modell:

$$\frac{Y_i}{X_i} = \beta_0 \frac{1}{X_i} + \beta_1 + \frac{\varepsilon_i}{X_i},\ i = 1,..,N \tag{11}$$

Im Vergleich zur oben angegebenen allgemeinen Form $\frac{1}{\sqrt{\mathrm{Var}(\varepsilon_i)}}$ konnten wir den Gewichtungsfaktor leicht vereinfachen, weil der konstante Anteil $\sqrt{\omega}$ gekürzt werden darf. In der neuen Regressionsgleichung tritt $\frac{Y_i}{X_i}$ als Kriterium, $\frac{1}{X_i}$ als Regressor und $\frac{\varepsilon_i}{X_i}$ als Residuum auf. Die Regressionskoeffizienten haben ihre üblichen Rollen getauscht, und die Fehlervarianzen im neuen Modell sind homogen:

$$\mathrm{Var}\left(\frac{\varepsilon_i}{X_i}\right) = \frac{1}{X_i^2}\mathrm{Var}(\varepsilon_i) = \frac{1}{X_i^2}\omega X_i^2 = \omega,\ i = 1,..,N$$

---

[1] Gute Darstellungen der WLS-Regression finden sich z.B. bei Draper & Smith (1981, S. 108ff) und bei Ryan (1997, S. 60ff).

[2] Bei der im Abschnitt 1.7.3.1 beschriebenen Simulationsstudie wurde übrigens exakt dieses Heteroskedastizitätsmuster angenommen.

Folglich kann man die Koeffizienten des neuen Modells mit der Kleinst-Quadrat - Methode effizient schätzen, und die Inferenzstatistik ist korrekt.

Im Allgemeinen führt der WLS-Lösungsansatz für ein einfaches Regressionsmodell zu einem multiplen Regressionsmodell (vgl. Abschnitt 2), das die beiden Regressoren

$$\frac{1}{W_i^{h/2}}, \quad \frac{X_i}{W_i^{h/2}}$$

enthält, aber *keinen* Ordinatenabschnitt (konstanten Term) besitzt.

Die Regression mit Fehlervarianzhomogenisierung durch eine Gewichtungsvariable mit geschätztem Exponenten ist in SPSS als eigene Prozedur realisiert, erreichbar über den Menübefehl

Analysieren > Regression > Gewichtungsschätzung

Um die WLS-Methode und das Verhalten der SPSS-Prozedur näher zu erkunden, führen wir eine Simulationsstudie passend zum eben diskutierten Beispiel mit

$$Y_i = X_i + \varepsilon_i,\ \mathrm{Var}(\varepsilon_i) = X_i^2,\ i = 1,..,N$$

durch. Dazu werden in einem Datenblatt mit der geöffneten Datendatei **Kockl.sav** die Variablen XS und YS durch die folgenden Kommandos ergänzt:[1]

```
compute xs = uniform(1).
compute ys = xs + normal(xs).
execute.
```

Im Plot der ausgelassen-studentisierten **OLS**-Residuen (*Ordinary Least Squares*) gegen die vorhergesagten Werte zeigt sich ein ausgeprägter Trichter:

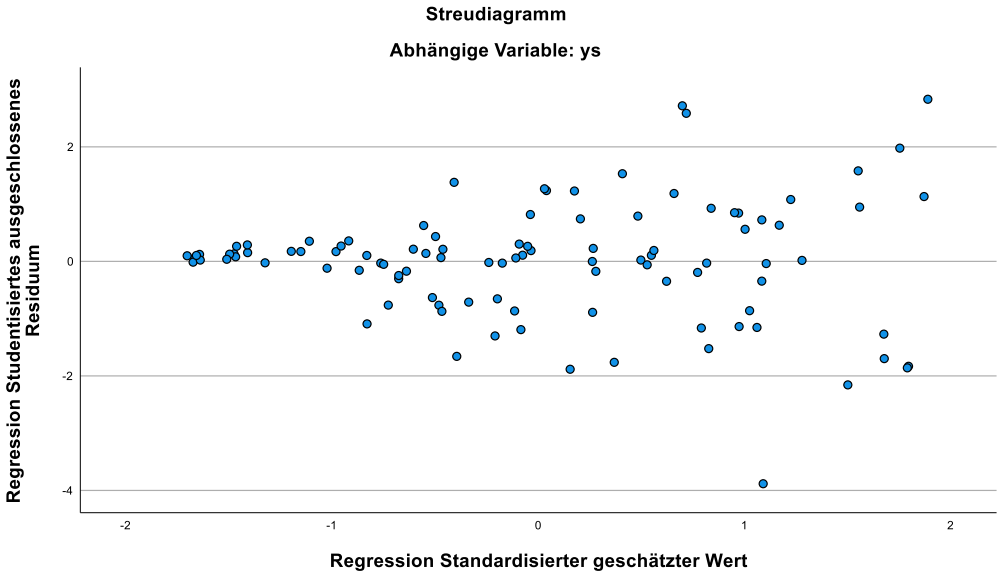


Abbildung 20: Ausgelassen-studentisierte Residuen vs. vorhergesagte Werte für eine OLS-Regression bei Varianzheterogenität mit $\mathrm{Var}(\varepsilon_i) = X_i^2,\ i = 1,..,N$

Auch das Testergebnis nach Breusch und Pagan (vgl. Abschnitt 1.7.3.2.4) spricht deutlich gegen die Homoskedastizitätsannahme:

[1] Die SPSS-Syntax zu allen anschließend beschriebenen Analysen befindet sich in der Datei **WLS-Sim.sps** an der im Vorwort verabredeten Stelle.

Modifizierter Breusch-Pagan-Test auf Heteroskedastizität[a,b,c]

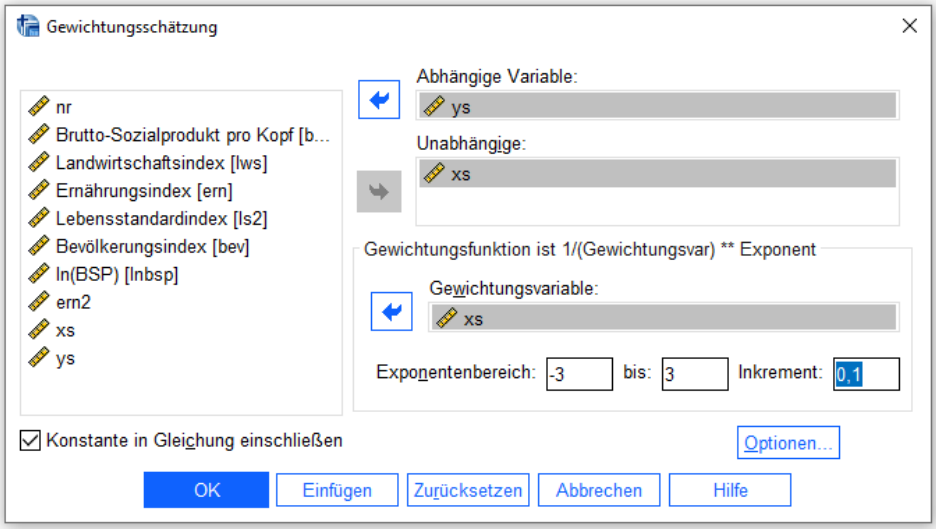
Chi-Quadrat	df	Sig.
21,175	1	,000

a. Abhängige Variable: ys

b. Testet die Nullhypothese, dass die Varianz der Fehler nicht von den Werte der unabhängigen Variablen abhängt.

c. Vorhergesagte Werte aus Design: Konstanter Term + xs

Die mit folgender Dialogbox



beauftragte WLS-Regressionsprozedur schätzt zunächst die optimale Potenz zur Gewichtungsvariablen und kommt dabei zum korrekten Ergebnis 2 (betragsmäßig kleinster **Log-Likelihood-Wert**):

Log-Likelihood-Werte[b]

Exponent	-3,000	-191,148
	-2,900	-186,861
	.	.
	.	.
	1,800	-31,963
	1,900	-31,226
	2,000	**-30,894**[a]
	2,100	-31,048
	2,200	-31,779
	.	.
	.	.
	2,900	-61,107
	3,000	-69,418

a. Der entsprechende Exponent wird für die weitere Analyse ausgewählt, weil dieser Wert den Wert der Log-Likelihood-Funktion maximiert.

b. Abhängige Variable: ys, Quellvariable: xs

Anschließend läuft der Algorithmus gemäß Formel (11) weiter, d.h. die ursprüngliche Regressionsgleichung wird durch XS dividiert:

Modellbeschreibung

Abhängige Variable		ys
Unabhängige Variablen	1	xs
Gewichtung	Quelle	xs
	Exponentenwert	2,000

Modell: MOD_1.

Modellzusammenfassung

Multiples R	,722
R-Quadrat	,521
Korrigiertes R-Quadrat	,517
Standardfehler des Schätzers	,946
Wert der Log-Likelihood-Funktion	-30,894

ANOVA

	Quadratsumme	Freiheitsgrade	Mittel der Quadrate	F	Sig.
Regression	97,438	1	97,438	108,924	,000
Residuum	89,455	100	,895		
Insgesamt	186,893	101			

Koeffizienten

	Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			
	B	Std.-Fehler	Beta	Std.-Fehler	t	Sig.
(Konstante)	,009	,009			1,052	,295
xs	1,102	,106	,722	,069	10,437	,000

In einer zur Probe gerechneten OLS-Regression mit dem Kriterium:

$$Y_W := \frac{YS}{XS}$$

und dem Regressor

$$C_W := \frac{1}{XS}$$

erhält man als *Ordinatenabschnitt* (siehe Gleichung (11)) exakt den geschätzten Steigungskoeffizienten aus der WLS-Ausgabe:

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	1,102	,106		10,437	,000	,893	1,312
	c_w	,009	,009	,105	1,052	,295	-,008	,026

a. Abhängige Variable: y_w

Der Test nach Breusch & Pagan (vgl. Abschnitt 1.7.3.2.4) plädiert nun mit einer Überschreitungswahrscheinlichkeit von 0,777 *für* die Homoskedastizität:

Modifizierter Breusch-Pagan-Test auf Heteroskedastizität[a,b,c]

Chi-Quadrat	df	Sig.
,080	1	,777

a. Abhängige Variable: y_w

b. Testet die Nullhypothese, dass die Varianz der Fehler nicht von den Werte der unabhängigen Variablen abhängt.

c. Vorhergesagte Werte aus Design: Konstanter Term + c_w

Der Standardfehler zum Steigungskoeffizienten ist in der WLS-Regression nur ca. halb so groß (!) wie der korrespondierende Standardfehler aus einer OLS-Regression unter Vernachlässigung der Varianzheterogenität:

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	-,031	,103		-,304	,762	-,235	,172
	xs	1,200	,191	,532	6,277	,000	,821	1,579

a. Abhängige Variable: ys

Tatsächlich liegt die WLS-Schätzung mit 1,10 näher am korrekten Wert 1,0 als die OLS-Schätzung (1,20).

Die UNIANOVA-Prozedur (siehe Abschnitt 1.7.3.4.1) liefert als Schätzer den ungenauen OLS-Wert und berechnet optional einen Heteroskedastizitäts-korrigierten Standardfehler von 0,224:

Parameterschätzungen mit robusten Standardfehlern

Abhängige Variable: ys

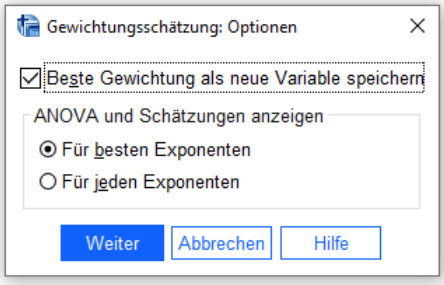
Parameter	Regressions-koeffizient B	Robuster Standardfehler[a]	T	Sig.	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
Konstanter Term	-,031	,074	-,421	,675	-,178	,116
xs	1,200	,224	5,354	,000	,755	1,644

a. HC3-Methode

Im Vergleich zu UNIANOVA liefert die WLS-Methode ...

- eine genauere Schätzung des Regressionskoeffizienten
- einen kleineren Standardfehler, also ein kleineres Vertrauensintervall und eine größere Power beim Signifikanztest.

Wenn man bereits eine Variable mit geeigneten Gewichten besitzt, kann man diese übrigens auch in der SPSS-Prozedur REGRESSION zur **WLS-Gewichtung** verwenden. Die eben vorgestellte WLS-Prozedur erlaubt in der **Optionen**-Subdialogbox ein Speichern der von ihr ermittelten Gewichte:



1.7.4 Normalverteilung der Residuen

Wir haben bei den Untersuchungen zur Modellgültigkeit die Reihenfolge mit Bedacht gewählt, weil bei verletzter Linearität oder Varianzhomogenität eine Untersuchung der Residuen auf Normalverteilung nicht sinnvoll ist.

1.7.4.1 Relevanz der Normalverteilungsvoraussetzung und zentraler Grenzwertsatz

Abweichungen von der Residual-Normalverteilung haben oft keine gravierenden Auswirkungen auf die Ergebnisse einer Regressionsanalyse (Norušis 2005, S. 231). Mit wachsender Stichprobengröße neutralisiert die segensreiche Wirkung des zentralen Grenzwertsatzes eine Störung der Inferenzstatistik durch Abweichungen von der Normalverteilung der Residuen (siehe z. B. Fahrmeir et al. 2007, S. 105). Nach Bühner & Ziegler (2009, S. 674) ist bei *Abwesenheit von Ausreißern* ab 100 Fällen eine Verletzung der Normalverteilung der Residuen „weniger problematisch".

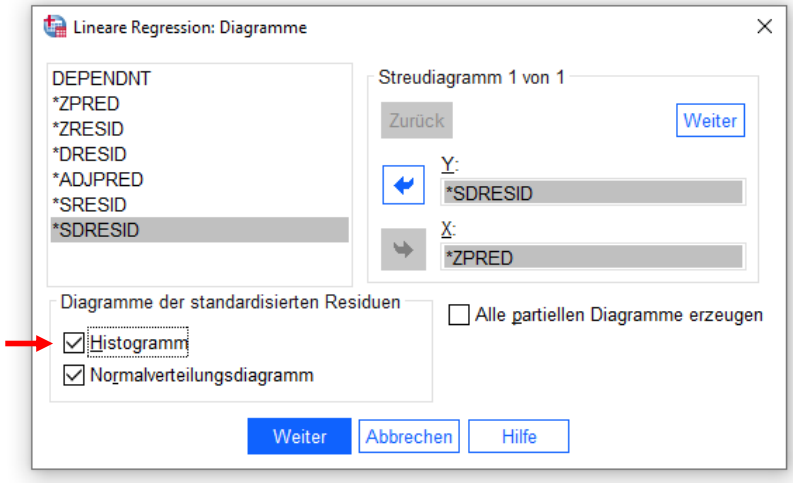
1.7.4.2 Gemeinsame Prüfung für alle Ausprägungen des Regressors

Eigentlich muss die Normalität der Residuen für jeden in die Studie einbezogenen Wert des Regressors separat überprüft werden. In der Regel (speziell bei Anwesenheit eines *metrischen* Regressors) sind aber die zugehörigen Teilstichproben zu klein, um eine Normalitätsbeurteilung zu ermöglichen. Daher wird meist eine gemeinsame Beurteilung *aller* Residuen vorgenommen (vgl. Urban & Mayerl 2018, S. 187).

Damit beschränkt man sich auf die Prüfung einer *notwendigen* Bedingung für das Bestehen der Normalitätsvoraussetzung. Es ist möglich, aber unwahrscheinlich, dass sich „abnorme" Verteilungen in den Teilpopulationen zu einer gemeinsamen Normalverteilung kombinieren.

1.7.4.3 Auswahl der zu prüfenden Residuen

In der Regressions-Subdialogbox **Diagramme** kann man für die *standardisierten* Residuen Diagramme zur Normalitätsbeurteilung anfordern:



Die standardisierten Residuen sind jedoch auch bei Gültigkeit aller Modellannahmen *nicht* perfekt varianzhomogen (vgl. Abschnitt 1.7.2.2) und damit auch zur Prüfung der Normalitätsvoraussetzung nicht ideal geeignet. Allerdings wirken sich die theoretischen Mängel der standardisierten Residuen in der Praxis kaum aus. Wer jedoch auf maximale Präzision Wert legt und den Umgang mit SPSS-Syntax nicht scheut, kann sehr einfach das von der SPSS-Regressionsdialogbox über den Schalter **Einfügen** erzeugte REGRESSION-Kommando so abändern, dass statt der standardisierten Residuen (interner Variablenname

ZRESID) die nach den Ausführungen auf Seite 39 zu bevorzugenden ausgelassen-studentisierten Residuen untersucht werden (interner Variablenname SDRESID):

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT gewicht
  /METHOD=ENTER groesse
  /RESIDUALS HIST(ZRESID) NORM(ZRESID) .
```

→

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT gewicht
  /METHOD=ENTER groesse
  /RESIDUALS HIST(SDRESID) NORM(SDRESID) .
```

Alternativ kann man die ausgelassen-studentisierten Residuen per Regressions-Dialogbox als neue Variable in die Arbeitsdatei schreiben lassen (Schalter bzw. Subdialog **Speichern**) und dann mit der Prozedur zur explorativen Statistik (Menübefehl: **Analysieren > Deskriptive Statistiken > Explorative Datenanalyse**) auf Normalität untersuchen. Dabei werden auch Signifikanztests zur Normalverteilungs-Nullhypothese durchgeführt (siehe Abschnitt 1.7.4.5).

1.7.4.4 Grafische Diagnosemethoden

Die anschließend präsentierten Diagramme zu den *ausgelassen-studentisierten* Residuen aus der Regression von GEWICHT auf GROESSE (für $N = 159$ Männer) wurden durch das im Abschnitt 1.7.4.3 beschriebene REGRESSION-Kommando erstellt.

Das **Histogramm** mit eingezeichneter Normalverteilungsdichte zeigt sich auf den ersten Blick ein akzeptables Bild, das nur von drei relativ großen Residuen am rechten Rand leicht gestört wird:

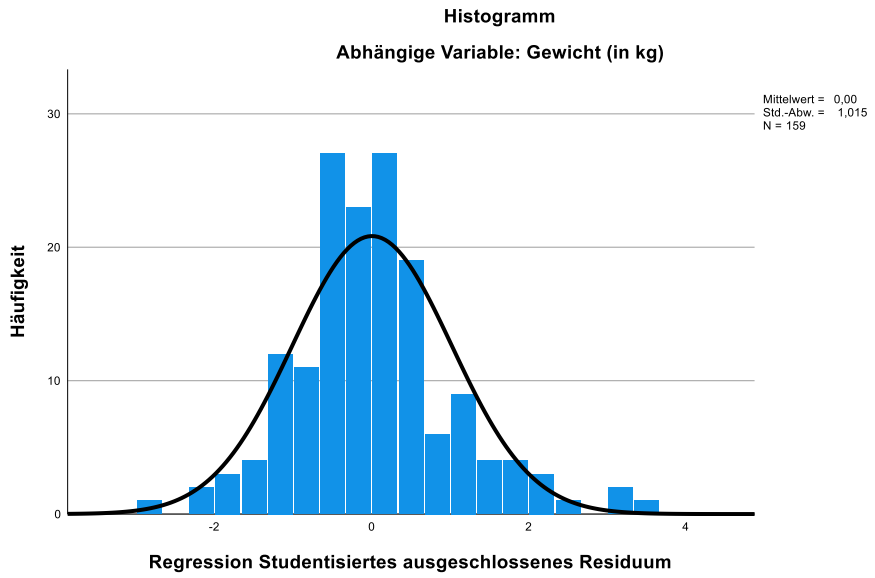


Abbildung 21: Histogramm der ausgelassen-studentisierten Residuen
aus der Regression von GEWICHT auf GROESSE (bei $N = 159$ Männern)

Das **Normalverteilungsdiagramm** kommt folgendermaßen zu Stande: SPSS berechnet zunächst zu jedem geschätzten Residuum r_i , das bei gültigem Modell den Erwartungswert 0 und die Varianz 1 hat, den Wert der empirischen Verteilungsfunktion ev_i nach folgender Formel (Blom-Anteilschätzung):

$$ev_i := \frac{\mathrm{rg}(r_i) - 3/8}{N + 1/4}$$

Dabei ist $\mathrm{rg}(r_i)$ der Rang von r_i in der Stichprobe. Außerdem wird der Wert der Standardnormalverteilungsfunktion an der Stelle r_i berechnet:

$$tv_i := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r_i} e^{-\frac{x^2}{2}} dx$$

Wenn etwa bei N = 100 ein Residuum r_i = -1,3 den Rangplatz 10 einnimmt, d.h. genau 9 andere Residuen übertrifft, denn resultiert

$$ev_i = \frac{10 - 3/8}{100 + 1/4} = 0{,}0960$$

Die Standardnormalverteilung verteilt unterhalb von -1,3 die Masse

$$tv_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-1{,}3} e^{-\frac{x^2}{2}} dx = 0{,}0968$$

Sind die Residuen modellgemäß standardnormalverteilt, dann gilt bis auf zufällige Stichprobenschwankungen:

$$ev_i = tv_i$$

Die geplotteten Wertepaare (ev_i, tv_i) sollten also annähernd auf einer Ursprungsgeraden mit Steigung 1 liegen. Bei der Interpretation eines Wahrscheinlichkeitsdiagramms brauchen Sie nur diese Regel zu kennen. Für unser Beispiel erhalten wir ein Bild, das erkennbar vom Ideal abweicht:

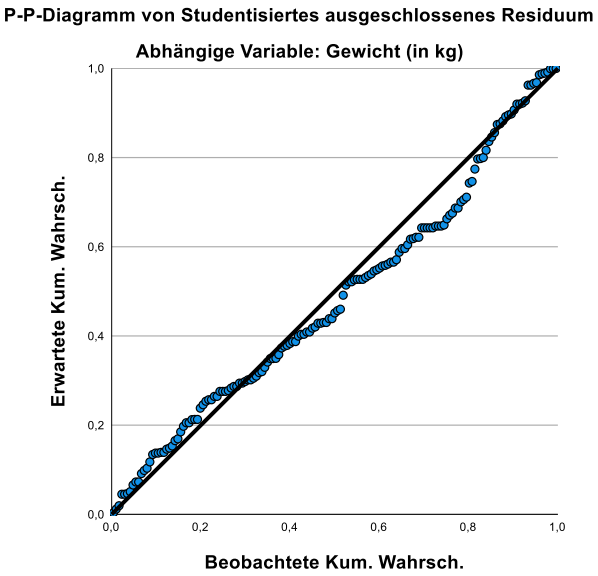
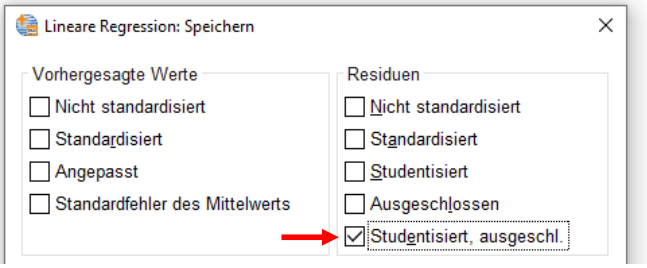


Abbildung 22: P-P - Diagramm der ausgelassen-studentisierten Residuen aus der Regression von GEWICHT auf GROESSE (bei N = 159 Männern)

Nach Abschnitt 1.7.2.3 folgen die ausgelassen-studentisierten Residuen bei einem gültigen Modell mit k Regressoren einer t-Verteilung mit N - k - 2 Freiheitsgraden, sodass im Wahrscheinlichkeitsplot statt der Standardnormalverteilung eigentlich diese t-Verteilung als Vergleichsmaßstab dienen müsste. Bei nicht allzu kleinen Stichproben (N - k - 2 ≥ 30) stimmen die beiden Verteilungen allerdings nahezu perfekt überein (Fox 1997, S. 296).

1.7.4.5 Statistische Diagnosemethoden

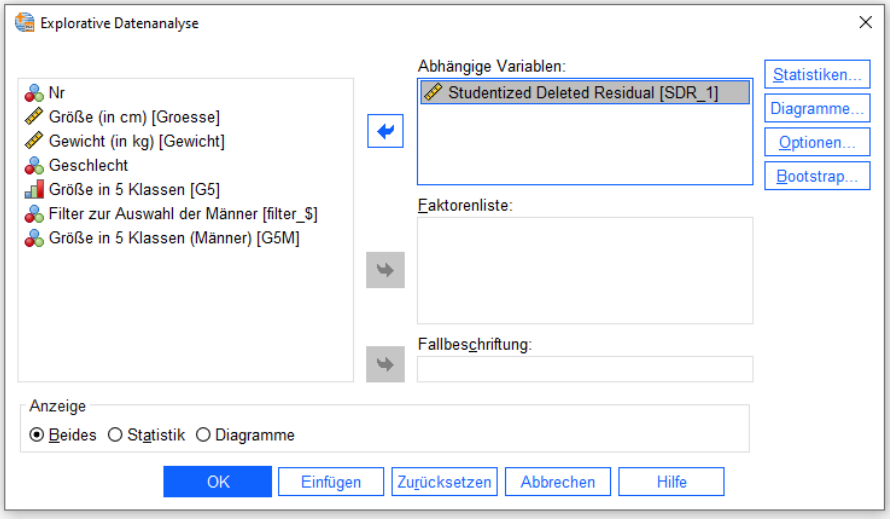
Mit der Prozedur zur explorativen Datenanalyse können in SPSS auch formale Tests der Normalitätsannahme durchgeführt werden. Man lässt zunächst die gewünschten Residuen von der Regressionsprozedur über die Subdialogbox **Speichern** als neue Variable in die Arbeitsdatei schreiben, wobei in der Regel die ausgelassen-studentisierten Residuen zu bevorzugen sind:



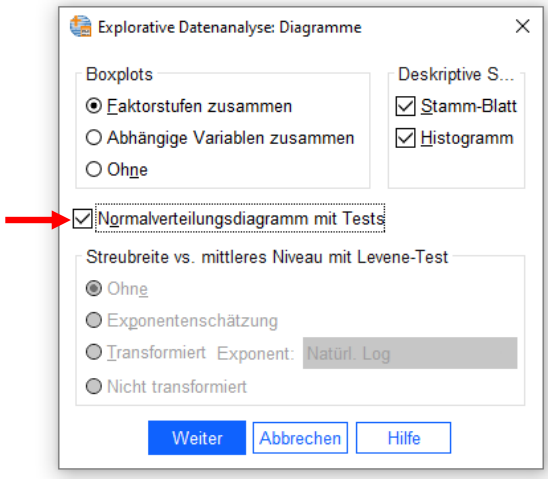
Nach dem Start der Dialogbox zur explorativen Datenanalyse mit

Analysieren > Deskriptive Statistik > Explorative Datenanalyse

wählt man die zu untersuchende Variable



und fordert die Normalverteilungsbeurteilung in der Subdialogbox **Diagramme** an:



Für die Residuen aus der Regression von Gewicht auf Größe in der männlichen Teilstichprobe wird die Nullhypothese der Normalverteilung von den beiden durchgeführten Tests (Kolmogorov-Smirnov und Shapiro-Wilk) abgelehnt:

Tests auf Normalverteilung

	Kolmogorov-Smirnov [a]			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Studentized Deleted Residual	,103	159	,000	,966	159	,001

a. Signifikanzkorrektur nach Lilliefors

Die beiden Normalverteilungsanpassungstests kommen nicht selten zu unterschiedlichen Entscheidungen (siehe Abschnitt 1.7.4.6.1). Nach Brosius (2013, S. 405) ist vor allem bei kleineren Stichproben ($N < 50$) das Verfahren von Shapiro-Wilk zu bevorzugen.

Bei der Interpretation eines Normalverteilungsanpassungstests ist generell zu berücksichtigen (vgl. Abschnitt 1.2.4):

- Die Normalverteilungsannahme ist streng genommen fast immer verletzt.
- Ob der Defekt zu einem signifikanten Ergebnis in einem Normalverteilungsanpassungstest führt, hängt wesentlich von der Stichprobengröße ab, aber auch von der Defektstärke und vom gewählten Signifikanztest.

In den meisten empirischen Studien sind für die mit hoher Wahrscheinlichkeit bestehende Verletzung der Normalverteilungsannahme die folgenden Fragen zu klären:

- Wie gravierend ist die Abweichung von der Normalität?
- Kann auf die Robustheit der Inferenzstatistik gegenüber der vorliegenden Verletzung vertraut werden?

Zur Klärung der Frage nach dem Ausmaß einer Normalverteilungsverletzung kann ein Signifikanztest trotz seiner Abhängigkeit von der Stichprobengröße einen Beitrag leisten:

- Eine signifikante Entscheidung gegen die Nullhypothese der Normalität ist umso gravierender, je kleiner die Stichprobe und damit die Power des Tests ist.
- Eine ausbleibende Signifikanz gewinnt an Bedeutung mit der Stichprobengröße.

Oft helfen die Statistiken zur Schiefe und Wölbung der empirischen Residualverteilung beim Vergleich mit der Normalverteilung. SPSS liefert sie z.B. in der Prozedur zur explorativen Datenanalyse zusammen mit ihren Standardfehlern:

Deskriptive Statistik

			Statistik	Std.-Fehler
Studentized Deleted Residual	Mittelwert		,0022727	,08047684
	95% Konfidenzintervall des Mittelwerts	Untergrenze	-,1566764	
		Obergrenze	,1612219	
	5% getrimmtes Mittel		-,0359083	
	Median		-,1218938	
	Varianz		1,030	
	Std.-Abweichung		1,01477430	
	Minimum		-2,70062	
	Maximum		3,50708	
	Spannweite		6,20770	
	Interquartilbereich		1,01515	
	Schiefe		,676	,192
	Kurtosis		1,334	,383

Beide Statistiken sind bei einer Normalverteilung gleich null. Bei einem α-Fehlerrisiko von 5% ist die *ungerichtete* Nullhypothese, dass die Schiefe in der Population gleich 0 sei, zu verwerfen, wenn gilt:

$$\frac{|\text{Schiefe}|}{\text{SF(Schiefe)}} > 1{,}96$$

Analog lässt sich ein Test zur Wölbung (von SPSS als *Kurtosis* bezeichnet) in der Population durchführen. In unserem Beispiel sind beide Tests signifikant. Wir sind allerdings aktuell an Kriterien zur Verteilungsbeurteilung interessiert, die *nicht* von der Stichprobengröße abhängig sind. Nach verbreiteter Auffassung kann eine Abweichung von der Normalverteilung als geringfügig eingeschätzt werden, wenn die Beträge der Schiefe- und der Wölbungsstatistik unter 1 bleiben (siehe z.B. Urban & Mayerl 2018, S. 190). In unserem Beispiel liegt die Wölbungsstatistik über diesem Grenzwert.

Nach dem zentralen Grenzwertsatz der Statistik folgen die Stichprobenverteilungen der geschätzten Regressionskoeffizienten auch ohne Normalverteilung der Residuen approximativ der erwarteten t-Verteilung, wenn die Stichprobe hinreichend groß ist (siehe Abschnitt 1.7.4.1). Bühner & Ziegler (2009, S. 674) rechnen ab 100 Fällen für die Inferenzstatistik der Regressionsanalyse mit einer Robustheit gegenüber einer Normalverteilungsverletzung, sofern keine Ausreißer vorhanden sind.

Im Abschnitt 1.7.4.6.1 wird sich zeigen, dass es in unserem Beispiel *keine* kluge Strategie wäre, die Fälle mit den großen ausgelassen-studentisierten Residuen zu ignorieren und auf den zentralen Grenzwertsatz zu vertrauen. Ausreißer sollten im Manuskript eigentlich aus didaktischen Gründen erst im Abschnitt 3.1 behandelt werden. Bei der Normalverteilungsbeurteilung im Gewicht/Größe - Beispiel ist aber ein Vorgriff auf die Ausreißerbehandlung nicht zu vermeiden.

1.7.4.6 Maßnahmen bei einer inakzeptablen Verletzung der Normalität

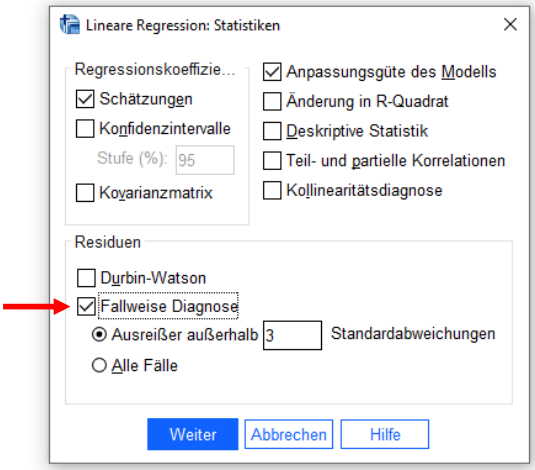
Wenn die Annahme normalverteilter Residuen wesentlich verletzt ist, und ein Vertrauen auf den zentralen Grenzwertsatz der Statistik nicht gerechtfertigt erscheint, kommen die folgenden Maßnahmen in Frage:

- Elimination von Ausreißern (siehe Abschnitte 1.7.4.6.1 und 3.1.1)
- Transformation des Kriteriums (siehe Abschnitt 1.7.4.6.2)
- Verwendung der Bootstrap-Technik (siehe Abschnitt 1.7.3.4.2 sowie Fox 2008, Kap. 21)
- Verwendung eines generalisierten linearen Modells
  Bei einer Zählvariablen als Kriterium (z. B. Anzahl der im letzten Jahr gelesenen Bücher) ist oft ein generalisiertes lineares Modell mit der Annahme einer Poisson-Verteilung oder einer negativen Binomialverteilung für die Residuen besser geeignet als ein lineares Modell, das von normalverteilten Residuen ausgeht (siehe Baltes-Götz 2016a). Eventuell muss man auf ein parametrisches Verteilungsmodell für die Residuen verzichten und die ordinale logistische Regression verwenden (siehe Baltes-Götz 2012).

1.7.4.6.1 Elimination von Ausreißern

In der Verteilung der ausgelassen-studentisierten Residuen aus der Regression von Gewicht auf Größe sind uns 3 Fälle mit großen Residuen (> 3) aufgefallen. Wie die ausführliche Behandlung der bisher aus didaktischen Gründen ausgesparten Ausreißerthematik im Abschnitt 3.1.1 ergeben wird, kann man bei einem ausgelassen-studentisierten Residuum > 3 von einem Ausreißer ausgehen.

Um per SPSS-Regressionsprozedur eine Liste der ausgelassen-studentisierten Residuen mit Beträgen größer 3 zu erhalten, fordert man zunächst in der **Statistiken**-Subdialogbox eine **fallweise Diagnose** mit dem gewünschten **Ausreißer**-Kriterium an:



Anschließend lässt man sich in der Hauptdialogbox per **Einfügen**-Schalter die korrespondierende Syntax erzeugen und fordert im Subkommando **CASEWISE** ausgelassen-studentisierte Residuen (interner Name: SDRESID) an Stelle der standardisierten Residuen (interner Name: ZRESID) an:

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS CI(95)
   R ANOVA COLLIN TOL ZPP
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT Gewicht
 /METHOD=ENTER Groesse
/CASEWISE PLOT(ZRESID) OUTLIERS(3).
```

→

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS CI(95)
   R ANOVA COLLIN TOL ZPP
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT Gewicht
/METHOD=ENTER Groesse
/CASEWISE PLOT(SDRESID) OUTLIERS(3).
```

Im Beispiel resultiert die folgende Liste:

Fallweise Diagnose[a]

Fallnummer	Stud. gelöschtes Residuum	Gewicht (in kg)	Nicht standardisierter vorhergesagter Wert	Nicht standardisierte Residuen
19	3,150	90	67,14	22,858
85	3,122	85	62,55	22,450
314	3,507	110	84,59	25,406

a. Abhängige Variable: Gewicht (in kg)

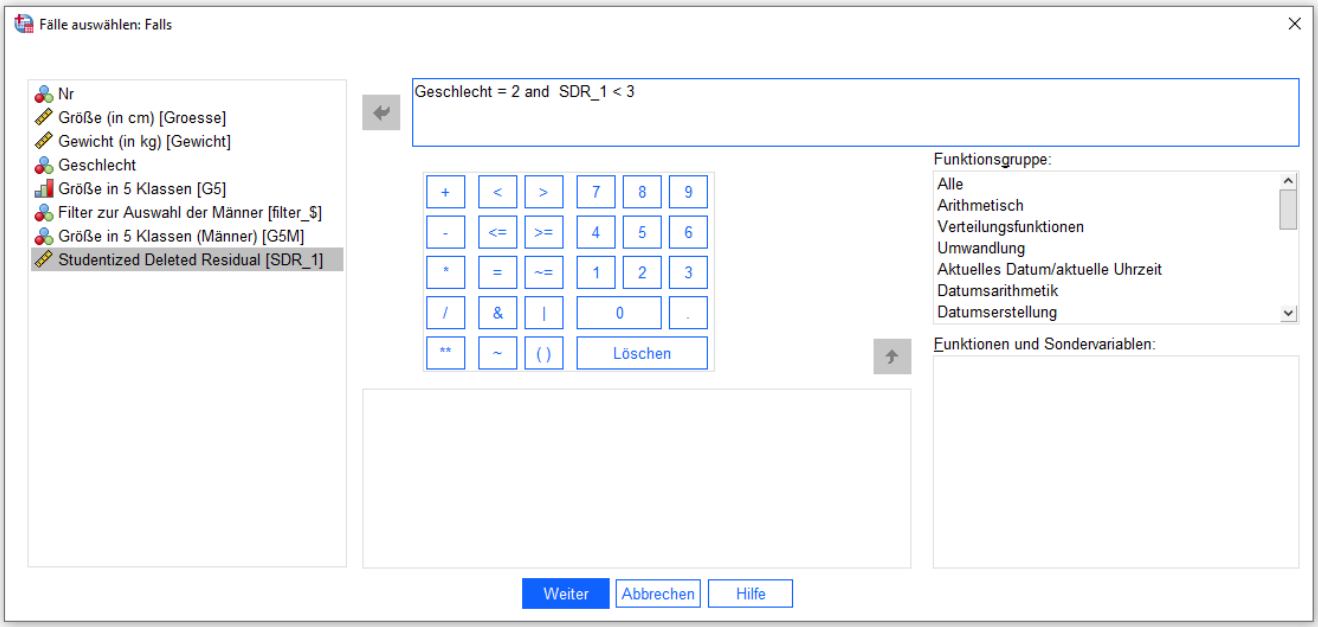
Im Streudiagramm von Gewicht und Größe fallen die drei Personen, die relativ zu ihrer Größe sehr schwer sind, deutlich auf:

Abbildung 23: Extreme ausgelassen-studentisierte Residuen in der Regression von GEWICHT auf GROESSE (bei N = 159 Männern)

Schließt man diese Personen nach dem Menübefehl

Daten > Fälle auswählen

über die folgende Teilnahmebedingung



aus, dann wächst der geschätzte Steigungskoeffizient von Größe erheblich von 0,92 auf 0,98:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	-101,343	14,876		-6,812	,000	-130,730	-71,955
	Größe (in cm)	,984	,082	,696	12,027	,000	,822	1,146

a. Abhängige Variable: Gewicht (in kg)

Außerdem steigt der korrigierte Determinationskoeffizient von 0,41 auf 0,48:

Modellzusammenfassung[b]

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,696 [a]	,484	,481	6,847

a. Einflußvariablen : (Konstante), Größe (in cm)

b. Abhängige Variable: Gewicht (in kg)

Die Anwesenheit der 3 Fälle hat den geschätzten Steigungskoeffizienten und die geschätzte Erklärungsleistung des Modells drastisch gemindert. Das Problem zu ignorieren und auf den zentralen Grenzwertsatz zu vertrauen, wäre hier *keine* kluge Strategie.

Eine nähere Inspektion ergibt, dass die drei betroffenen Fälle (und nur diese) einen Body Mass Index über 30 besitzen, sodass von einer Adipositas auszugehen ist.[1] Wir haben die Ausreißer über ein statistisches Kriterium identifiziert und können nun über ein inhaltliches Kriterium begründen, dass sie aus der Studie ausgeschlossen werden sollten, weil es in der Studie um die Regression von Gewicht auf Größe in der normalgewichtigen Population geht. Im Forschungsbericht müssen solche Maßnahmen dokumentiert und begründet werden.

Das Histogramm der ausgelassen-studentisierten Residuen aus einer Regression *ohne* die mutmaßlichen Adipositas-Fälle zeigt eine verbesserte Normalverteilungs-Approximation:

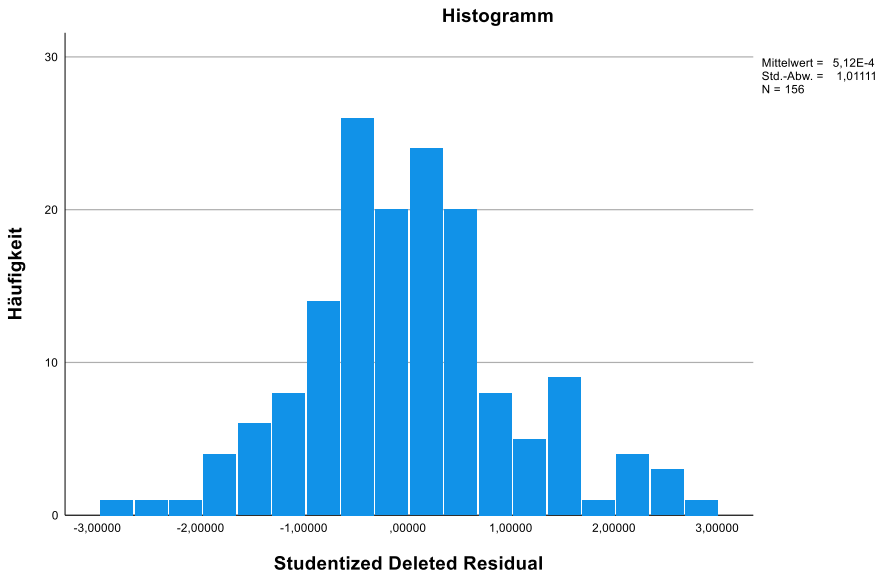


Abbildung 24: Histogramm der ausgelassen-studentisierten Residuen im Größe/Gewicht - Beispiel nach der Elemination von 3 Ausreißern (bei N = 156 Männern)

Die Beträge der Schiefe- und der Wölbungsstatistik sind deutlich kleiner als 1,

[1] https://www.adipositas-gesellschaft.de/index.php?id=39

Deskriptive Statistik

			Statistik	Std.-Fehler
Studentized Deleted Residual	Mittelwert		,0005122	,08095345
	95% Konfidenzintervall des Mittelwerts	Untergrenze	-,1594022	
		Obergrenze	,1604266	
	5% getrimmtes Mittel		-,0155738	
	Median		-,1065290	
	Varianz		1,022	
	Std.-Abweichung		1,01110832	
	Minimum		-2,95747	
	Maximum		2,73008	
	Spannweite		5,68756	
	Interquartilbereich		1,03939	
	Schiefe		,251	,194
	Kurtosis		,512	,386

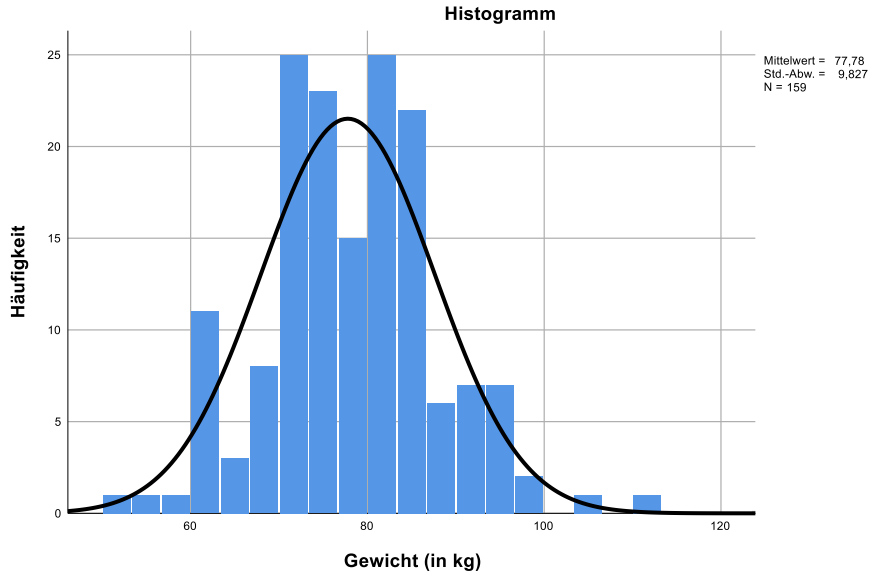
und der Normalverteilungsanpassungstest nach Shapiro-Wilk verwirft seine Nullhypothese *nicht* mehr:

Tests auf Normalverteilung

	Kolmogorov-Smirnov [a]			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Studentized Deleted Residual	,082	156	,012	,984	156	,070

a. Signifikanzkorrektur nach Lilliefors

Wir haben bei der Normalitätsprüfung 3 Ausreißer hinsichtlich der *Modellresiduen* entdeckt (die Fälle mit den Nummern 19, 85 und 314). Bei der univariaten Verteilungsanalyse für die Kriteriumsvariable fällt nur der Fall 314 auf:



Es gilt generell, dass eine univariate Ausreißeranalyse für die abhängige Variable von begrenztem Wert ist. Auf diese Weise gelingt es keinesfalls, das Problem der Ausreißer *vor* Beginn der eigentlichen Modellprüfung oder Modellsuche aus der Welt zu schaffen.

1.7.4.6.2 Box-Cox - Transformation zur Normalisierung der Residualverteilung

Im volkswirtschaftlichen Beispiel zeigt der P-P - Plot für die ausgelassen-studentisierten Residuen aus der Regression von BSP (Pro-Kopf-Bruttosozialprodukt) auf ERN (Ernährungsindex) kein überzeugendes Bild:[1]

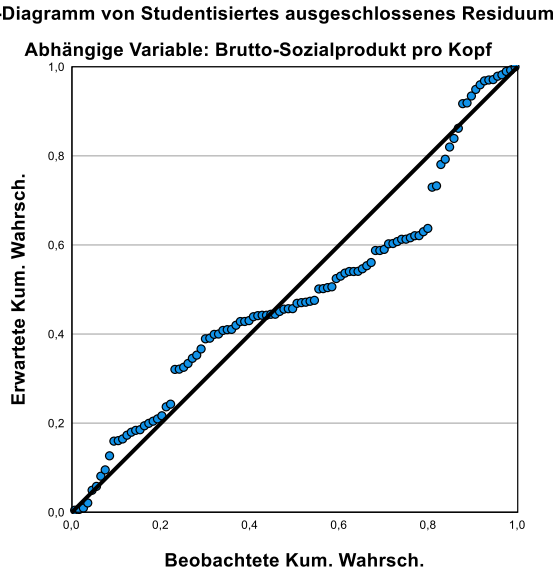


Abbildung 25: P-P - Diagramm der ausgelassen-studentisierten Residuen aus der Regression von BSP auf ERN im volkswirtschaftlichen Beispiel

Die mit den abgespeicherten ausgelassen-studentisierten Residuen durchgeführten Normalverteilungsanpassungstests sprechen sich deutlich gegen die Nullhypothese aus:

Tests auf Normalverteilung

	Kolmogorov-Smirnov [a]			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Studentized Deleted Residual	,171	102	,000	,943	102	,000

a. Signifikanzkorrektur nach Lilliefors

Außerdem ist der von Bühner & Ziegler (2009, S. 674) angegebene minimale Stichprobenumfang von $N = 100$ für die Robustheit der linearen Regression gegen Verletzungen der Normalverteilungsvoraussetzung nur knapp erreicht.

Es soll daher versucht werden, durch eine Box-Cox - Transformation des Kriteriums (siehe z. B. Fox 1997, S. 322f) die Verteilung der Residuen zu normalisieren. Dabei wird ein Exponent λ so bestimmt, dass die neue Kriteriumsvariable

$$\begin{cases} \frac{Y_i^{\lambda}}{\boldsymbol{\lambda}} & \text{für } \lambda \neq 0 \\ \ln(Y_i) & \text{für } \lambda = 0 \end{cases}$$

zu einer Residualverteilung mit möglichst großer Ähnlichkeit zur Normalverteilung führt.

Leider ist die Schätzung von λ in SPSS nur auf Umwegen möglich. Wer sich nicht als Makro-Programmierer betätigen möchte, installiert am besten die **R-Integration** für SPSS (siehe Anhang) und

[1] Für das volkswirtschaftliche Beispiel ist zur Homogenisierung der Fehlervarianzen bereits eine logarithmische Transformation des Kriteriums vorgesehen (vgl. Abschnitt 1.7.3.3). Wir ignorieren diesen Fortschritt der Modellierung vorübergehend, um den Nutzen der Box-Cox - Transformation demonstrieren zu können.

nutzt das von John Fox (siehe Fox & Weisberg 2011) erstellte **car** - Paket der freien Programmiersprache **R**. Wer über den Menübefehl

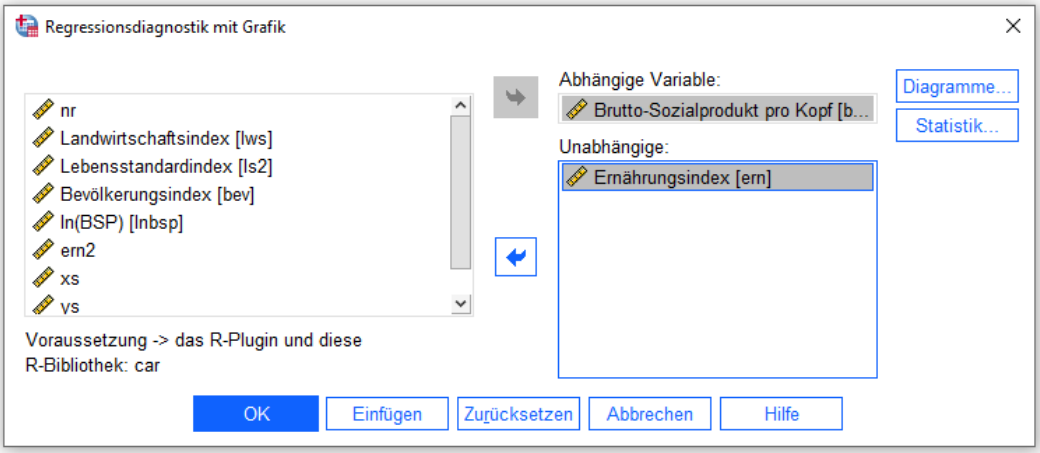
Erweiterungen > Erweiterungshub

die Erweiterung **RRegDiagGraph** von Hans Grüner installiert (siehe Anhang), der kann diverse von John Fox programmierte Diagnoseverfahren für Regressionsmodelle bequem über den Menübefehl

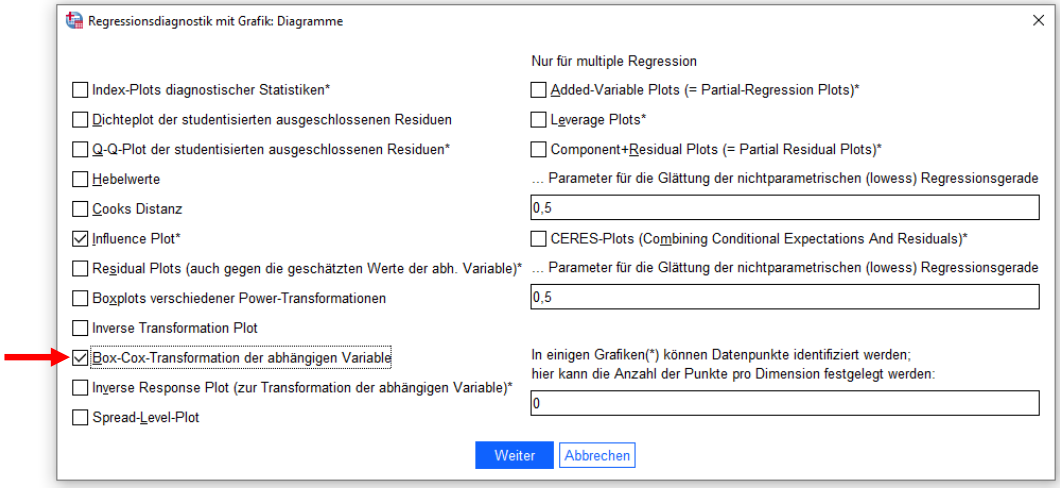
Analysieren > Regression > Regressionsdiagnostik mit Grafik

anfordern.

Wir betrachten das volkswirtschaftliche Modell in der ursprünglichen Form (Regression von BSP auf ERN), dessen Residuen schon als ausgeprägt heterogen aufgefallen sind (vgl. Abschnitt 1.7.3.2). Im Hauptdialog der **Regressionsdiagnostik** wird das Modell auf gewohnte Weise spezifiziert:



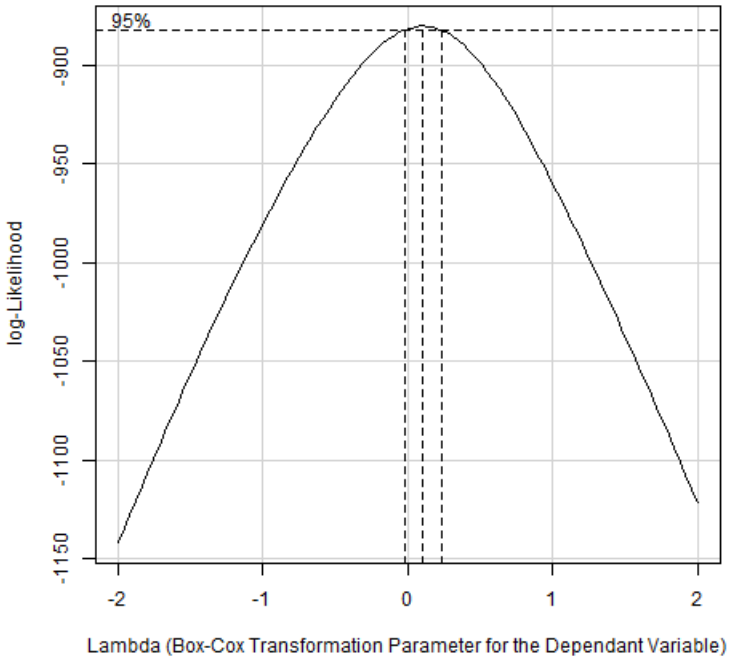
Im **Diagramme**-Subdialog kann u.a. die **Box-Cox - Transformation der abhängigen Variablen** angefordert werden:



Wir erhalten einen Schätzwert für λ samt Vertrauensintervall

```
bcPower Transformation to Normality
   Est Power Rounded Pwr Wald Lwr bnd Wald Upr Bnd
Y1    0.1055           0       -0.019       0.2299
```

und ein Diagramm, das den optimalen λ-Wert samt Vertrauensintervall zeigt:



Das ermittelte Vertrauensintervall [-0,019; 0,2299] enthält auch den Wert 0, der für eine logarithmische Transformation steht, für die wir uns schon zur Fehlervarianzharmonisierung entschieden haben (vgl. Abschnitt 1.7.3.3). Im Hinblick auf die beiden gleichzeitig anzustrebenden Ziele der Varianzhomogenität und der Normalität scheint die logarithmische Transformation des Kriteriums eine gute Wahl zu sein.

Zur numerischen Ausgabe gehören auch zwei Likelihood-Quotienten - Tests zu den Nullhypothesen

$$\lambda = 0 \text{ bzw. } \lambda = 1$$

Im Beispiel wird die erste Nullhypothese, die einer logarithmischen Transformation entspricht, beibehalten ($p = 0{,}100$) und die zweite Nullhypothese, die den Verzicht auf eine Transformation empfiehlt, abgelehnt ($p < 0{,}001$):

```
Likelihood ratio tests about transformation parameters
                              LRT df       pval
LR test, lambda = (0)   2.709587  1 0.09974683
LR test, lambda = (1) 159.372693  1 < 2.22e-16
```

Im Zusammenhang mit der Varianzhomogenitätsbeurteilung (siehe Abschnitt 1.7.3.3) haben wir uns entschieden, für das Kriterium LNBSP eine quadratische Regression auf ERN und ERN2 durchzuführen. Die Ergebnisse werden im Abschnitt 4 über die polynomische Regression berichtet. Dort sind auch die (erfreulichen) Ergebnisse zur Normalitätsprüfung für das revidierte Modell zu finden.

1.8 Stichprobenumfangsplanung

Für einen geplanten F- oder t-Test zur ungerichteten Nullhypothese über den Steigungskoeffizienten β_1

$$H_0: \beta_1 = 0 \text{ versus } H_1: \beta_1 \neq 0$$

oder für den t-Test zu einer gerichteten Nullhypothese

$$H_0: \beta_1 \leq 0 \text{ versus } H_1: \beta_1 > 0$$
$$\text{bzw.}$$
$$H_0: \beta_1 \geq 0 \text{ versus } H_1: \beta_1 < 0$$

lässt sich …

- zu einer bekannten (oder angenommenen) Effektstärke in der Population, die letztlich über den Determinationskoeffizienten (siehe Abschnitt 1.5) definiert ist,
- und ein akzeptiertes α-Fehlerrisiko

berechnen, welche Stichprobengröße benötigt wird, um mit einer gewünschten Wahrscheinlichkeit (Teststärke, Power) ein signifikantes (den vorhandenen Effekt aufdeckendes) Ergebnis zu erzielen. In der modernen Forschungsmethodik verschiedener Fächer (siehe z. B. Eid et al 2013, S. 191ff für die Psychologie oder Urban & Mayerl 2018, S. 129ff für die Soziologie) ist es selbstverständlich geworden, Effektstärken zu beachten und Stichprobenumfänge zu begründen.

1.8.1 Analysen mit G*Power

Wir verwenden zur Stichprobenumfangsplanung zunächst das exzellente Power-Analyse-Programm **GPower 3.1** (Faul et al. 2009, GPower 2017), das für macOS und MS-Windows kostenlos über die folgende Webseite zu beziehen ist:

http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/

Auf den ZIMK-Pool-PCs an der Universität Trier unter dem Betriebssystem Windows befindet sich G*Power 3.1 in der Startmenü-Gruppe **Statistik**:

Statistik > GPower

1.8.1.1 Modell mit einem fixierten Regressor

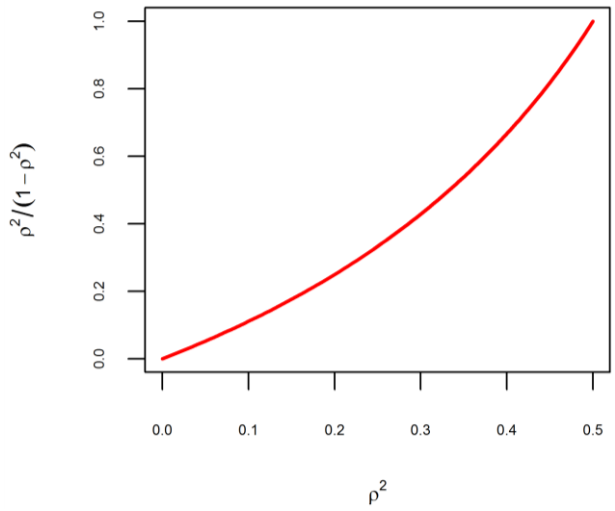
Wir verwenden zunächst das der SPSS-Prozedur REGRESSION zugrundeliegende Modell mit einem fixierten Regressor und wählen nach dem GPower-Programmstart den folgenden Aufgabentyp:

- **Test family**: **t-Tests**
- **Statistical test**: **Linear Multiple Regression: Fixed model, single regression coefficient**
- **Type of power analysis**: **A priori**

GPower 3.1 verwendet das von Cohen (1988, S. 410) vorgeschlagene Effektstärkemaß f^2, das in folgender Beziehung zum Determinationskoeffizienten ρ^2 in der Population (Anteil der erklärten Kriteriumsvarianz, siehe Abschnitt 1.5) steht:

$$f^2 = \frac{\rho^2}{1-\rho^2}$$

Wie das folgende Diagramm zeigt, stehen ρ^2 und f^2 im interessanten Wertebereich von 0 bis 0,5 fast in einer linearen Beziehung zueinander:

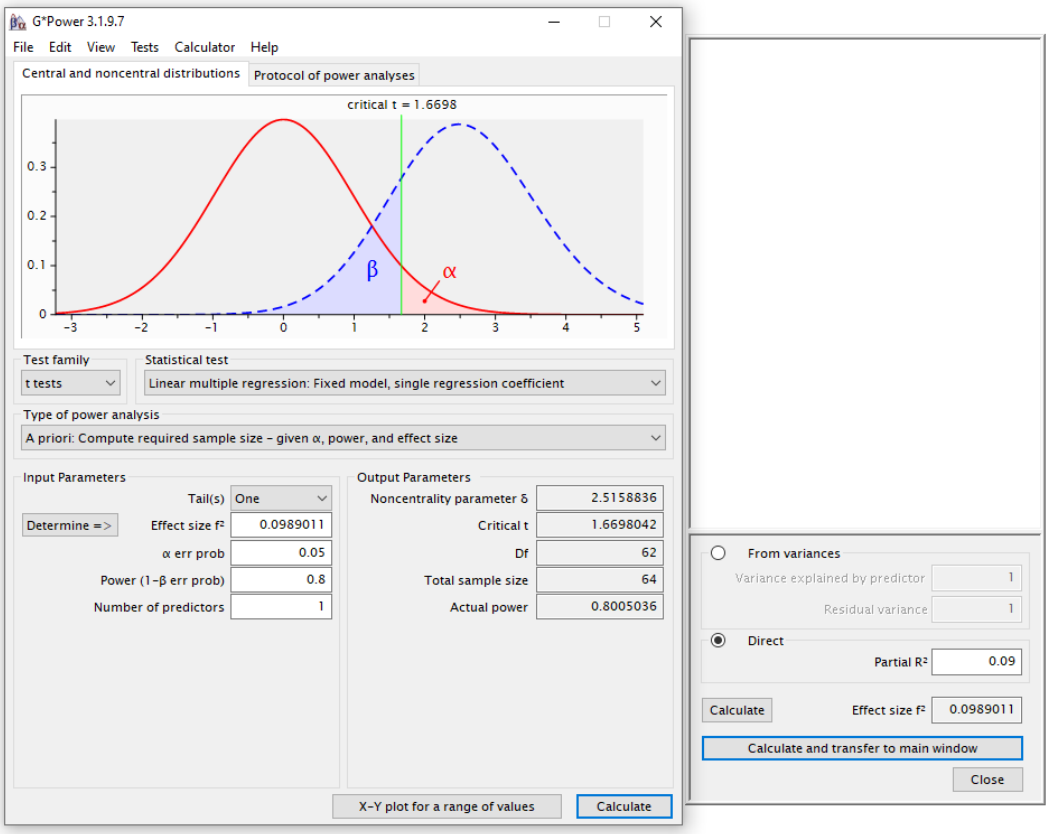


Da wir keine konkrete Stichprobenumfangsplanung vornehmen, sondern **GPower 3.1** erkunden, arbeiten wir beispielhaft mit einer mittleren Effektstärke sensu Cohen (1988, S. 79ff). Die schon aus Abschnitt 1.5 bekannte Tabelle mit Cohens Orientierungswerten wird um eine Spalte für den Effektstärkeindex f^2 erweitert:

Effektstärke in der Population	erklärter Varianzanteil	Effektstärke f^2
klein	0,01	0,01
mittel	0,09	0,10
groß	0,25	0,33

Einem mittleren Effekt entspricht also in der bivariaten Regression der Determinationskoeffizient 0,09 bzw. der f^2 - Wert 0,10.

Nach einem Klick auf den Schalter **Determine** tragen wir im erscheinenden Effektstärkenfenster zur Option **Direct** den Wert 0,09 als **Partial R2** ein:



Bei der von Cohen (1988, S. 56) als Standardwert empfohlenen Power (Entdeckungswahrscheinlichkeit) von 0,8 (β-Fehler: 0,2) resultiert für den in der Regel empfehlenswerten *einseitigen* Test ein erforderlicher Stichprobenumfang von 64 Fällen.

Für einen *zweiseitigen* Test werden bei ansonsten identischen Bedingungen 82 Fälle benötigt. Wer den Unterschied zwischen gerichteten und ungerichteten Hypothesen ignoriert und mit dem bei Computer-Programmen für die bivariate Regression üblicherweise voreingestellten zweiseitigen Test arbeitet, muss also einen erhöhten Aufwand bei der Datenerhebung betreiben bzw. verliert bei identischem Stichprobenumfang an Teststärke.

1.8.1.2 Modell mit einem stochastischen Regressor

Werden die Algorithmen des bivariaten Regressionsmodells für einen fixierten Regressor auf Daten mit einem stochastischen Regressor angewendet, dann bleiben die Schätzungen erwartungstreu und die Signifikanztests korrekt (vgl. Abschnitt 1.2.3). Allerdings führt in dieser Situation die im Abschnitt 1.8.1.1 beschriebene A-priori - Power-Analyse zu einem systematisch unterschätzten Stichprobenbedarf. Um dies an einem Beispiel zu demonstrieren, wählen wir in GPower den folgenden Aufgabentyp:

- **Test family**: **Exact**
- **Statistical test**: **Linear Multiple Regression: Random model**
- **Type of power analysis**: **A priori**

Diesmal wird die bivariate Normalverteilung von Kriterium und Regressor vorausgesetzt (GPower 2017, S. 18).

Zur Spezifikation der Effektstärke ist der Determinationskoeffizient auf Populationsebene unter der Alternativ- und unter der Nullhypothese anzugeben. Wir verwenden analog zu Abschnitt 1.8.1.1 folgende Werte:

- H_1: 0,09
  Dies ist nach Cohen (1988, S. 79ff; vgl. Tabelle im Abschnitt 1.8.1.1) eine mittlere Effektstärke in der bivariaten linearen Regression, wobei als Effektstärkebegriff diesmal nicht f^2 verwendet wird, sondern ρ^2.
- H_0: 0,0

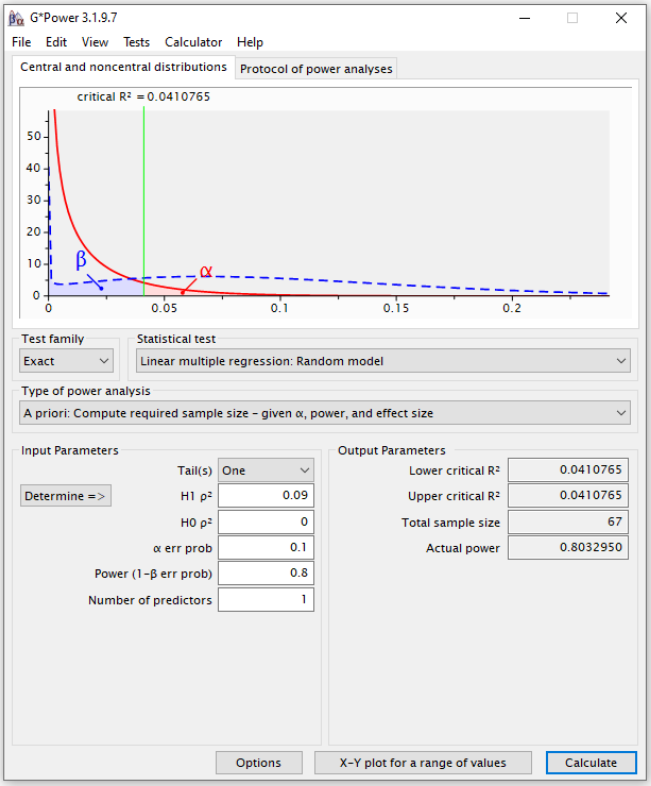
Bei einem einseitigen Testproblem bzgl. des Determinationskoeffizienten ρ^2

$$H_0: \rho^2 = 0 \text{ vs. } H_1: \rho^2 > 0$$

liegt bzgl. des Steigungsparameters β_1 ein *zweiseitiges* Testproblem

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0$$

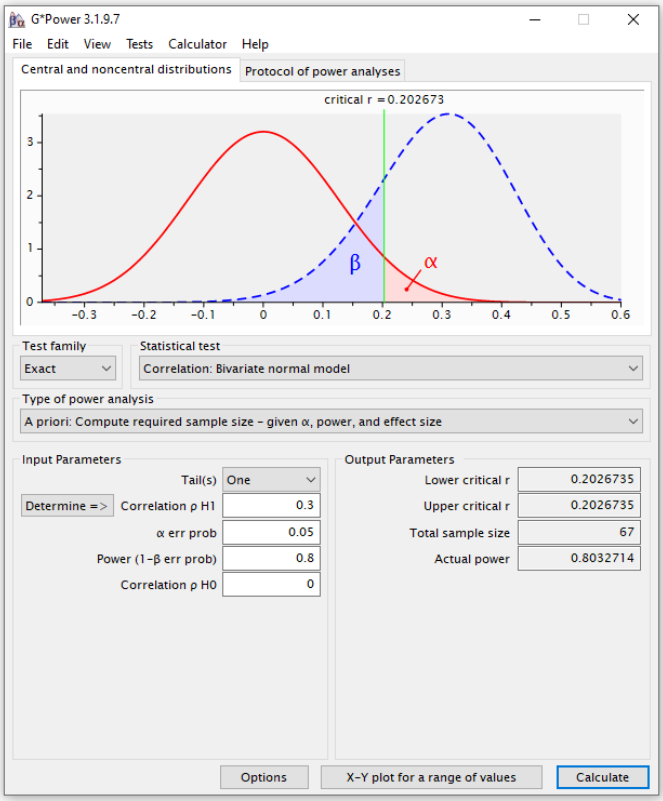
vor, weil positive und negative Steigungsparameter den Determinationskoeffizienten gleichermaßen erhöhen. Um zu einem *einseitigen* Test zum Niveau $\alpha = 0{,}05$ bzgl. des Steigungsparameters zu gelangen, verdoppeln wir das α-Niveau auf 0,10. Bei einer gewünschten Power von 0,8 resultiert eine Stichprobengröße von 67, die erwartungsgemäß den Wert 64 für das Modell mit einem fixierten Regressor (vgl. Abschnitt 1.8.1.1) leicht übertrifft:



Zum selben Ergebnis kommt man bei einer Stichprobenumfangsplanung für eine Korrelation korrespondierenden Größe ($\rho = 0{,}3$). In GPower ist der folgende Aufgabentyp zu wählen:

- **Test family**: **Exact**
- **Statistical test**: **Correlation: Bivariate normal model**
- **Type of power analysis**: **A priori**

Diesmal gelangen wir ohne den „Trick“ mit der α-Fehlerrate zur Stichprobengröße $N = 67$:



1.8.2 Analysen mit SPSS

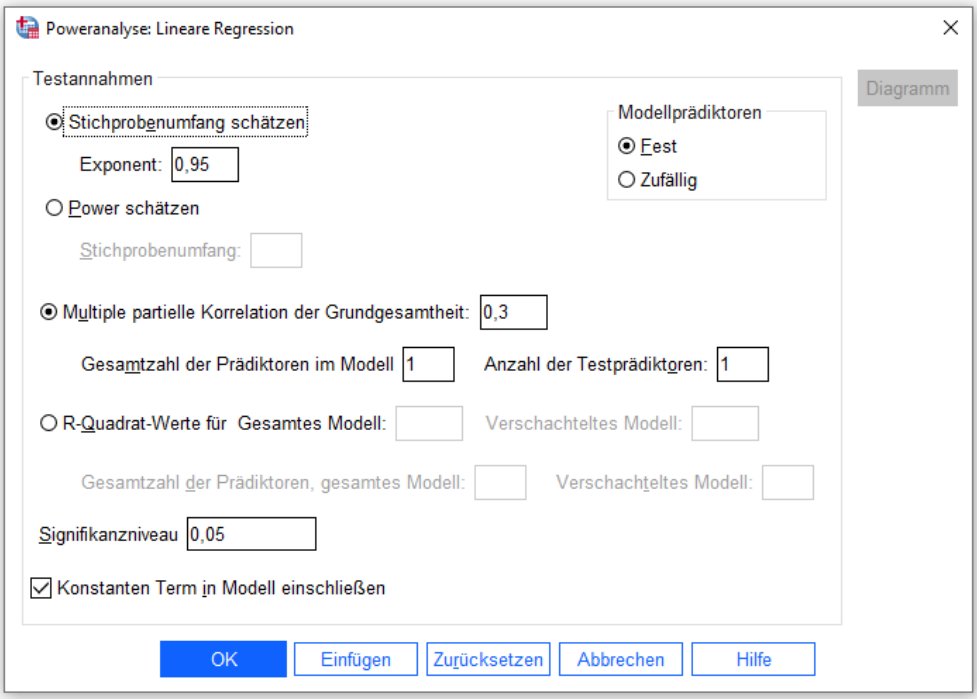
Seit der Version 27 besitzt SPSS Statistics eine Prozedur zur Power-Analyse. Weil diese zu begrüßende Erweiterung nicht ignoriert werden soll, wiederholen wir mit SPSS nach

Analysieren > Poweranalyse > Regression > Univariat linear

die Power-Analyse für die bivariate lineare Regression. Wir wählen als Aufgabenstellung im Power-Analyse - Dialog:

- die Schätzung des **Stichprobenumfangs**
- ein Modell mit **festen** Effekten (= Voreinstellung)
- 0,8 als gewünschte Power (falsch beschriftet mit **Exponent**)
- 0,3 als **Korrelation**
- ein Modell mit *einem* Regressor, der folglich auch getestet werden soll
- als **Signifikanzniveau** (akzeptiertes α-Fehlerrisiko) den Wert 0,05
- ein Modell mit **konstantem Term** (= Voreinstellung)

Dass die Power-Analyse in SPSS beim t-Test zum Steigungskoeffizienten in der bivariaten linearen Regression keine gerichteten Tests unterstützt, ist als Mangel zu bezeichnen:



Es ist ein leider immer noch oft anzutreffender Irrglaube, dass die wichtige Unterscheidung zwischen ein- und zweiseitigen Tests in der linearen Regression nicht anwendbar sei. G*Power ist von diesem Problem nicht betroffen.

SPSS liefert exakt dasselbe Ergebnis ($N = 82$), das wir auch von G*Power für die zweiseitige Testung im Modell mit einem fixierten Regressor erhalten haben:

Poweranalysetabelle.

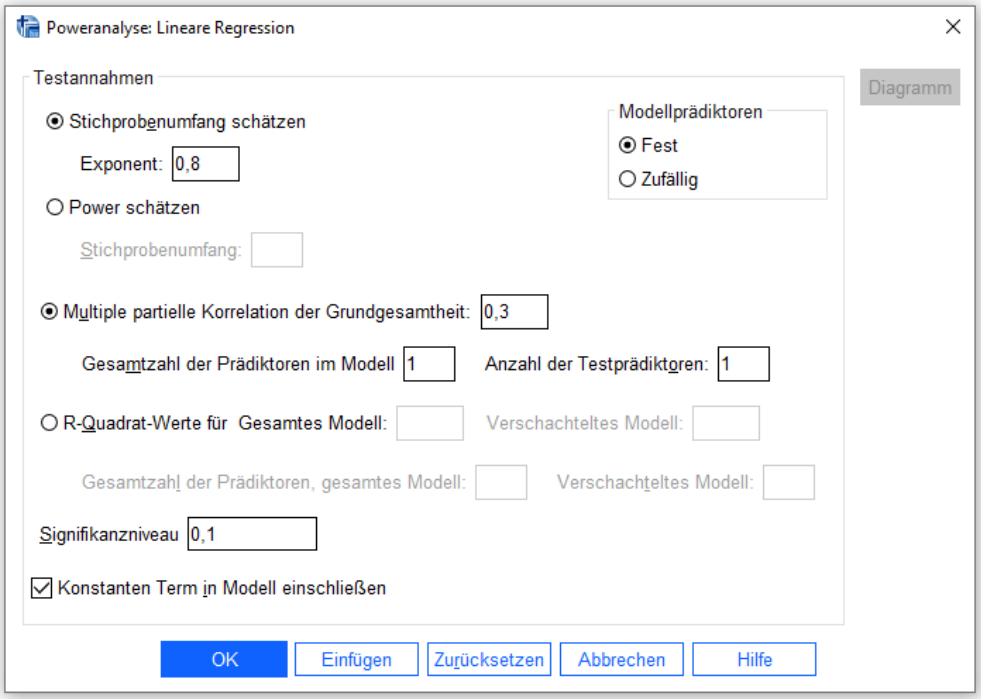
	N	Tatsächliche Leistung [b]	Prädiktoren		Testannahmen		
			Gesamt	Test	Exponent	Partiell [c]	Sig.
F-Test vom Typ III [a]	82	,803	1	1	,8	,3	,05

a. Ein konstanter Term wird einbezogen.

b. Es wird davon ausgegangen, dass Prädiktoren festgelegt werden.

c. Multipler partieller Korrelationskoeffizient.

Um mit SPSS eine Stichprobenumfangsplanung für den *einseitigen* Test durchzuführen, ist die α-Fehlerrate zu verdoppeln:



Dann resultiert dasselbe Ergebnis ($N = 64$), das wir von G*Power für die einseitige Testung im Modell mit einem fixierten Regressor erhalten haben:

Poweranalysetabelle.

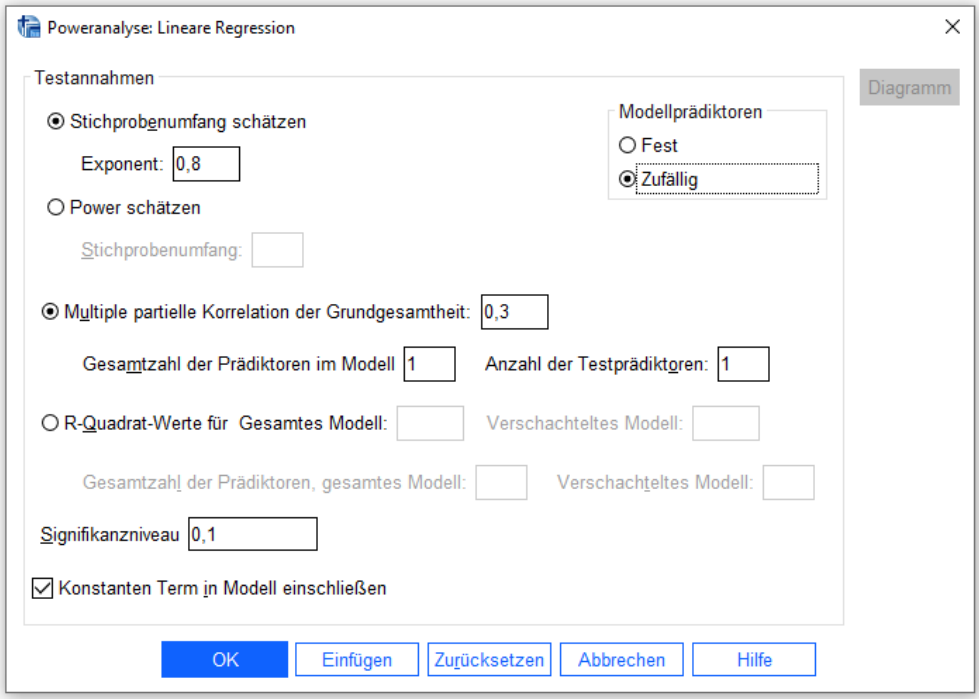
	N	Tatsächliche Leistung [b]	Prädiktoren		Testannahmen		
			Gesamt	Test	Exponent	Partiell [c]	Sig.
F-Test vom Typ III [a]	64	,801	1	1	,8	,3	,1

a. Ein konstanter Term wird einbezogen.

b. Es wird davon ausgegangen, dass Prädiktoren festgelegt werden.

c. Multipler partieller Korrelationskoeffizient.

Um die Stichprobenumfangsplanung zum Modell für einen stochastischen Regressor anzufordern, ist im SPSS-Dialog zur Poweranalyse die Option **Zufällig** im Rahmen **Modellprädiktoren** zu wählen:



Es resultiert erneut dasselbe Ergebnis (N = 67), das wir von G*Power erhalten haben:

Poweranalysetabelle.

	N	Tatsächliche Leistung [b]	Prädiktoren		Testannahmen		
			Gesamt	Test	Exponent	Partiell [c]	Sig.
F-Test vom Typ III [a]	67	,803	1	1	,8	,3	,1

a. Ein konstanter Term wird einbezogen.

b. Es wird davon ausgegangen, dass Prädiktoren beliebig sind.

c. Multipler partieller Korrelationskoeffizient.

1.8.3 Simulationsstudie zur Verlässlichkeit von Stichprobenumfangsempfehlungen

Um die Verlässlichkeit der Stichprobenumfangsempfehlungen für das Modell mit einem fixierten Regressor (siehe Abschnitt 1.8.1.1) und für das Modell mit einem stochastischen Regressor (siehe Abschnitt 1.8.1.2) vergleichen zu können, wurde eine Simulationsstudie mit dem folgenden wahren Modell

$$Y = 0{,}5 \cdot X + \varepsilon,\ \ X \sim \mathrm{N}(0;\ 1),\ \varepsilon \sim \mathrm{N}(0;\ 1{,}6) \qquad (12)$$

durchgeführt. Hier besteht annähernd die im Abschnitt 1.8.1.1 angenommene Effektstärke $f^2 = 0{,}1$ bzw. die im Abschnitt 1.8.1.2 angenommene Effektstärke $\rho^2 = 0{,}09$. Es wird der üblichen Praxis folgend ein stochastischer Regressor verwendet. Mit den im Abschnitt 1.8.1.1 für das unrealistische Modell mit einem fixierten Regressor bzw. im Abschnitt 1.8.1.2 für das Modell mit einem stochastischen Regressor ermittelten Stichprobenumfängen (N = 64 bzw. N = 67) wurden jeweils 500 Stichproben gezogen. Mit dem Stichprobenumfang N = 64 lag die empirische Rate signifikanter Ergebnisse mit einer relativen Häufigkeit von 0,78 *unter* der geplanten Power von 0,8:

Signifikant

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Nein	110	22,0	22,0	22,0
	Ja	390	78,0	78,0	100,0
	Gesamt	500	100,0	100,0	

Demgegenüber wurde mit dem Stichprobenumfang $N = 67$ annähernd die erwartete Rate signifikanter Ergebnisse erzielt:

Signifikant

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Nein	102	20,4	20,4	20,4
	Ja	398	79,6	79,6	100,0
	Gesamt	500	100,0	100,0	

Wer bei einer a-priori - Power-Analyse mit dem Modell fixierter Regressoren arbeitet, sollte also etwas über den empfohlenen Stichprobenumfang hinausgehen.

Zudem müssen auch die Effekte von Messfehlern in den Regressoren auf die Stichprobenumfangsplanung berücksichtigt werden. Wird z.B. in der Situation von Gleichung (12) das angemessene Modell mit einem stochastischen Regressor unterstellt (also $N = 67$ verwendet), und kommt dabei ein Regressor mit der Reliabilität 0,8 zum Einsatz (entgegen der Annahme einer perfekten Reliabilität von 1), dann liegt die Rate signifikanter Ergebnisse erheblich unter der geplanten Power von 0,8:

Signifikant

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Nein	146	29,2	29,2	29,2
	Ja	354	70,8	70,8	100,0
	Gesamt	500	100,0	100,0	

Aus den berichteten Simulationsergebnissen entsteht der Eindruck, dass sich die falsche Annahme eines fixierten Regressors weniger stark verfälschend auf die Stichprobenumfangsempfehlung auswirkt als die falsche Annahme eines fehlerfrei gemessenen Regressors.

2 Multiple lineare Regression

Bei der *multiplen* linearen Regression wird der simultane Einfluss von mehreren Regressoren auf ein Kriterium analysiert. Da in den Sozial-, Wirtschafts- oder Geowissenschaften nur selten monokausale Phänomene zu untersuchen sind, ist die multiple Regressionsanalyse eine unverzichtbare Forschungsmethode und kann in der Regel *nicht* durch eine Serie von bivariaten Regressionen ersetzt werden. Es finden sich mit Leichtigkeit Systeme mit mehreren Einflussgrößen, über die bivariate Analysen unvollständige oder irreführende Aussagen liefern:

- Im Abschnitt 6.1 wird das bekannte Beispiel mit dem bivariaten **Scheineffekt** der Storchenzahl (*X*) in schwedischen Landkreisen auf die Geburtenrate (*Y*) aufgegriffen. Nach Erweiterung des Modells um den Regressor Industrialisierungsgrad (*Z*) verschwindet der „signifikante" Effekt von *X*.
- Bei der **Vulnerabilitäts-Kompensation** (siehe Abschnitt 2.7.3) erweist sich umgekehrt eine Variable in der multiplen Regression als bedeutsam, die praktisch keine bivariate Beziehung zum Kriterium besitzt.

Das Modell der multiplen linearen Regression eignet sich für Merkmale mit folgender Skalenqualität:

- Bei der **abhängigen Variablen** wird **metrisches Skalenniveau** vorausgesetzt.[1] Außerdem sollte die Verteilung der abhängigen Variablen von der angenommenen Stetigkeit nicht zu weit entfernt sein.[2]
- Bei den **unabhängigen Variablen** wird das **metrische** und das **kategoriale Skalenniveau** unterstützt. Eine kategoriale Variable mit *w* Ausprägungen wird im Design durch (*w* - 1) Kodiervariablen repräsentiert (siehe z. B. Cohen et al. 2003, S. 302ff). Ordinale Regressoren müssen auf kategoriales Niveau herabgestuft oder als metrische Regressoren behandelt werden.

2.1 Beispiel und Anforderung einer multiplen linearen Regression in SPSS

2.1.1 Einflussfaktoren auf die Mortalität in amerikanischen Städten

Als Beispiel betrachten wir eine Studie zum Einfluss von Luftschadstoffen, klimatischen Bedingungen und soziodemografischen Faktoren auf die Mortalität in amerikanischen Städten (siehe Dixon 1992, S. 637f). Als Kriterium (SPSS-Variablenname MORTAL) wird die Anzahl der jährlichen Todesfälle pro 100.000 Einwohner betrachtet, und als Regressoren werden einbezogen:

Merkmal	**Variablenname**
Mittlere jährliche Niederschlagsmenge in Zoll	REGEN
Mittlere Anzahl abgeschlossener Schuljahre bei Erwachsenen	BILDUNG
Prozentanteil der nicht-weißen Bevölkerung	FARBIG
Belastung der Luft mit Stickstoffoxyden (NO_x)	NOX
Belastung der Luft mit Schwefeldioxyd (SO_2)	SO2

Die SPSS-Datendatei **Airpoll.sav** mit diesen Variablen befindet sich an der im Vorwort verabredeten Stelle.

---

[1] Liegt ein geordnet-kategoriales (ordinales) oder ein ungeordnet-kategoriales (nominales) Messniveau vor, dann bietet die ordinale oder die multinomiale logistische Regression weitgehend analoge Analysemöglichkeiten (siehe z.B. Baltes-Götz 2012).

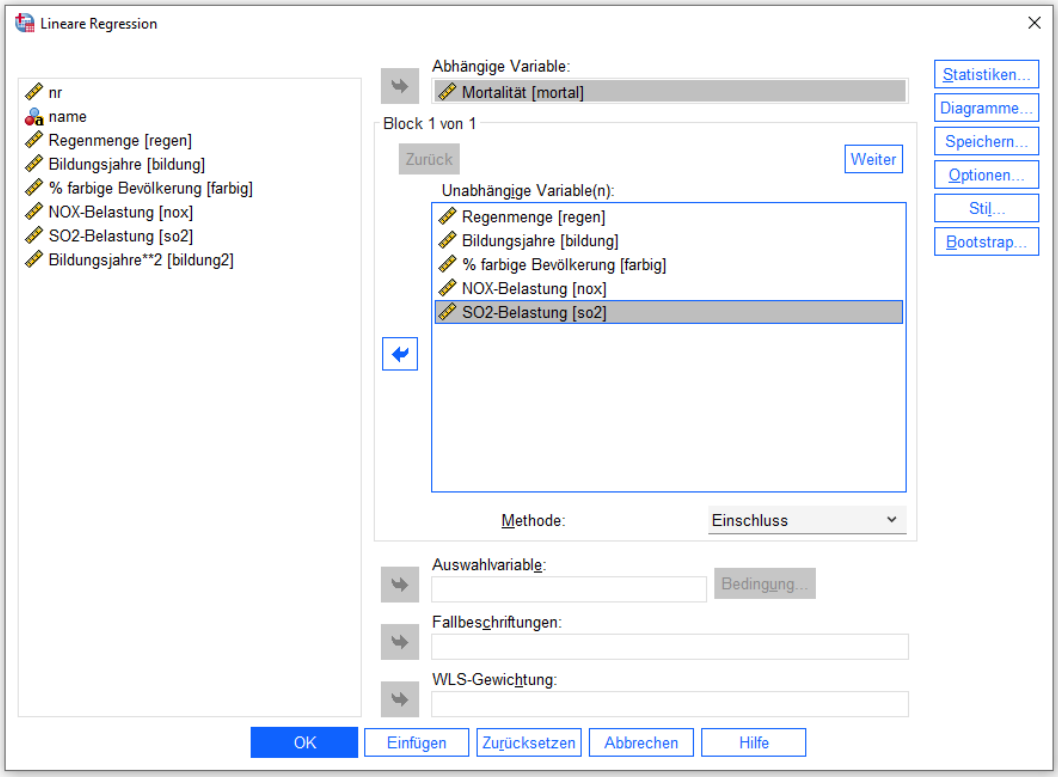
[2] Zähldaten besitzen zwar interpretierbare Intervalle, doch können nur wenige, positive Werte auftreten, sodass für die Modellresiduen die Unterstellung der stetigen Normalverteilung mit homogenen Varianzen grob unplausibel ist. Wenn eine parametrische diskrete Verteilung für die Residuen approximativ passt (z. B. die Poisson-Verteilung oder die negative Binomialverteilung), dann kommt ein *verallgemeinertes lineares Modell* in Frage (siehe z.B. Baltes-Götz 2016).

2.1.2 Anforderung einer multiplen linearen Regression

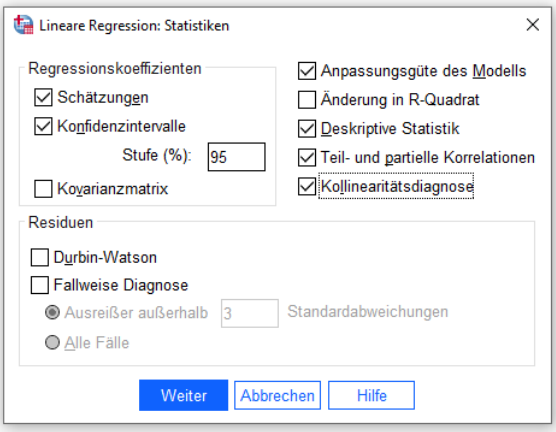
Wir fordern per

Analysieren > Regression > Linear

die Ergebnisse an, mit denen im weiteren Verlauf von Abschnitt 2 die Grundzüge der multiplen linearen Regressionsanalyse erläutert werden sollen:



Verlangen Sie in der **Statistiken** - Subdialogbox über die Voreinstellung hinausgehend noch **Konfidenzintervalle**, **deskriptive Statistiken**, **(semi-)partielle Korrelationen** und die **Kollinearitätsdiagnose**:



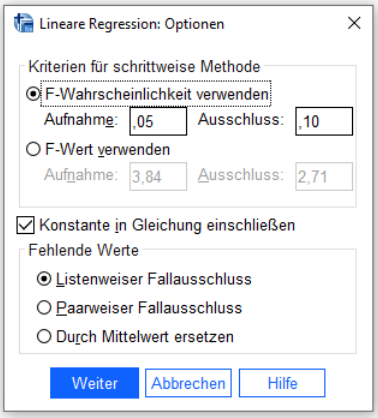
Aufgrund der Anforderung von deskriptiven Statistiken erhalten wir u.a. die folgende Tabelle, die neben den Mittelwerten und Standardabweichungen der Variablen auch den bei fallweiser Behandlung fehlender Werte verbleibenden Stichprobenumfang zeigt:

Deskriptive Statistiken

	Mittelwert	Std.-Abweichung	N
Mortalität	940,382	62,2124	60
Bildungsjahre	10,973	,8453	60
Regenmenge	37,37	9,985	60
% farbige Bevölkerung	11,870	8,9211	60
NOX-Belastung	15,98	16,730	60
SO2-Belastung	53,77	63,390	60

Im Beispiel sind die Daten komplett vorhanden, sodass bei univariaten Häufigkeitsanalysen für alle Variablen identische Stichprobenumfänge und Statistiken resultieren würden.

Fehlende Werte behandelt die Regressionsprozedur per Voreinstellung durch **listenweisen Fallausschluss**, d.h. es verbleiben nur die Fälle mit einem vollständigen Datensatz. Im **Optionen**-Subdialog der Regressionsprozedur werden zwei Alternativen zur fallweisen Behandlung fehlender Werte angeboten:



Beim **paarweisen Fallausschluss** nutzt man zum Schätzen der Korrelationen, die im Algorithmus der linearen Regression als Zwischenergebnisse eine entscheidende Rolle spielen, alle Fälle mit Werten bei den beiden jeweils beteiligten Variablen. Folglich basieren die einzelnen Schätzwerte in der Korrelationsmatrix im Allgemeinen auf unterschiedlichen Teilstichproben. Dies kann zu einer defekten (indefiniten) Korrelationsmatrix führen, die unsinnige Regressionsergebnisse liefert (z. B. einen extrem hohen Determinationskoeffizienten).

Lässt man die fehlenden Werte durch die **Mittelwerte** der jeweiligen Variablen **ersetzen**, dann resultieren verzerrte Schätzer für Varianzen, Korrelationen und Regressionskoeffizienten.

Von den drei Verfahren zur Behandlung fehlender Werte, die von der Regressionsprozedur angeboten werden, ist der voreingestellte fallweise Ausschluss in den meisten Fällen das kleinste Übel. Aber auch hier kann es zu verzerrten Schätzungen der Regressionskoeffizienten kommen, wenn für das Muster fehlender Werte die MCAR-Bedingung (*Missing Completely At Random*) nicht erfüllt ist. In einem solchen kann man ...

- die von SPSS unterstützte *multiple Imputation* einsetzen
- oder ein Strukturgleichungsanalyseprogramm wie Amos oder Mplus verwenden, das die FIML-Schätzmethode geherrscht (*Full Information Maximum Likelihood*)

Eine ausführliche Beschreibung der diversen Methoden zur Behandlung fehlender Werte ist in Baltes-Götz (2013) zu finden.

Nach diesen knappen Bemerkungen zum Problem fehlender Werte kehren wir zur multiplen Regressionsanalyse zurück. Die aus dem obigen Anforderungsdialog resultierenden Schätz- und Testergebnisse zum

Mortalitätsbeispiel dürfen vor einer Prüfung der Regressionsvoraussetzungen noch nicht interpretiert werden:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,816 [a]	,666	,635	37,6014

a. Einflußvariablen : (Konstante), SO2-Belastung, Regenmenge, % farbige Bevölkerung, Bildungsjahre, NOX-Belastung

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	152003,893	5	30400,779	21,502	,000 [b]
	Nicht standardisierte Residuen	76348,876	54	1413,868		
	Gesamt	228352,770	59			

a. Abhängige Variable: Mortalität

b. Einflußvariablen : (Konstante), SO2-Belastung, Regenmenge, % farbige Bevölkerung, Bildungsjahre, NOX-Belastung

Koeffizienten[a]

		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B		Korrelationen			Kollinearitätsstatistik	
Modell		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze	Nullter Ordnung	Partiell	Teil	Toleranz	VIF
1	(Konstante)	998,966	93,028		10,738	,000	812,457	1185,476					
	Regenmenge	1,658	,686	,266	2,415	,019	,282	3,034	,509	,312	,190	,510	1,961
	Bildungsjahre	-15,912	7,160	-,216	-2,223	,030	-30,266	-1,558	-,510	-,289	-,175	,654	1,528
	% farbige Bevölkerung	3,026	,650	,434	4,658	,000	1,723	4,328	,644	,535	,367	,713	1,402
	NOX-Belastung	,085	,487	,023	,174	,862	-,891	1,060	,216	,024	,014	,362	2,764
	SO2-Belastung	,312	,122	,318	2,570	,013	,069	,556	,426	,330	,202	,404	2,478

a. Abhängige Variable: Mortalität

Anschließend werden wir das Modell der multiplen Regression und wichtige Methoden zur Überprüfung der Modellvoraussetzungen diskutieren. Bei der Behandlung der multiplen Regression beschränken wir uns generell auf das inhomogene Modell (*inklusive* Achsenabschnitt β_0).

2.2 Modell und Annahmen

Wir notieren das multiple lineare Regressionsmodell mit dem Kriterium Y und k unabhängigen Variablen X_1 bis X_k analog zum bivariaten Fall (vgl. Abschnitt 1.2). Die beim Fall i realisierten Regressorwerte

$$X_{1_i} = x_{1_i}, .., X_{k_i} = x_{k_i}$$

gelten im Modell fixierter Regressoren als vorab geplant (vgl. Abschnitt 1.2.1). Demgegenüber realisiert sich bei jedem Fall i eine beobachtbare Zufallsvariable Y_i.

Das Modell enthält eine Behauptung über die Erwartungswerte $\mathrm{E}(Y_i)$ der Zufallsvariablen Y_i und Annahmen über die Residuen ε_i (die Abweichungen der tatsächlichen Y_i - Beobachtungswerte von den Erwartungswerten $\mathrm{E}(Y_i)$):

$$\begin{aligned} \mathrm{E}(Y_i) &= \beta_0 + \beta_1 x_{1_i} + \ldots + \beta_k x_{k_i} \\ \varepsilon_i &:= Y_i - \mathrm{E}(Y_i) \qquad\qquad i = 1, .., N \\ \varepsilon_i &\sim \mathrm{N}(0, \sigma_\varepsilon^2), \mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathrm{I}_N \end{aligned} \tag{13}$$

Der Regressionskoeffizient β_j quantifiziert den Einfluss des Regressors x_j auf den Erwartungswert der Kriteriumsvariablen. Aus Gründen, die im Abschnitt 2.9.1 erläutert werden, bezeichnet man β_j auch als *partiellen Regressionskoeffizienten* oder (mit gleicher Berechtigung) als *semipartiellen Regressionskoeffizienten*.

Für die unabhängigen „Variablen“ x_j, $j = 1,.., k$, wird angenommen:

- Sie sind *fehlerfrei* gemessen.
- Ihre Ausprägungen resultieren *nicht* aus einem Zufallsprozess, sondern sind *fixiert*. Folglich enthält das multiple Regressionsmodell auch *keine* Verteilungsannahme zu den Regressoren.

Über die Verteilungen der Zufallsvariablen ε_i und Y_i macht das Modell der multiplen linearen Regression die anschließend diskutierten Annahmen.

2.2.1 Linearität

Der Erwartungswert (Mittelwert) der Variablen Y_i zu einen festen Wertevektor $(x_{1_i}, \ldots, x_{k_i})$ hängt linear von den Regressorwerten ab. Erhöht man z. B. x_{1_i} um eine Einheit und hält gleichzeitig die übrigen Regressoren konstant, so steigt der Erwartungswert von Y_i um β_1 Einheiten. Das „Erhöhen“ ist nicht unbedingt im Sinne einer Manipulation bei einem konkreten Fall zu verstehen, sondern bedeutet oft den Übergang zu einem Fall mit einer höheren x_j - Ausprägung und identischen Werten bei den anderen Regressoren. In der Terminologie von Darlington & Hayes (2017, S. 88) ist dies die *primäre* Annahme des Modells.

Denkt man sich im Fall von zwei Regressoren X_1 und X_2 die (x_{1_i}, x_{2_i}) - Werte in der (X_1, X_2) - Ebene eines dreidimensionalen Koordinatensystems mit X_1 -, X_2 - und Y-Achse angeordnet, so liegen die vom Modell behaupteten Erwartungswerte der zugehörigen Y_i - Variablen auf einer *Ebene*, welche die Y-Achse im Punkt β_0 schneidet, in Richtung der X_1 - Achse die Steigung β_1 hat und in Richtung der X_2 - Achse die Steigung β_2, z. B.:

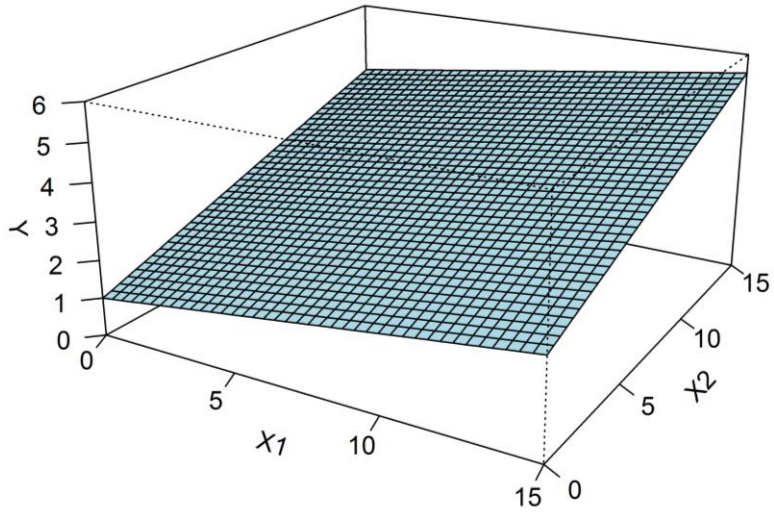


Abbildung 26: Regressionsebene im Modell

$$E(Y) = 1 + 0{,}1\, x_1 + 0{,}2\, x_2$$

Aus der Definition des Residuums ε_i zur Beobachtung *i* mit der Regressorwertekombination $(x_{1_i}, \ldots, x_{k_i})$

$$\varepsilon_i := Y_i - E(Y_i),\ i = 1,.., N$$

folgt unmittelbar, dass ε_i den Erwartungswert 0 besitzt.

$$E(\varepsilon_i) = E(Y_i - E(Y_i)) = E(Y_i) - E(Y_i) = 0$$

2.2.2 Normalverteilung und Varianzhomogenität der Residuen

Um die Eigenschaften von Schätz- und Testergebnissen im Regressionsmodell begründen zu können, wird angenommen:

- Die Residuen ε_i sind normalverteilt.
- Alle Residualverteilungen besitzen dieselbe Varianz, die im Modell durch den Parameter σ_ε^2 vertreten ist.

2.2.3 Unkorreliertheit der Residuen

Die *N* Residualvariablen ε_i sind modellgemäß unkorreliert. Ihre Kovarianzmatrix ist eine Diagonalmatrix der Dimension *N* mit identischen Einträgen σ_ε^2 auf der Hauptdiagonalen (für die als identisch angenommenen Fehlervarianzen) und Nullen an allen anderen Positionen (für die Kovarianzen). Weil die Residuen normalverteilt sind, folgt aus der Unkorreliertheit die stochastische Unabhängigkeit.

Zu möglichen Ursachen und zu den Konsequenzen einer verletzten Unabhängigkeitsannahme siehe Abschnitt 1.2.2.6.

2.2.4 Technische Voraussetzungen

In einem Modell mit *k* Regressoren muss *N* größer als (*k* + 1) sein, damit Signifikanztests zu den Regressoren durchgeführt werden können. Für Signifikanztests mit einer angemessenen Power und für Vertrauensintervalle mit einer akzeptablen Präzision werden allerdings erheblich mehr Beobachtungen benötigt (siehe Abschnitt 2.10).

Die (*k* + 1) Vektoren

$$\begin{pmatrix}1\\1\\.\\.\\1\end{pmatrix}, \begin{pmatrix}x_{1_1}\\x_{1_2}\\.\\.\\x_{1_N}\end{pmatrix}, \begin{pmatrix}x_{2_1}\\x_{2_2}\\.\\.\\x_{2_N}\end{pmatrix}, \quad \dots \quad , \begin{pmatrix}x_{k_1}\\x_{k_2}\\.\\.\\x_{k_N}\end{pmatrix}$$

müssen *linear unabhängig* sein. Es darf also kein Vektor als Linearkombination der übrigen darstellbar sein. Zur Vermeidung von Multikollinearitätsproblemen wird allerdings eine weitergehende Eigenständigkeit der Regressoren benötigt (siehe Abschnitt 3.2).

2.3 Parameterschätzung

Im Folgenden werden durch x_{j_i} bzw. y_i (j = 1,.., k; i = 1,.., N) die Werte der unabhängigen bzw. abhängigen Variablen in der Stichprobe sowie durch $\bar{x}_1$, ..., $\bar{x}_k$ bzw. $\bar{y}$ die zugehörigen Stichprobenmittelwerte bezeichnet. Aus den empirischen Daten werden Schätzer $b_0, b_1, ..., b_k$ für die Regressionsparameter $\beta_0, \beta_1, ..., \beta_k$ analog zu Abschnitt 1.3 nach der Methode der kleinsten Quadrate so bestimmt, dass die Summe der quadrierten geschätzten Residuen (Abweichungen der Modellprognosen von den beobachteten Werten) minimal wird, wobei alle Beobachtungen mit gleichem Gewicht eingehen (OLS, *Ordinary Least Squares*):

$$\sum_{i=1}^{N}(y_i-(b_0+b_1x_{1_i}+\ldots+b_kx_{k_i}))^2 \quad \overset{!}{=} \quad \min$$

Die Bestimmung der geschätzten Regressionskoeffizienten erfolgt mit den Methoden der Differentialrechnung (Nullsetzen der partiellen Ableitungen nach $b_0, b_1, \ldots, b_k$ sowie Auflösen der entstehenden Normalgleichungen).

Bei der multiplen Regressionsrechnung bedient man sich meist der Matrixschreibweise, um zu überschaubaren Ausdrücken zu gelangen. Mit den Definitionen:

$$\mathbf{y} := \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_N \end{bmatrix}, \quad \mathbf{X} := \begin{bmatrix} 1 & x_{1_1} & . & . & x_{k_1} \\ 1 & x_{1_2} & . & . & x_{k_2} \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & x_{1_N} & . & . & x_{k_N} \end{bmatrix}, \quad \mathbf{b} := \begin{bmatrix} b_0 \\ b_1 \\ . \\ . \\ b_k \end{bmatrix}$$

erhält man als Kleinstquadratschätzer **b**:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Die Matrix **X** mit den Werten der Regressoren und einer führenden Einserspalte nennt man **Designmatrix**, weil ihre Einträge (zumindest im klassischen Modell für fixierte Regressoren) als vor Beginn der Datenerhebung festgelegt gelten.

Für die Residualvarianz σ_ε^2 wird folgender Schätzer berechnet:

$$\hat{\sigma}_\varepsilon^2 := \frac{1}{N-k-1}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$$

Bei erfüllter Linearitätsannahme sind die angegebenen Schätzer erwartungstreu. Sind außerdem die Varianzhomogenität und die Unabhängigkeit der Residuen gegeben, dann resultieren nach dem Satz von Gauß-Markov sogar **BLUE-Schätzer** (*Best Linear Unbiased Estimators*) (vgl. Abschnitt 1.3).

Ist außerdem die Normalverteilung der Residuen gegeben, dann sind die im Abschnitt 2.5 beschriebenen Signifikanztests und Konfidenzintervalle gerechtfertigt.

In unserem Anwendungsbeispiel, für das die Überprüfung der Voraussetzungen noch aussteht, erhalten wir die folgende geschätzte Regressionsgleichung (siehe Spalte **Regressionskoeffizient B** in der **Koeffizienten**-Tabelle):

MORTAL = 998,966
+ 1,658 · REGEN
– 15,912 · BILDUNG
+ 3,026 · FARBIG
+ 0,085 · NOX
+ 0,312 · SO2
+ Residuum

Es zeigt sich z. B., dass mit einem Jahr zusätzlicher Schulbildung bei konstanten Werten für die anderen Regressoren eine Abnahme der Mortalität um ca. 16 Todesfälle pro 100.000 Einwohner verbunden ist.

2.4 Standardisierte Regressionskoeffizienten

Weil die Regressoren in der Regel verschiedene Maßeinheiten besitzen, kann man ihre Koeffizienten nicht direkt miteinander vergleichen, um Aussagen über die Einflussstärken der Regressoren zu gewinnen. In unserem Beispiel ist der Regressionskoeffizient zur Bildung betragsmäßig erheblich größer als der Koeffizient zur Schwefeldioxydbelastung (15,912 gegen 0,312). Diese Werte erlauben keinen direkten Vergleich der beiden Regressoren hinsichtlich der Wirkung auf das Kriterium.

Um eine Vergleichbarkeit der Regressionskoeffizienten herzustellen, ersetzt man alle *metrischen* Variablen (Kriterium und Regressoren) durch **standardisierte Varianten** (mit Mittelwert 0 und Varianz 1):

$$x_{j_i} \longrightarrow \frac{x_{j_i} - \bar{x}_j}{s_{X_j}}, \text{ mit } s_{X_j} := \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_{j_i} - \bar{x}_j\right)^2}$$

$$y_i \longrightarrow \frac{y_i - \bar{y}}{s_Y}, \text{ mit } s_Y := \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2}$$

Bei *kategorialen* Regressoren ist das Standardisieren *nicht* sinnvoll (siehe Darlington & Hayes 2017, S. 132).

Die Koeffizienten aus der Regression des standardisierten Kriteriums auf die standardisierten Regressoren werden von SPSS in der **Koeffizienten**-Tabelle als **Beta**-Koeffizienten bezeichnet, was uns ein Bezeichnungsproblem beschert, weil wir den griechischen Buchstaben Beta bereits für die Populationsparameter zu den unstandardisierten Variablen verwenden. Wir behelfen uns mit dem folgenden Symbol für die Stichprobenschätzung des Beta-Koeffizienten zum Regressor X_j:

$$\tilde{b}_j$$

Er steht in folgender Beziehung zum korrespondierenden Regressionskoeffizienten b_j für die unstandardisierten Variablen:

$$\tilde{b}_j = b_j \frac{s_{X_j}}{s_Y}$$

In unserem Beispiel ist für alle Regressoren die Berechnung eines standardisierten Regressionskoeffizienten sinnvoll, weil alle das metrische Skalenniveau besitzen. Es resultiert die folgende geschätzte Regressionsgleichung für die standardisierten Variablen:

$$\begin{aligned}\text{MORTAL} \;=\; & 0{,}266 \cdot \text{REGEN} \\ & - 0{,}216 \cdot \text{BILDUNG} \\ & + 0{,}434 \cdot \text{FARBIG} \\ & + 0{,}023 \cdot \text{NOX} \\ & + 0{,}318 \cdot \text{SO2} \\ & + \text{Residuum}\end{aligned}$$

Ein konstanter Term ist in dieser Gleichung *nicht* vorhanden, weil alle Variablen den Mittelwert 0 haben.[1]

---

[1] Im Modell der multiplen linearen Regression liefern generell die Mittelwerte der Regressoren als Prognosewert den Mittelwert des Kriteriums:

$$\bar{y} = b_0 + \sum_{j=1}^{k} b_j \bar{x}_j$$

Eine Auflösung nach b_0 liefert die folgende Schätzgleichung:

$$b_0 = \bar{y} - \sum_{j=1}^{k} b_j \bar{x}_j$$

Im Beispiel zeigt sich, dass mit dem Zuwachs der Bildungsvariablen um eine Standardabweichungseinheit eine Abnahme der Mortalität um das 0,216 - fache einer Standardabweichungseinheit verbunden ist, wenn alle anderen Regressoren konstant gehalten werden. Demgegenüber bewirkt die Steigerung der Schwefeldioxydbelastung um eine Standardabweichungseinheit im Kriterium einen betragsmäßig deutlich stärkeren Effekt von 0,318 Standardabweichungseinheiten, wenn alle anderen Regressoren konstant gehalten werden. Die Effekte der verschiedenen Regressoren in einer Gleichung können nun scheinbar besser verglichen werden.

Urban & Mayerl (2018, S. 96ff) zweifeln allerdings daran, dass standardisierte Regressionskoeffizienten sinnvoll zu interpretieren sind. Ihrer Argumentation zufolge wäre es z. B. sehr gewagt, aus den Ergebnissen für die Mortalität in amerikanischen Städten zu folgern, dass die Niederschlagsmenge einen stärkeren Effekt hätte als die Bildung. Beim Standardisieren werden reale Maßeinheiten (z. B. Bildungsjahre, Regenmenge) durch statistische Maßeinheiten ersetzt, die von den Streuungsverhältnissen in einer Stichprobe bzw. Population abhängen. Dadurch wird die Interpretation der Koeffizienten nicht unbedingt erleichtert.

Besonders kritisch ist die Abhängigkeit der Beta-Koeffizienten von den Standardabweichungen der untersuchten Stichprobe bzw. Population, wenn Ergebnisse zum *selben* Regressor aus verschiedenen Studien (und damit eventuell aus verschiedenen Populationen) verglichen werden. In dieser Situation betrachtet man besser die Koeffizienten zu den *unstandardisierten* Variablen.

Beim Bildungsregressor ist zudem der standardisierte Regressionskoeffizient (Anstieg der Mortalität um 0,216 Standardabweichungseinheiten bei einem Anstieg der Bildung um eine Standardabweichungseinheit) weniger informativ als der unstandardisierte Regressionskoeffizient (Reduktion der Mortalität um ca. 16 Fälle pro 100.000 Einwohner für jedes Jahr zusätzlicher Schulbildung).

Trotz der berechtigten Skepsis gegenüber den standardisierten Koeffizienten ist doch festzustellen, dass sie in der Regel zum Vergleich der Effektstärken der Regressoren besser geeignet sind als die unstandardisierten Koeffizienten.

In einen Forschungsbericht sollten *beide* Varianten der Regressionskoeffizienten aufgenommen werden.

Während in der bivariaten Regression der standardisierte Regressionskoeffizient mit der Korrelation identisch ist, also den Wertebereich [-1; 1] besitzt, können die standardisierten Koeffizienten in der multiplen Regression Werte im Intervall (-∞;∞) annehmen. Man kann den standardisierten Regressionskoeffizienten $\tilde{b}_j$ von Regressor *j* mit Hilfe der semipartiellen Korrelation sr_j von Regressor *j* mit dem Kriterium (siehe Abschnitt 2.8.1, Wertebereich [-1; 1]) und der Toleranz (Eigenständigkeit) des Regressors (siehe Abschnitt 3.2, Wertebereich (0; 1]) ausdrücken (Disabato 2016):

$$\tilde{b}_j = \frac{sr_i}{\sqrt{Tol_j}}$$

Wenn ein Regressor eine kleine Toleranz besitzt, wenn also seine Varianz in hohem Maß durch die anderen Regressoren aufgeklärt wird, dann kann der Betrag von $\tilde{b}_j$ beliebig über den maximalen Betrag einer Korrelation hinauswachsen.

Das hat Konsequenzen für die Verwendbarkeit der standardisierten Regressionskoeffizienten zur Effektstärkenbeschreibung. Während zur Beschreibung von Korrelationsstärken weithin anerkannte Orientierungswerte existieren (siehe Tabelle im Abschnitt 1.8.1.1), sind vergleichbare Werte für die standardisierten Koeffizienten in der multiplen Regression *nicht* möglich.

Die quadrierte partielle Korrelation pr_j von Regressor *j* mit dem Kriterium bei statistischer Kontrolle der restlichen Regressoren bietet aber eine Möglichkeit zur Beschreibung des Effekts von Regressor *j* samt Orientierungswerten (siehe Abschnitt 2.8.3).

2.5 Signifikanztests und Konfidenzintervalle

2.5.1 Quadratsummenzerlegung und F-Test zur globalen Nullhypothese

Die totale Quadratsumme SQT (mit N - 1 Freiheitsgraden):

$$\text{QST} := \sum_{i=1}^{N} (y_i - \bar{y})^2$$

lässt sich auch im Modell der multiplen linearen Regression additiv zerlegen in einen durch die Regressoren erklärten Anteil QSR (mit k Freiheitsgraden):

$$\text{QSR} := \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2$$

und die unerklärte Fehlerquadratsumme QSF (mit N - k - 1 Freiheitsgraden):

$$\text{QSF} := \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Es gilt also:

$$\sum_{i=1}^{N} (y_i - \bar{y})^2 = \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Unter der globalen Nullhypothese

$$H_0: \beta_1 = ... = \beta_k = 0 \qquad (14)$$

ist bei gültigem Modell der Quotient

$$F := \frac{\frac{\text{QSR}}{k}}{\frac{\text{QSF}}{N-k-1}}$$

F-verteilt mit k Zählerfreiheitsgraden sowie N - k - 1 Nennerfreiheitsgraden und erlaubt damit eine Beurteilung der globalen Nullhypothese.

Für unser Beispiel liefert SPSS in der **ANOVA**-Tabelle den F-Wert 21,502 mit einer Überschreitungswahrscheinlichkeit kleiner 0,001. Falls die Prüfung der Voraussetzungen positiv endet, kann die globale Nullhypothese also verworfen werden.

2.5.2 Signifikanztests zu den einzelnen Regressionskoeffizienten

Für das zweiseitige Testproblem zum Regressionskoeffizienten β_j

$$H_0: \beta_j = 0 \text{ versus } H_1: \beta_j \neq 0$$

kann analog zum bivariaten Fall ein t-Test hergeleitet werden. Ferner kann ein Vertrauensintervall für β_j bestimmt werden.

Unter den Annahmen der multiplen Regression gilt für die Stichprobenverteilung von b_j:

$$b_j \sim N(\beta_j, \sigma_\varepsilon^2 (\mathbf{X}'\mathbf{X})_{jj}^{-1})$$

Dabei ist $(\mathbf{X}'\mathbf{X})_{jj}^{-1}$ das j-te Hauptdiagonalelement der Matrix $(\mathbf{X}'\mathbf{X})^{-1}$. Als geschätzte Standardabweichung $\hat{\sigma}_{b_j}$ der Stichprobenverteilung von b_j ergibt sich

$$\hat{\sigma}_{b_j} := \sqrt{\hat{\sigma}_\varepsilon^2 (\mathbf{X}'\mathbf{X})^{-1}_{jj}}$$

SPSS präsentiert diese Größe in der Spalte **Std.-Fehler** der **Koeffizienten**-Tabelle. In derselben Tabelle (siehe Spalte **T**) findet sich auch das Stichprobenergebnis zur folgenden Prüfgröße:

$$t := \frac{b_j}{\hat{\sigma}_{b_j}}$$

Sie folgt unter der Nullhypothese ($\beta_j = 0$) einer t-Verteilung mit $N - k - 1$ Freiheitsgraden.

In unserem Beispiel, dessen Annahmen noch zu prüfen sind, werden alle Regressoren mit Ausnahme der Stickstoffoxydbelastung vom zweiseitigen t-Test als signifikant beurteilt (siehe Spalte **Sig.** der **Koeffizienten**-Tabelle auf Seite 90).

In vielen Lehrbüchern zur Regressionsanalyse findet man ausschließlich die gerade beschriebenen *zweiseitigen* Hypothesen bzw. Signifikanztests zu den Regressionskoeffizienten (siehe z. B. Darlington & Hayes 2017, S. 105f). M.E. sollte in der Regel zum Regressionskoeffizienten β_j aber ein *einseitiger* Test bevorzugt werden, z. B. mit dem Hypothesenpaar:

$$H_o: \beta_j \leq 0 \text{ versus } H_1: \beta_j > 0$$

Die Ungleichungen im Hypothesenpaar sind problemadäquat zu wählen, wobei die mutmaßliche Richtung eines Effekts (z. B. von Luftverschmutzung auf Mortalität) wohl in der Regel a-priori bekannt sein dürfte.

Wer sich auf das Vorzeichen eines Regressionskoeffizienten festlegt, wird durch eine bessere Power beim einseitigen Signifikanztest belohnt: Wegen der Symmetrie der t-Verteilung kann die Überschreitungswahrscheinlichkeit eines einseitigen Tests durch Halbieren der von SPSS ausschließlich mitgeteilten Überschreitungswahrscheinlichkeit des zweiseitigen Tests ermittelt werden. Da man beim einseitigen Testen nur einen einseitigen Ablehnungsbereich verwendet, dessen Wahrscheinlichkeit unter der Nullhypothese maximal 5% beträgt, hält die beschriebene Testprozedur das α-Niveau ein. Die Wahrscheinlichkeit, einen vorhandenen Effekt zu entdecken, ist beim einseitigen Test größer. Dementsprechend wird sich im Abschnitt 2.10 zur Stichprobenumfangsplanung zeigen, dass beim einseitigen Test weniger Fälle erforderlich sind, um eine gewünschte Power zu erzielen.

Für die Beschränkung auf den zweiseitigen Test zu Regressionskoeffizienten werden gelegentlich folgende Argumente vorgebracht:

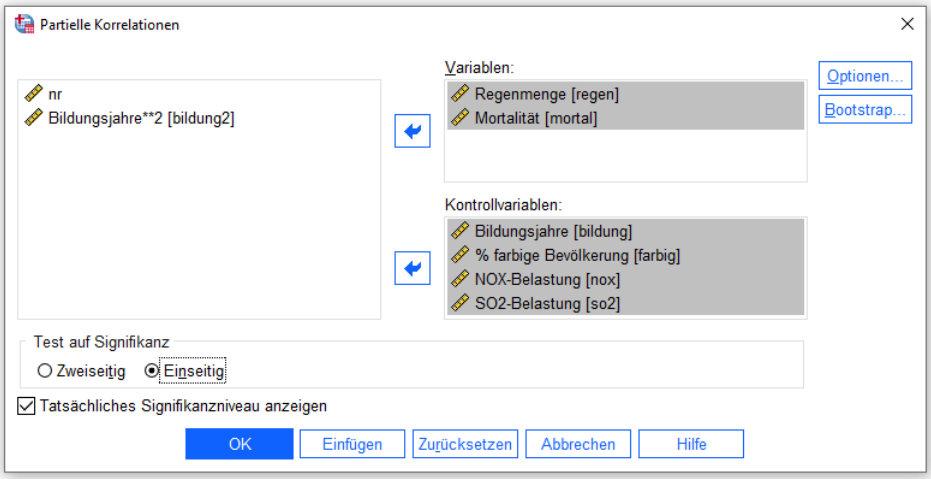
- Diese Praxis sei *üblich*, und ein einseitiger Test könne als Täuschungsversuch ausgelegt werden.
- Der zweiseitige Test sei konservativer (vorsichtiger beim Verwerfen seiner H_0).

Beide Argumente sind aber m. E. wenig überzeugend.

Äquivalent zum Test für einen Regressionskoeffizienten β_j ist übrigens der Test zur **Partialkorrelation** zwischen Y und X_j bei Kontrolle der restlichen Prädiktoren (vgl. Abschnitt 2.8.3). In SPSS können Sie die Partialkorrelation, mit der wir uns im Abschnitt 2.6 noch beschäftigen werden, samt Signifikanztest über den Menübefehl

Analysieren > Korrelation > Partiell

anfordern, z. B. für die Variablen MORTAL und REGEN:



Diesmal dürfen Sie zwischen ein- und zweiseitiger Testung wählen. Wie nach obigen Überlegungen zu erwarten, ist die von SPSS ausgegebene empirische Überschreitungswahrscheinlichkeit

Korrelationen

Kontrollvariablen			Regenmenge	Mortalität
Bildungsjahre & % farbige Bevölkerung & NOX-Belastung & SO2-Belastung	Regenmenge	Korrelation	1,000	,312
		Signifikanz (einseitig)	.	,010
		Freiheitsgrade	0	54
	Mortalität	Korrelation	,312	1,000
		Signifikanz (einseitig)	,010	.
		Freiheitsgrade	54	0

beim einseitigen Test genau halb so groß wie die im Rahmen der multiplen Regressionsanalyse berechnete zweiseitige Variante:

$$0{,}009568 \approx \frac{0{,}019136}{2}$$

Während man sich in der Regel nur für *eine* partielle Korrelation interessiert, treten t-Tests zu den Koeffizienten in einer multiplen Regression gruppenweise auf, sodass eine α-Fehler - Kumulierung zu befürchten ist. Im Abschnitt 2.5.4 wird darüber spekuliert, dass es sich bei ...

- der zweiseitigen Testung von Regressionskoeffizienten trotz vorhandener Richtungshypothesen
- und dem Verzicht auf Maßnahmen gegen eine α-Fehler - Kumulierung

um zwei Fehler handelt, die sich in ihre Wirkung auf die α- bzw. β-Fehler in den Tests zu den Regressionskoeffizienten teilweise gegenseitig kompensieren.

2.5.3 Konfidenzintervalle zu den Regressionskoeffizienten

Analog zum bivariaten Fall lässt sich mit Hilfe des geschätzten Standardfehlers $\hat{\sigma}_{b_j}$ zu b_j (siehe Abschnitt 2.5.2) ein Konfidenzintervall für β_j bestimmen. Das folgende Intervall

$$[b_j - t_{1-\alpha/2;N-k-1}\,\hat{\sigma}_{b_j};\quad b_j + t_{1-\alpha/2;N-k-1}\,\hat{\sigma}_{b_j}]$$

enthält bei Gültigkeit aller Voraussetzungen mit einer Wahrscheinlichkeit von (1 - α) den wahren Parameter β_j. Dabei ist $t_{1-\alpha/2;N-k-1}$ das (1 - α/2) - Quantil der t-Verteilung mit $N - k - 1$ Freiheitsgraden (= kritischer t-Wert für das zweiseitige Testen zum Niveau α).

SPSS liefert die 95% - Vertrauensintervalle zu den Regressionsgewichten (aufgrund unserer Zusatzanforderung in der **Statistiken**-Subdialogbox, siehe Abschnitt 2.1.2) in der **Koeffizienten**-Tabelle. Zum Regressionsgewicht der Bildungsvariablen erhalten wir z. B. das Intervall:

$$[-30{,}266;\ -1{,}558]$$

Auch bei der multiplen linearen Regression gilt, dass der *zweiseitige* Signifikanztest zum Regressionskoeffizienten β_j beim α-Fehlerrisiko von 5% seine Nullhypothese ($\beta_j = 0$) genau dann beibehält, wenn das 95% - Vertrauensintervall zu β_j den Wert 0 enthält.

Passend zum *einseitigen* Signifikanztest lässt sich auch ein *einseitiges* Konfidenzintervall konstruieren. Bei

$$H_1: \beta_j < 0$$

resultiert zu b_j z. B. das folgende linksseitig offene einseitige Vertrauensintervall:

$$(-\infty;\ b_1 + t_{1-\alpha;N-k-1}\,\hat{\sigma}_{b_j}\,]$$

In unserem Beispiel ($N = 60$, $k = 5$, $\alpha = 0{,}05$) erhalten wir als relevantes Quantil der t-Verteilung

$$t_{0,95;54} \approx 1{,}674$$

und als linksseitig offenes, einseitiges Vertrauensintervall für das Regressionsgewicht zur Bildungsvariablen:

$$(-\infty;\ -15{,}912 + 1{,}674\cdot\ \cdot 7{,}160] \approx (-\infty;\ -3{,}926]$$

Durch den Verzicht auf eine linke Schranke erhalten wir eine präzisere rechte Schranke.

Der *einseitige* Signifikanztest zum Regressionskoeffizienten β_j (siehe Abschnitt 2.5.2) verwirft seine Nullhypothese übrigens genau dann, wenn das zugehörige einseitige Konfidenzintervall zu β_j den Wert 0 *nicht* enthält.

2.5.4 Multiples Testen

In einer multiplen Regressionsanalyse werden mehrere Signifikanztests ausgeführt, sodass Überlegungen zur **α-Fehler - Kumulierung** angebracht sind.

2.5.4.1 Scheinbar geschützte t-Tests zu den Regressionskoeffizienten

Sind in einer multiplen Regressionsanalyse alle Modellannahmen erfolgreich geprüft worden, dann wird bei den nun zulässigen Signifikanztests in der Regel so verfahren:

- Zuerst wird F-Test zur globalen Nullhypothese (siehe Formel (14)) durchgeführt (meist mit einem α-Fehlerrisiko von 5%).
- Kann die globale Nullhypothese verworfen werden, dann werden zu den einzelnen Regressoren t-Tests auf dem 5% - Niveau durchgeführt.

Cohen et al. (2003, S. 187ff) sehen in dieser Teststrategie die erfolgreiche Übertragung der von Sir Ronald Fisher für die einfaktorielle Varianzanalyse vorgeschlagenen *geschützten t-Tests* auf die multiple Regression. Gemäß Fishers Idee dürfen bei der einfaktoriellen Varianzanalyse nach einem signifikanten F-Test zur globalen Nullhypothese paarweise Einzelvergleiche über t-Tests auf dem 5% - Niveau durchgeführt werden. Allerdings hält Fishers Vorschlag das multiple Niveau α *nicht* ein, schützt also *nicht* vor einer α-Fehler - Kumulierung (siehe z. B. Hsu 1996).

Dementsprechend kann in der multiplen Regressionsanalyse der vorgeschaltete Globaltest eine α-Fehler - Kumulierung bei den t-Tests zu den einzelnen Regressionsparametern *nicht* verhindern. Bei Gültigkeit der

globalen Nullhypothese (*alle* Koeffizienten sind gleich 0) erfüllt der F-Test zweifellos eine Schutzfunktion, doch handelt es sich hier um eine eher seltene Konstellation.

Zur weiteren Illustration des Problems soll an Stelle einer anstrengenden mathematischen Argumentation von einer Simulationsstudie berichtet werden. Es liegt das folgende wahre Modell zugrunde:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \quad \text{mit } \beta_1 = 0{,}7 \text{ und } \beta_0 = \beta_2 = \beta_3 = ... = \beta_6 = 0$$

Wir betrachten die folgende Hypothesenfamilie:[1]

$$H_0 : \beta_1 = \beta_2 = ... = \beta_6 = 0$$

$$H_0^{(j)} : \beta_j = 0,\ j = 1,...,6$$

Nach Konstruktion der künstlichen Population sind H_0 und $H_0^{(1)}$ falsch, alle anderen Nullhypothesen sind wahr. Es soll demonstriert werden, dass die oben beschriebene Teststrategie in der Tradition von Sir Ronald Fisher zu einer erheblichen α-Fehler - Kumulierung führt.

Mit der folgenden SPSS-Syntax wurde die Teststrategie in 100 Stichproben durchgeführt:[2]

```
input program.
+ loop #i = 1 to 50.
-  compute x1 = normal(1).
-  compute y = 0.7 * x1 + normal(1).
-  end case.
+ end loop.
+ end file.
end input program.

do repeat  zuf = x2 to x6.
  compute  zuf = normal(1).
end repeat.

regression  dependent = y /enter = x1, x2 to x6.
```

Dabei wurde in **27** Stichproben mindestens eine wahre Nullhypothese abgelehnt, z. B.:

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	34,728	6	5,788	5,212	,000 [b]
	Nicht standardisierte Residuen	47,750	43	1,110		
	Gesamt	82,478	49			

a. Abhängige Variable: y

b. Einflußvariablen : (Konstante), x6, x1, x2, x5, x4, x3

[1] Weil es sich um *künstliche* Variablen handelt, verwenden wir trotz der Diskussion im Abschnitt 2.5.2 hier *zweiseitige* Hypothesentests.

[2] Die SPSS-Syntax ist in der Datei **LSDemo.sps** an der im Vorwort vereinbarten Stelle zu finden.

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.
1	(Konstante)	-,051	,165		-,309	,759
	x1	,672	,164	,502	4,111	,000
	x2	,093	,144	,079	,646	,522
	x3	-,443	,166	-,329	-2,665	,011
	x4	,190	,171	,137	1,115	,271
	x5	,177	,195	,110	,903	,371
	x6	-,113	,176	-,078	-,644	,523

a. Abhängige Variable: y

Wer eine α-Fehler - Kumulierung vermeiden will, muss eine korrekte multiple Testprozedur zum globalen Niveau α für die Familie $\{ H_0^{(j)} : \beta_j = 0,\ j = 1,\ldots,6 \}$ durchführen, sodass gilt: Die Wahrscheinlichkeit, bei Durchführung *aller* Tests aus der Familie einen oder mehrere Fehler erster Art zu begehen, ist kleiner als α.

Im Abschnitt 2.5.4.2 werden multiple Tests zum globalen Niveau α beschrieben, ohne eine Empfehlung für diese Teststrategie auszusprechen. Maßnahmen gegen die α-Fehler - Kumulierung führen bei konstantem Stichprobenumfang zu einer reduzierten Power, also zu einem erhöhten Risiko für β-Fehler.

Offenbar sind bei der üblichen Testpraxis in der multiplen Regressionsanalyse zwei Fehler mit gegenläufigen Wirkungen beteiligt:

- Nach einem erfolgreichen Globaltest für die Regressoren *k* Einzeltests zum Niveau α durchzuführen, führt in der Regel zu einer α-Fehlerkumulierung, wobei die Power der *k* Einzeltests profitiert.
- Die Einzeltests zu den Regressoren grundsätzlich zweiseitig durchzuführen (vgl. Abschnitt 2.5.2), obwohl in der Regel a-priori - Hypothesen über die Wirkrichtungen vorhanden sind, reduziert als konservative Teststrategie den α-Fehler und erhöht gleichzeitig den β-Fehler.

In vielen Fällen kann man die Entscheidung für α-Fehler - reduzierende zweiseitige Tests, die dann aber ohne Berücksichtigung der α-Fehlerkumulierung durchgeführt werden, als Bonferroni-Prozedur interpretieren, wobei aber nicht α/*k* als Risiko für Fehler erster Art verwendet wird, sondern α/2. Der vorgeschaltete Globaltest wird bei dieser Betrachtung ignoriert in der Annahme, dass nur selten ausschließlich irrelevante Regressoren in eine Studie einbezogen werden. Folglich wird die globale Nullhypothese wohl in den meisten Fällen verworfen.

Trotz der Möglichkeit der partiellen Aufhebung von zwei entgegengesetzt wirkenden Fehlern kann die übliche Teststrategie im Vergleich zu korrekten multiplen Tests zum globalen Niveau α nicht überzeugen:

- Es resultiert eine erhöhte Rate von Fehlern erster Art, was vor allem *bei einer großen Anzahl* von Regressoren ins Gewicht fällt (siehe obige Simulationsstudie).
- Vor allem *bei einer kleinen Anzahl* von Regressoren (z.B. zwei) ist die Power reduziert.

2.5.4.2 Multiple Tests mit α-Fehler - Kumulierungskontrolle nach Bonferroni-Holm

Das Bonferroni-Holm - Verfahren für multiple Tests zum globalen Niveau α ist einfach anwendbar, erlaubt sowohl zwei- als auch einseitige Tests und ist bei einer kleinen Anzahl von Regressoren auch effektiv (siehe z. B. Sonnemann 1982). Man ordnet die t-Tests zu den *k* Regressoren aufsteigend nach den erzielten Überschreitungswahrscheinlichkeiten (p-Levels) und geht dann so vor:

- Der Test mit dem kleinsten p-Level wird zum Niveau $\frac{\alpha}{k}$ durchgeführt. Akzeptiert er seine Nullhypothese, dann stoppt das Verfahren, und *alle* Nullhypothesen werden beibehalten.
- Verwirft der erste Test seine Nullhypothese, dann wird der nächste Test (mit dem nächst-größeren p-Level) zum Niveau $\frac{\alpha}{k-1}$ durchgeführt usw.
- Sobald ein Test seine Nullhypothesen beibehält, stoppt das Verfahren, und *alle* weiteren Nullhypothesen werden ebenfalls beibehalten. Wenn die ersten *k* - 1 Tests ihre Nullhypothese verwerfen, wird der letzte Test zum Niveau $\frac{\alpha}{k-(k-1)} = \alpha$ durchgeführt.

Für die im Abschnitt 2.5.4.1 berichtete Realisation einer Simulationsstudie zur α-Fehler - Kumulierung in der multiplen linearen Regressionsanalyse ergibt sich bei diesem Verfahren:

- Die Nullhypothese zu X_1 wird (zu Recht) verworfen:
$$0{,}000174 < \frac{0{,}05}{6} = 0{,}008\overline{3}$$
- Die Nullhypothese zu X_3 wird (ebenfalls zu Recht) akzeptiert:
$$0{,}010812 > \frac{0{,}05}{6-1} = 0{,}01$$
- Damit werden auch die Nullhypothesen zu den restlichen Regressoren (wiederum zu Recht) akzeptiert.

2.6 Determinationskoeffizient

Als Maß für die Erklärungsleistung eines multiplen Regressionsmodells wird wie im bivariaten Fall der Determinationskoeffizient R^2 verwendet (vgl. Abschnitt 1.5):

$$R^2 := \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = \frac{\mathrm{QSR}}{\mathrm{QST}} = \frac{\mathrm{QST} - \mathrm{QSF}}{\mathrm{QST}} = 1 - \frac{\mathrm{QSF}}{\mathrm{QST}}$$

Die Stichprobenstatistik überschätzt den determinierten Varianzanteil in der Population, weil die R^2 - Definition die in den Quadratsummen enthaltenen Freiheitsgrade nicht berücksichtigt. Man kann z. B. die Fehlerquadratsumme beliebig nahe an den „idealen" Wert 0 und damit R^2 beliebig nahe an den „idealen" Wert 1 bringen, indem man die Anzahl der Regressoren in Relation zum Stichprobenumfang hinreichend steigert. Um die artifizielle Inflationierung zu verhindern und zu einem nahezu unverzerrten Schätzwert zu gelangen, werden im **korrigierten** bzw. **adjustierten** R_a^2 alle Quadratsummen durch die Anzahl der enthaltenen Freiheitsgrade dividiert, sodass sich im Fall der multiplen Regression ergibt:

$$R_a^2 := 1 - \frac{\frac{\sum_{i=1}^{N} e_i^2}{N-k-1}}{\frac{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2}{N-1}} = 1 - \frac{\frac{\mathrm{QSF}}{N-k-1}}{\frac{\mathrm{QST}}{N-1}} \quad (k: \text{Anzahl der Regressoren})$$

Diesem adjustierten R_a^2 sollte zur Beurteilung der Erklärungsleistung eines Modells regelmäßig der Vorzug gegeben werden.

Für die Beziehung zwischen R_a^2 und R^2 gilt bei der multiplen Regression:

$$R_a^2 = 1-(1-R^2)\frac{N-1}{N-k-1} = R^2-(1-R^2)\frac{k}{N-k-1}$$

Wegen $N > (k + 1)$ (technische Voraussetzung, siehe Abschnitt 2.2.4) und $R^2 \in [0; 1]$ gilt:

- $R_a^2 \leq R^2$
- $R_a^2 < R^2$, falls $R^2 < 1$

Der minimale Wert von R_a^2 ist *negativ* ($-\frac{k}{N-k-1}$) und wird erreicht, wenn $R^2 = 0$ ist. Negative R_a^2 - Werte sollten auf 0 gesetzt werden.

Der Korrekturterm im Vergleich zu R^2

$$-(1-R^2)\frac{k}{N-k-1}$$

zeigt die „Bestrafung“ für eine wachsende Anzahl k von Regressoren.

Wie Wooldridge (2013, S. 194) ausführt, stimmen die Empfehlungen des korrigierte Determinationskoeffizient zur Aufnahme neuer Regressoren in ein Modell nicht unbedingt mit den Empfehlungen der zugehörigen Signifikanztests überein:

- Bei Aufnahme eines neuen Regressors steigt der korrigierte Determinationskoeffizient genau dann, wenn die t-Prüfgröße des zugehörigen Signifikanztests betragsmäßig größer als 1 ist. Der korrigierte Determinationskoeffizient kann also auch bei der Aufnahme eines neuen Regressors steigen, der als *insignifikant* beurteilt wird.
- Bei der Aufnahme eines Blocks von Regressoren (vgl. Abschnitt 5.1) steigt der korrigierte Determinationskoeffizient genau dann, wenn die F-Prüfgröße des zugehörigen Signifikanztests größer als 1 ist.

SPSS liefert in der Tabelle **Modellzusammenfassung** für das Mortalitätsbeispiel ($k = 5$, $N = 60$) einen korrigierten R_a^2 - Wert von 0,635, den wir der Übung halber nachrechnen wollen:

$$R_a^2 = 1-\frac{\frac{76348,876}{54}}{\frac{228352,77}{59}} = 0,635$$

Derart hohe Werte sind in den Bio-, Geo-, Sozial- und Wirtschaftswissenschaften selten. Es ist fast als didaktischer Mangel des Manuskripts zu werten, dass in allen Beispielen sehr große Determinationskoeffizienten auftreten, während die meisten Leser in eigenen Studien vermutlich mit deutlich kleineren Werten zufrieden sein müssen (und dürfen).

Zur Beschreibung von Populations-Effektstärken von *multiplen* Regressionsmodellen nennen Cohen et al. (2003, S. 93; siehe auch Cohen 1988, S. 413f) folgende Orientierungsgrößen für die Verhaltens- und Sozialwissenschaften:

Effektstärke in der Population	erklärter Varianzanteil	Effektstärkeindex f^2
klein	0,02	0,02
mittel	0,13	0,15
groß	0,26	0,35

In der dritten Spalte befindet sich der zur Stichprobenumfangsplanung benötigte Effektstärkeindex f^2 (vgl. Abschnitt 2.10). Im Vergleich zur bivariaten Regression fallen die Effektstärkenorientierungswerte für die *multiple* linearen Regression etwas höher aus (vgl. Abschnitt 1.5).

Cohen et al. (2003, S. 93) nennen die Orientierungswerte für beliebige multiple Regressionsmodelle (unabhängig von der Anzahl der Regressoren), merken aber an:

> These values should probably be adjusted upward by the researcher who intends to use more than a few IVs (*independent variables*).

Wie schon im Abschnitt 1.5 erläutert wurde, dienen die Effektstärken-Orientierungswerte zur Beschreibung von Modellen, nicht zur Bewertung.

Das *nicht*-korrigierte R^2 ist identisch mit dem Quadrat der multiplen Stichprobenkorrelation zwischen dem Kriterium und den Regressoren, und diese multiple Korrelation stimmt wiederum mit der einfachen Korrelation zwischen dem Kriterium und der Modellprognose überein.[1]

Zur Berechnung eines Vertrauensintervalls zum Determinationskoeffizienten für das Modell mit fixierten Regressoren oder für das Modell mit stochastischen Regressoren kann man exakt nach der Beschreibung im Abschnitt 1.5.3 vorgehen. Im Mortalitätsbeispiel erhält man vom R-Paket **MBESS** für das (streng genommen unangemessene) Modell mit fixierten Regressoren das 95% - Vertrauensintervall [0,473; 0,736]:

```
$Lower.Conf.Limit.R2
[1] 0.4725262

$Upper.Conf.Limit.R2
[1] 0.7363877
```

Für das Modell mit stochastischen Regressoren resultiert das erwartungsgemäß größere 95% - Vertrauensintervall [0,559; 0,767]:

```
$Lower.Conf.Limit.R2
[1] 0.4591981

$Upper.Conf.Limit.R2
[1] 0.7667689
```

2.7 Phänomene der multiplen Regression

In diesem Abschnitt werden Phänomene behandelt, die in einer *multiplen* Regressionsanalyse im Unterscheid zu bivariaten Zusammenhangsanalysen auftreten und dabei unser Verständnis von einem empirischen System verbessern können. Dabei beschränken wir uns auf

- partielle Redundanz
- Vulnerabilitäts-Kompensation
- Suppression

Zwei verwandte Themen werden später behandelt:

- Multikollinearität (siehe Abschnitt 3.2)
- Kausale Fehlschlüsse durch vergessene Regressoren (siehe Abschnitt 6.1)

[1] Diese Aussagen gelten für die üblicherweise verwendete inhomogene Regression (*inklusive* Achsenabschnitt β_0), auf die wir uns im Abschnitt 2 beschränken.

2.7.1 Einfacher und partieller Regressionskoeffizient

Die Formeln für den geschätzten Koeffizienten b_1 zum Regressor X_1 (vgl. Darlington & Hayes 2017, S. 83)

a) in der bivariaten Regression von Y auf X_1

$$b_1 = r_{YX_1} \frac{s_Y}{s_{X_1}}$$

b) sowie in der multiplen Regression von Y auf X_1 und X_2

$$b_1 = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{1 - r^2_{X_1X_2}} \frac{s_Y}{s_{X_1}}$$

zeigen, wovon die Änderung des Koeffizienten zu X_1 bei der Aufnahme des Regressors X_2 in das Modell abhängt: Der Koeffizient zu X_1 bleibt unverändert, wenn X_2 nicht mit X_1 korreliert.

2.7.2 Partielle Redundanz

Im eher seltenen Fall *unkorrelierter* Regressoren ist das (*nicht*-adjustierte) R^2 identisch mit der Summe der quadrierten Einzelkorrelationen r_j der Regressoren mit dem Kriterium, also mit der Summe der bivariaten Determinationskoeffizienten:

$$R^2 = \sum_{j=1}^{k} r_j^2$$

Das folgende Pfaddiagramm beschreibt ein Modell mit zwei unkorrelierten Regressoren X_1 und X_2, die beide einen Effekt auf das Kriterium Y ausüben:

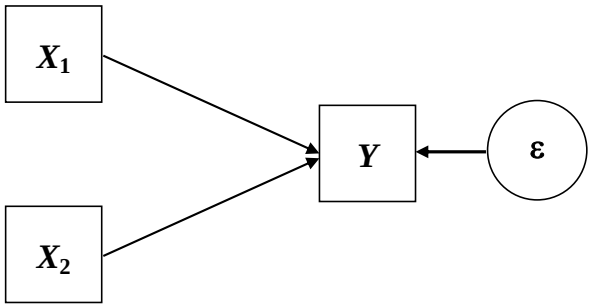


Abbildung 27: Modell mit 2 unkorrelierten Regressoren

Speziell in Beobachtungsstudien sind die Regressoren oft untereinander korreliert, sodass sich bei zwei Regressoren mit Effekten auf das Kriterium das folgende Pfadmodell ergibt:

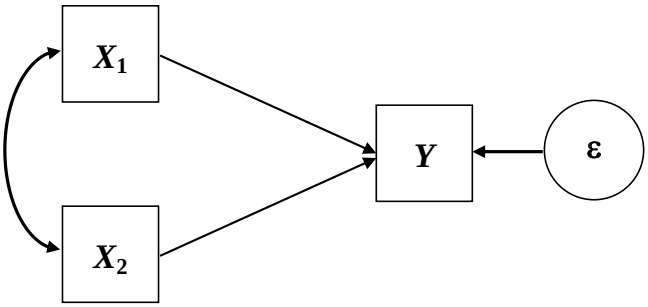


Abbildung 28: Modell mit zwei korrelierten Regressoren (partielle Redundanz)

Häufig liegt eine partielle Überlappung der in beiden Regressoren über das Kriterium enthaltenen Informationen vor. In dieser Situation bleibt das R^2 der multiplen Regression hinter der Summe der quadrierten Einzelkorrelationen zurück, weil die Erklärungsleistungen der Regressoren partiell **redundant** sind:

$$R^2 < \sum_{j=1}^{k} r_j^2$$

Im Abschnitt 3.2 über die **Multikollinearität** wird eine Konstellation mit *hochgradiger* Redundanz behandelt, wobei ein Regressor fast vollständig durch andere Regressoren erklärt werden kann und somit fast keine eigenständige Information in das Modell einbringt.

Um ein multiples Regressionsmodell Modell mit partieller Redundanz besser zu verstehen, sollte der Grund für die Korrelation zwischen den Regressoren aufgeklärt werden. Zwei Regressoren sind z.B. dann korreliert und (sogar vollständig) redundant, wenn beide als Manifestationen derselben, kausal wirksamen latenten Variablen betrachtet werden können:

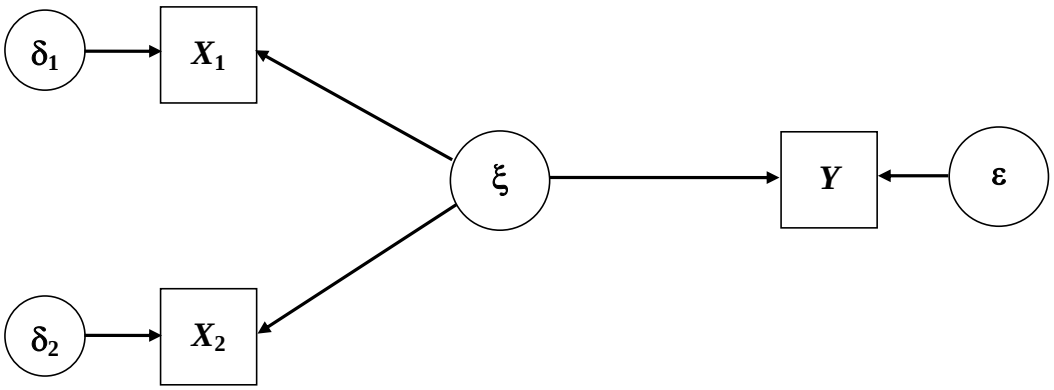


Abbildung 29 In ein multiples Regressionsmodell sollten nicht mehrere Manifestationen derselben, kausal wirksamen latenten Variablen aufgenommen werden.

2.7.3 Vulnerabilitäts-Kompensation

Der Determinationskoeffizient in einer multiplen Regression kann auch *größer* ausfallen als die Summe der quadrierten Kriteriumskorrelationen:

$$R^2 > \sum_{j=1}^{k} r_j^2$$

Das passiert z. B., wenn ...

- das Kriterium *Y* eine Schadensbilanz quantifiziert (z.B. Schäden durch Waldbrand in den als Fälle fungierenden Regionen),
- ein Regressor *X* bei zunehmender Ausprägung die Schadensbilanz im Kriterium *Y* erhöht (z.B. Waldbrandrisiko einer Region)
- und gleichzeitig den Einsatz eines kompensatorisch wirkenden Regressors *Z* steigert (z.B. Waldbrandprävention in einer Region).

Im folgenden extremen Beispiel heben sich der direkte Effekt von *X* auf *Y* und der durch *Z* vermittelte indirekte Effekt von *X* auf *Y* gegenseitig auf:

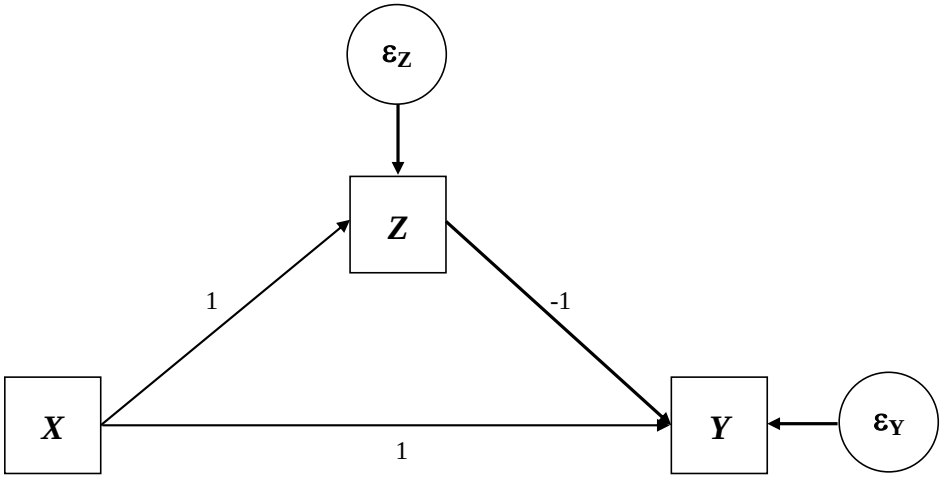


Abbildung 30 Mediationsmodell mit Vulnerabilitäts-Kompensation
(direkter und indirekter Effekt mit unterschiedlichen Vorzeichen)

Mit simulierten Daten (N = 500) nach dem Muster von Abbildung 30 zeigen sich in der multiplen Regression von Y auf X und Z zwei signifikante Regressionskoeffizienten, während fast keine Korrelation zwischen X und Y besteht (0,021):

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B		Korrelationen		
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze	Nullter Ordnung	Partiell	Teil
1	(Konstante)	,059	,046		1,290	,198	-,031	,150			
	x	1,044	,067	,721	15,572	,000	,912	1,175	,021	,573	,506
	z	-1,002	,047	-,982	-21,218	,000	-1,095	-,910	-,469	-,689	-,689

a. Abhängige Variable: y

Die quadrierte multiple Korrelation (unkorrigiert) beträgt 0,476 und ist erheblich größer als die Summe der beiden quadrierten Kriteriumskorrelationen:

$$0{,}4755 > 0{,}0005 + 0{,}2196$$

Wir haben gerade zur Erläuterung von Phänomenen der multiplen Regression ein sogenanntes *Mediationsmodell* betrachtet (siehe Abschnitt 6.2).

2.7.4 Suppressoreffekt

Auch im interessanten (aber seltenen) Fall einer Suppression übertrifft das multiple R^2 die Summe der bivariaten Determinationskoeffizienten:

$$R^2 > \sum_{j=1}^{k} r_j^2$$

Zu einer Suppression kommt es z.B. dann, wenn ein Kriterium Y sowie zwei Regressoren X_1 und X_2 folgendermaßen aus den beiden latenten Variablen ξ_1 und ξ_2 hervorgehen:

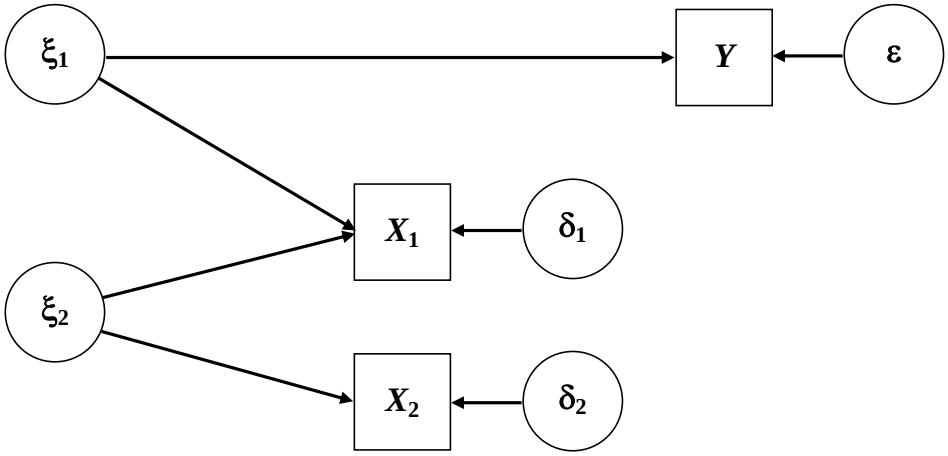


Abbildung 31: Bedingungshintergrund für einen Supressoreffekt in der Regression von Y auf X_1 und X_2

Zwar enthält X_2 keine Information über Y, doch ist dieser Regressor im Stande, den von ξ_2 stammenden, störenden Anteil in X_1 zu unterdrücken.

In einer Simulationsstudie mit $N = 300$ Fällen wurde das Modell aus dem obigen Pfaddiagramm realisiert, wobei der Einfachheit halber alle Regressionskoeffizienten auf 1 gesetzt wurden. Erwartungsgemäß besteht keine Korrelation von X_2 und Y (siehe Spalte mit den **Korrelationen nullter Ordnung**), während X_1 einen wesentlichen Beitrag zur Vorhersage des Kriteriums leistet:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B		Korrelationen		
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze	Nullter Ordnung	Partiell	Teil
1	(Konstante)	2,878	,063		46,035	,000	2,755	3,001			
	x1	,968	,060	,942	16,056	,000	,850	1,087	,514	,682	,681
	x2	-,949	,090	-,620	-10,568	,000	-1,126	-,773	,031	-,523	-,448

a. Abhängige Variable: y

Für X_1 wird im multiplen Regressionsmodell ein Koeffizient im Bereich des tatsächlichen Effekts (von ξ_1) geschätzt, während in der bivariaten Regression dieses realistische Ergebnis weit verfehlt wird:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B		Korrelationen		
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze	Nullter Ordnung	Partiell	Teil
1	(Konstante)	2,911	,073		39,811	,000	2,767	3,055			
	x1	,528	,051	,514	10,341	,000	,428	,629	,514	,514	,514

a. Abhängige Variable: y

Aus der multiplen Regressionsgleichung mit dem signifikanten Gewicht von X_2, das auf seiner Suppressorwirkung beruht, auf einen kausalen Effekt von X_2 zu schließen, wäre ein Fehler. Daraus abgeleitet Interventionsprogramme (z.B. X_2 reduzieren, um den Kriteriumswert zu verbessern) hätten keine Aussicht auf Erfolg.

In vielen Regressionsmodellen (z. B. mit einer partiellen Redundanz, einer Suppression oder mit einer Mediation) liefern die einfachen Kriteriumskorrelationen wichtige Beiträge zur Interpretation, sodass sie im Ergebnisbericht erscheinen sollten. Man erhält sie von der SPSS-Regressionsprozedur bei Anforderung der (semi)partiellen Korrelationen (vgl. Abschnitt 2.1). Neben der Bequemlichkeit spricht für diese Methode zur Anforderung von Kriteriumskorrelationen die Konsistenz mit der fallweisen Behandlung fehlender Werte durch die Regressionsprozedur. Bei einer separaten Berechnung (z. B. per **Analysieren > Korrelation > Bivariat**) mit paarweiser Behandlung fehlender Werte ist mit einer variablen Stichprobenzusammensetzung und dementsprechend abweichenden Ergebnissen zu rechnen.

2.8 Individuelle Erklärungsbeiträge der Regressoren

In diesem Abschnitt geht es darum, wie sich in einer multiplen Regressionsgleichung die (direkten) Erklärungsbeiträge der einzelnen Regressoren quantifizieren lassen.

2.8.1 Quadrierte semipartielle Kriteriumskorrelationen

Zur Beurteilung des eigenständigen Erklärungsbeitrags eines Regressors X_j betrachtet man den von ihm bewirkten R^2 - *Anstieg*. Man erhält ihn als Differenz der R^2 - Werte aus dem vollständigen und dem reduzierten Modell, wobei letzteres durch Streichen des zu beurteilenden Regressors entsteht. Dieser R^2 - Anstieg ist identisch mit dem Quadrat der semipartiellen Korrelation sr_j des Kriteriums mit dem fraglichen Regressor, aus dem alle anderen Regressoren auspartialisiert worden sind. Bei der semipartiellen Korrelation von Y und X_1 wird also zunächst von X_1 die beste X_1 - Prognose aufgrund der übrigen Regressoren (X_2, X_3, ...) subtrahiert. Dieses Residuum wird dann mit Y korreliert.

Weil die im eben beschriebenen Sinn bereinigten Regressoren im Manuskript später erneut auftauchen, soll ihre Definition präzisiert werden: Wenn man mit $\mathbf{X}^{(j)}$ die Designmatrix ohne die Spalte zu X_j bezeichnet und mit $\tilde{\beta}^{(j)}$ den Parametervektor aus der Regression von X_j auf die restlichen Regressoren, dann kann man den j-ten bereinigten Regressor $\delta^{(j)}$ so definieren:

$$\delta^{(j)} := X_j - \mathbf{X}^{(j)} \cdot \tilde{\beta}^{(j)}$$

SPSS liefert die (unquadrierten) semipariellen Korrelationen des Kriteriums mit den Regressoren aufgrund unserer Zusatzanforderung (vgl. Abschnitt 2.1) in der mit **Teil** beschrifteten Spalte der **Koeffizienten**-Tabelle:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B		Korrelationen			Kollinearitätsstatistik	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze	Nullter Ordnung	Partiell	Teil	Toleranz	VIF
1	(Konstante)	998,966	93,028		10,738	,000	812,457	1185,476					
	Regenmenge	1,658	,686	,266	2,415	,019	,282	3,034	,509	,312	,190	,510	1,961
	Bildungsjahre	-15,912	7,160	-,216	-2,223	,030	-30,266	-1,558	-,510	-,289	-,175	,654	1,528
	% farbige Bevölkerung	3,026	,650	,434	4,658	,000	1,723	4,328	,644	,535	,367	,713	1,402
	NOX-Belastung	,085	,487	,023	,174	,862	-,891	1,060	,216	,024	,014	,362	2,764
	SO2-Belastung	,312	,122	,318	2,570	,013	,069	,556	,426	,330	,202	,404	2,478

a. Abhängige Variable: Mortalität

Daraus lässt sich z. B. für den Regressor FARBIG ein eigenständiger R^2 - Beitrag von ca. 13,5% ermitteln:

$$0{,}367^2 \approx 0{,}135$$

Bei unkorrelierten Regressoren sind die semipartiellen Korrelationen identisch mit den einfachen Kriteriumskorrelationen r_j und ihre quadrierten Werte addieren sich zum (nicht-korrigierten) Determinationskoeffizienten:

$$R^2 = \sum_{j=1}^{k} r_j^2$$

Im weitaus häufigeren Fall *korrelierter* Regressoren besteht eine solche additive Zerlegung des Determinationskoeffizienten in die eigenständigen Erklärungsbeiträge der einzelnen Regressoren *nicht*. Die Summe der quadrierten semipariellen Korrelationen ist im typischen Fall wechselseitiger Redundanz (vgl. Abschnitt 2.7.2) *kleiner* als der unkorrigierte Determinationskoeffizient. Im Beispiel erhalten wir:

$$R^2 = 0{,}666 > \sum_{j=1}^{k} sr_j^2 = 0{,}242$$

Die Summe der quadrierten semipariellen Korrelationen kann den unkorrigierten Determinationskoeffizienten aber auch übertreffen (z.B. bei einem Suppressoreffekt, siehe Abschnitt 2.7.4).

2.8.2 R^2 - Anstiege aus einer Serie geschachtelter Modelle

Sofern unter den Regressoren X_1, X_2, …, X_k eine **kausale Ordnung** derart besteht, dass kein Regressor X_j einen Effekt auf einen anderen Regressor mit kleinerem Index ausübt, dann kann trotz korrelierter Regressoren eine additive R^2 - Zerlegung über eine Serie von geschachtelten Modellen vorgenommen werden. Man startet mit einem Modell, das lediglich den Regressor mit der höchsten kausalen Priorität (also X_1) enthält und nimmt sukzessive entlang der Kausalitätsordnung weitere Regressoren auf. Die dabei festgestellten R^2 - Zuwächse werden dem jeweils aufgenommenen Regressor als Erklärungsleistung zugeschrieben. Die so über R^2 - Anstiege in sukzessiv erweiterten Modellen definierten Individualbeiträge der Regressoren addieren sich trivialerweise zum R^2 des Komplettmodells. Selbstverständlich hängt die Höhe der so ermittelten Individualbeiträge entscheidend von der angenommenen kausalen Anordnung der Regressoren ab.

2.8.3 Quadrierte partielle Kriteriumskorrelationen (partielles Eta-Quadrat)

Bei der Power-Analyse (siehe Abschnitt 2.10) wird als Maß für den individuellen Erklärungsbeitrag eines Regressors X_j seine quadrierte *partielle* Korrelation mit dem Kriterium (bei statistischer Kontrolle der restlichen Regressoren) verwendet. Bei der *semipartiellen* Korrelation sr_j eines Regressors X_j mit dem Kriterium Y werden die restlichen Regressoren nur aus X_j, nicht jedoch aus dem Kriterium auspartialisiert. Bei der *partiellen* Korrelation pr_j eines Regressors X_j mit dem Kriterium Y werden die restlichen Regressoren hingegen aus X_j *und* Y auspartialisiert.

Neben dem im Abschnitt 2.8.1 definierten Residuum $\delta^{(j)}$ (X_j bereinigt um die restlichen Regressoren) ist bei der partiellen Kriteriumskorrelation auch das um die anderen Regressoren bereinigte Kriterium beteiligt. Wenn man mit $\mathbf{X}^{(j)}$ die Designmatrix des Modells ohne die Spalte zu X_j bezeichnet und mit $\boldsymbol{\beta}^{(j)}$ den Parametervektor aus der Regression von Y auf den um X_j reduzierten Prädiktorensatz, dann stellt folgendes Residuum das bereinigte Kriterium dar:

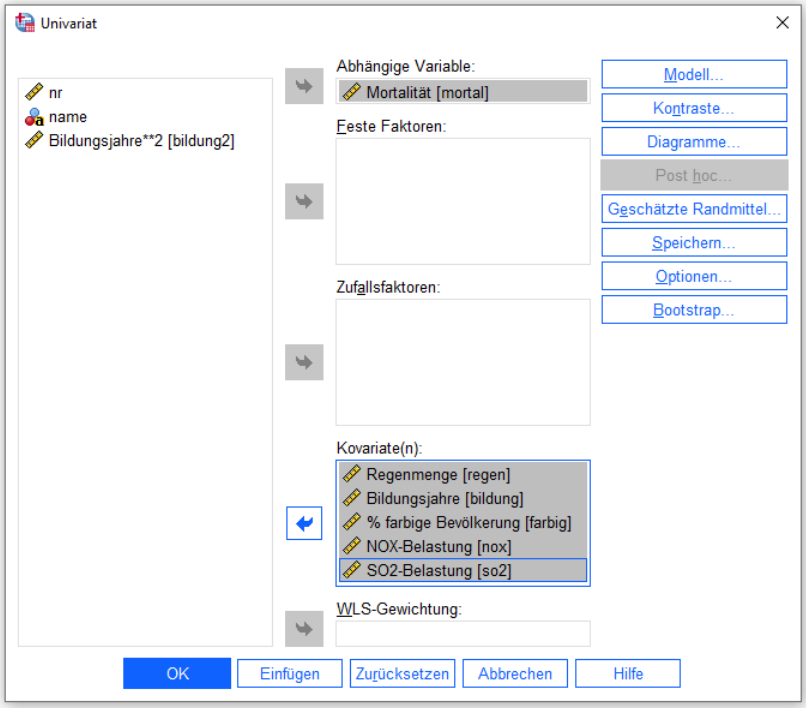
$$\varepsilon^{(j)} := Y - \mathbf{X}^{(j)} \cdot \boldsymbol{\beta}^{(j)}$$

Die partielle Kriteriumskorrelation pr_j zu X_j ist gerade die Korrelation von $\delta^{(j)}$ und $\varepsilon^{(j)}$. Durch ihr Quadrat erfährt man den von X_j aufgeklärten Anteil an demjenigen Teil der Kriteriumsvarianz, den die anderen Regressoren „übrig lassen“.

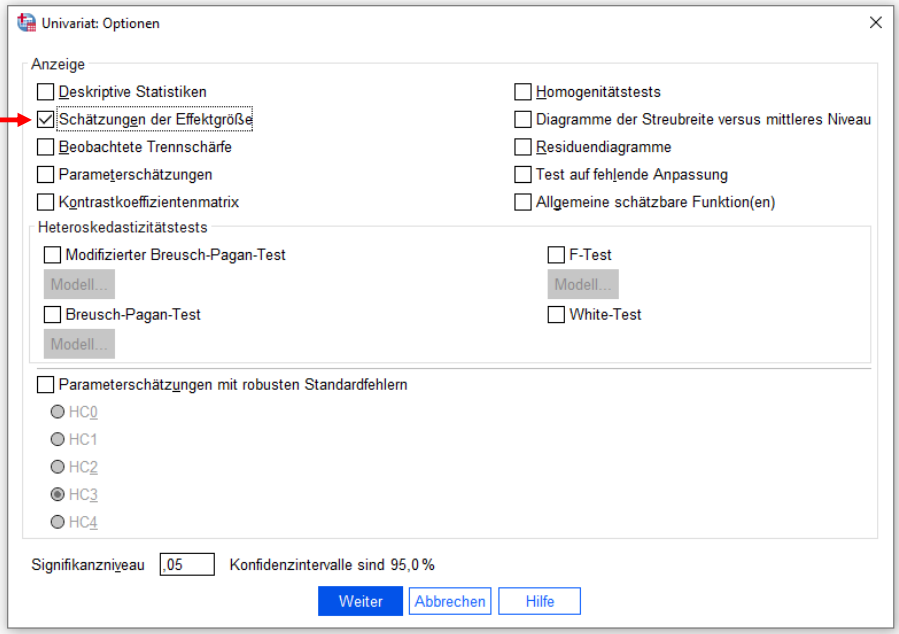
Man muss nicht unbedingt die von der SPSS-Prozedur zur linearen Regression auf Wunsch gelieferten partiellen Kriteriumskorrelationen manuell quadrieren. Alternativ kann man nach dem folgenden Menübefehl

Analysieren > Allgemeines lineares Modell > Univariat

bei der SPSS-Prozedur zum allgemeinen linearen Modell eine Regressionsanalyse anfordern



und per **Optionen**-Dialog den Ausgabeumfang um die **Schätzungen der Effektgröße** erweitern:



Dann erhält man die quadrierten partiellen Kriteriumskorrelationen in der mit **Partielles Eta-Quadrat** beschrifteten Spalte der Tabelle mit den **Tests der Zwischensubjekteffekte**. Im Mortalitätsbeispiel erzielt z. B. der Regressor FARBIG ein partielles Eta-Quadrat von 0,287:

Tests der Zwischensubjekteffekte

Abhängige Variable: Mortalität

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	152003,893 [a]	5	30400,779	21,502	,000	,666
Konstanter Term	163037,397	1	163037,397	115,313	,000	,681
regen	8248,233	1	8248,233	5,834	,019	,098
bildung	6983,968	1	6983,968	4,940	,030	,084
farbig	30674,409	1	30674,409	21,695	,000	,287
nox	42,873	1	42,873	,030	,862	,001
so2	9338,562	1	9338,562	6,605	,013	,109
Fehler	76348,876	54	1413,868			
Gesamt	53287413,510	60				
Korrigierte Gesamtvariation	228352,770	59				

a. R-Quadrat = ,666 (korrigiertes R-Quadrat = ,635)

Dieser Wert ist identisch mit der quadrierten partiellen Korrelation (siehe Tabelle im Abschnitt 2.8.1)

$$0{,}5354^2 \approx 0{,}287$$

und deutlich größer als die quadrierte semipartielle Korrelation (0,135).

Bei seinen Empfehlungswerten zur Beurteilung von Effektstärken in der multiplen Regression differenziert Cohen (1988, S. 413) *nicht* zwischen dem Determinationskoeffizienten für ein komplettes Modell und der quadrierten partiellen Korrelation für einen einzelnen Regressor, sodass auch zur Beurteilung eines einzelnen Regressors die Effektstärkentabelle aus Abschnitt 2.6 zu verwenden ist.

2.9 Überprüfung der Modellannahmen

Nun wollen wir für das Modell zur Mortalität in amerikanischen Städten die Linearitäts-, die Homoskedastizitäts- und die Normalitätsannahme prüfen.

2.9.1 Linearität

Bei der bivariaten Regression haben wir u.a. den Plot der ausgelassen-studentisierten Residuen gegen die prognostizierten Werte verwendet, um Verletzungen der Linearitätsannahme zu diagnostizieren. Im Modell der multiplen Regression ist dieses Instrument zu grob, weil wir für jeden einzelnen Regressor wissen möchten, ob sein eigenständiger Beitrag im linearen Sinn erfolgt. Den isolierten Effekt eines Regressors X_j kann man theoretisch dadurch analysieren, dass man alle anderen Regressoren konstant hält und dann das Kriterium für verschiedene X_j - Werte beobachtet. Allerdings liegen für jede konkrete Wertekombination der anderen Regressoren in der Regel nur wenige X_j - Beobachtungen vor, sodass die Konstanthaltungstechnik in der Regel nicht anwendbar ist. Da man die restlichen Regressoren nicht konstant halten kann, kontrolliert man sie auf statistischem Weg, indem man vom Kriterium und von X_j jeweils die lineare Prognose aufgrund der restlichen Regressoren subtrahiert und anschließend eine bivariate Regression des residualisierten Kriteriums auf den residualisierten Regressor durchführt.

Die residualisierten Regressoren $\delta^{(j)}$ haben wir schon im Abschnitt 2.8.1 im Zusammenhang mit der semipartiellen Kriteriumskorrelation definiert,

$$\delta^{(j)} := X_j - \mathbf{X}^{(j)} \cdot \tilde{\beta}^{(j)}$$

und die residualisierten Kriteriumsvarianten $\varepsilon^{(j)}$ wurden im Abschnitt 2.8.3 im Zusammenhang mit der partiellen Kriteriumskorrelation eingeführt:

$$\varepsilon^{(j)} := Y - \mathbf{X}^{(j)} \cdot \beta^{(j)}$$

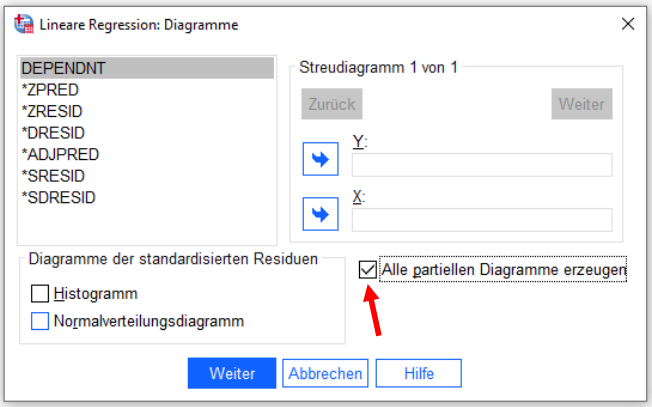
Anhand des Streudiagramms zu den Variablen $\varepsilon^{(j)}$ und $\delta^{(j)}$ lässt sich die residualisierte Kovariation von X_j und Y beurteilen. Ein solches Streudiagramm bezeichnet man als **partielles Regressionsdiagramm**, weil die restlichen Prädiktoren aus Y *und* X_j auspartialisiert wurden.

Es lässt sich zeigen (siehe z. B. Darlington & Hayes 2017, S. 72f; Kockläuner 1988, S. 120), dass der Regressionskoeffizient β_j aus dem eigentlich interessierenden multiplen Regressionsmodell identisch ist ...

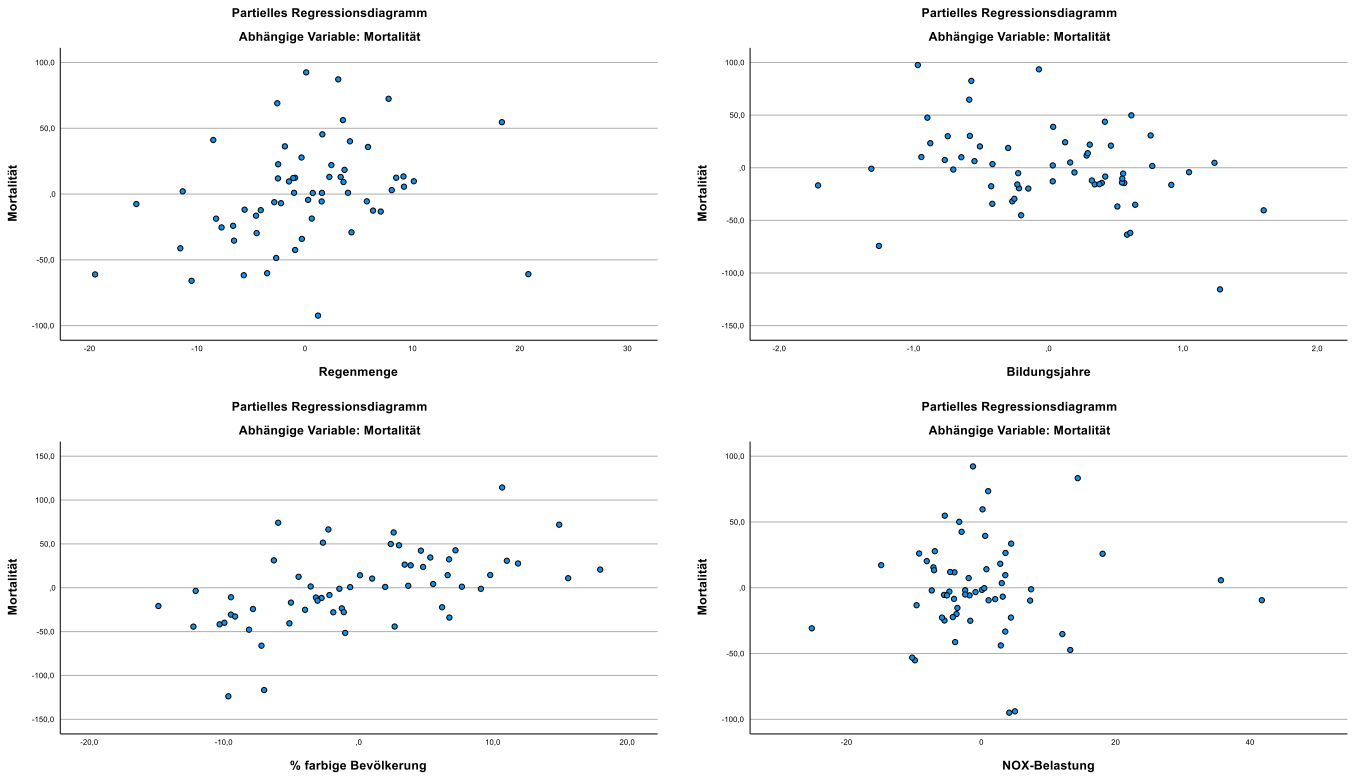
(1) mit dem Steigungskoeffizienten aus der bivariaten Regression von $\varepsilon^{(j)}$ auf $\delta^{(j)}$
(2) mit dem Steigungskoeffizienten aus der bivariaten Regression von Y auf $\delta^{(j)}$

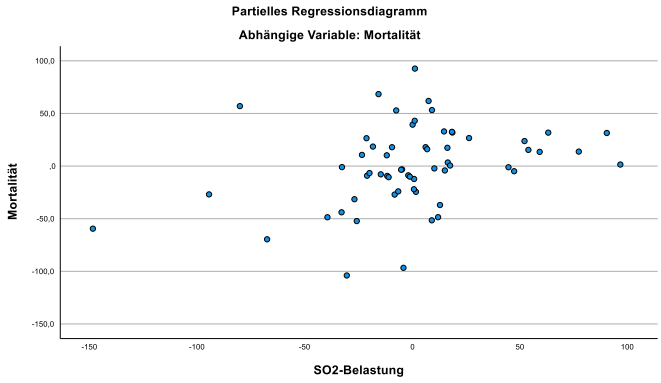
Die Identität (1) liefert die Rechtfertigung dafür, die Koeffizienten β_j aus der multiplen Regression als *partielle Regressionskoeffizienten* zu bezeichnen (vgl. Abschnitt 2.2). Wegen der Identität (2) kann man sie aber auch als *semipartielle Regressionskoeffizienten* bezeichnen.

In SPSS werden die Partialdiagramme in der Regressions-Subdialogbox **Diagramme** angefordert:



Für das Mortalitätsbeispiel erhalten wir:





Das partielle Regressionsdiagramm zur Anzahl der Bildungsjahre zeigt am linken Rand eine relativ schlechte Anpassung des linearen Modells, wobei aber auch ein Ausreißerproblem vorliegen könnte (vgl. Abschnitt 3.1). Bei der folgenden Darstellung wurde der Punktauswahlmodus im SPSS-Diagrammeditor (zu aktivieren mit dem Symbol) dazu benutzt, um zu den kritischen Fällen ihre laufende Nummer in der Arbeitsdatei anzeigen zu lassen:

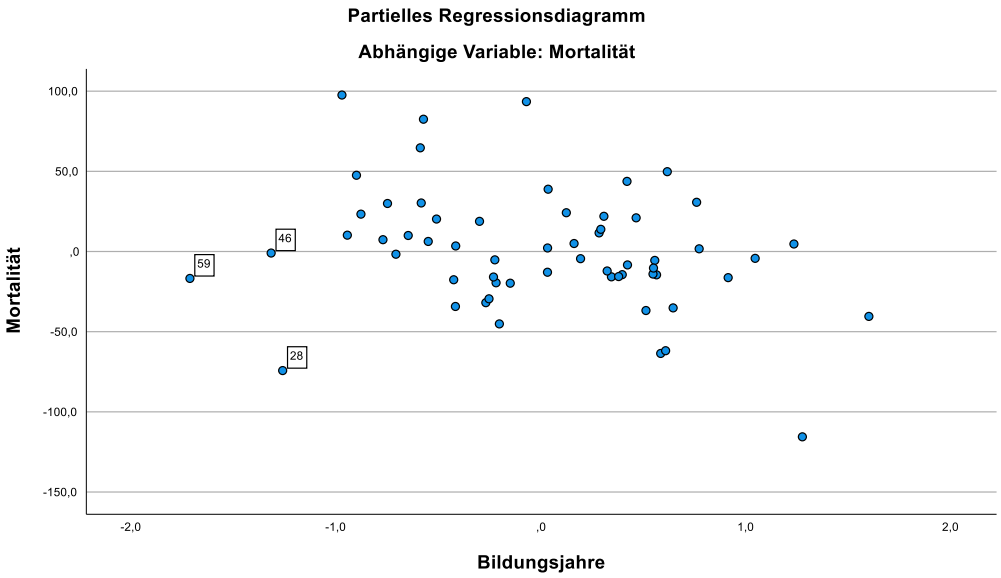


Abbildung 32: Partielles Regressionsdiagramm für BILDUNG im Mortalitätsbeispiel mit Markierung für die schlecht erklärbaren Fälle

Bei Aufnahme des quadratischen Terms BILDUNG^2 im Sinne einer polynomischen Regressionsgleichung (vgl. Abschnitt 4) verfehlt der zweiseitige t-Test knapp die Signifikanzgrenze:[1]

[1] Hier erscheint es trotz der Argumentation im Abschnitt 2.5.2 weniger sinnvoll, den einseitigen Test (mit der größeren Po wer) zu verwenden. Zwar hat der geschätzte Regressionskoeffizient das erwartete negative Vorzeichen (für eine nach unten offene Parabel), doch resultierte diese „Erwartung“ aus einer Inspektion der Daten.

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.
1	(Konstante)	-476,984	764,437		-,624	,535
	Regenmenge	1,439	,679	,231	2,119	,039
	Bildungsjahre	261,167	142,664	3,549	1,831	,073
	Bildungsjahre**2	-12,846	6,606	-3,802	-1,945	,057
	% farbige Bevölkerung	2,956	,635	,424	4,658	,000
	NOX-Belastung	,267	,484	,072	,552	,583
	SO2-Belastung	,266	,121	,271	2,197	,032

a. Abhängige Variable: Mortalität

Bei der endgültigen Entscheidung über die Erweiterung des Modells um einen quadratischen Term für die Bildung sind die Ergebnisse anderer Tests (z. B. auf Ausreißer oder einflussreiche Fälle, siehe Abschnitt 3.1.3) zu berücksichtigen.

Nicht zuletzt sind **theoretische Überlegungen** relevant. Weil nur wenige Argumente dafür sprechen, dass sich eine besonders niedrige Bildung *positiv* auf die Lebenserwartung auswirkt, sollte man aus theoretischer Sicht auf den quadratischen Term eher verzichten.

Zur Verbesserung der Modellanpassung kommen neben einer Änderung der funktionalen Form für die Wirkung einzelner Regressoren noch andere Maßnahmen in Frage, z. B.

- die Aufnahme von weiteren Regressoren (z. B. Einwohnerzahl)
  Möglicherweise haben kleinere Städte eine eher ländliche Lage mit gesünderen Lebensverhältnissen und gleichzeitig eine kürzere durchschnittliche Ausbildungszeit.
- Aufnahme von Moderatoreffekten (siehe z. B. Baltes-Götz 2020a)

Das partielle Regressionsdiagramm zur Regenmenge zeigt einen schlecht erklärbaren Fall, mit dem wir uns im Abschnitt 3 noch beschäftigen werden. Ein Anlass zum Zweifel an der Linearität besteht hier nicht:

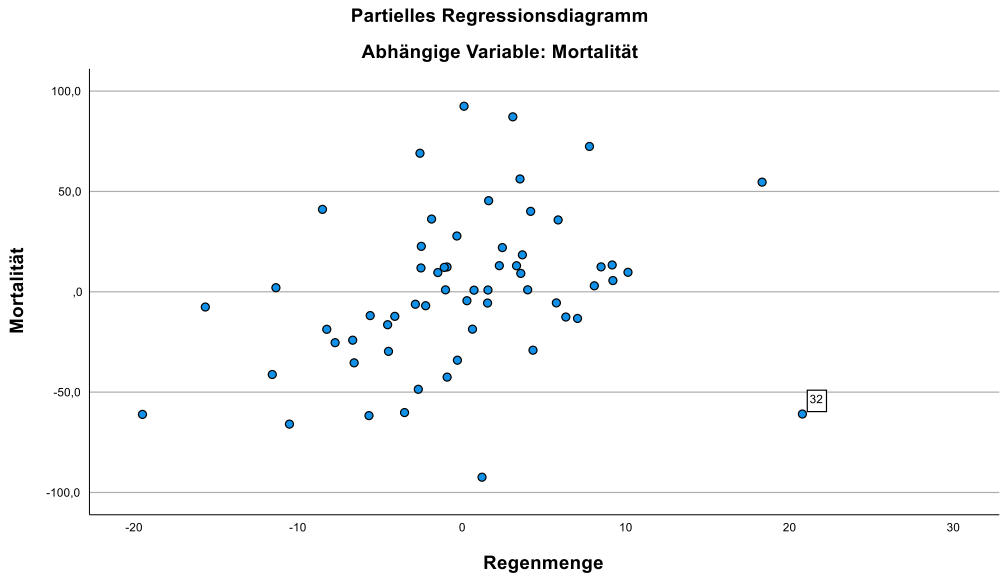
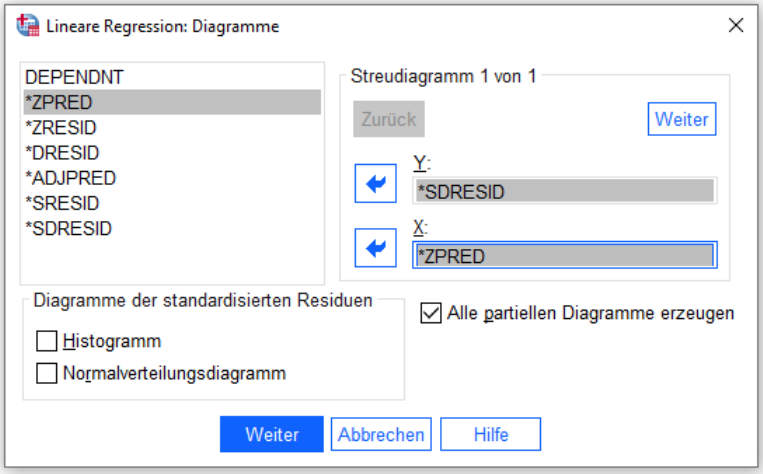


Abbildung 33: Partielles Regressionsdiagramm für REGENMENGE im Mortalitätsbeispiel mit Markierung für einen schlecht erklärbaren Fall

2.9.2 Homoskedastizität

Zur grafischen Beurteilung der Homoskedastizität fordern wir einen Plot der ausgelassen-studentisierten Residuen gegen die standardisierten prognostizierten Werte an:



Im Mortalitätsbeispiel ergeben sich dabei keine offensichtlichen Hinweise zu Ungunsten der Homoskedastizitätsannahme:

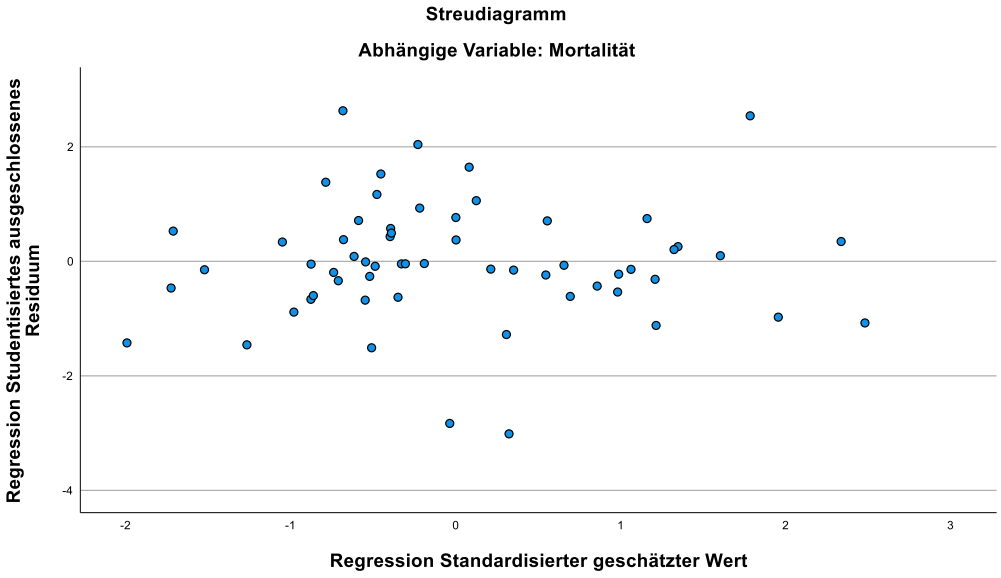


Abbildung 34: Plot der studentisierten Residuen gegen die standardisierten Prognosewerte im Mortalitätsbeispiel

Der Spread & Level - Plot (vgl. Abschnitt 1.7.3.2) zeigt praktisch *keine* Änderung der Variabilität in Abhängigkeit von der Modellprognose:

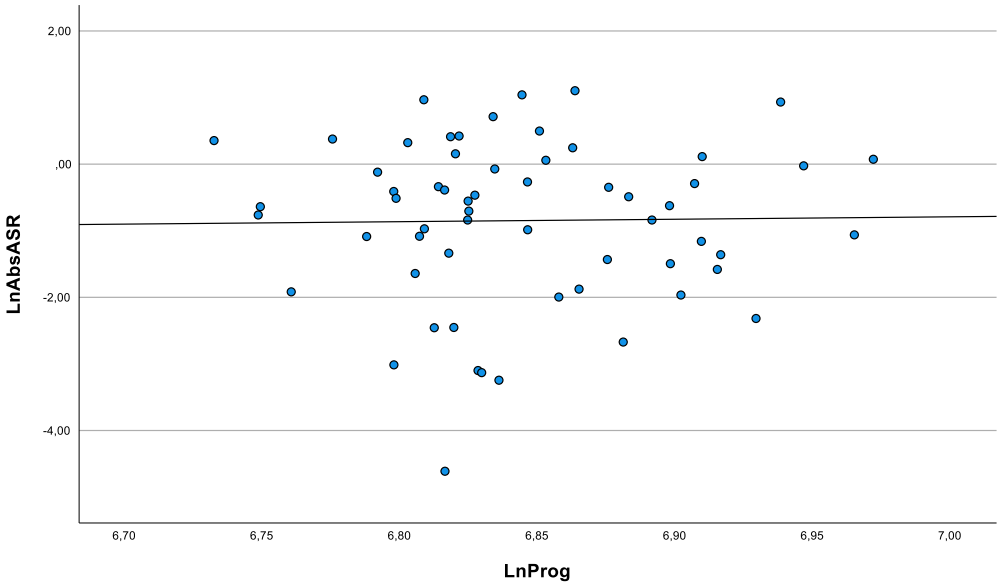


Abbildung 35: Plot der logarithmierten Beträge der ausgelassen-studentisierten Residuen gegen die logarithmierten Modellprognosen im Mortalitätsbeispiel

Auch der Breusch-Pagan - Test (vgl. Abschnitt 1.7.3.2.4) äußert keine Kritik an der Varianzhomogenität:

Modifizierter Breusch-Pagan-Test auf Heteroskedastizität[a,b,c]

Chi-Quadrat	df	Sig.
,001	1	,980

a. Abhängige Variable: Mortalität

b. Testet die Nullhypothese, dass die Varianz der Fehler nicht von den Werte der unabhängigen Variablen abhängt.

c. Vorhergesagte Werte aus Design: Konstanter Term + regen + bildung + farbig + nox + so2

In der dem Breusch-Pagan - Test zugrunde liegenden Regression der quadrierten einfachen Residuen auf die Prädiktoren des Modells findet sich kein signifikantes Regressionsgewicht:

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.
1	(Konstante)	3473,491	5321,413		,653	,517
	Regenmenge	50,964	39,269	,232	1,298	,200
	Bildungsjahre	-316,368	409,545	-,122	-,772	,443
	% farbige Bevölkerung	-27,956	37,160	-,114	-,752	,455
	NOX-Belastung	17,642	27,830	,135	,634	,529
	SO2-Belastung	-10,860	6,953	-,314	-1,562	,124

a. Abhängige Variable: Quadrierte Residuen

Trotz der im Abschnitt 2.9.1 beobachteten Ausreißer-Verdachtsfälle besteht eine klare Befundlage zugunsten der Homoskedastizität.

Ist die Homoskedastizität wesentlich verletzt, kommen die folgenden im Zusammenhang mit der bivariaten Regression beschriebenen Maßnahmen in Frage:

- Transformation des Kriteriums (siehe Abschnitt 1.7.3.3)
- Verwendung einer heteroskedastizitäts-robusten Inferenzstatistik (siehe Abschnitt 1.7.3.4)
- WLS-Regression (siehe Abschnitt 1.7.3.5)
- Auch die Erweiterung des Modells (z. B. um eine Wechselwirkung) kann die richtige Antwort auf eine ausgeprägte Heteroskedastizität sein.

2.9.3 Normalverteilung

Auch zur Beurteilung der Normalität sind die ausgelassen-studentisierten Residuen gut geeignet. Im Abschnitt 1.7.4.3 wird beschrieben, wie man von der Regressionsprozedur per Syntax das Histogramm und den Wahrscheinlichkeitsplot für die ausgelassen-studentisierten Residuen anfordert. Im Mortalitäts-Beispiel zeigt sich eine symmetrische, aber ausgeprägt spitzgipflige Verteilung:

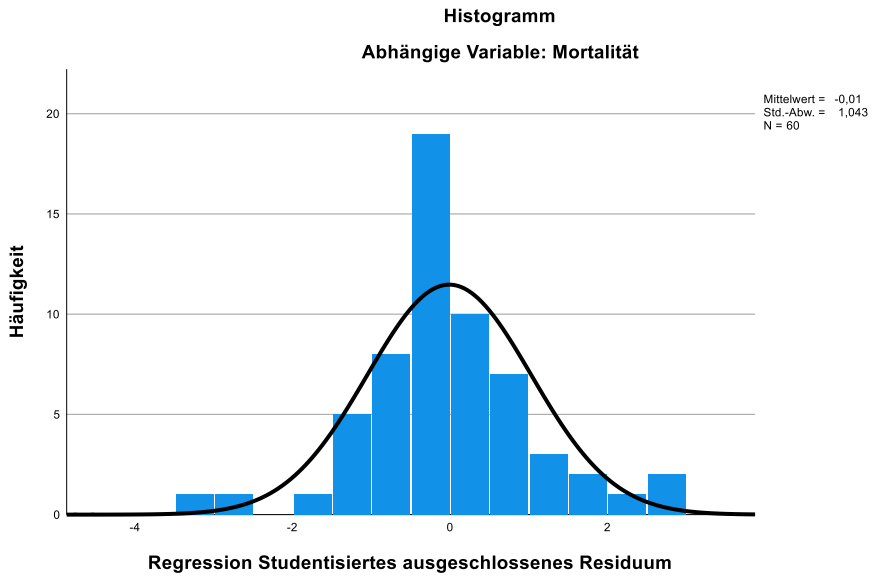


Abbildung 36: Histogramm zu den ausgelassen-studentisierten Residuen im Mortalitätsbeispiel

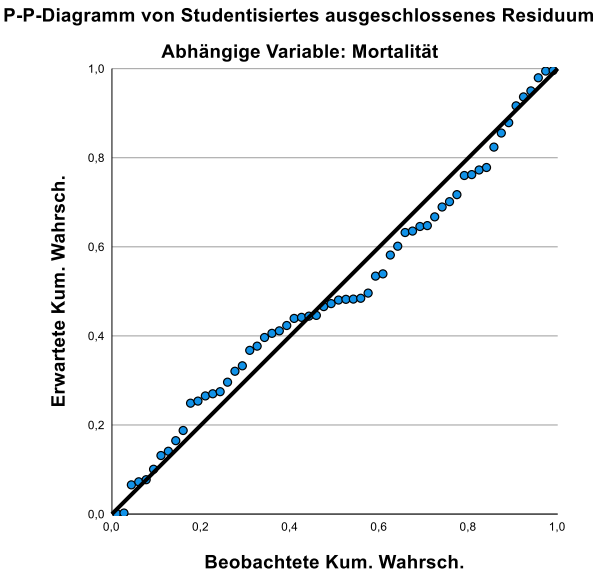


Abbildung 37: Wahrscheinlichkeitsplot zum Vergleich der empirischen Verteilung der ausgelassen-studentisierten Residuen mit der Standardnormalverteilung im Mortalitätsbeispiel

Nach dem Abspeichern der ausgelassen-studentisierten Residuen als neue Variable (siehe Abschnitt 1.7.2.3) kann ihre Verteilung über die mit

Analysieren > Deskriptive Statistik > Explorative Datenanalyse

zu startende Prozedur zur explorativen Datenanalyse auf Normalität getestet werden (vgl. Abschnitt 1.7.4.5). Im Beispiel wird die Normalverteilungs-Nullhypothese von den beiden durchgeführten Tests (Kolmogorov-Smirnov und Shapiro-Wilk) beibehalten:

Tests auf Normalverteilung

	Kolmogorov-Smirnov [a]			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Studentized Deleted Residual	,095	60	,200 *	,963	60	,068

*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

Allerdings hat die Wölbungsstatistik (Kurtosis) einen Betrag größer 1 (vgl. Abschnitt 1.7.4.5):

Deskriptive Statistik

			Statistik	Standard Fehler
Studentized Deleted Residual	Mittelwert		-,0119096	,13466189
	95% Konfidenzintervall des Mittelwerts	Untergrenze	-,2813674	
		Obergrenze	,2575482	
	5% getrimmtes Mittel		-,0105974	
	Median		-,0591135	
	Varianz		1,088	
	Standard Abweichung		1,04308655	
	Minimum		-3,01243	
	Maximum		2,62926	
	Spannweite		5,64169	
	Interquartilbereich		1,10233	
	Schiefe		-,087	,309
	Kurtosis		1,545	,608

Im Vergleich zur Normalitätsprüfung für die Residuen aus der Regression von Gewicht auf Größe mit dem Stichprobenumfang $N = 159$ (vgl. Abschnitt 1.7.4.5) kommt uns hier die relativ kleine Stichprobe „zu Gute“. Bei $N = 60$ haben die Normalitätstests eine relativ geringe Power und zeigen sich daher „großzügig“. Bei einer größeren Stichprobe werden die (praktisch immer vorhandenen, aber oft unkritischen) Abweichungen von der Normalverteilung eher aufgedeckt.

Wenn die Annahme normalverteilter Residuen wesentlich verletzt ist, und ein Vertrauen auf den zentralen Grenzwertsatz der Statistik (siehe Abschnitt 1.7.4) nicht gerechtfertigt erscheint, dann kommen die folgenden Maßnahmen in Frage:

- Elimination von Ausreißern (siehe Abschnitte 1.7.4.5 und 3.1.1)
- Transformation des Kriteriums (siehe Abschnitt 1.7.4.6.2)
- Verwendung der Bootstrap-Technik (siehe Abschnitt 1.7.3.4.2 sowie Fox 2008, Kap. 21)
- Verwendung eines generalisierten linearen Modells
  Bei einer Zählvariablen als Kriterium (z. B. Anzahl der im letzten Jahr gelesenen Bücher) ist oft ein generalisiertes lineares Modell mit der Annahme einer Poisson-Verteilung oder einer negativen Binomialverteilung für die Residuen besser geeignet als ein lineares Modell, das von normalverteilten Residuen ausgeht (siehe Baltes-Götz 2016a). Eventuell muss man auf ein parametrisches Verteilungsmodell für die Residuen verzichten und die ordinale logistische Regression verwenden (siehe Baltes-Götz 2012).

2.10 Power-Analyse

Eine Power-Analyse kann *vor* der Datenerhebung oder *nach* der Hypothesenprüfung durchgeführt werden:

- Bei der **A-priori - Power-Analyse** wird die erforderliche Stichprobengröße ermittelt, die zur Aufdeckung eines Effekts mit bestimmter Stärke eine gewünschte Wahrscheinlichkeit garantiert.
- Bei der **Post-hoc - Power-Analyse** wird z. B. für verschiedene hypothetische Effektstärken ermittelt, mit welcher Entdeckungswahrscheinlichkeit sie in der durchgeführten Studie ausgestattet waren. Bestand für eine Effektstärke eine angemessene Entdeckungswahrscheinlichkeit (z. B. ≥ 0,95), die aber „ungenutzt“ blieb, dann spricht das Untersuchungsergebnis *gegen* die Existenz eines solchen oder eines noch stärkeren Effekts.

Anschließend werden beide Varianten unter Verwendung der schon im Abschnitt 1.8 vorgestellten Software **GPower 3.1** (Faul et al. 2009, GPower 2017) beschrieben. Auf die Wiederholung der Analysen mit der schlechter ausgestatteten Power-Analyse - Prozedur von SPSS verzichten wir diesmal.

2.10.1 Modelle mit fixierten Regressoren

2.10.1.1 A priori - Poweranalyse

2.10.1.1.1 Globaler F-Test

Die Effektstärke ist analog zum bivariaten Fall (siehe Abschnitt 1.8) definiert durch:

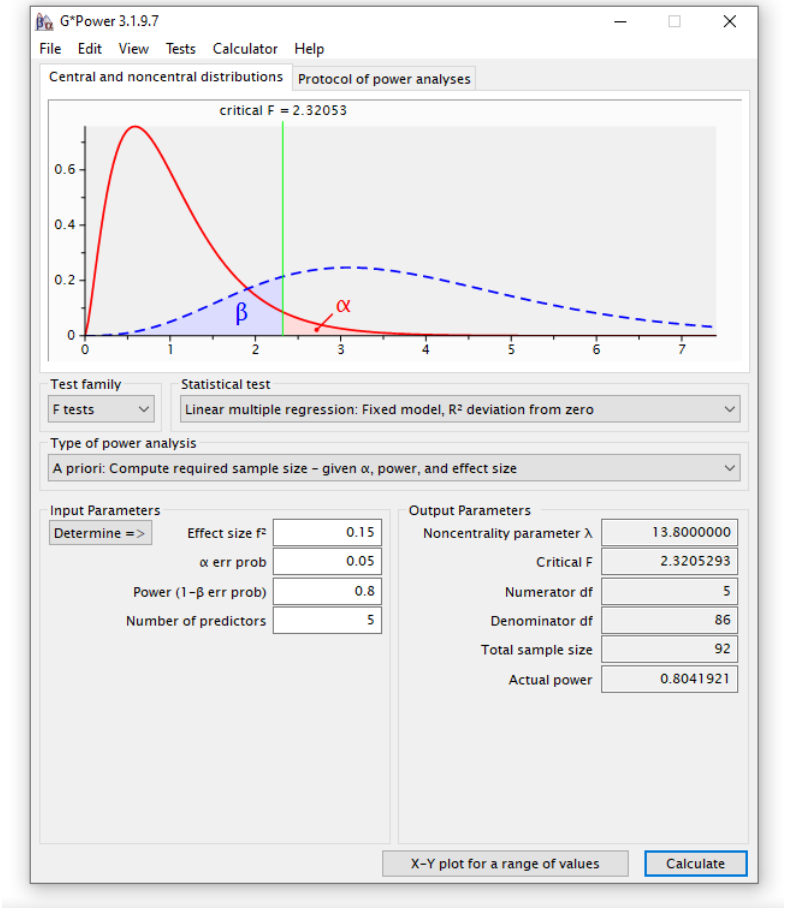
$$f^2 = \frac{\rho^2}{1-\rho^2}$$

Dabei steht ρ^2 für den Determinationskoeffizienten in der Population. Da es momentan um kein konkretes Modell geht, nehmen wir für die Effektstärke einen mittleren Wert von $f^2 = 0{,}15$ an (vgl. Tabelle im Abschnitt 2.6).

Wir wählen in **GPower 3.1** den folgenden Aufgabentyp:

Test family: **F-Tests**
Statistical test: **Linear mult. Regr.: Fixed model, R^2 dev. from zero**
Type of power analysis: **A priori**

Um bei der Prüfung eines Modells mit 5 Regressoren in einem Test zum Niveau $\alpha = 0{,}05$ eine Teststärke von 0,8 zu erreichen, werden 92 Fälle benötigt:



2.10.1.1.2 T-Test zu einem einzelnen Regressor

Bei der Prüfung eines Modells mit 5 Regressoren soll für einen *einzelnen* Regressor in einem zwei- oder einseitigen t-Test zum α-Niveau 0,05 für eine Teststärke von 0,8 der erforderliche Stichprobenumfang ermittelt werden. In Anlehnung an Cohen (1988, S. 410ff) bezeichnen wir mit

J	eine Menge, die genau den Regressor X_j enthält
A	eine Menge, die alle anderen Regressoren enthält
$\rho^2_{YJ\cdot A}$	die quadrierte partielle Korrelation von X_j mit Y, wobei aus X_j *und* Y die linearen Prognosen der anderen Regressoren entfernt worden sind (vgl. Abschnitt 2.8.3)

Die Effektstärke f_j^2 für einen einzelnen Regressors X_j im Rahmen eines multiplen Regressionsmodells lässt sich nach Cohen (1988, S. 412) durch die quadrierte partielle Korrelation zwischen X_j und Y bei Kontrolle der restlichen Regressoren ausdrücken:

$$f_j^2 = \frac{\rho^2_{YJ\cdot A}}{1-\rho^2_{YJ\cdot A}}$$

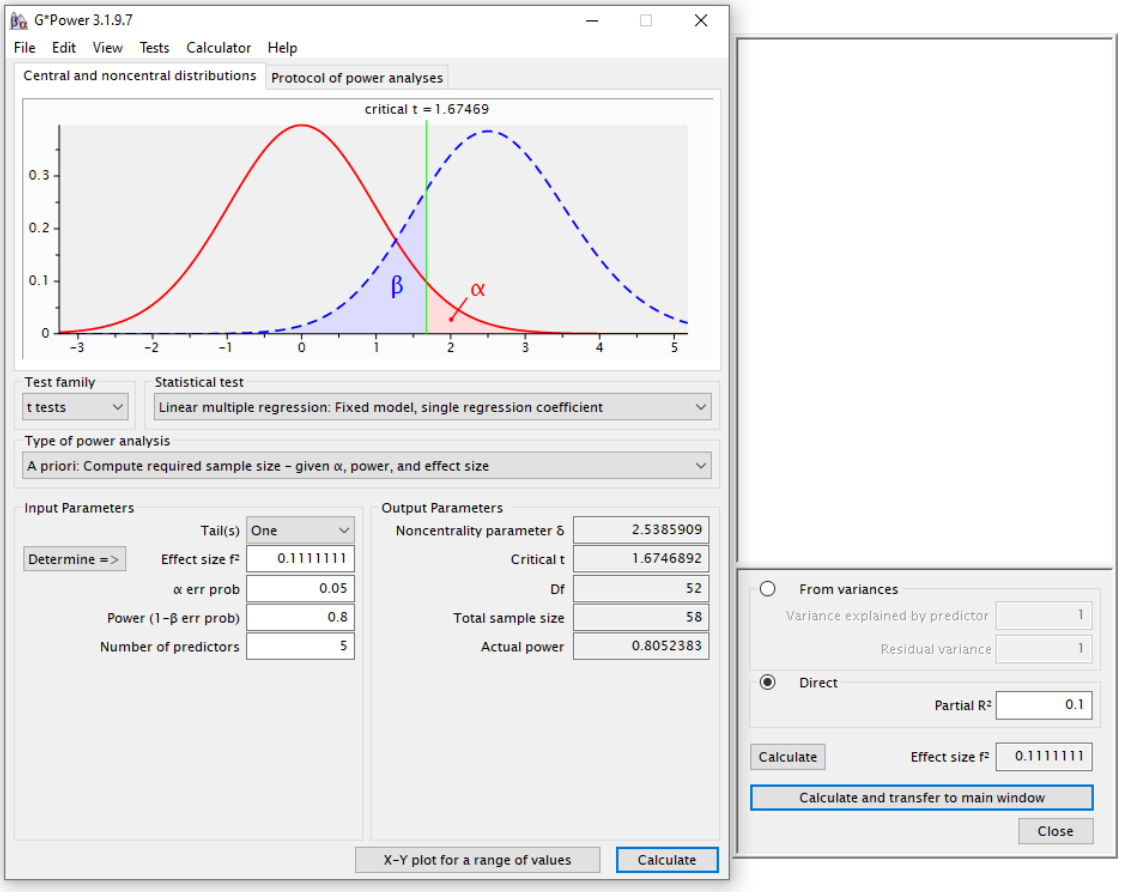
Wir nehmen für den (beliebigen) Regressor X_j eine quadrierte partielle Korrelation $\rho^2_{YJ\cdot A}$ von 0,1 an, wobei die Effektstärke $f_j^2 = 0{,}\overline{1}$ resultiert.

Bei seinen Empfehlungswerten zur Beurteilung von Effektstärken in der multiplen Regression (siehe Tabelle im Abschnitt 2.6) differenziert Cohen (1988, S. 413) *nicht* zwischen dem ρ^2 für ein komplettes Modell (vgl. Abschnitt 2.10.1.1.1) und der quadrierten *partiellen* Korrelation $\rho^2_{YJ\cdot A}$ für eine Teilmenge J von Regressoren. Wir betrachten gerade den Spezialfall mit einer einelementigen Menge J.

Wählen Sie in GPower 3.1 den folgenden Aufgabentyp:

Test family: **t-Tests**
Statistical test: **Linear mult. Regr.: Fixed model, single regr. coefficient**
Type of power analysis: **A priori**

Bei einem *zweiseitigen* Test zum Niveau α = 0,05 werden 73 Fälle benötigt, um eine Teststärke von 0,8 zu erreichen. Bei einem *einseitigen* Test sind dazu lediglich 58 Fälle erforderlich:



2.10.1.1.3 Mehrere t-Tests zu einzelnen Regressoren

Wenn ein Modell 5 Regressoren mit derselben, von null verschiedenen Effektstärke enthält, dann kommt es bei den erforderlichen 5 Tests zu einer **β-Fehler** - Kumulierung. Unter der vereinfachenden (sehr unrealistischen) Annahme unabhängiger Regressoren resultiert bei einer Power von 0,8 pro Einzeltest als Wahrscheinlichkeit für 5 korrekte Testausgänge:

$$0{,}8^5 \approx 0{,}33$$

Um die Wahrscheinlichkeit für 5 korrekte Testausgänge auf 0,8 zu erhöhen, muss die Power p_s für einen Einzeltest auf ca. 0,9564 verbessert werden:

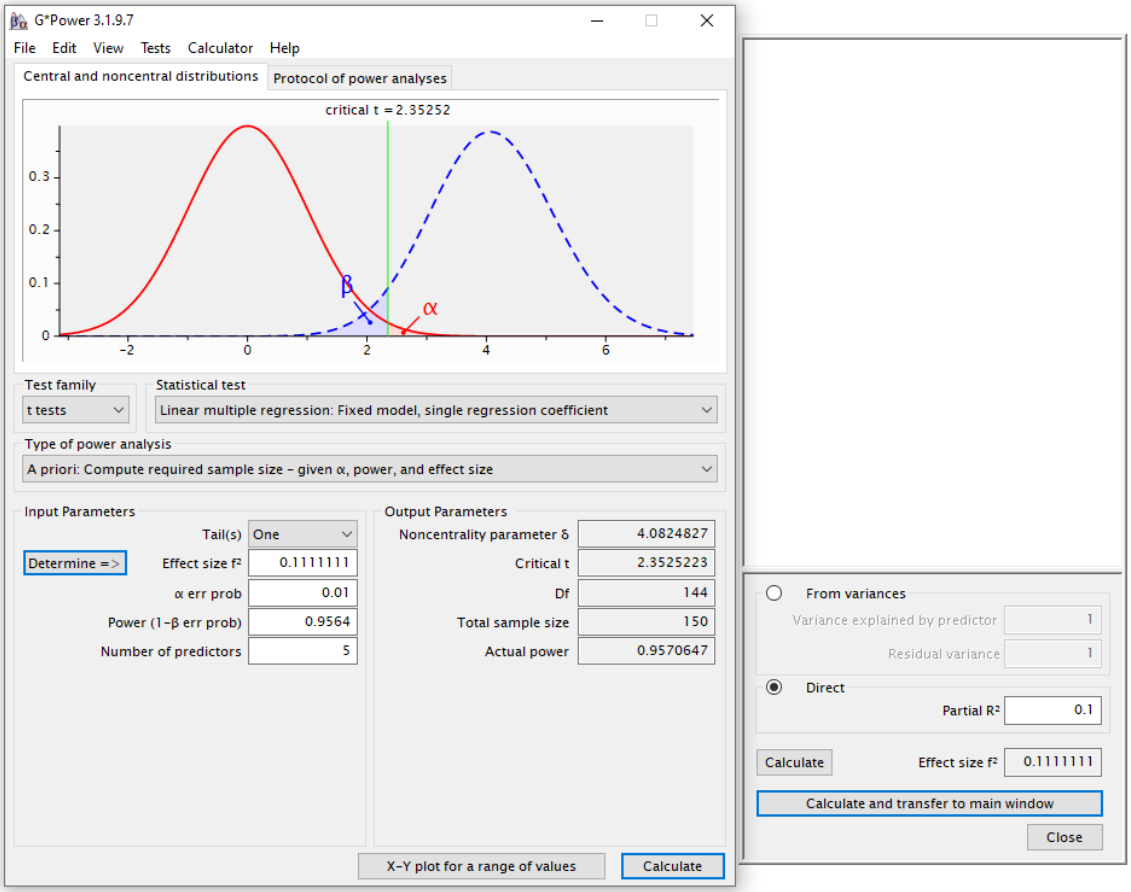
$$p_s^{\,5} = 0{,}8 \Leftrightarrow p_s = 0{,}8^{0{,}2} \approx 0{,}9564$$

Eine A-priori - GPower-Analyse ergibt, dass dazu in der folgenden Testkonstellation

- einseitige Tests
- Effektstärke $f_j^2 = 0{,}\overline{1}$
- α = 0,05 für einen Einzeltest

103 Fälle erforderlich sind.

Wenn alle Tests zum Niveau $\alpha = 0{,}05$ durchgeführt werden, kommt es zu einer Kumulation des α-Fehlerrisikos (siehe Abschnitt 2.5.4.2 zum multiplen Testen). Dies lässt sich vermeiden, indem das Risiko eines Einzeltests nach dem Bonferroni-Prinzip auf $\frac{\alpha}{5} = 0{,}01$ gesenkt wird. Zur gleichzeitigen Vermeidung einer β-Fehler - Kumulierung sind dann bei einseitigen Tests 150 Fälle erforderlich:



Diese Zahl ist in vielen Studien (z.B. mit Hilfe der Online-Forschung, siehe Baltes-Götz 2020c) problemlos realisierbar.

Die im Beispiel für jeden Regressor angenommene Effektstärke (quadrierte partielle Korrelation $\rho^2_{YJ \cdot A}$ von 0,1) ist allerdings sehr groß. In der aktuell analysierten Stichprobe mit Mortalitätsdaten aus amerikanischen Städten erreichen immerhin vier von fünf Regressoren annähernd diese Effektstärke (siehe Tabelle mit den partiellen Eta-Quadraten im Abschnitt 2.8.3). Nimmt man für die Regressoren jeweils eine *kleine* Effektstärke an, dann steigt der Stichprobenbedarf in der beschriebenen Konstellation auf 818 Fälle.

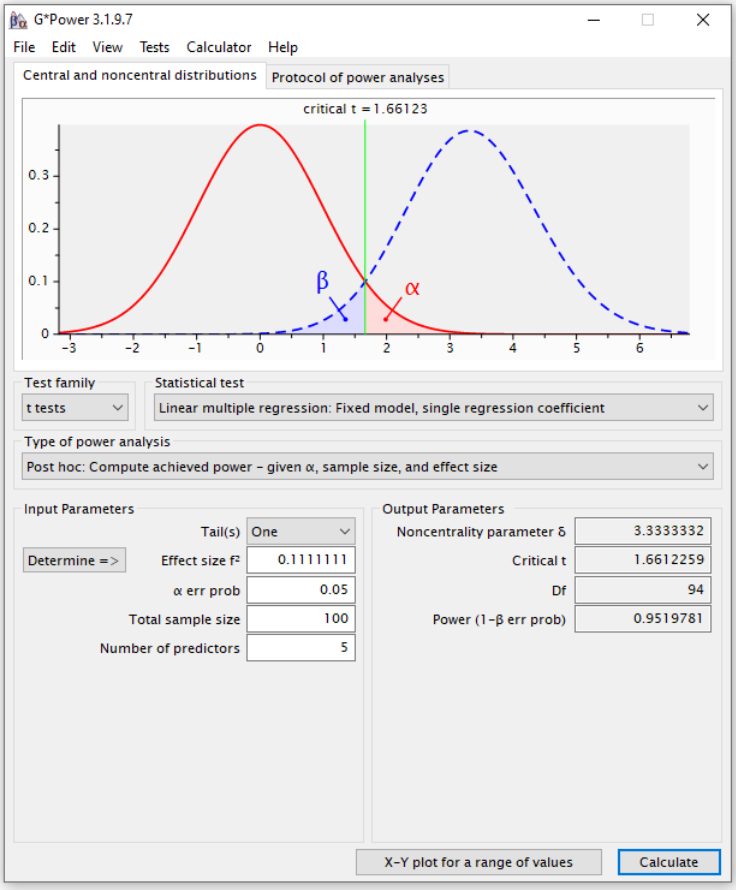
Um die Power der multiplen Tests im Vergleich zum simplen Bonferroni-Verfahren zu verbessern, sollte das im Abschnitt 2.5.4.2 beschriebene Bonferroni-Holm - Verfahren eingesetzt werden.

2.10.1.2 Post hoc - Poweranalyse

In einer Post-Hoc - Poweranalyse soll beispielhaft ermittelt werden, welche Power beim t-Test zu einem *einzelnen* Regressor vorliegt, wenn in einer Studie mit 5 Regressoren eine Stichprobengröße von $N = 100$ realisiert worden ist. Wählen Sie in GPower 3.1 den folgenden Aufgabentyp:

Test family: **t-Tests**
Statistical test: **Linear mult. Regr.: Fixed model, single regr. coefficient**
Type of power analysis: **Post hoc: Compute achieved power**

Bei einer Effektstärke von $f_j^2 = 0{,}\overline{1}$, $\alpha = 0{,}05$ und $N = 100$ resultiert für den einseitigen Test eine Power von ca. 0,95:



Ergibt sich in einer konkreten Studie *kein* signifikantes Ergebnis, dann kann man davon ausgehen, dass kein Effekt mit der angenommenen oder einer größeren Stärke vorliegt.

2.10.2 Modelle mit stochastischen Regressoren

Werden die Algorithmen des Regressionsmodells für fixierte Regressoren auf Daten mit stochastischen Regressoren angewendet, dann bleiben die Schätzungen erwartungstreu und die Signifikanztests korrekt (vgl. Abschnitt 1.2.3). Allerdings führt in dieser Situation die A-priori - Power-Analyse gemäß Abschnitt 2.10.1 zu einem leicht unterschätzten Stichprobenbedarf. Um für den globalen F-Test eine verbesserte Schätzung des Stichprobenbedarfs zu erhalten, wählen wir in GPower den folgenden Aufgabentyp:

- **Test family**: **Exact**
- **Statistical test**: **Linear Multiple Regression: Random model**
- **Type of power analysis**: **A priori**

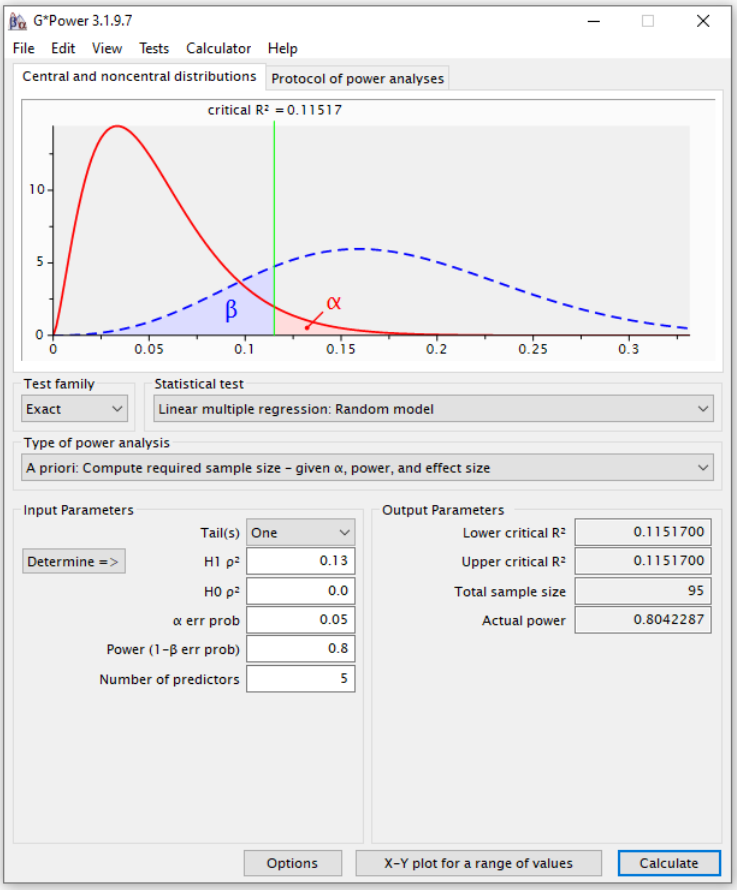
Zur Spezifikation der Effektstärke sind die Determinationskoeffizienten auf Populationsebene unter der Alternativ- und der Nullhypothese anzugeben. Wir verwenden folgende Werte:

- H_1: 0,13 (mittlere Effektstärke sensu Cohen et al. (2003, S. 93), vgl. Tabelle im Abschnitt 2.6)
- H_0: 0

Wir betrachten das folgende einseitige Testproblem zum Determinationskoeffizienten ρ^2:

$$H_0: \rho^2 = 0 \text{ vs. } H_1: \rho^2 > 0$$

Bei 5 Regressoren und einer gewünschten Power von 0,8 resultiert eine Stichprobengröße von 95, die den Wert 92 für das Modell mit fixierten Regressoren (vgl. Abschnitt 2.10.1) leicht übertrifft:



3 Gefahren für eine erfolgreiche Modellierung

In diesem Abschnitt behandeln wir zwei Gefahren für eine erfolgreiche Modellierung:

- Von Fällen mit unangemessen starkem Einfluss auf die Ergebnisse kann jedes Regressionsmodell betroffen sein. Bisher haben wir das Thema aus didaktischen Gründen weitgehend ausgespart.
- Das Multikollinearitätsproblem betrifft nur die multiple Regression. Dabei kommt es zu Schwierigkeiten beim Schätzen und Testen, weil die Varianz einzelner Regressoren nahezu perfekt durch die übrigen Regressoren aufgeklärt wird.

3.1 Fälle mit großem Einfluss auf die Schätzung oder Beurteilung des Modells

Je weniger Fälle verfügbar sind, umso größer ist das Risiko, dass einzelne Fälle einen unangemessen großen Einfluss auf die Parameterschätzungen und/oder die Modellbeurteilung besitzen. In dieser Situation resultiert eine Unsicherheit über die Interpretierbarkeit der Ergebnisse und die angemessene Behandlung des Problems (Fälle ausschließen oder beibehalten?). Um das Risiko allzu einflussreicher Fälle gering zu halten, sollten Stichproben mit weniger als 100 Fällen möglichst vermieden werden.

3.1.1 Schlecht prognostizierbare Fälle

Schlecht prognostizierbare Fälle können auch als *Ausreißer bzgl. der Residuen* bezeichnet werden. Ihre Kriteriumswerte können vom betrachteten Regressionsmodell schlecht erklärt bzw. vorhergesagt werden, sodass betragsmäßig große Residuen resultierten. Solche Fälle müssen individuell analysiert werden:

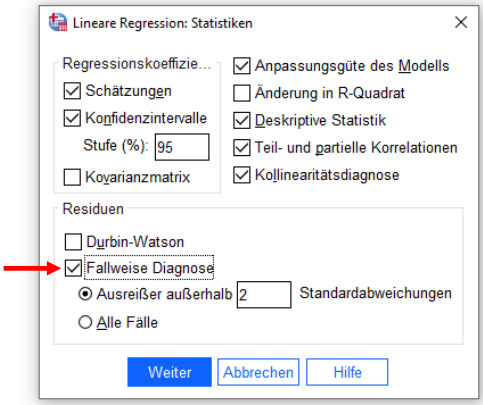
- Stellt sich ein technischer Fehler bei der Datenerhebung oder -erfassung als Ursache heraus, muss der Beobachtungswert korrigiert oder eliminiert werden.
- Findet sich keine technische Panne, ist die Vorgehensweise problematisch, weil natürlich jede Form von „Datenkosmetik“ zu unterlassen ist.
    - Gelegentlich führt die Ausreißeranalyse zu einer verbesserten Modellspezifikation, z. B. durch Aufnahme zusätzlicher Regressoren.
    - Wenn der auffällige Merkmalsträger nicht zu der interessierenden Population gehört, sollte er ausgeschlossen werden.

Bei Gültigkeit eines Modells mit k Regressoren folgen die ausgelassen-studentisierten Residuen (zur Definition siehe Abschnitt 1.7.2.3) einer t-Verteilung mit $N - k - 2$ Freiheitsgraden (Fox & Weisberg 2011, S. 287). Damit kann für ein einzelnes Residuum durch Vergleich mit dem $(1 - \alpha/2)$ - Quantil dieser t-Verteilung in einem zweiseitigen Test geprüft werden, ob es sich modellkonform verhält. Dabei ist zu beachten, dass auch bei perfekter Modellgültigkeit betragsmäßige Residuen jenseits des gerade begründeten kritischen Wertes mit der Wahrscheinlichkeit α zu erwarten sind.

Für unser Mortalitätsbeispiel sollen die Tests zum Niveau $\alpha = 0{,}05$ durchgeführt werden, sodass sich bei $k = 5$ und $N = 60$ der kritische Wert $t_{0{,}975;\,53} \approx 2{,}01$ ergibt, mit dem die Beträge der ausgelassen-studentisierten Residuen zu vergleichen sind. Zur Berechnung des kritischen t-Wertes taugt das folgende SPSS-Kommando COMPUTE, das die inverse Verteilungsfunktion der t-Verteilung mit 53 Freiheitsgraden an der Stelle 0,995 mit Hilfe der Funktion **IDF.T** ermittelt:

```
compute tkrit = idf.t(0.975, 53).
execute.
```

Um per SPSS-Regressionsprozedur z.B. eine Liste der ausgelassen-studentisierten Residuen mit Beträgen größer 2 zu erhalten, fordert man zunächst in der **Statistiken**-Subdialogbox eine **fallweise Diagnose** mit dem gewünschten **Ausreißer**-Kriterium an:



Anschließend lässt man sich in der Hauptdialogbox per **Einfügen**-Schalter die korrespondierende Syntax erzeugen und fordert im Subkommando **CASEWISE** ausgelassen-studentisierte Residuen (interner Name: SDRESID) an Stelle der standardisierten Residuen (interner Name: ZRESID) an:

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS CI(95)
   R ANOVA COLLIN TOL ZPP
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT mortal
 /METHOD=ENTER regen bildung
              farbig nox so2
 /CASEWISE PLOT(ZRESID) OUTLIERS(2).
```

→

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS CI(95)
   R ANOVA COLLIN TOL ZPP
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT mortal
 /METHOD=ENTER regen bildung
              farbig nox so2
 /CASEWISE PLOT(SDRESID) OUTLIERS(2).
```

Es resultiert die folgende Liste mit 6 Werten außerhalb der Signifikanzgrenzen für einen einzelnen Test:

Fallweise Diagnose[a]

Fallnummer	Stud. gelöschtes Residuum	Mortalität	Nicht standardisierter vorhergesagter Wert	Nicht standardisierte Residuen
2	2,629	997,9	905,587	92,3133
9	2,040	1002,0	928,664	73,3355
28	-2,831	844,1	938,464	-94,3636
32	-3,012	861,4	956,708	-95,3078
37	2,541	1113,0	1030,919	82,0808

a. Abhängige Variable: Mortalität

Dies stellt keine gravierende Abweichung von der erwarteten Häufigkeit von 3 zufälligen Signifikanzen bei 60 Tests zum Niveau $\alpha = 0{,}05$ unter jeweils gültiger Nullhypothese dar.

Um zu einer genaueren Bewertung zu kommen, müssen wir uns um das Problem der multiplen Signifikanztests kümmern, auf das wir bei der Residuenbeurteilung treffen. Soll bei *N* unter gültiger Nullhypothese durchgeführten Tests die Wahrscheinlichkeit für einen oder mehrere α-Fehler unter 0,05 gehalten werden, bietet sich die Bonferroni-Adjustierung an. Fox & Weisberg (2011, S. 295) schlagen dementsprechend vor, das p-Level zum betragsmäßig größten ausgelassen-studentisierten Residuum mit *N* zu multiplizieren und dann mit α zu vergleichen. Zur Berechnung des gesuchten p-Wertes taugt in SPSS das folgende COMPUTE-Kommando, das die kumulative Verteilungsfunktion der t-Verteilung mit 53 Freiheitsgraden an der Stelle 3,012 mit Hilfe der Funktion **CDF.T** ermittelt:

```
compute mpv = (1 - cdf.t(3.012, 53)) * 2 * 60.
execute.
```

Die am rechten Verteilungsrand ermittelte Überschreitungswahrscheinlichkeit (1 - cdf.t(3.012, 53)) muss verdoppelt werden, weil zweiseitige Tests durchzuführen sind. Weil im Beispiel das Ergebnis 0,238 deutlich über 0,05 liegt, kann die Nullhypothese, dass sich alle Residuen modellkonform verhalten, beibehal-

ten werden, und es sind keine Maßnahmen zur Neutralisierung von Ausreißern bzgl. der Residuen erforderlich.

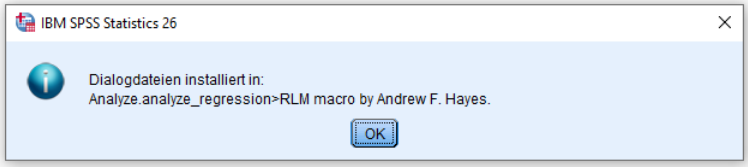
Wer sich die vorgeführte manuelle Beurteilung des betragsmäßig größten ausgelassen-studentisierten Residuums ersparen möchte, kann das Makro **RLM** von Andrew Hayes mit statistischen Methoden zur Unterstützung der linearen Regression über die folgende Webseite

http://afhayes.com/regression-analysis-and-linear-models.html

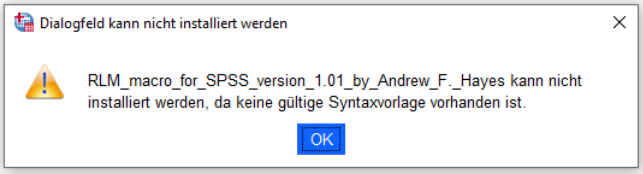
herunterladen. Leider klappt die Installation des benutzerdefinierten Dialogfelds über den Menübefehl

Erweiterungen > Extras > Benutzerdefiniertes Dialogfeld installieren (Kompatibilitätsmodus)

seit SPSS 26 *mit* FixPack 1 nicht mehr. Statt wie in SPSS 26 *ohne* FixPack 1 die Erfolgsmeldung



zu sehen, erhält man in SPSS 27 (und in SPSS 26 *mit* FixPack 1) die Fehlermeldung:



Das RLM-Makro ist aber trotzdem nutzbar:

- Nach dem Auspacken des heruntergeladenen ZIP-Archivs findet sich im Ordner **RLM macro** die Datei **rlm.sps**.
- Öffnen Sie diese Syntaxdatei mit SPSS (z.B. mit dem Menübefehl **Datei > Öffnen > Syntax**), und lassen Sie die Syntax ausführen (z.B. mit dem Menübefehl **Ausführen > Alle**).
- Lassen Sie in einem SPSS-Syntaxfenster für unser Mortalitätsbeispiel das folgende Kommando ausführen:

```
rlm y=mortal /x=regen bildung farbig nox so2 /diagnose=1.
```

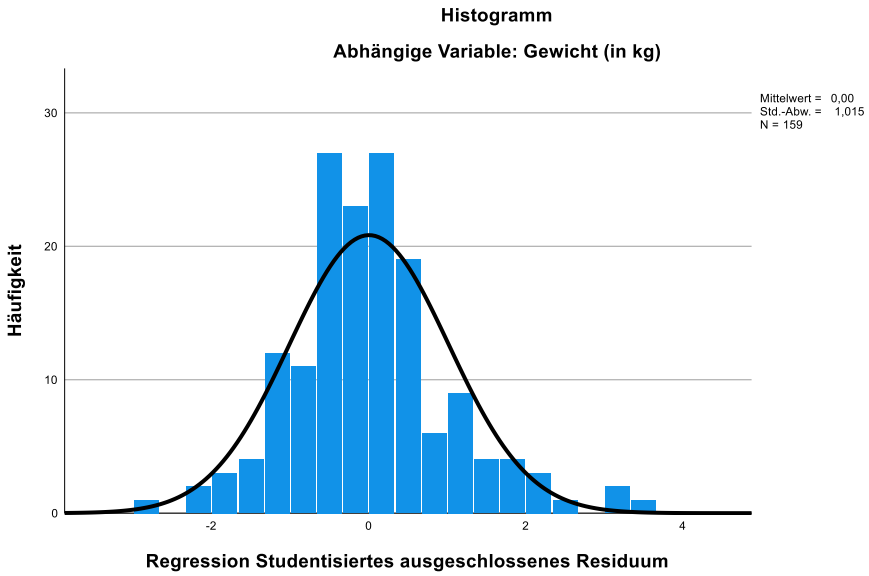
Anschließend erscheint u.a. die folgende Ausgabe, die unsere Berechnungen bestätigt:

```
Bonferroni-corrected p for largest t-residual
  t-resid  p-value  Casenum
  -3,0124    ,2380  32,0000
```

Als alternative Kriterien zur Identifikation von extremen Residuen kommen in Frage

- Der Betrag des ausgelassen-studentisierten Residuums überschreitet einen kritischen Wert, wobei die Stichprobengröße berücksichtigt werden sollte (siehe in Cohen et al. 2003, S. 401; Gordon 2011, S. 366; Urban & Mayerl 2018, S. 183). Als kritische Werte können gelten:
  - 3, 3,5 oder 4 bei großen Stichproben
  - 2,5 bei kleinen Stichproben
- Extremwertdiagnose per Boxplot (siehe Baltes-Götz 2022, Kapitel 8)

Bei der bivariaten Regression von Gewicht auf Größe in der männlichen Teilpopulation ($N = 159$) haben wir im Histogramm mit den ausgelassen-studentisierten Residuen am rechten Rand drei auffällige Werte festgestellt und schlussendlich als Ausreißer identifiziert (siehe Abschnitte 1.7.4.5 und 1.7.4.6.1):



Aufgrund der Bonferroni-korrigierten Beurteilung des betragsmäßig größten ausgelassen-studentisierten Residuums könnte im Gewicht-Größe - Beispiel die Nullhypothese eines modellkonformen Verhaltens beibehalten werden:

```
Bonferroni-corrected p for largest t-residual
  t-resid  p-value  Casenum
   3,5071    ,0941 151,0000
```

Hier zeigt sich, ...

- dass die Bonferroni-korrigierte Ausreißerbeurteilung eventuell etwas zu großzügig ist,
- dass statistische Kriterien allein nicht immer eine gute Entscheidung über Ausreißer ermöglichen.

Im Gewicht-Größe - Beispiel stellte sich für die 3 fraglichen Fälle ein BMI-Wert (*Body Mass Index*) über 30 heraus, während bei allen anderen Fällen der BMI-Wert unter diesem kritischen Wert für die Adipositas-Diagnose blieb. Diese inhaltliche Beobachtung brachte uns zur Entscheidung, dass die 3 Fälle nicht zum Anwendungsbereich des Modells gehören und daher ausgeschlossen werden sollten. Dabei haben sich die Schätzungen für den Steigungskoeffizienten und den Determinationskoeffizienten erheblich verändert.

3.1.2 Fälle mit starker Hebelwirkung

Fälle mit extremen Werten bei *un*abhängigen Variablen verfügen über einen starken Einfluss auf die Schätz- und Testergebnisse, wenn ihr Residuum relevant von null abweicht. Man spricht hier von einer *starken Hebelwirkung*. Die im Abschnitt 1.7.2.1 im Zusammenhang mit der bivariaten Regression beschriebenen Hebelwerte sind analog auch bei der multiplen Regression definiert.

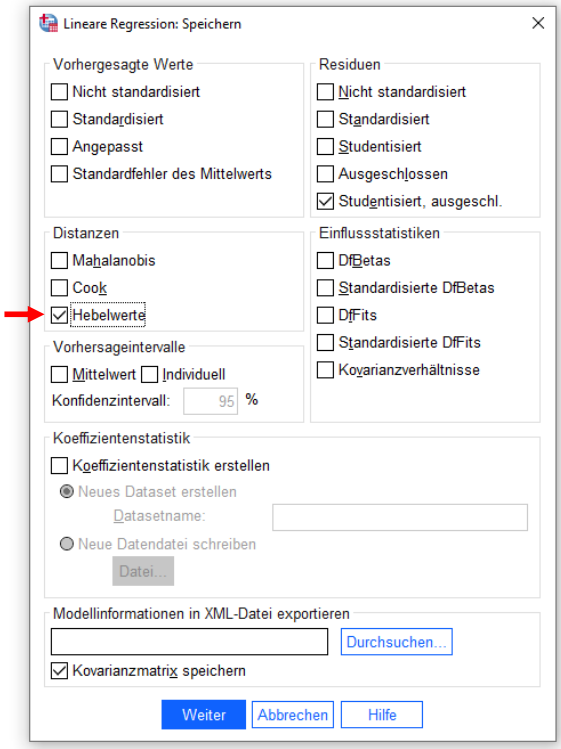
Die von SPSS berechneten *zentrierten* Hebelwerte nehmen Werte im Intervall von 0 bis 1 - 1/*N* an. Andere Programme und viele Autoren (z. B. Fox & Weisberg 2011, S. 296) verwenden eine leicht abweichende Hebel-Definition, sodass ein Wertebereich von 1/*N* bis 1 resultiert.

Urban & Mayerl (2018, S. 181) schlagen für die von SPSS berechneten zentrierten Hebelwerte als kritische Schwellenwerte vor (*k* = Anzahl der Regressoren):[1]

[1] Wie Urban & Mayerl (2018) in ihrem Buch an anderer Stelle (S. 178) erläutern, gehen sie bei einem Stichprobenumfang unter 200 von einer kleinen Stichprobe aus und halten Stichproben mit mehr als 500 Fällen für groß. Dazwischen liegt die mittlere Stichprobengröße.

- $\frac{2k}{N}$ bei großen Stichproben
- $\frac{3k}{N}$ bei kleinen Stichproben

Über den **Speichern**-Dialog der SPSS-Regressionsprozedur kann man die zentrierten Hebelwerte als neue Variable in die Arbeitsdatei schreiben lassen:



Über das folgende REGRESSION-Subkommando

```
/RESIDUALS outliers(lever)
```

erhält man im Ausgabefenster eine Tabelle mit den 10 größten Hebelwerten. Hier ist das Ergebnis für das Mortalitätsbeispiel zu sehen:

Ausreißerstatistik[a]

		Fallnummer	Statistik
Zentrierter Hebelwert	1	47	,441
	2	48	,304
	3	29	,246
	4	12	,227
	5	40	,224
	6	6	,178
	7	37	,171
	8	32	,170
	9	7	,169
	10	49	,151

a. Abhängige Variable: Mortalität

Wenn wir die Mortalitäts-Stichrobe als klein ansehen (N = 60), resultiert nach Urban & Mayerl (2018, S. 181) bei k = 5 Regressoren ein kritischer zentrierter Hebelwert von

$$\frac{15}{60} \approx 0{,}25$$

Er wird von zwei Städten aus Kalifornien in ernst zu nehmender Weise übertroffen:

- San Diego (Fallnummer 47, Hebelwert = 0,441)
- San Fernando (Fallnummer 48, Hebelwert = 0,304)

Ob diese Städte tatsächlich einen unangemessen starken Einfluss auf die Schätzergebnisse haben, hängt von ihrem Residuum ab. Im nächsten Abschnitt wird die sogenannte *Cook-Distanz* vorgestellt, die den aus der Hebelwirkung *und* dem Residualbetrag resultierenden Einfluss eines Falles auf die Schätzergebnisse quantifiziert.

Im **Speichern**-Dialog werden auch die **Mahalanobis-Distanzen** angeboten, die aus den zentrierten Hebelwerten durch eine Multiplikation mit dem Faktor (N - 1) entstehen und folglich dieselbe Information enthalten.

Über das folgende REGRESSION-Subkommando

```
/RESIDUALS outliers(mahal)
```

erhält man im Ausgabefenster eine Tabelle mit den 10 größten Mahalanobis-Distanzen. Hier ist das Ergebnis für das Mortalitätsbeispiel zu sehen:

Ausreißerstatistik[a]

		Fallnummer	Statistik
Mahalanobis-Abstand	1	47	26,016
	2	48	17,922
	3	29	14,497
	4	12	13,366
	5	40	13,231
	6	6	10,526
	7	37	10,081
	8	32	10,001
	9	7	9,952
	10	49	8,892

a. Abhängige Variable: Mortalität

Unter der Annahme einer multivariaten Normalverteilung der Regressoren folgen die Mahalanobis-Distanzen einer χ^2 - Verteilung mit k Freiheitsgraden. Es hat sich eingebürgert, für Mahalanobis-Distanzen die χ^2 - Verteilung auch unabhängig von der Normalverteilungsannahme für die Regressoren zu unterstellen. Bei der Verwendung der Mahalanobis-Distanzen zur Identifikation von Ausreißern bzgl. der Regressoren schlagen Tabachnick & Fidell (2013, S. 74) als kritische Grenze den χ^2 - Wert mit der Überschreitungswahrscheinlichkeit 0,001 vor. In unserem Fall (k = 5) resultiert der kritische Wert 20,52, den man von SPSS über das folgende Kommando ermitteln lassen kann:

```
compute k = idf.chisq(0.999, 5).
execute.
```

Urban & Mayerl (2018, S. 182) schlagen als kritische Grenze den χ^2 - Wert mit der Überschreitungswahrscheinlichkeit 0,01 vor, wobei für unser Beispiel der Wert 15,09 resultiert und (wie schon bei der Beurteilung der Hebelwerte) die Städte mit den Fallnummern 47 und 48 auffallen.

3.1.3 Gesamtbeurteilung der Einflussstärke

Einflussreiche Fälle sind dadurch gekennzeichnet, dass sie einen sehr hohen Einfluss auf den Vektor **b** mit den geschätzten Regressionskoeffizienten haben. Wenn einige wenige Fälle das Gesamtergebnis zu stark prägen, ist Skepsis hinsichtlich der externen Validität (Generalisierbarkeit) der Ergebnisse angebracht.

Eventuell wird man die Studie wiederholen mit einer verbesserten Designmatrix, in der bestimmte *X*-Wertkonstellationen mit anderen Häufigkeiten realisiert sind. Mit instabilen, von Einzelfällen stark abhängenden Ergebnissen ist vor allem in kleinen Stichproben zu rechnen (< 100 Fälle).

Wie stark ein Fall die Modellschätzung beeinflusst, hängt von folgenden Faktoren ab, die multiplikativ zusammenwirken:

- Betrag des Residuums
- Hebelwert

In der **Cook-Distanz** wird der Einfluss des *i*-ten Falles folgendermaßen definiert:

$$D_i := \frac{\tilde{e}_i^2}{k+1} \cdot \frac{\tilde{h}_i}{1-\tilde{h}_i}, \quad i = 1,..,N$$

Dabei sind:

$\tilde{e}_i^2$ das quadrierte studentisierte Residuum sensu SPSS (vgl. Abschnitt 1.7.2.2), das z. B. bei Fox & Weisberg (2011) als *standardisiertes Residuum* bezeichnet wird

$\tilde{h}_i$ der Hebelwert sensu Fox & Weisberg (2011), der sich aus dem von SPSS ermittelten zentrieren Hebelwert durch Addieren von 1/*N* ergibt (vgl. Abschnitt 1.7.2.1)

Einige Autoren nehmen für die Cook-Distanzen bei gültigen Modellannahmen eine $F_{(k+1,\ N-k-1)}$ - Verteilung an und konstruieren einen Signifikanztest zum Einfluss eines Falls auf die Schätzung der Regressionsgewichte (z.B. Kockläuner 1988, S. 85ff). Allerdings fallen die Cook-Distanzen meist kleiner als F-Werte aus, und Darlington & Hayes (2017, S. 489f) zweifeln zu Recht an der F-Verteilungsannahme.

Für unser Beispiel liefert die Regressionsprozedur bei Verwendung des REGRESSION-Subkommandos

```
/RESIDUALS outliers(cook)
```

die folgende Tabelle mit den 10 größten Cook-Abständen:

Ausreißerstatistik[a]

		Fallnummer	Statistik	Sig. F
Cook-Distanz	1	32	,301	,934
	2	37	,226	,967
	3	28	,149	,989
	4	49	,067	,999
	5	29	,047	1,000
	6	59	,047	1,000
	7	6	,047	1,000
	8	2	,035	1,000
	9	47	,031	1,000
	10	31	,031	1,000

a. Abhängige Variable: Mortalität

Zu allen Cook-Distanzen resultiert ein F-Test mit sehr großem p-Level, dem vermutlich *keine* Bedeutung beizumessen ist.

Zu den kritischen Werten der Cook-Distanz sind stark abweichende Vorschläge gemacht worden:

- Weisberg (1985, S. 120) empfiehlt, Fälle ab einer Cook-Distanz von 1 als einflussreich einzustufen.
- Oft wird der 50% - Punkt der $F_{(k+1,\ N-k-1)}$ - Verteilung genannt. In unserem Beispiel resultiert dabei 0,903.[1]
- Andere Autoren (z. B. Gordon 2010, S. 367) nennen mit $4/N$ einen in der Regel deutlich kleineren Grenzwert. In unserem Beispiel ergibt sich der Grenzwert 0,067.

Schließt man die beiden Fälle aus, die den Wert $4/N$ erheblich überschreiten (Fallnummern 32, 37), dann ändern sich die Regressionsgewichte deutlich, und der Regressor BILDUNG verfehlt Test die Signifikanzgrenze:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	883,999	87,449		10,109	,000	708,521	1059,478
	Regenmenge	2,341	,661	,380	3,540	,001	1,014	3,668
	Bildungsjahre	-7,297	6,682	-,106	-1,092	,280	-20,705	6,111
	% farbige Bevölkerung	2,482	,592	,371	4,194	,000	1,295	3,670
	NOX-Belastung	-,068	,441	-,020	-,153	,879	-,952	,817
	SO2-Belastung	,388	,112	,426	3,462	,001	,163	,613

a. Abhängige Variable: Mortalität

Oft interessiert, welche Regressionskoeffizienten ein Fall unangemessen stark beeinflusst. Entsprechende Informationen liefert die sogenannte **DFBETA-Statistik**, deren standardisierte Variante nach Urban & Mayerl (2018, S. 183) für einen Fall i und einen Regressor X_j folgendermaßen definiert ist:

$$DFBETAS_j := \frac{b_j - b_j^{(i)}}{\hat{\sigma}_{b_j}},\ i = 1,..,N,\ j = 0,..,k$$

Darin sind:

$b_j^{(i)}$ der bei Ausschluss von Fall i geschätzte Regressionskoeffizient

$\hat{\sigma}_{b_j}$ der Standardfehler des Regressionskoeffizienten (vgl. Abschnitt 1.4.1.3)

Urban & Mayerl (2018, S. 184) schlagen für die standardisierten DFBETA-Statistiken folgende Schwellenwerte vor:[2]

- $\frac{\pm 2}{\sqrt{N}}$ bei großen Stichproben
- ± 1 bei kleinen Stichproben

Wenn wir die Mortalitäts-Stichrobe als klein ansehen (N = 60), resultiert also ein Betrag von 1 als Schwellenwert.

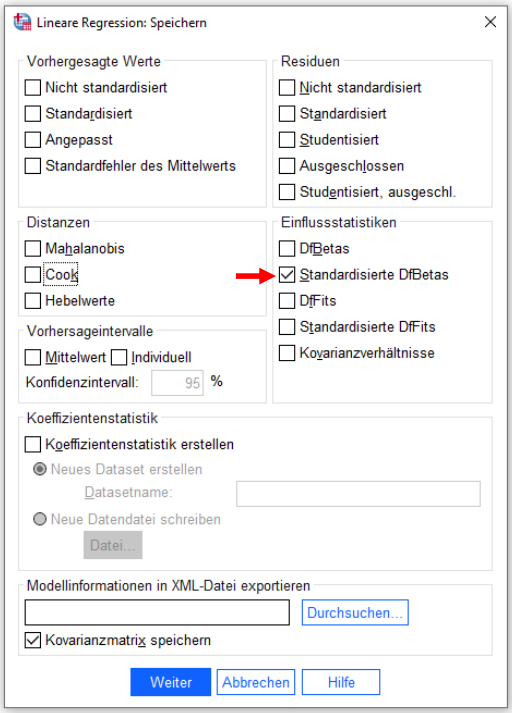
Über den **Speichern**-Dialog der SPSS-Regressionsprozedur kann man die standardisierten DFBETA-Statistiken als neue Variablen in die Arbeitsdatei schreiben lassen:

---

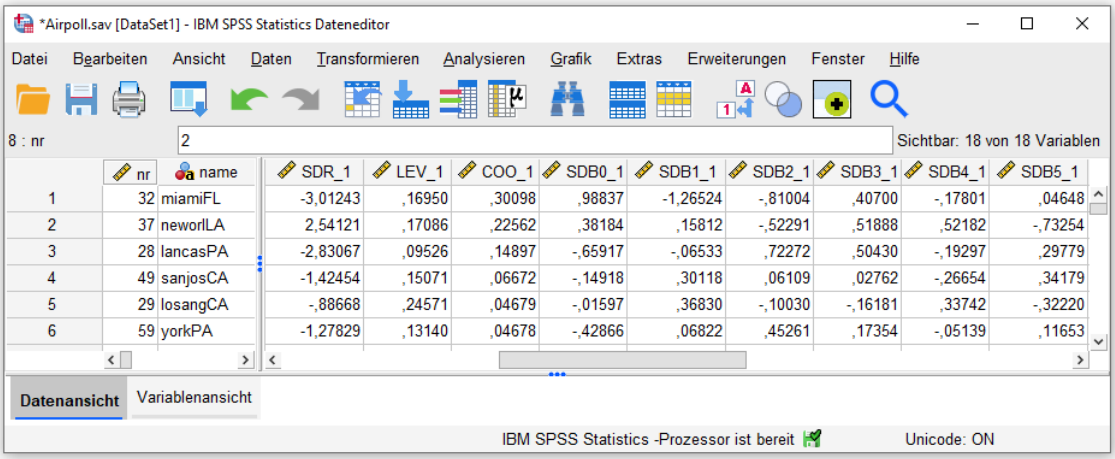
[1] Zur Berechnung kann die folgende SPSS-Syntax verwendet werden:

```
compute k = idf.f(0.5, 6, 54).
execute.
```

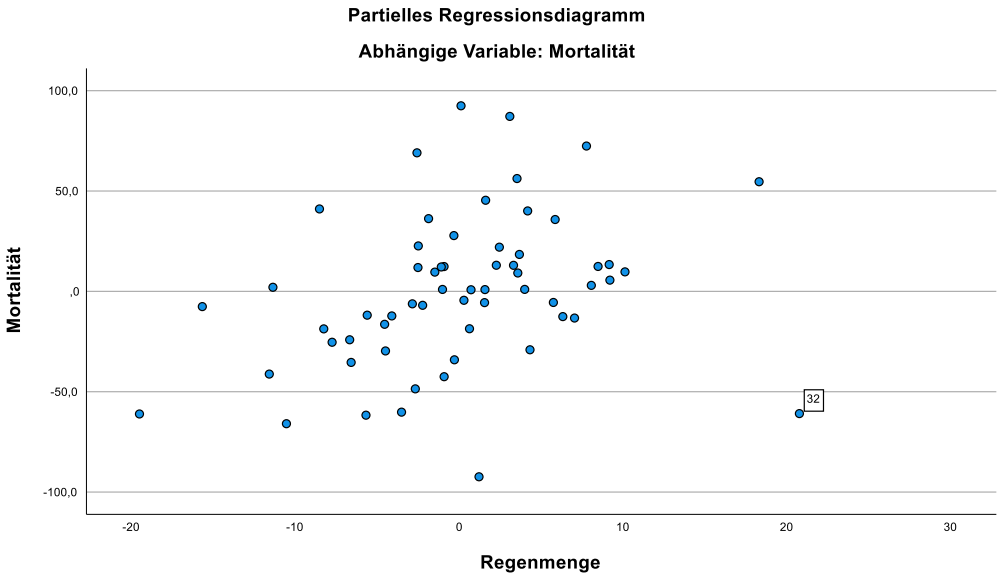
[2] Wie Urban & Mayerl (2018) in ihrem Buch an anderer Stelle (S. 178) erläutern, gehen sie bei einem Stichprobenumfang unter 200 von einer kleinen Stichprobe aus und halten Stichproben mit mehr als 500 Fällen für groß. Dazwischen liegt die mittlere Stichprobengröße.



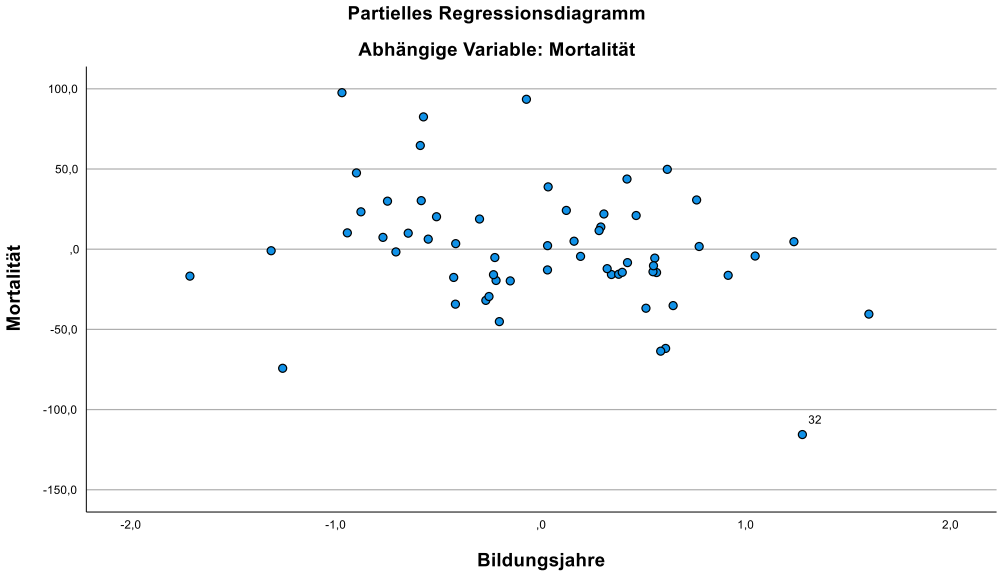
Im folgenden Dateneditorfenster sind die Fälle nach der Cook-Distanz absteigend sortiert. Wir erhalten für den Fall mit der größten Cook-Distanz (Fallnummer 32, Miami in Florida) beim Regressor REGEN einen standardisierten DFBETA-Wert oberhalb der betragsmäßigen Schwelle von 1 (Variable SDB1_1), und beim Regressor BILDUNG liegen ähnliche Verhältnisse vor (Variable SDB2_1):



Beim REGEN drückt der kritische Fall den Regressionskoeffizienten von 2,468 auf 1,658. Im partiellen Regressionsdiagramm (vgl. Abschnitt 2.9.1) ist der Grund dafür gut zu erkennen:



Bei der BILDUNG hat derselbe Fall einen (betragsmäßig) steigernden Effekt auf den Regressionskoeffizienten:



Wie die Listen extremer Ausreißer- bzw. Hebelwerte in den Abschnitten 3.1.1 bzw. 3.1.2 belegen, wirkt der einflussreiche Fall 32 vor allem über sein betragsmäßig großes Residuum. Im Idealfall kann ein modifiziertes (z. B. erweitertes) Modell diesen Fall besser erklären, sodass er keinen bedenklichen Einfluss auf die Schätzergebnisse mehr hat. Im Beispiel sollte vor allem geklärt werden, warum in amerikanischen Städten mit der Niederschlagsmenge die Mortalität ansteigt. Eventuell kann der Fall Miami einen Beitrag dazu leisten, die tatsächlichen Wirkmechanismen besser zu verstehen.

Ohne den Fall 32 ändern sich die geschätzten Koeffizienten samt Signifikanzbeurteilung bei den Regressoren REGEN und BILDUNG deutlich:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	913,210	91,317		10,000	,000	730,051	1096,368
	Regenmenge	2,468	,694	,384	3,554	,001	1,075	3,861
	Bildungsjahre	-10,503	6,915	-,144	-1,519	,135	-24,373	3,366
	% farbige Bevölkerung	2,779	,611	,404	4,546	,000	1,553	4,006
	NOX-Belastung	,165	,455	,045	,364	,717	-,746	1,077
	SO2-Belastung	,307	,113	,315	2,709	,009	,080	,535

a. Abhängige Variable: Mortalität

Man muss wohl feststellen, dass die Beispieldaten nur einen begrenzten Beitrag zur Aufklärung der Mortalitätsursachen in amerikanischen Städten leisten können:

- Zur Regenmenge liegt ein signifikantes Regressionsgewicht vor, doch es fehlt eine überzeugende theoretische Erklärung für den Effekt.
- Bei der Bildung ist die Wirkung theoretisch plausibel, doch fehlt eine statistische Bestätigung. Dafür ist einerseits die kleine Stichprobe (schlechte Power) verantwortlich. Andererseits ist uns schon bei der Linearitätsbeurteilung (siehe Abschnitt 2.9.1) eine Modellschwäche bei Städten mit einem niedrigen Bildungsindex aufgefallen.

Es besteht der Verdacht, dass im Modell relevante Regressoren fehlen (z. B. die Einwohnerzahl), sodass die geschätzten Koeffizienten nur eingeschränkt kausal interpretiert werden können.

Schließlich soll noch der Einfluss von Fällen mit extremer Hebelwirkung anhand von simulierten Daten demonstriert werden. Bei einer bivariaten Regression in einer Stichprobe mit 102 Fällen besteht *kein* Effekt des Regressors *X* auf das Kriterium *Y* (p-Level für β_1: 0,71):

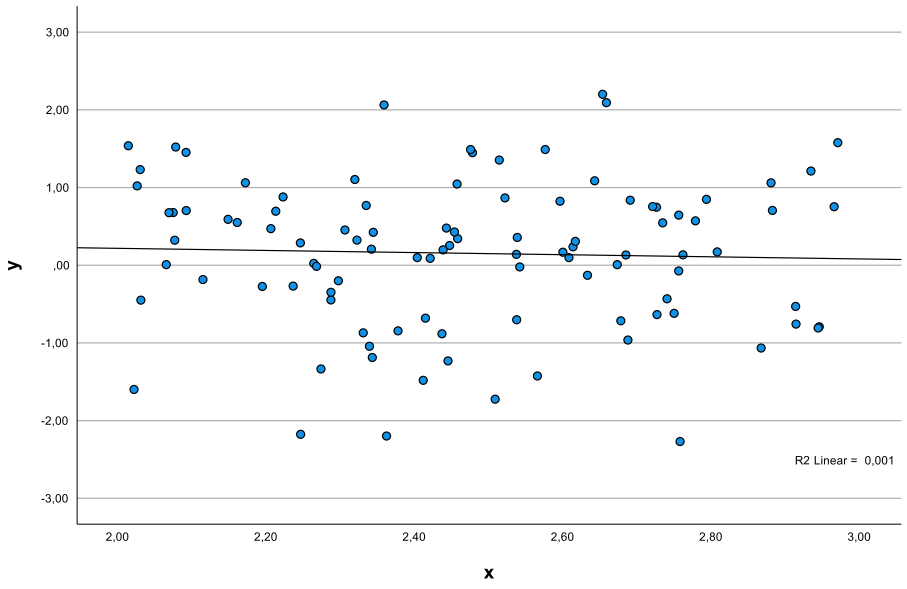


Abbildung 38: Bivariate Regression ohne Effekt

Ein zusätzlicher Fall mit extremer (*x*, *y*) - Kombination zieht die Regressionsgerade deutlich nach oben:

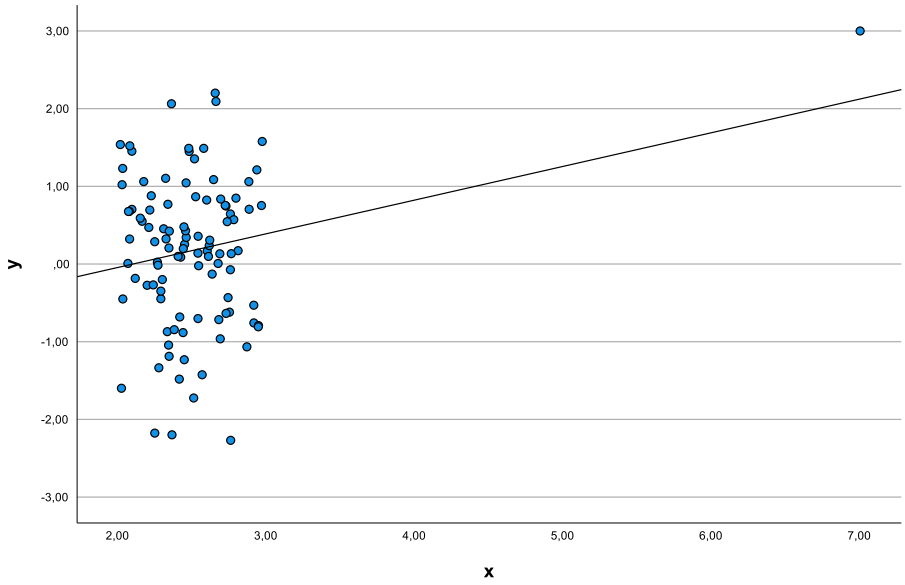


Abbildung 39: Einzelfall mit starkem Einfluss auf b_1

Der Regressionskoeffizient β_1 wird als signifikant beurteilt:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.
1	(Konstante)	-,910	,475		-1,914	,058
	x	,433	,185	,228	2,343	,021

a. Abhängige Variable: y

Für den kritischen Fall ergibt sich eine sehr große Cook-Distanz:

Ausreißerstatistik [a]

		Fallnummer	Statistik	Sig. F
Cook-Distanz	1	1	4,805	,010
	2	89	,043	,958
	3	37	,034	,966
	4	2	,031	,969
	5	39	,026	,974
	6	4	,026	,975
	7	48	,022	,978
	8	50	,022	,978
	9	77	,022	,978
	10	15	,020	,980

a. Abhängige Variable: y

Der Fall zeigt ein *unauffälliges* ausgelassen-studentisiertes Residuum von 1,823 und wird daher *nicht* als Ausreißer klassifiziert. Sein Hebelwert liegt jedoch weit über der kritischen Schwelle ($3/102 \approx 0{,}03$):

Ausreißerstatistik[a]

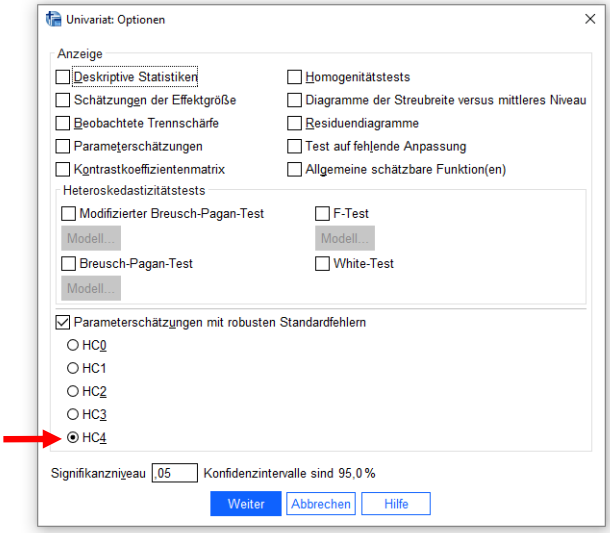
		Fallnummer	Statistik
Zentrierter Hebelwert	1	1	,737
	2	39	,009
	3	4	,009
	4	78	,009
	5	51	,009
	6	76	,009
	7	17	,008
	8	80	,007
	9	95	,007
	10	69	,007

a. Abhängige Variable: y

Bei Anwesenheit von Fällen mit großer Hebelwirkung werden für die Regressionskoeffizienten Signifikanztests auf der Basis von robusten Standardfehlern in der Variante HC4 empfohlen (siehe Abschnitt 1.7.3.4.1). Man kann sie von der SPSS-Prozedur für das allgemeine lineare Modell nach dem folgenden Menübefehl

Analysieren > Allgemeines lineares Modell > Univariat

in der **Optionen**-Dialogbox anfordern:



Der HC4-Signifikanztest spricht sich klar dafür aus, die Nullhypothese über den Regressor beizubehalten:

Parameterschätzungen mit robusten Standardfehlern

Abhängige Variable: y

Parameter	Regressions-koeffizient B	Robuster Standardfehler[a]	T	Sig.	95% Konfidenzintervall Untergrenze	95% Konfidenzintervall Obergrenze
Konstanter Term	-,910	5,545	-,164	,870	-11,911	10,092
x	,433	2,257	,192	,848	-4,044	4,911

a. HC4-Methode

Am Schätzwert für den Regressionskoeffizienten ändert sich durch die HC4-Inferenzstatistik nichts. Daher ist im Beispiel eine Regressionsanalyse *ohne* den einflussreichen Fall sinnvoller, weil man eine realistische Signifikanzbeurteilung *und* eine realistische Schätzung für den Regressionskoeffizienten erhält:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.
1	(Konstante)	,490	,900		,544	,587
	x	-,136	,362	-,038	-,377	,707

a. Abhängige Variable: y

3.2 Multikollinearität

Starke lineare Beziehungen zwischen den Regressoren verursachen bei der Parameterschätzung große Standardfehler, in extremen Fällen auch numerische Probleme. In der folgenden Darstellung für den geschätzten Standardfehler $\hat{\sigma}_{b_j}$ zum Regressionskoeffizienten b_j (vgl. Cohen et al. 2003, S. 86) zeigt sich, dass eine hohe quadrierte multiple Korrelation R_j^2 zwischen X_j und den übrigen Regressoren den Standardfehler erhöht, was zu einem großen Konfidenzintervall und einer geringen Power von Hypothesentests führt:

$$\hat{\sigma}_{b_j} = \frac{s_Y}{s_{X_j}} \sqrt{\frac{1}{1-R_j^2}} \sqrt{\frac{1-R^2}{N-k-1}}$$

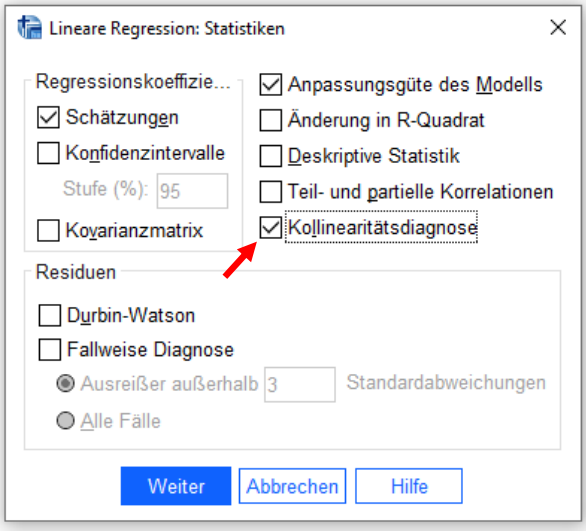
$$s_Y := \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})^2}\ ,\ s_{X_j} := \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_{j_i}-\bar{x}_j)^2}$$

Folglich sollte R_j^2 möglichst nahe beim idealen Wert 0 und die sogenannte **Toleranz** eines Regressors

$$Tol_j := 1-R_j^2$$

demzufolge möglichst nahe am idealen Wert 1 liegen. Als bedenklich werden meist Toleranzwerte kleiner als 0,10 eingeschätzt. Urban & Mayerl (2018, S. 238) empfehlen sogar den strengeren Grenzwert 0,25.

SPSS berichtet die Toleranzwerte zu den Regressoren bei entsprechender Anforderung in der Regressions-Subdialogbox **Statistiken**:



Im Mortalitätsbeispiel (ohne die Fälle 32, 37) zeigen sich unbedenkliche Werte:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			Kollinearitätsstatistik	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Toleranz	VIF
1	(Konstante)	883,999	87,449		10,109	,000		
	Regenmenge	2,341	,661	,380	3,540	,001	,501	1,998
	Bildungsjahre	-7,297	6,682	-,106	-1,092	,280	,617	1,620
	% farbige Bevölkerung	2,482	,592	,371	4,194	,000	,736	1,359
	NOX-Belastung	-,068	,441	-,020	-,153	,879	,350	2,854
	SO2-Belastung	,388	,112	,426	3,462	,001	,382	2,620

a. Abhängige Variable: Mortalität

Den ebenfalls berichteten Kehrwert der Toleranz

$$VIF_j := \frac{1}{Tol_j} = \frac{1}{1-R_j^2}$$

bezeichnet man verständlicherweise als **Varianzinflationierungsfaktor** (siehe obige Formel für $\hat{\sigma}_{b_j}$).

Während korrelierte Regressoren in Beobachtungsstudien eine alltägliche Aufgabenstellung der Datenanalyse darstellen, sind Toleranzprobleme ($R_j^2 > 0{,}75$) eher als vermeidbare Schwäche in der Forschungsmethodik zu betrachten. Sie treten z. B. dann auf, wenn in eine Regressionsgleichung mehrere Indikatoren derselben, kausal wirksamen latenten Variablen aufgenommen werden, was mit künstlichen Daten demonstriert werden soll:

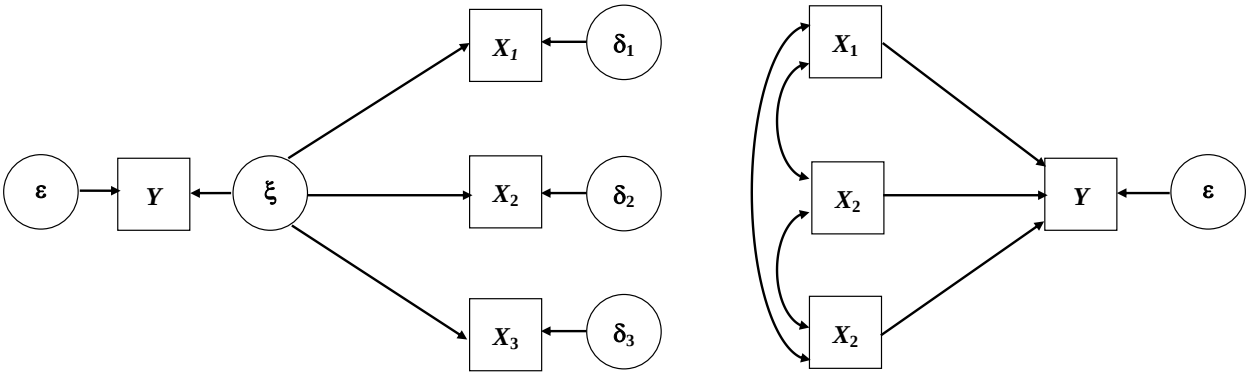


Abbildung 40: Drei äquivalente Indikatoren als Regressoren

Das linke Pfaddiagramm zeigt das wahre Modell und insbesondere die Herkunft der drei Variablen X_1, X_2 und X_3, die im rechten Pfaddiagramm als Regressoren-Trio agieren. Weil es sich um äquivalente Indikatoren desselben Konstrukts handelt, resultieren irritierende Regressionsergebnisse mit einem hochsignifikanten F-Test zur globalen Nullhypothese,

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	47,022	3	15,674	14,654	,000 [b]
	Nicht standardisierte Residuen	263,122	246	1,070		
	Gesamt	310,144	249			

a. Abhängige Variable: y

b. Einflußvariablen : (Konstante), x3, x1, x2

aber *ohne* signifikante t-Tests zu den einzelnen Regressionskoeffizienten:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			Kollinearitätsstatistik	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Toleranz	VIF
1	(Konstante)	3,001	,066		45,708	,000		
	x1	,131	,130	,118	1,003	,317	,250	4,006
	x2	,190	,128	,181	1,484	,139	,233	4,297
	x3	,128	,133	,114	,966	,335	,248	4,031

a. Abhängige Variable: y

In der **Toleranz**-Spalte der **Koeffizienten**-Tabelle wird deutlich, worin das Problem besteht (Toleranz ≤ 0,25). Hier bestätigt sich die Empfehlung von Urban & Mayerl (2018, S. 238), bei der Toleranz den Mindestwert 0,25 zu fordern.

Als Solisten erreichen alle Regressoren eine günstige Bewertung, z. B.:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			Kollinearitätsstatistik	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Toleranz	VIF
1	(Konstante)	2,990	,066		45,367	,000		
	x1	,403	,066	,363	6,129	,000	1,000	1,000

a. Abhängige Variable: y

Allerdings wird der wahre Koeffizient aus der Regression von Y auf ξ (= 0,5) aufgrund der Reliabilitätsmängel in X_1 deutlich unterschätzt. Wird statt *eines* Indikators der Mittelwert aus den *drei* Indikatoren als Regressor verwendet, dann verbessert sich die Schätzung aufgrund des Reliabilitätsgewinns:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			Kollinearitätsstatistik	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Toleranz	VIF
1	(Konstante)	2,999	,065		46,010	,000		
	Mittelwert	,452	,068	,389	6,649	,000	1,000	1,000

a. Abhängige Variable: y

In einer explorativen Faktorenanalyse mit den drei Indikatoren wird die eindimensionale Struktur klar bestätigt. Werden die geschätzten Faktorwerte abgespeichert und als Regressor verwendet, ergibt sich kein Gewinn im Vergleich zur Verwendung des Mittelwerts aus den Indikatoren:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			Kollinearitätsstatistik	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Toleranz	VIF
1	(Konstante)	2,986	,065		45,830	,000		
	Faltir aus der EFA	,449	,067	,389	6,652	,000	1,000	1,000

a. Abhängige Variable: y

Das war zu erwarten, weil in den simulierten Daten die drei Indikatoren dieselbe Ladung auf dem Faktor aufweisen.

Optimal ist der skizzierten Situation eine Strukturgleichungsanalyse (vgl. Abschnitt 6.3) unter Verwendung das wahren Modells (siehe linkes Pfaddiagramm in Abbildung 40). Man vermeidet Multikollineari-

täts- bzw. Redundanz-Probleme und steigert zudem die Präzision der Schätzung, weil der Regressionskoeffizient ohne Minderung durch Messfehler geschätzt werden kann (vgl. Abschnitt 6.3). Anschließend sind die vom Strukturgleichungsanalyseprogramm IBM SPSS Amos ermittelten Schätzungen für die Beispieldaten zu sehen:

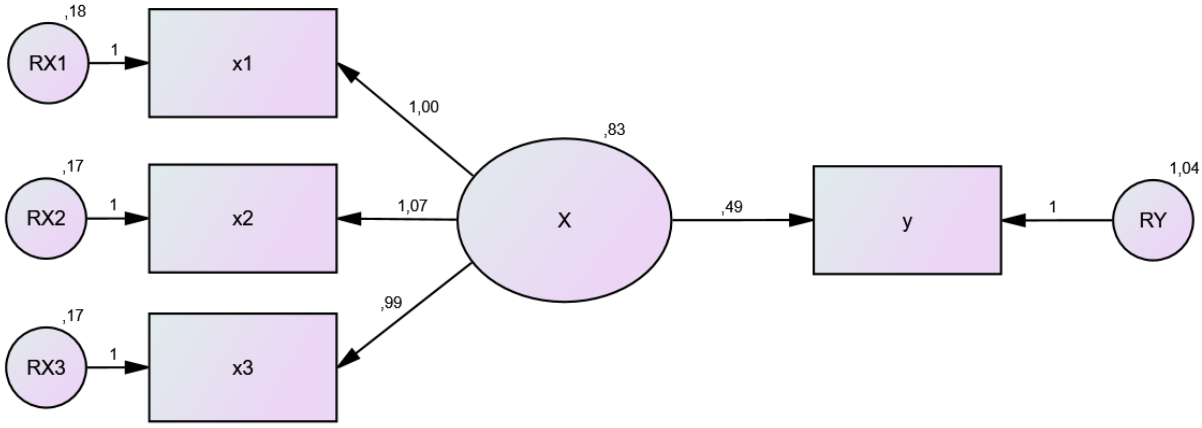


Abbildung 41 Mit Amos geschätztes Strukturgleichungsmodell

4 Polynomische Regressionsmodelle

4.1 Ein quadratisches Modell für das volkswirtschaftliche Beispiel

Bei unserem volkswirtschaftlichen Beispiel (siehe Abschnitt 1.7.3.2.2) sind wir im Rahmen der Homogenisierung der Fehlervarianzen (siehe Abschnitt 1.7.3.3) an eine multiple Regressionsgleichung geraten (mit Y als Bezeichnung für *LNBSP* und X als Bezeichnung für *ERN*):

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

Mit der eingezeichneten quadratischen Regressionsfunktion sieht die bivariate Verteilung von LNBSP und ERN folgendermaßen aus:

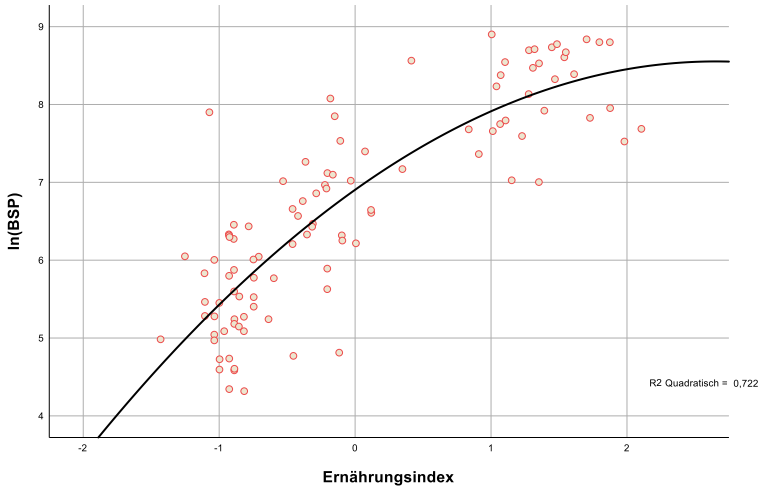


Abbildung 42: Regression von LNBSP auf ERN und ERN2

In der Regression von Y auf X und X^2 ist der zweite Regressor eine perfekte Funktion des ersten. Dies ist aber *kein* Verstoß gegen die im Abschnitt 2.2 genannte Voraussetzung *linear* unabhängiger Spalten in der Designmatrix.

Um komplexere Abhängigkeitsmuster zu modellieren, dürfen neben X und X^2 auch noch Terme höherer Ordnung auftreten. Da auf der Regressorenseite jeweils ein Polynom in X vorliegt, spricht man von **polynomischen Regressionsgleichungen**. Auf Besonderheiten im Vergleich zur normalen multiplen Regression trifft man z. B. bei der Interpretation der Regressionskoeffizienten. In unserem Modell kann man sich z. B. nicht vorstellen, den ERN-Wert zu erhöhen und gleichzeitig den ERN2-Wert konstant zu halten.

Die im Abschnitt 1.7.3.3 aus didaktischen Gründen zurückgehaltenen Ergebnisse zum quadratischen Modell für die Regression von LNBSP auf ERN sollen nun nachgeliefert werden:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,850 [a]	,722	,716	,69800

a. Einflußvariablen : (Konstante), ern2, Ernährungsindex

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	125,300	2	62,650	128,592	,000 [b]
	Nicht standardisierte Residuen	48,233	99	,487		
	Gesamt	173,533	101			

a. Abhängige Variable: ln(BSP)

b. Einflußvariablen : (Konstante), ern2, Ernährungsindex

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	6,904	,117		59,016	,000	6,672	7,136
	Ernährungsindex	1,242	,090	,948	13,756	,000	1,063	1,422
	ern2	-,234	,095	-,169	-2,455	,016	-,423	-,045

a. Abhängige Variable: ln(BSP)

In der **Koeffizienten**-Tabelle ist vor allem der t-Test zum Prädiktor ERN2 (X^2 in der Modellgleichung) von Interesse. Er beantwortet die Frage, ob das quadratische Modell im Vergleich zum einfach-linearen eine signifikant bessere Varianzaufklärung leistet. Aufgrund der empirischen Überschreitungswahrscheinlichkeit von 0,016 (im zweiseitigen Test) ist in unserem Beispiel das quadratische Modell zu bevorzugen.

Weil die Idee einer quadratischen Beziehung durch eine Inspektion der Stichprobendaten gewonnen wurde, liegt aber *kein* Signifikanztest im strengen Sinn vor. Eine unabhängige Stichprobe zur Prüfung der neu gewonnenen Hypothese ist allerdings im konkreten Fall (mit Nationen als Beobachtungseinheiten) schwer zu finden.

Das adjustierte R_a^2 steigt im Vergleich zum einfach-linearen Modell für das Kriterium LNBSP merklich, wenngleich nicht überwältigend, von 0,702 auf 0,716.

Die ausgelassen-studentisierten Residuen aus der Regression von LNBSP auf ERN und ERN2 bewegen sich im Plot gegen die vorhergesagten Werte bis auf einen Ausreißer (siehe unten) mit homogener Varianz um die Null-Lage, was für die Linearität und die Homoskedastizität spricht:

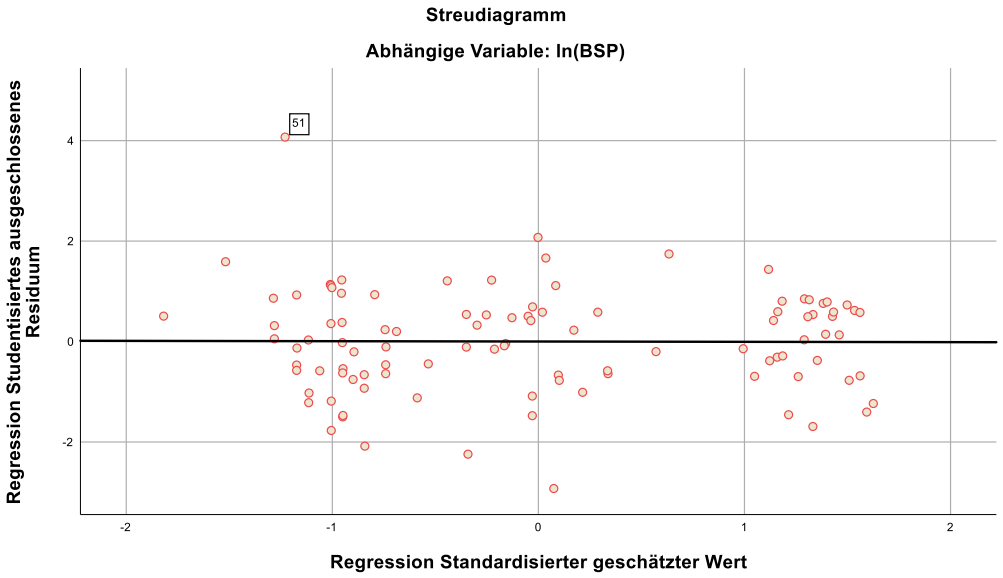


Abbildung 43: Ausgelassen-studentisierte Residuen gegen prognostizierte Werte aus der Regression von LNBSP auf ERN und ERN2

Trotz des Ausreißers wird die Varianzhomogenitätsannahme vom modifizierten Breusch-Pagan - Test akzeptiert (vgl. Abschnitt 1.7.3.2.4):

Modifizierter Breusch-Pagan-Test auf Heteroskedastizität[a,b,c]

Chi-Quadrat	df	Sig.
1,937	1	,164

a. Abhängige Variable: ln(BSP)

b. Testet die Nullhypothese, dass die Varianz der Fehler nicht von den Werte der unabhängigen Variablen abhängt.

c. Vorhergesagte Werte aus Design: Konstanter Term + ern + ern2

Im Histogramm der ausgelassen-studentisierten Residuen stört der Ausreißer durch seinen inakzeptabel hohen Wert von 4,07 jedoch erheblich:

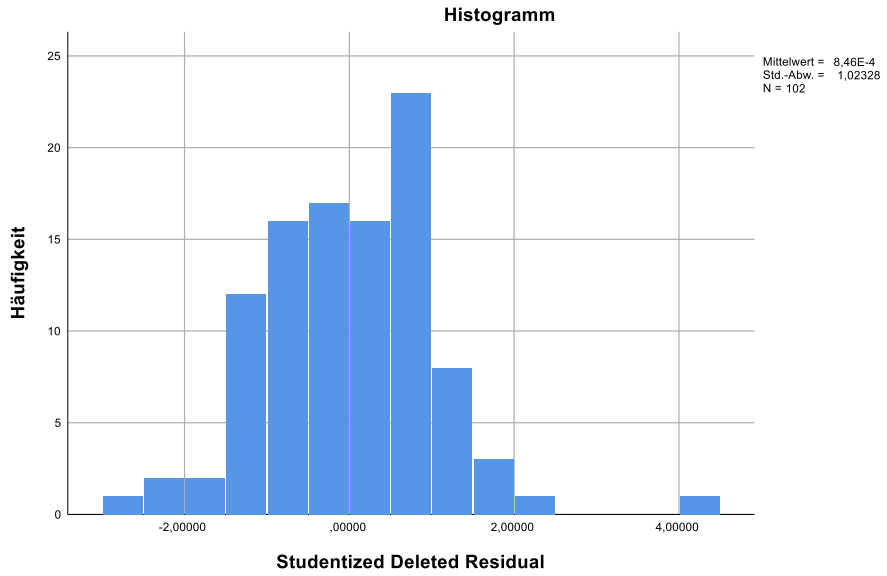


Abbildung 44 Histogramm der ausgelassen-studentisierten Residuen aus der Regression von LNBSP auf ERN und ERN2

Beim betroffenen Fall Nr. 51 handelt sich um Puerto Rico, ein karibisches Außengebiet der USA. Sein BSP liegt weit über dem Niveau der Nachbarschaft, doch sein Ernährungsindex bewegt sich auf dem Niveau von Entwicklungsländern wie Bangladesch oder Haiti.

Bei einer Ausreißerbeurteilung unter Berücksichtigung der multiplen Testung (siehe Abschnitt 3.1.1) erhält man durch das folgende SPSS-Kommando COMPUTE

```
compute mpv = (1 - cdf.t(4.07, 98)) * 2 * 102.
execute.
```

das signifikante p-Level 0,010.

In der mit dem folgenden REGRESSION-Subkommando

```
/RESIDUALS outliers(cook)
```

angeforderten Tabelle mit den zehn größten Cook-Einflussgrößen (vgl. Abschnitt 3.1.3) belegt Puerto Rico einen einsamen Spitzenplatz:

Ausreißerstatistik[a]

		Fallnummer	Statistik	Sig. F
Cook-Distanz	1	51	,165	,920
	2	97	,083	,969
	3	101	,076	,973
	4	9	,069	,976
	5	10	,052	,984
	6	11	,034	,992
	7	60	,033	,992
	8	13	,027	,994
	9	99	,027	,994
	10	18	,024	,995

a. Abhängige Variable: ln(BSP)

Der von Gordon (2010, S. 367) vorgeschlagene Schwellenwert für die Cook-Distanz (4/102 = 0,039) wird deutlich übertroffen (siehe Abschnitt 3.1.3).

Puerto Rico gehört offenbar nicht zum Erklärungsbereich des Modells und wird daher von der Schätzung ausgeschlossen, was zu deutlich verbesserten Ergebnissen führt. Das adjustierte R_a^2 steigt von 0,716 auf 0,755:

Modellzusammenfassung[b]

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,872 [a]	,760	,755	,64885

a. Einflußvariablen : (Konstante), ern2, Ernährungsindex

b. Abhängige Variable: ln(BSP)

Die Regressoren erreichen prägnantere Schätzwerte und eine günstigere Signifikanzbeurteilung:

Koeffizienten[a]

		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
Modell		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	6,918	,109		63,582	,000	6,702	7,134
	Ernährungsindex	1,296	,085	,987	15,248	,000	1,127	1,464
	ern2	-,275	,089	-,199	-3,079	,003	-,452	-,098

a. Abhängige Variable: ln(BSP)

Für das schlussendlich gewählte Modell zeigen die ausgelassen-studentisierten Residuen eine akzeptable Approximation an die Normalverteilung:

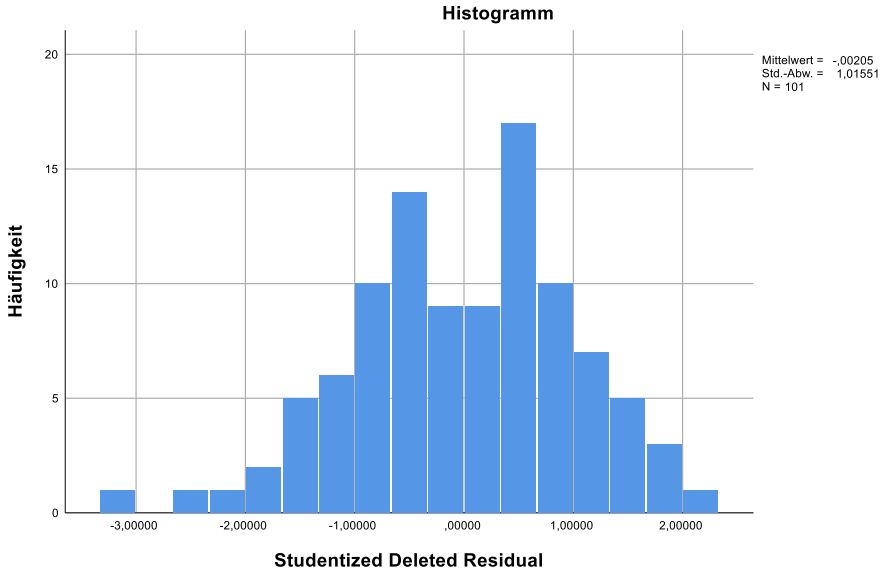


Abbildung 45 Histogramm der ausgelassen-studentisierten Residuen aus der Regression von LNBSP auf ERN und ERN2 (ohne Puerto Rico)

In der inferenzstatistischen Beurteilung wird die Normalverteilungs-Nullhypothese von beiden Tests (Shapiro-Wilk und Kolmogorov-Smirnov) beibehalten:

Tests auf Normalverteilung

	Kolmogorov-Smirnov [a]			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Studentized Deleted Residual	,082	101	,093	,987	101	,456

a. Signifikanzkorrektur nach Lilliefors

Vom modifizierten Breusch-Pagan - Test (siehe Abschnitt 1.7.3.2.4) wird die Varianzhomogenitäts-Nullhypothese nach der Elimination des Ausreißers erwartungsgemäß noch günstiger beurteilt:

Modifizierter Breusch-Pagan-Test auf Heteroskedastizität[a,b,c]

Chi-Quadrat	df	Sig.
,611	1	,434

a. Abhängige Variable: ln(BSP)

b. Testet die Nullhypothese, dass die Varianz der Fehler nicht von den Werte der unabhängigen Variablen abhängt.

c. Vorhergesagte Werte aus Design: Konstanter Term + ern + ern2

4.2 Zentrieren und Multikollinearität bei der polynomischen Regression

Im volkswirtschaftlichen Beispiel ist der Regressor ERN als ehemalige Hauptkomponente (Ergebnis einer explorativen Faktorenanalyse) z-standardisiert und hat damit insbesondere den Mittelwert 0. Infolgedessen fällt die Korrelation zwischen ERN und ERN^2 niedrig aus, und wir bleiben von beunruhigenden Multikollinearitätswerten verschont.

Es wird vielfach empfohlen, in einem polynomischen Regressionsmodell den Regressor zur Vermeidung von Multikollinearitätsproblemen grundsätzlich zu zentrieren. Um einen Eindruck vom Effekt der Zentrierung zu gewinnen, sollen simulierte Daten (N = 100) aus einer Population mit dem folgenden wahren Modell

$$Y = 1 + 0{,}5 \cdot X + 0{,}5 \cdot X^2 + \varepsilon,\ X \sim \mathrm{N}(3,\ 1)$$

betrachtet werden. Verwendet man den unzentrierten Prädiktor (normalverteilt mit Erwartungswert 3 und Varianz 1) und sein Quadrat, dann resultieren schlechte Toleranzwerte, und die Beurteilung des Parameters zum einfachen Regressor leidet unter einem großen Standardfehler:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			Kollinearitätsstatistik	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Toleranz	VIF
1	(Konstante)	1,047	,510		2,053	,043		
	x	,384	,325	,099	1,182	,240	,066	15,052
	x2	,525	,050	,881	10,512	,000	,066	15,052

a. Abhängige Variable: y

Nach dem Zentrieren zeigt sich das erwartete Bild mit hohen Toleranzen und kleinen Standardfehlern zu allen Parametern:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			Kollinearitätsstatistik	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Toleranz	VIF
1	(Konstante)	6,921	,109		63,641	,000		
	xz	3,532	,085	,911	41,580	,000	,971	1,029
	xz2	,525	,050	,230	10,512	,000	,971	1,029

a. Abhängige Variable: y

Beim primär relevanten Regressor X^2 bleiben allerdings der geschätzte Regressionskoeffizient und dessen inferenzstatistische Beurteilung (Standardfehler, Signifikanztest) *völlig identisch*. Hier wird der verkleinerte Varianzinflationierungsfaktor komplett neutralisiert durch die Verkleinerung der Standardabweichung des Regressors (Hayes 2018, S. 309).

Außerdem ist zu beachten, dass sich die Bedeutung der Regressionskoeffizienten β_0 und β_1 durch das Zentrieren ändert, z. B.:

$$\begin{aligned} Y &= \tilde{\beta}_0 + \tilde{\beta}_1(X - \mathrm{E}(X)) + \tilde{\beta}_2(X - \mathrm{E}(X))^2 + \varepsilon \\ &= \tilde{\beta}_0 + \tilde{\beta}_1 X - \tilde{\beta}_1 \mathrm{E}(X) + \tilde{\beta}_2(X^2 - 2\,\mathrm{E}(X)X + \mathrm{E}(X)^2) + \varepsilon \\ &= \tilde{\beta}_0 - \tilde{\beta}_1 \mathrm{E}(X) + \tilde{\beta}_2 \mathrm{E}(X)^2 \quad + \quad (\tilde{\beta}_1 - 2\tilde{\beta}_2 \mathrm{E}(X))X \quad + \quad \tilde{\beta}_2 X^2 \quad + \quad \varepsilon \end{aligned}$$

Die Koeffizienten des ursprünglichen Modells stehen in den folgenden Beziehungen zu den Koeffizienten des Modells für den zentrierten Regressor und sein Quadrat:

$$\begin{aligned} \beta_0 &= \tilde{\beta}_0 - \tilde{\beta}_1 \mathrm{E}(X) + \tilde{\beta}_2 \mathrm{E}(X)^2 \\ \beta_1 &= \tilde{\beta}_1 - 2\tilde{\beta}_2 \mathrm{E}(X) \\ \beta_2 &= \tilde{\beta}_2 \end{aligned}$$

Andersherum betrachtet stehen die Koeffizienten des Modells für den zentrierten Regressor und sein Quadrat in den folgenden Beziehungen zu den Koeffizienten des ursprünglichen Modells:

$$\begin{aligned} \tilde{\beta}_0 &= \beta_0 + (\beta_1 + 2\beta_2 \mathrm{E}(X))\mathrm{E}(X) - \beta_2 \mathrm{E}(X)^2 \\ &= \beta_0 + \beta_1 \mathrm{E}(X) + \beta_2 \mathrm{E}(X)^2 \\ \tilde{\beta}_1 &= \beta_1 + 2\beta_2 \mathrm{E}(X) \\ \tilde{\beta}_2 &= \beta_2 \end{aligned}$$

Die im Beispiel zu beobachtende Veränderung der inferenzstatistische Beurteilung des zentrierten Regressors *X* ist also *nicht* auf eine Reduktion des Kollinearität zurückzuführen, sondern auf eine veränderte Bedeutung des Regressionskoeffizienten.

Andrew Hayes (2018, S. 312) beschreibt die Rolle der Zentrierung sehr prägnant in einem Abschnitt über „einige Mythen“:[1]

> To be sure, there are interpretational advantages associated with mean-centering, but the differences in model coefficients and standard errors have nothing to do with reduced collinearity that results from mean-centering.

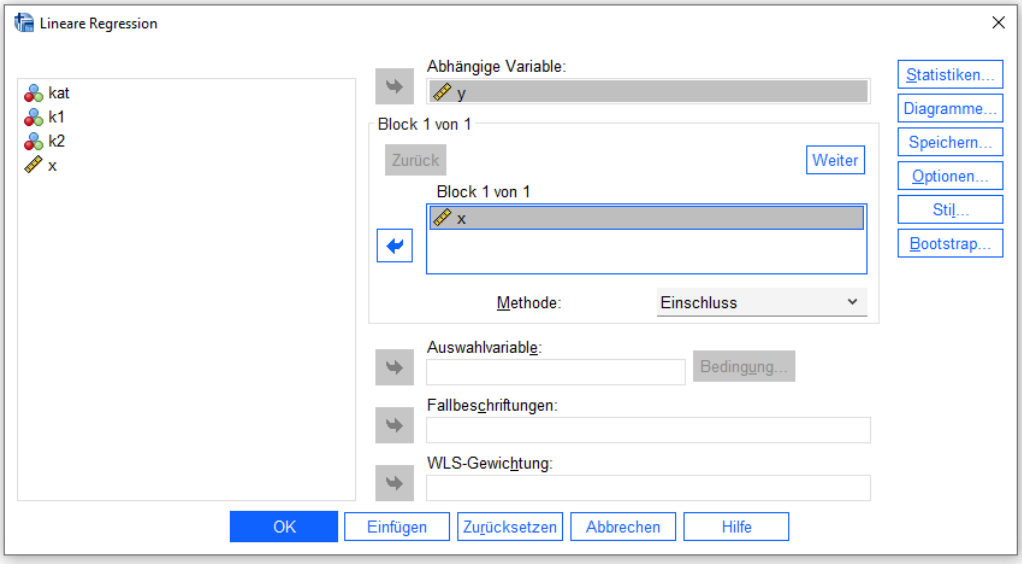
[1] Die zitierte Passagen in Hayes (2018) befassen sich zwar mit Moderatormodellen, doch lassen sich Aussagen über das Produkt *XW* von zwei Regressoren *X* und *W* auf das Produkt *XX* des Regressors *X* mit sich selbst übertragen.

5 Alternativen zum simultanen Einschluss aller Regressoren

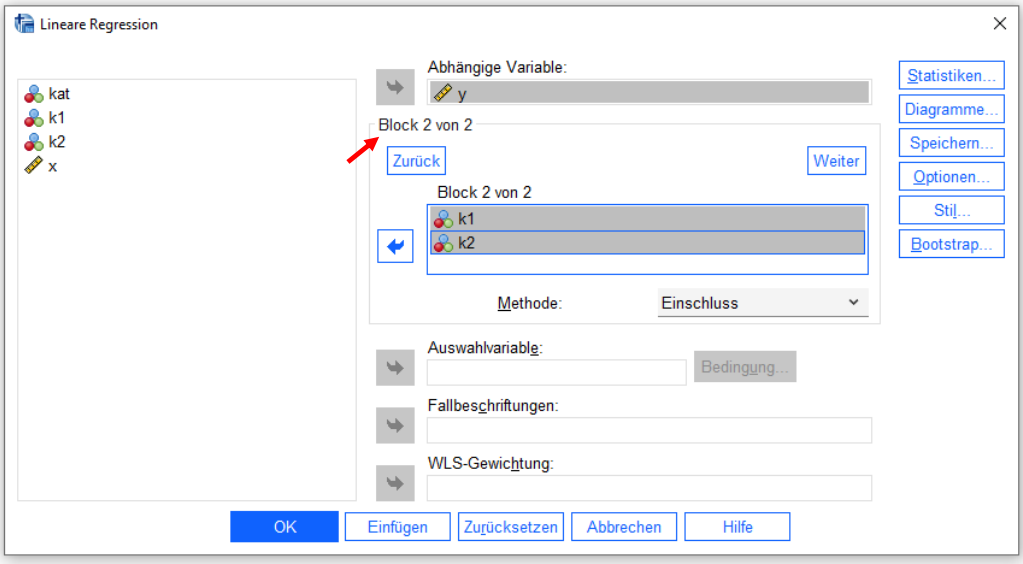
5.1 Blockbildung zur Untersuchung eines kategorialen Regressors

Die SPSS-Prozedur zur linearen Regression erlaubt eine sukzessive Modellerweiterung um *Blöcke* von Regressoren, wobei durch einen F-Test beurteilt werden kann, ob der zuletzt einbezogene Block von Regressoren eine signifikante R^2 - Verbesserung bewirkt hat. Man spricht dabei auch von einer *hierarchischen Regressionsanalyse*.

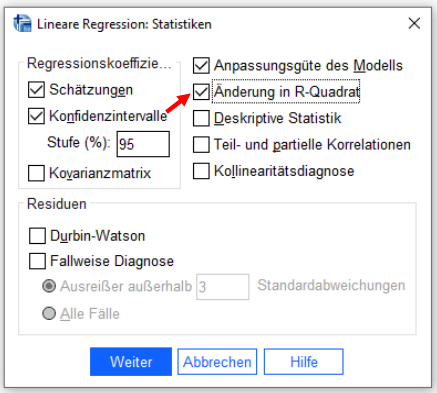
Eine Blockbildung bietet sich z. B. bei den (w - 1) Kodiervariablen zu einem kategorialen Regressor mit w Ausprägungen an. Zur Demonstration verwenden wir ein künstliches Beispiel mit einem metrischen Regressor X und einem kategorialen Regressor mit drei Ausprägungen, wobei im ersten Block der metrische Regressor aufgenommen werden soll:



Nach einem Mausklick auf den Schalter **Weiter** folgen in einem zweiten Block die beiden Kodiervariablen zum nominalskalierten Regressor:



Um einen Signifikanztest zur R^2 - Verbesserung durch die beiden sukzessive ins Modell aufgenommenen Blöcke zu erhalten, markiert man in der **Statistiken**-Subdialogbox das Kontrollkästchen **Änderung in R-Quadrat**:



In der folgenden Tabelle erfahren wir, dass beide Blöcke ausgehend vom jeweils bereits vorhandenen Modell eine signifikante R^2 - Verbesserung bewirkt haben:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Statistikwerte ändern: Änderung in R-Quadrat	Änderung in F	df1	df2	Sig. Änderung in F
1	,355 [a]	,126	,119	1,05205	,126	18,444	1	128	,000
2	,412 [b]	,170	,150	1,03324	,044	3,352	2	126	,038

a. Einflußvariablen : (Konstante), x

b. Einflußvariablen : (Konstante), x, k1, k2

Durch den Signifikanztest zum zweiten Block wird der kategoriale Regressor insgesamt auf Signifikanz beurteilt bzw. die folgende Nullhypothese (mit β_2 und β_3 als Koeffizienten zu den beiden Kodiervariablen des Faktors) geprüft:

$$H_0: \beta_2 = \beta_3 = 0$$

Im Beispiel wird die Nullhypothese abgelehnt (p = 0,038).

In der Koeffiziententabelle finden sich u.a. die Signifikanztests zu den beiden Kodiervariablen des Faktors. Verwendet man Indikator- bzw. Dummy-Variablen mit der dritten Kategorie als Referenz, dann stehen die Koeffizienten der beiden Kodiervariablen für die Abweichung der ersten bzw. zweiten Kategorie von der Referenzkategorie bei statistischer Kontrolle des metrischen Regressors:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten: Regressionskoeffizient B	Nicht standardisierte Koeffizienten: Std.-Fehler	Standardisierte Koeffizienten: Beta	T	Sig.	95,0% Konfidenzintervalle für B: Untergrenze	95,0% Konfidenzintervalle für B: Obergrenze
1	(Konstante)	,412	,092		4,467	,000	,230	,595
	x	,404	,094	,355	4,295	,000	,218	,589
2	(Konstante)	,082	,158		,518	,606	-,231	,394
	x	,391	,093	,344	4,216	,000	,207	,575
	k1	,540	,226	,225	2,395	,018	,094	,987
	k2	,452	,220	,193	2,052	,042	,016	,887

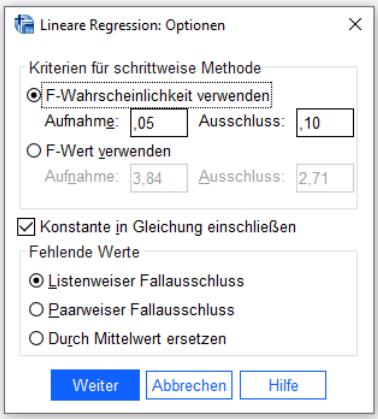
a. Abhängige Variable: y

5.2 Automatische Modellsuche

Innerhalb eines Blocks, der auch *alle* in Frage kommenden Regressoren umfassen darf, kann man SPSS automatisch nach einem guten Modell suchen lassen. Es wird schrittweise anhand von zweiseitigen Signifikanztests entschieden, ob ein Regressor aufgenommen oder entfernt werden soll, wobei drei Strategien zur Verfügung stehen:

- **Vorwärts**
  Ausgehend vom Modell *ohne* den fraglichen Block wird Schritt für Schritt darüber entschieden, ob ein (weiterer) Regressor aus dem Block aufgenommen werden soll.
- **Rückwärts**
  Zunächst wird der gesamte Block aufgenommen. Dann wird Schritt für Schritt geprüft, ob ein Regressor entfernt werden soll.
- **Schrittweise**
  Ausgehend vom Modell *ohne* den fraglichen Block wird Schritt für Schritt darüber entschieden, ob ein (weiterer) Regressor aus dem Block aufgenommen werden soll. Außerdem wird in jedem Schritt geprüft, ob ein vorhandener Regressor aus dem aktuellen Block mittlerweile überflüssig geworden ist.

In der **Optionen**-Subdialogbox kann man die Kriterien für die Aufnahme bzw. für den Ausschluss eines Regressors festlegen:



Gegen eine explorative *Modellsuche* ist nichts einzuwenden, solange der gravierende Unterschied zu einer *Modellprüfung* beachtet wird. Bietet man z. B. der schrittweise Modellsuche 100 Prädiktoren an, die nichts als puren Zufall enthalten, so kann man beim Einschlusskriterium $p = 0{,}05$ ein Modell mit ca. fünf „signifikanten“ Regressionskoeffizienten erwarten, z. B.:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.
1	(Konstante)	,045	,086		,526	,600
	zuf82	,422	,101	,386	4,182	,000
2	(Konstante)	,089	,084		1,056	,293
	zuf82	,395	,098	,361	4,033	,000
	zuf87	,216	,075	,258	2,885	,005
3	(Konstante)	,114	,083		1,381	,170
	zuf82	,403	,095	,369	4,233	,000
	zuf87	,231	,073	,276	3,161	,002
	zuf03	-,199	,078	-,224	-2,571	,012
4	(Konstante)	,120	,081		1,479	,142
	zuf82	,389	,094	,355	4,151	,000
	zuf87	,233	,072	,278	3,249	,002
	zuf03	-,210	,076	-,235	-2,753	,007
	zuf55	,176	,079	,191	2,239	,027
5	(Konstante)	,106	,080		1,332	,186
	zuf82	,426	,093	,389	4,565	,000
	zuf87	,228	,070	,272	3,243	,002
	zuf03	-,187	,075	-,210	-2,490	,014
	zuf55	,173	,077	,187	2,243	,027
	zuf34	,173	,077	,191	2,239	,027

a. Abhängige Variable: av

Hier findet offenbar eine Kapitalisierung von Zufall statt. Die allzu häufige Verwertung solcher „Forschungsergebnisse“ in der Literatur mag ein Grund für die neuerdings oft diskutierte Replikationskrise in vielen Wissenschaften sein. Für die Psychologie wurden in Studien zur Glaubwürdigkeit von berichteten Forschungsergebnissen bescheidene Replikationsraten zwischen 25% und 50% festgestellt.[1]

Während die automatische Modellsuche einerseits bei Verwendung von Zufallsdaten mit mathematischer Notwendigkeit zu Scheineffekten führt, ist sie andererseits aufgrund ihrer Orientierung an der Signifikanzbeurteilung zu *einzelnen* Regressoren blind für *Blöcke* von Regressoren, die nur gemeinsam stark sind. Zur Demonstration verwenden wir ein künstliches Beispiel mit Suppressoreffekt (vgl. Abschnitt 2.7.4), wobei die schrittweise Prozedur mit den Standardeinstellungen *keinen* Prädiktor aufnimmt:

Warnungen

Für die Gleichung wurden keine Variablen angegeben.

Bei der explizit angeordneten gemeinsamen Aufnahme erhalten Regressor und Suppressor hingegen ein signifikantes Gewicht:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			Korrelationen		
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Nullter Ordnung	Partiell	Teil
1	(Konstante)	-,044	,049		-,894	,373			
	x	,171	,039	,732	4,378	,000	,175	,406	,405
	sup	-,167	,042	-,669	-4,002	,000	-,060	-,376	-,371

a. Abhängige Variable: y

Der beschriebene Fehler wäre bei einer Rückwärts-Modellsuche *nicht* aufgetreten. Wer der Automatik eine Chance gibt, sollte sie also vorsichtshalber vorwärts und rückwärts suchen lassen.

[1] https://en.wikipedia.org/wiki/Replication_crisis

Man muss nicht unbedingt generell von der schrittweisen Regression abraten, was z. B. Urban & Mayerl (2018, S. 104ff) tun. Jedoch sollten die Vorschläge von sequentiellen Suchtechniken nur als zusätzliche Informationsquelle betrachtet werden, wobei die Modellsuche letztlich primär theoriegeleitet unter der Verantwortung des Forschers zu erfolgen hat (siehe Abschnitt 5.3).

Wenn trotz der Vorbehalte versucht werden soll, aus einer Menge von Regressoren über eine automatische Suche das optimale Modell (die beste Teilmenge von Regressoren) zu finden, dann empfiehlt sich eine Modellauswahl über ein modifiziertes Bootstrap-Verfahren. Statt einmalig sequentiell nach einem guten Modell zu suchen, wird die Suche für zahlreiche Sekundärstichproben durchgeführt. Abweichend vom üblichen Bootstrap-Verfahren müssen die Sekundärstichproben deutlich *kleiner* als die Primärstichprobe gewählt werden, um eine konsistente Modellauswahl zu erhalten, d.h. eine mit zunehmender Stichprobengröße gegen 1 strebende Wahrscheinlichkeit für die Wahl des korrekten Modells (siehe z.B. Shao 1996).

5.3 Empfehlungen zur Modellbildung

Eine automatische Modellsuche kann zwar als ergänzende, explorative Technik in bestimmten Situationen sinnvoll sein, doch sollte die Modellkonstruktion in der Regel durch einen kreativen und rationalen Forscher unter Verwendung von inhaltlichen und statistischen Informationen vorgenommen werden. Dazu einige Empfehlungen:

- Man startet mit *allen* Regressoren, die nach inhaltlichen Überlegungen relevant sind. Dazu gehören auch Kontrollvariablen, die potentiell Scheineffekte entlarven können. Es gilt, den *omitted variable error* zu vermeiden (vgl. Abschnitt 6.1).
- Bei einer explorativen Forschungsstrategie werden sukzessive irrelevante (insignifikante) Variablen aus dem Modell entfernt, sodass sich eine Top-Down - Modellierungsstrategie ergibt.
- Bei einer konfirmatorischen Forschungsstrategie sollten empirisch irrelevante Regressoren im Modell verbleiben, wenn diese aufgrund überzeugender inhaltlicher Überlegungen aufgenommen worden sind (siehe dazu Abschnitt 6.1). Das gilt auch für die Kontrollvariablen.
- Sind im Modell Wechselwirkungen vorgesehen (vertreten durch das Produkt der beteiligten Regressoren, siehe Baltes-Götz 2020a), dann werden diese zuerst geprüft.
- Statistisch irrelevante Wechselwirkungen (Produktterme) sollten aus dem Modell entfernt werden, weil sie die Interpretation der übrigen Regressionskoeffizienten erschweren.
- Bei einer signifikanten Wechselwirkung müssen alle beteiligten Regressoren in das Modell aufgenommen werden. Wer komplexe Überlegungen vermeiden will, sollte generell mit *hierarchisch wohlgeformten Modellen* arbeiten, die zu jedem Interaktionsterm auch alle Komponenten von niedrigerer Ordnung enthalten (Cohen et al. 2003, S. 284).

6 Spezielle Themen

In diesem Abschnitt haben recht unterschiedliche Themen einen vorläufigen Platz gefunden.

6.1 Kausale Interpretation von Regressionskoeffizienten

Zur methodologisch anspruchsvollen Theorie der kausalen Regressionsmodelle, die z. B. von Steyer (1992) ausführlich behandelt wird, sind in diesem Manuskript nur einige kurze Anmerkungen möglich.

Falls eine *kausale* Interpretierbarkeit der Regressionskoeffizienten methodologisch untermauert werden soll, dann sind bei der Untersuchungsplanung und -auswertung über die Liste mit statistischen bzw. technischen Voraussetzungen (siehe Abschnitte 1.2.2 bzw. 2.2) hinaus zusätzliche Bedingungen zu erfüllen (siehe z.B. Hayes 2018, S. 15f):

- Es muss sichergestellt sein, dass die behaupteten **Ursachen den Effekten zeitlich vorgeordnet** sind.
- Sprechen die zeitliche Anordnung sowie statistische Schätz- und Testergebnisse für ein Kausalmodell, dann müssen noch **Alternativerklärungen ausgeschlossen** werden. Dies gelingt besonders überzeugend durch experimentelle Manipulation und randomisierte Gruppeneinteilung. In Beobachtungsstudien muss man sich auf eine **statistische Kontrolle** von Kausalitätskonkurrenten beschränken.

6.1.1 Verzerrte Effektschätzungen durch vergessene relevante Regressoren

Um in der multiplen Regressionsanalyse eine statistische Kontrolle von Kausalitätskonkurrenten zu ermöglichen, dürfen keine relevanten Regressoren weggelassen werden, die mit im Modell vorhandenen Regressoren korrelieren. Im folgenden Modell übt die Variable *Z* einen Effekt auf die Variablen *X* und *Y* aus, während *X* ohne Wirkung auf *Y* bleibt:

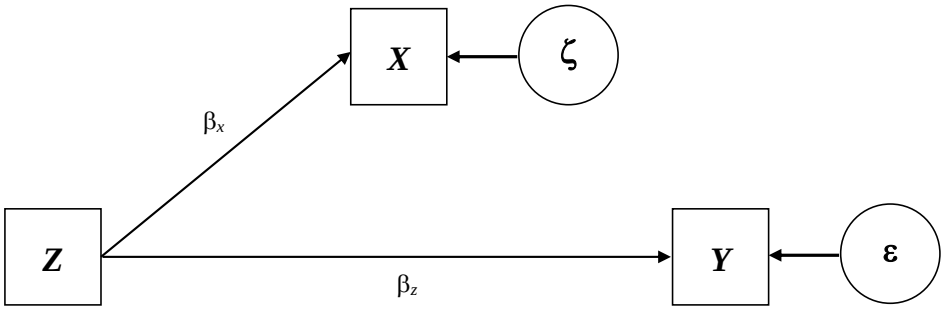


Abbildung 46: Wirkungsweise des wahren Regressors *Z*

Der Einfachheit halber soll für alle Variablen ein Mittelwert von 0 und eine Varianz von 1 angenommen werden. In diesem Fall sind der Regressionskoeffizient β_x sowie der „inverse“ Koeffizient aus der Regression von *Z* auf *X* identisch mit der Korrelation von *X* und *Z*.

Bei der bivariaten Regression von *Y* auf *X* tritt ein „Scheineffekt“ von *X* auf:

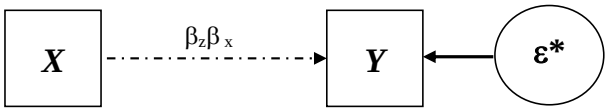


Abbildung 47: In diesem Regressionsmodell fehlt der eigentlich wirksame Regressor

Aus

$$Y = \beta_z Z + \varepsilon$$

folgt wegen

$$Z = \beta_x X + \delta$$

für die Regression von *Y* auf *X*:

$$\begin{aligned} Y &= \beta_z(\beta_x X + \delta) + \varepsilon \\ &= \beta_z\beta_x X + \beta_z\delta + \varepsilon \\ &= \beta_z\beta_x X + \varepsilon^* \quad \text{mit} \quad \varepsilon^* := \beta_z\delta + \varepsilon \end{aligned}$$

Generell kann bei der Regressionsanalyse ein Prädiktor *X* ein von 0 verschiedenes (und statistisch signifikantes) Regressionsgewicht auch *ohne* jeden kausalen Effekt erzielen, wenn er mit einer im Regressionsmodell *fehlenden* Variablen *Z* korreliert ist, die einen kausalen Effekt auf das Kriterium ausübt (*omitted variable error*, siehe z. B. Baltes-Götz 1994, S. 1-4).

Ein Beispiel mit der eben beschriebenen Struktur ist weithin bekannt:

- Beobachtungseinheiten: schwedische Landkreise
- *X*: Anzahl der Störche
- *Z*: Industrialisierungsgrad
- *Y*: Geburtenrate (der menschlichen Bewohner)

6.1.2 Auswirkungen überflüssiger Regressoren auf die Effektbeurteilungen

Im letzten Abschnitt haben wir uns damit beschäftigt, wie sich das Fehlen relevanter Regressoren auf die Schätzungen der Modellparameter auswirken kann. Nun geht es darum, wie sich im Modell vorhandene, aber vermutlich überflüssige Regressoren auswirken.

Bei der Erstellung eines Ergebnisberichts stellt sich oft die Frage, wie mit statistisch irrelevanten (insignifikanten) Regressoren verfahren werden soll. Wenn statistisch irrelevante Regressoren nur geringfügig mit relevanten Regressoren korrelieren, führt ihre Anwesenheit lediglich zu einer reduzierten Präzision beim Schätzen und Testen durch den Verlust von Freiheitsgraden. In großen Stichproben ist dieser Effekt vernachlässigbar. Insignifikante Regressoren (Kovariaten, Kontrollvariablen) von theoretischem Interesse sollten nach Möglichkeit im Modell bleiben, um zu demonstrieren, dass ihr potentieller Effekt statistisch kontrolliert worden ist.

Wenn irrelevante Regressoren erheblich mit relevanten Regressoren korrelieren, kommt es zu einer spürbaren Belastung der Inferenzstatistik (große Vertrauensintervalle, reduzierte Power bei Hypothesentests). Darlington & Hayes (2017, S. 121) plädieren jedoch dafür, auch Kontrollvariablen mit starken Beziehungen zu primär interessierenden Regressoren im Modell zu belassen.

Wenn allzu viele irrelevante Regressoren in ein Modell einbezogen werden, können einige zu unverdientem Ruhm gelangen, weil ihr Regressionskoeffizient durch zufällige Stichprobenschwankungen signifikant wird (erhöhte α-Fehlerrate, siehe Abschnitt 5.2).

6.2 Mediation

In dem durch das folgende Pfaddiagramm beschriebenen Modell wirkt sich das Weglassen der Variablen *Z* trotz ihres Effekts auf das Kriterium und ihrer Korrelation mit dem Regressor *X nicht* kritisch aus:

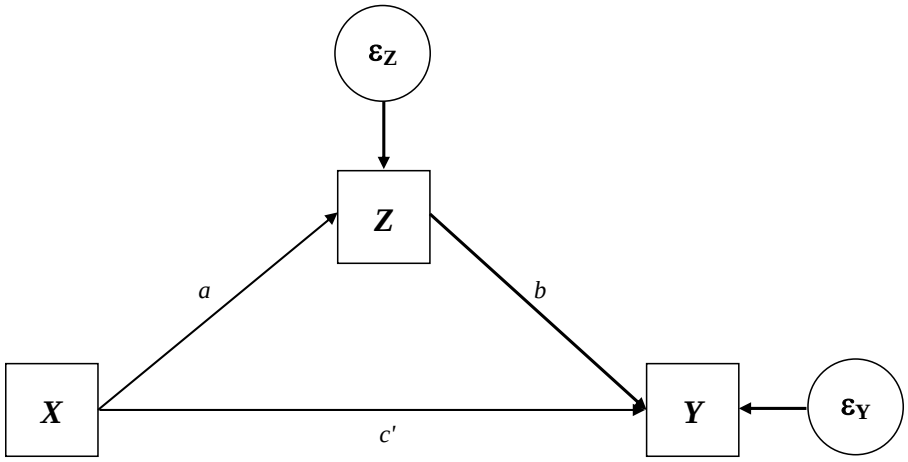


Abbildung 48: Regressionsmodell mit Mediator

Dieses Modell besteht aus *zwei* Regressionsgleichungen:

- Regression von *Y* auf *X* und *Z*
- Regression von *Z* auf *X*

Z vermittelt den Effekt von *X* auf *Y* partiell.

Wird der Mediator *Z* weggelassen und nur die Regression von *Y* auf *X* betrachtet,

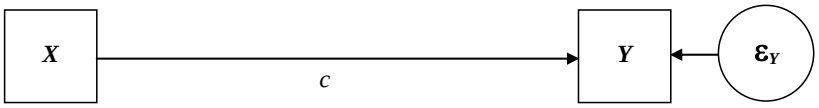


Abbildung 49: Totaler Effekt von *X* auf *Y*

dann kann man zwar weniger gut verstehen, *wie X* auf *Y* einwirkt, doch wird der *totale* Effekt von *X* auf *Y* korrekt geschätzt:

$$c = c' + a \cdot b$$

Zeigt sich in einer multiplen Regression von *Y* auf *X* und *Z* kein *direkter* Effekt von *X* (Koeffizient *c*' nicht signifikant von 0 verschieden), dann darf nicht voreilig geschlossen werden, dass *X* keinen Effekt auf *Y* habe. Eine sorgfältige theoretische Analyse und die Betrachtung der bivariaten Regression von *Y* auf *X* helfen, solche Fehlschlüsse zu vermeiden.

Der direkte und der indirekte Effekt können durchaus *unterschiedliche* Vorzeichen haben und sich im Extremfall sogar gegenseitig zu einem totalen Effekt von 0 aufheben. Das folgende Modell wurde schon im Abschnitt 2.7.3 im Zusammenhang mit Phänomenen der multiplen Regression beschrieben:

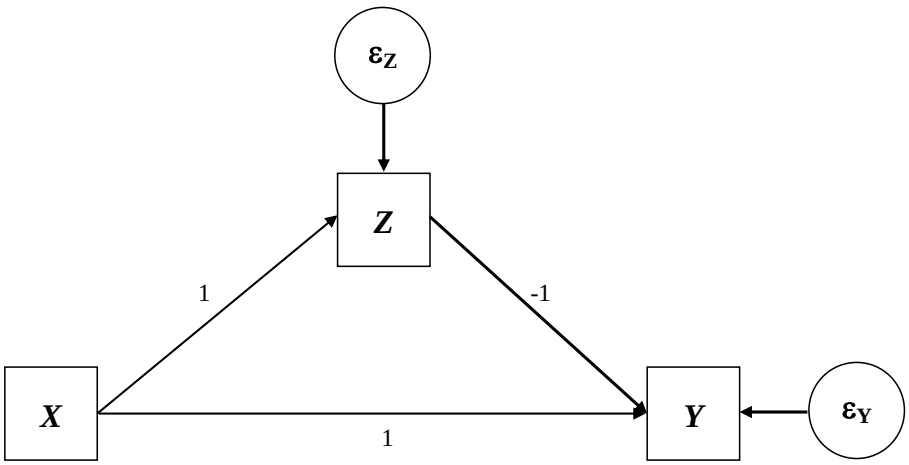


Abbildung 50: Der direkte und der indirekte Effekt heben sich auf

Das Beispiel zeigt, in welche Richtung wir uns nach der gründlichen Behandlung der multiplen Regression methodisch weiterentwickeln sollten: Es wird Zeit, simultan *mehrere* Regressionsgleichungen zu betrachten, um möglichst viele empirische Systeme angemessen modellieren zu können. Weitere Informationen zur Mediationsanalyse finden sich z. B. in Baltes-Götz (2020a).

6.3 Strukturgleichungsmodelle

Falls bei stochastischen Regressoren (vgl. Abschnitt 1.2.3) die Annahme fehlerfrei gemessener *X*-Werte nicht haltbar ist, dann liefert die Regressionsanalyse verzerrte Schätzergebnisse (vor allem geminderte Regressionskoeffizienten). Eine mögliche Lösung stellen dann **Strukturgleichungsmodelle** mit **latenten Variablen** dar, die von Programmen wie z. B. **AMOS**, **LISREL** und **MPlus** oder durch das **R**-Paket **lavaan** unterstützt werden.

Neben der Lösung des Messfehlerproblems durch latente Variablen haben Strukturgleichungsmodelle noch einen weiteren Vorteil gegenüber Regressionsmodellen. Sie erlauben die simultane Betrachtung *mehrerer* Gleichungen, wobei auch **nonrekursive Gleichungssysteme** zugelassen sind, mit denen etwa die wechselseitige Beeinflussung zweier Variablen modelliert werden kann, z. B.:

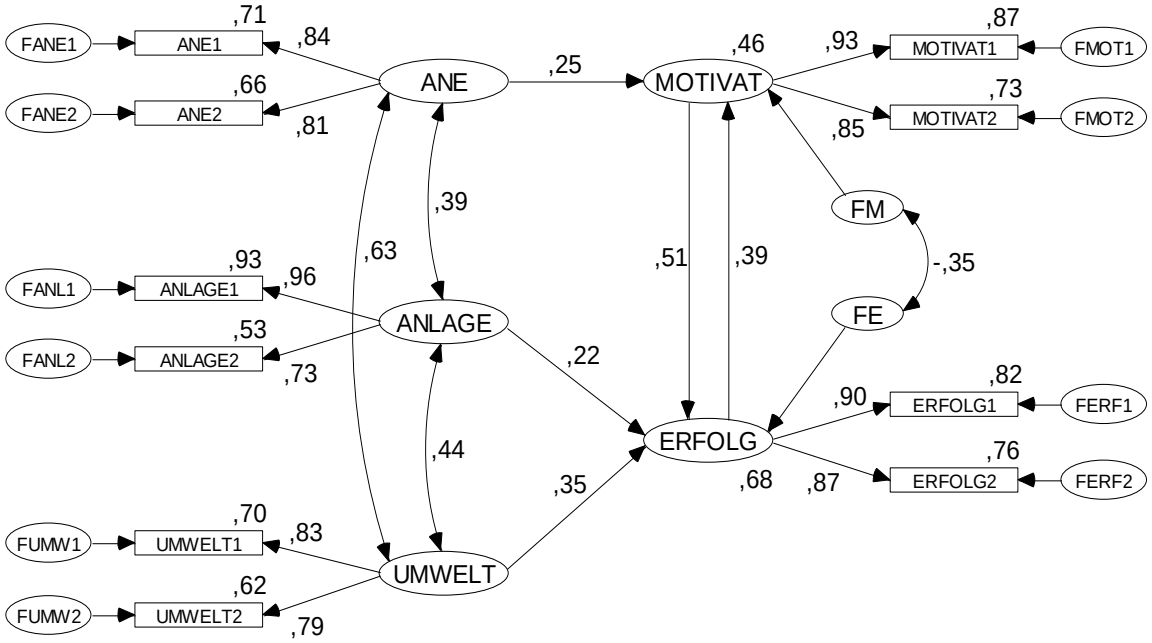


Abbildung 51: Strukturgleichungsmodell mit latenten Variablen und wechselseitiger Abhängigkeit

Man kann selbstverständlich mehrere (unabhängig voneinander geschätzte) Regressionsgleichungen zu einem sogenannten **Pfadmodell** kombinieren (siehe Abschnitt 6.2), doch lassen sich auf diese Weise nur *rekursive* Modelle bilden, die z. B. keine wechselseitigen Abhängigkeiten enthalten können. Vermutlich liegen in vielen von der Sozialwissenschaft zu untersuchenden empirischen Systemen wechselseitige Abhängigkeiten vor, während derzeit noch überwiegend rekursive (unidirektionale) Modelle eingesetzt werden.

Über das Arbeiten mit dem von IBM/SPSS vertriebenen Strukturgleichungsprogramm AMOS informiert ein ZIMK-Manuskript (Baltes-Götz 2015), das auf dem Webserver der Universität Trier von der Startseite (http://www.uni-trier.de/) ausgehend folgendermaßen zu finden:

IT-Services > Downloads & Broschüren > Downloads
Analyse von Strukturgleichungsmodellen mit Amos

6.4 Vergleich der Determinationskoeffizienten von nicht geschachtelten Modellen

Gelegentlich sollen zwei alternative, nicht geschachtelte Regressorensätze hinsichtlich ihrer Erklärungsleistung für *dasselbe* Kriterium verglichen werden, z. B. bei der Frage, ob die berufliche Bewährung von Bewerbern besser durch objektive Leistungstests oder durch Expertenurteile vorhergesagt werden kann. Für den Vergleich von zwei aus derselben Stichprobe ermittelten Determinationskoeffizienten, die jeweils identisch sind mit der quadrierten Korrelation zwischen dem Kriterium und der Modellprognose (vgl. Abschnitt 2.6), schlagen Tabachnick & Fidell (2013, S. 153f) den von Steiger (1980) beschriebenen Test zum Vergleich von abhängigen Korrelationen (r_{YX} versus r_{YZ}) vor. Die fehlende Berücksichtigung der beiden Prädiktorenzahlen wiegt nicht allzu schwer, weil der Test ohnehin nur in großen Stichproben gültig ist.

7 Regressionsmodelle für Zeitreihendaten

Die Inferenzstatistik in der üblichen OLS (*Ordinary Least Squares*) - Regressionsanalyse setzt u.a. voraus, dass die Modellresiduen unkorreliert sind. Stammt jede Y_i-Beobachtung von einem eigenständigen Merkmalsträger wie in unserem Beispiel mit der Regression von GEWICHT auf GROESSE (siehe Abschnitt 1.1), wobei keinerlei Beziehung zwischen den Merkmalsträgern besteht, dann kann die Unabhängigkeit der Residuen als gesichert gelten. Gelegentlich möchte man jedoch z. B. regressive Beziehungen anhand *zeitlich geordneter* Beobachtungen an einem *einzigen* Merkmalsträger untersuchen. Benachbarte Y_i-Variablen können dann modellfremde Einflüsse gemeinsam haben, sodass ihre Residualvariablen ε_i korrelieren. Das wirkt sich folgendermaßen auf die OLS-Schätz- und -Testergebnisse aus:

- Die Schätzungen für die Regressionskoeffizienten bleiben unverzerrt.
- Allerdings verlieren die Schätzungen ihre Effizienz, die von der Varianzhomogenität und der Unabhängigkeit der Residuen garantiert wird (vgl. Abschnitt 1.3). Seriell abhängige Residuen führen also potentiell zu *ungenau* geschätzten Regressionskoeffizienten.
- Im typischen Fall einer positiven Autokorrelation der Residuen erhält man *unterschätzte* Standardfehler zu den Regressionskoeffizienten und damit eine erhöhte Fehlerrate erster Art (zu liberale Tests)

Eine serielle Korrelation der *Kriteriumswerte* stellt noch keine Verletzung der OLS-Voraussetzungen dar, weil sich die Unkorreliertheitsforderung auf die *Residuen* bezieht. Grundsätzlich lässt sich eine serielle Korrelation der Residuen trotz Zeitreihendesign durch ein Modell mit geeigneten Regressoren vermeiden.

7.1 Beispiel

Als Beispiel für die Regressionsanalyse mit Zeitreihendaten betrachten wir einen von Durbin & Watson (1951) berichteten Datensatz, der drei logarithmisch transformierte Zeitreihen mit jährlich in England vorgenommenen Messungen aus dem Beobachtungszeitraum von 1870 bis 1938 enthält:[1]

- Alkoholverbrauch (Variablenname CONSUMP)
- Pro-Kopf-Einkommen (Variablenname INCOME)
- Inflationsbereinigter Preisindex (Variablenname PRICE)

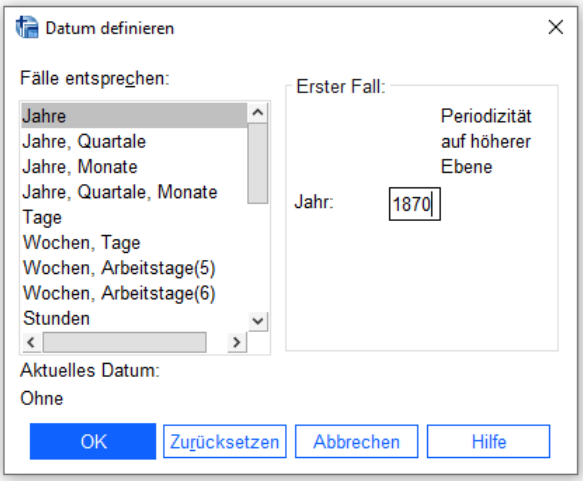
Es soll versucht werden, den logarithmierten Alkoholverbrauch durch ein lineares Regressionsmodell mit dem logarithmierten Einkommen und dem logarithmierten Preis als unabhängigen Variablen zu erklären. Dabei ist zu befürchten, dass zahlreiche im Modell nicht berücksichtigte Einflüsse auf den Alkoholverbrauch in benachbarten Jahren relativ ähnlich ausgeprägt waren, sodass zwischen zeitlich benachbarten Residuen Korrelationen bestehen, die im OLS-Regressionsmodell verboten sind.

Zunächst wollen wir uns einen optischen Eindruck vom Verlauf der drei Zeitreihen verschaffen. Dazu öffnen wir die SAV-Datei mit den Beispieldaten und vereinbaren nach dem Menübefehl

Daten > Datum und Uhrzeit definieren

in der folgenden Dialogbox eine neue Variable YEAR_ mit den Jahreszahlen ab 1870:

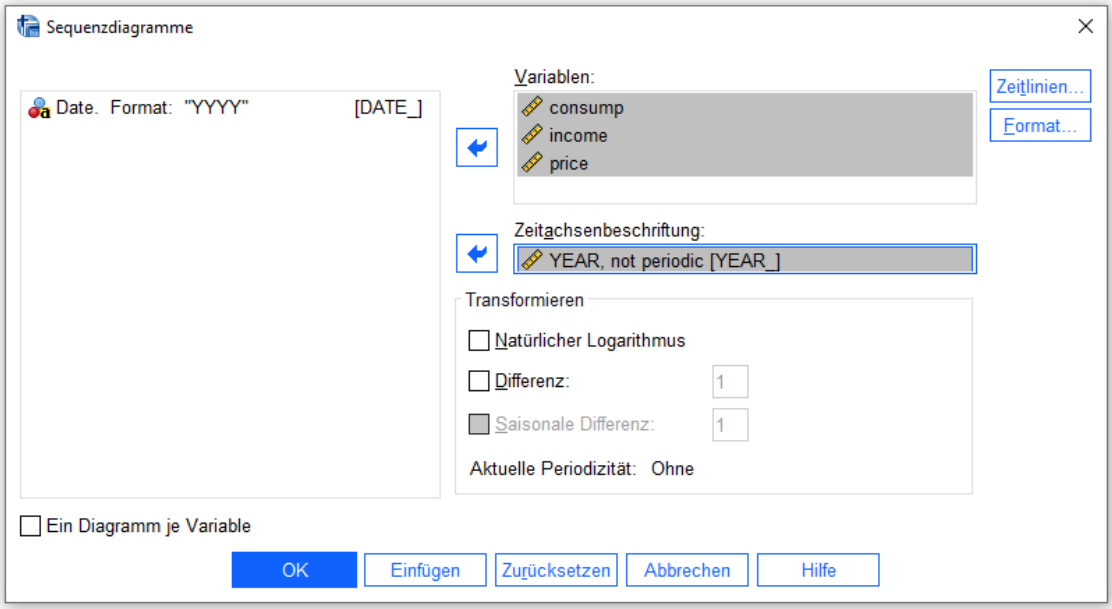
[1] Sie finden die Daten an der im Vorwort genannten Stelle in der Datei **DW.sav**.



Über den Menübefehl

Analysieren > Vorhersage > Sequenzdiagramme

fordern wir im folgenden Dialog



ein Verlaufsdiagramm mit den drei Variablen an:

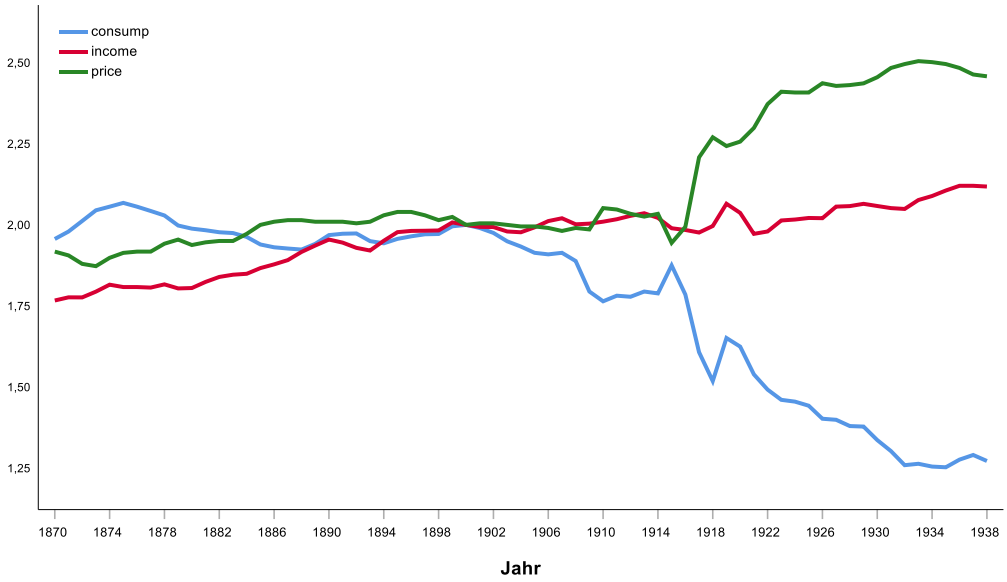


Abbildung 52: Verlauf der Zeitreihen CONSUMP, INCOME und PRICE aus dem Datensatz von Durbin & Watson (1951)

Im Abschnitt 7 geht es darum,

- bei der Analyse solcher Zeitreihen die kritische Voraussetzung unkorrelierter Residuen zu prüfen
- und gegebenenfalls die OLS-Regression durch ein adäquates Verfahren zu ersetzten.

7.2 Autokorrelation der Residuen

Der im Abschnitt 7.3 vorzustellende Durbin-Watson - Test beurteilt die einfachen (nicht-standardisierten) Residuen des Modells. Diese bewegen sich bei erfüllter Unabhängigkeitsannahme im Sequenzdiagramm zufällig (als weißes Rauschen) um die Null-Lage, wie es im folgenden Beispiel mit künstlichen Daten (ohne serielle Abhängigkeit) zu sehen ist:

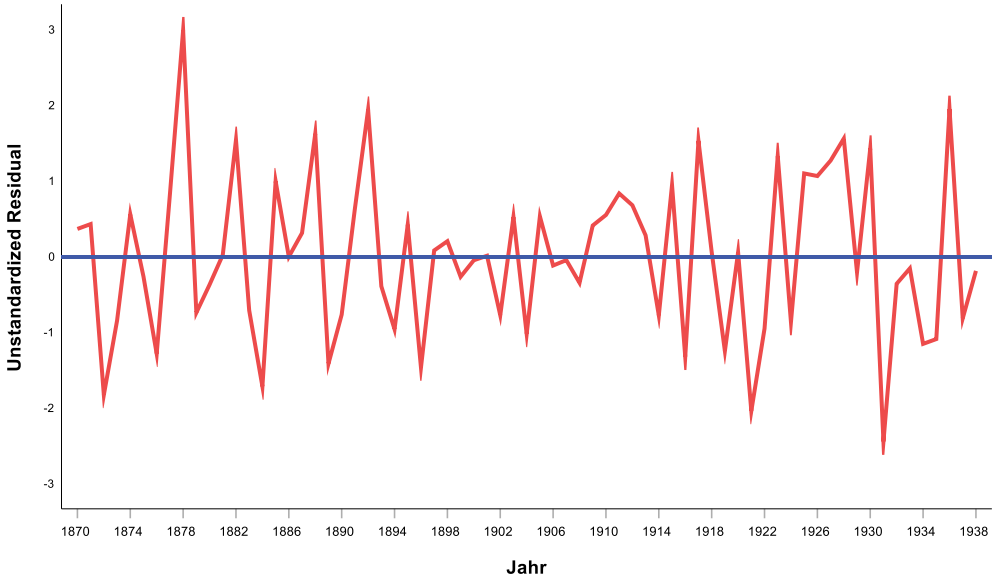


Abbildung 53: Sequenzdiagramm mit OLS-Residuen bei *gültiger* Unabhängigkeitsannahme

Bei den Residuen zu der OLS-Regressionsgleichung für die Durbin-Watson - Daten zeigt sich ein deutlich anderes Bild:

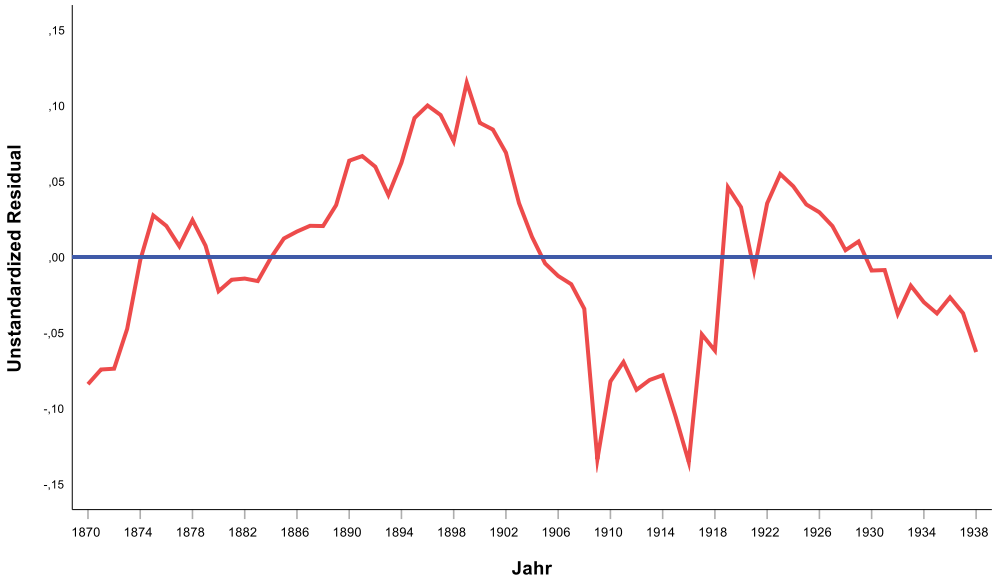


Abbildung 54: Sequenzdiagramm für die OLS-Residuen aus dem Durbin-Watson - Beispiel

Diese Residualzeitreihe schlängelt sich gemächlich um die Null-Lage, d.h. jeder Wert liegt relativ nahe bei seinem Vorgänger. Er kann folglich durch seinen Vorgänger gut vorhergesagt werden, und wir erwarten einen hohen positiven Wert für die sogenannte **Autokorrelation erster Ordnung** zwischen der Residualzeitreihe und ihrer um einen Zeittakt verschobenen Variante. Als Bezeichnung verwenden wir ein griechisches ρ (Rho). Aus den Stichprobendaten wird die erste Autokorrelation ρ der Residuen folgendermaßen geschätzt:

$$\hat{\rho} = \frac{\sum_{t=2}^{N} e_t e_{t-1}}{\sum_{t=1}^{N} e_t^2} \qquad (15)$$

Bei der Analyse von Zeitreihendaten ist häufig eine positive Residual-Autokorrelation erster Ordnung zu beobachten. Negative Residual-Autokorrelationen erster Ordnung sind eher selten.

Besteht eine positive Residual-Autokorrelation, dann bleiben die OLS-Schätzer der Regressionskoeffizienten erwartungstreu, doch werden ihre Standardfehler unterschätzt. Infolgedessen sind die Signifikanztests zu liberal und die Vertrauensintervalle zu klein.

7.3 Durbin-Watson-Test

Bei einer Regressionsanalyse mit zeitlich (oder auch räumlich) geordneten Daten ist damit zu rechnen, dass benachbarte Werte ähnlichen modellfremden Einflüssen ausgesetzt waren. Es resultiert eine positive Autokorrelation erster Ordnung, die routinemäßig überprüft werden sollte. Da wir die *generelle* Unkorreliertheit der Residuen voraussetzen müssen, ist das Fehlen einer Autokorrelation erster Ordnung nur eine notwendige, jedoch keine hinreichende Anwendungsvoraussetzung.

Weil in der Regel eine *positive* Residual-Autokorrelation zu erwarten ist, sollte der Signifikanztest *einseitig* durchgeführt werden:

H_0: $\rho \leq 0$ (Die Residual-Autokorrelation erster Ordnung ist *nicht* positiv.)

versus

H_1: $\rho > 0$ (Die Residual-Autokorrelation erster Ordnung ist positiv.)

Ein für dieses Testproblem geeignetes Verfahren wurde von Durbin &Watson (1951) entwickelt. Die DW-Prüfstatistik für eine Stichprobe der Größe *N* ist folgendermaßen definiert (vgl. Hartung 1989):

$$DW := \frac{\sum_{t=2}^{N}(e_t - e_{t-1})^2}{\sum_{t=1}^{N} e_t^2}$$

Durch Umformungen des DW-Koeffizienten wird seine enge Beziehung zur Stichproben-Autokorrelation erster Ordnung deutlich (vgl. Formel (15)):

$$\frac{\sum_{t=2}^{N}(e_t - e_{t-1})^2}{\sum_{i=1}^{N} e_t^2} = \frac{\sum_{t=2}^{N} e_t^2 - 2\sum_{t=2}^{N} e_t e_{t-1} + \sum_{t=2}^{N} e_{t-1}^2}{\sum_{t=1}^{N} e_t^2}$$

$$= \frac{2\sum_{t=2}^{N-1} e_t^2 + e_1^2 + e_N^2}{\sum_{t=1}^{N} e_t^2} - \frac{2\sum_{t=2}^{N} e_t e_{t-1}}{\sum_{t=1}^{N} e_t^2}$$

$$= 2(1-\hat{\rho}) - \frac{e_1^2 + e_N^2}{\sum_{t=1}^{N} e_t^2}$$

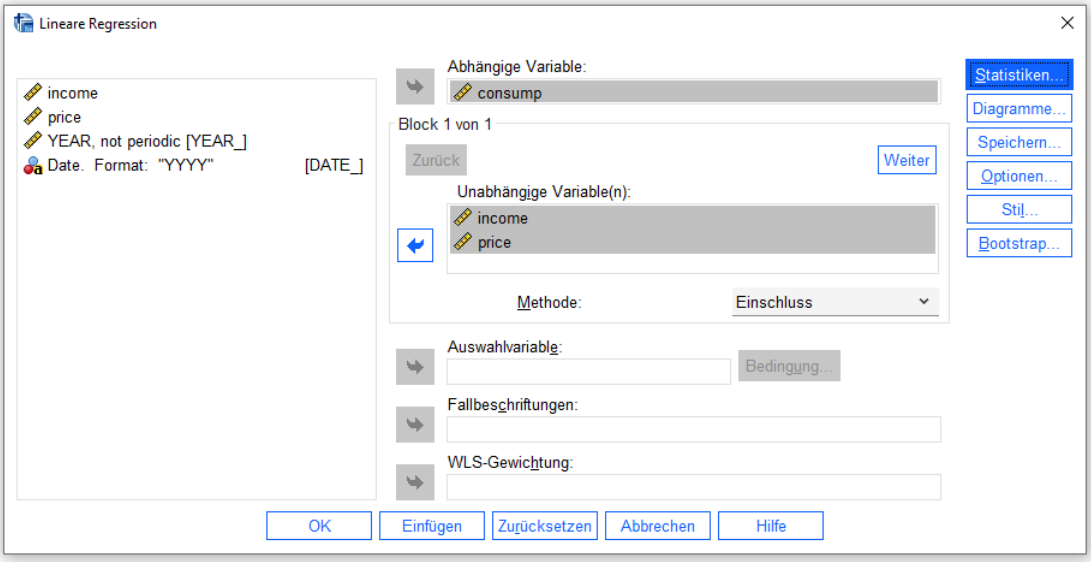
Für großes *N* erhalten wir approximativ:

$$2(1-\hat{\rho})$$

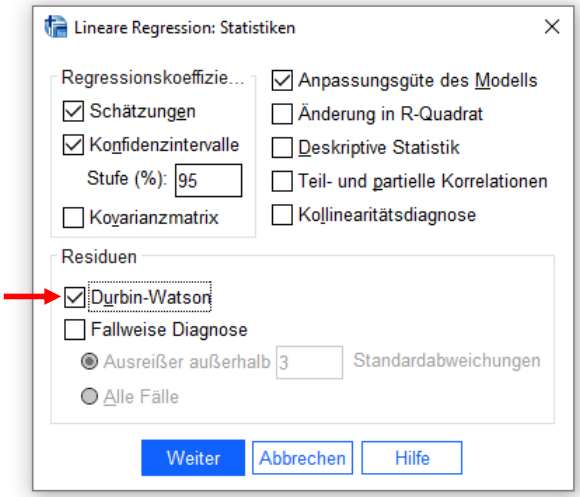
Damit ist klar, dass der Durbin-Watson-Koeffizient zwischen 0 und 4 variiert:

- Wenn $\hat{\rho}$ gegen 1 strebt, dann tendiert DW gegen 0.
- Wenn $\hat{\rho}$ gegen -1 strebt, dann tendiert DW gegen 4.
- Wenn $\hat{\rho}$ gegen 0 strebt, dann tendiert DW gegen 2.

Um in SPSS für die Regression von CONSUMP auf INCOME und PRICE



den **Durbin-Watson** - Test anzufordern, besuchen wir die Dialogbox **Lineare Regression: Statistiken**:



Das Ergebnis spricht klar für eine positive Autokorrelation erster Ordnung:

Modellzusammenfassung[b]

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik
1	,978 [a]	,956	,954	,0578598	,249

a. Einflußvariablen : (Konstante), price, income

b. Abhängige Variable: consump

Zur Durchführung des Durbin-Watson-*Tests* müssen Tabellen herangezogen werden. Im Handbuch zum Zeitreihenmodul der SPSS-Version 13 (SPSS 2004, Anhang A) sind Tabellen nach Savin & White (1977) enthalten, die in SPSS-Handbüchern zu späteren Versionen leider fehlen. Erfreulicherweise bieten einige Web-Seiten einen guten Ersatz, z. B.:

- https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/regression/supporting-topics/model-assumptions/test-for-autocorrelation-by-using-the-durbin-watson-statistic/
  Hier werden Stichprobenumfänge von 6 bis 200 unterstützt.
- https://www.real-statistics.com/statistics-tables/durbin-watson-table/
  Hier werden Stichprobenumfänge von 6 bis 2000 unterstützt, ab 40 mit wachsenden Lücken.

Leider kann beim Durbin-Watson - Test kein kritischer Wert bestimmt werden, der durch Vergleich mit einer empirisch ermittelten Prüfgröße zur Testentscheidung führt. Es ist lediglich möglich, eine untere (dL) und eine obere Schranke (dU) für den kritischen Wert in Abhängigkeit von folgenden Bedingungen zu ermitteln:

- inhomogene versus homogene Regression
- Stichprobengröße N bzw. T
- Prädiktorenzahl k
- gewünschtes Signifikanzniveau α

Dann sind beim einseitigen Testproblem

$$H_0: \rho \leq 0 \qquad \text{versus} \qquad H_1: \rho > 0$$

folgende Entscheidungsregeln anzuwenden:

DW-Wert					Testentscheidung
		DW	≤	*dL*	H_0 ablehnen, d.h. positive Autokorrelation
dL	<	*DW*	<	*dU*	keine Entscheidung möglich
dU	≤	*DW*			H_0 beibehalten

Auf der oben angegebenen Webseite der Software-Firma Minitab ermitteln wir für unsere Stichprobengröße ($N = 69$) sowie 2 Regressoren und $\alpha = 0{,}05$ die folgenden Grenzwerte:

$$dL = 1{,}551 \quad \text{und} \quad dU = 1{,}670$$

Aufgrund unseres empirischen *DW*-Werts von 0,249 wird die Nullhypothese deutlich verworfen. Wir gehen also davon aus, dass bei den Modellresiduen der OLS-Regression eine positive Autokorrelation erster Ordnung vorliegt, sodass die Voraussetzungen der OLS-Regression verletzt sind. In dieser Situation sind u.a. die Signifikanzbeurteilungen zu den Regressionskoeffizienten verfälscht.

Beim alternativen einseitigen Testproblem:

$$H_0: \rho \geq 0 \qquad \text{versus} \qquad H_1: \rho < 0$$

sowie beim zweiseitigen Testproblem:

$$H_0: \rho = 0 \qquad \text{versus} \qquad H_1: \rho \neq 0$$

gelten andere Entscheidungsregeln (siehe Hartung 1989, S. 740f).

Der Durbin-Watson-Test ist *nicht* anwendbar, wenn zeitversetzte Varianten der abhängigen Variablen als Regressoren verwendet werden (vgl. SPSS 2004, S. 177).

Im Abschnitt 7.4.3 wird mit dem Box-Ljung - Test eine Alternative zum Durbin-Watson - Test vorgestellt, wobei zur Durchführung mit SPSS keine externe Tabelle benötigt wird.

7.4 Regression mit integriertem AR(1) - Modell für die Residuen

7.4.1 Modell

Um eine Autokorrelation erster Ordnung in den Residuen explizit zu berücksichtigen, formulieren wir für die abhängige Variable Y_t zum Zeitpunkt t das folgende Modell:

$$Y_t = \beta_0 + \beta_1 X_{1_t} + \beta_2 X_{2_t} + ... + \beta_k X_{k_t} + U_t \tag{16}$$

$$U_t = \rho U_{t-1} + \varepsilon_t, \quad |\rho| < 1 \tag{17}$$

Es besteht aus einem normalen multiplen Regressionsmodell (16), das durch einen AR(1) - Fehlerprozess (17) ergänzt wird. Im Einzelnen bedeuten:

- X_{j_t} Der j-te Regressor zum Zeitpunkt t
- U_t Der Anteil von Y_t, der durch die Regressoren nicht erklärt werden kann.
- ρ Der autoregressive Parameter erster Ordnung zum Fehlerprozess. Er ist mit der ersten Autokorrelation der (U_t) - Zeitreihe identisch.
- ε_t Für die Epsilon-Residuen in der Zeitreihe (ε_t) wird angenommen, dass sie ein sogenanntes **weißes Rauschen** bilden, d.h.:
  - Die Epsilon-Residuen sind unkorreliert
  - und besitzen eine identische Normalverteilung mit dem Erwartungswert 0.

Für den AR(1) - Prozess (U_t) muss außerdem die **Stationarität** angenommen werden, was auf die folgende Forderung hinausläuft (vgl. Schlittgen & Streitberg 1989, S. 100):

$$|\rho| < 1$$

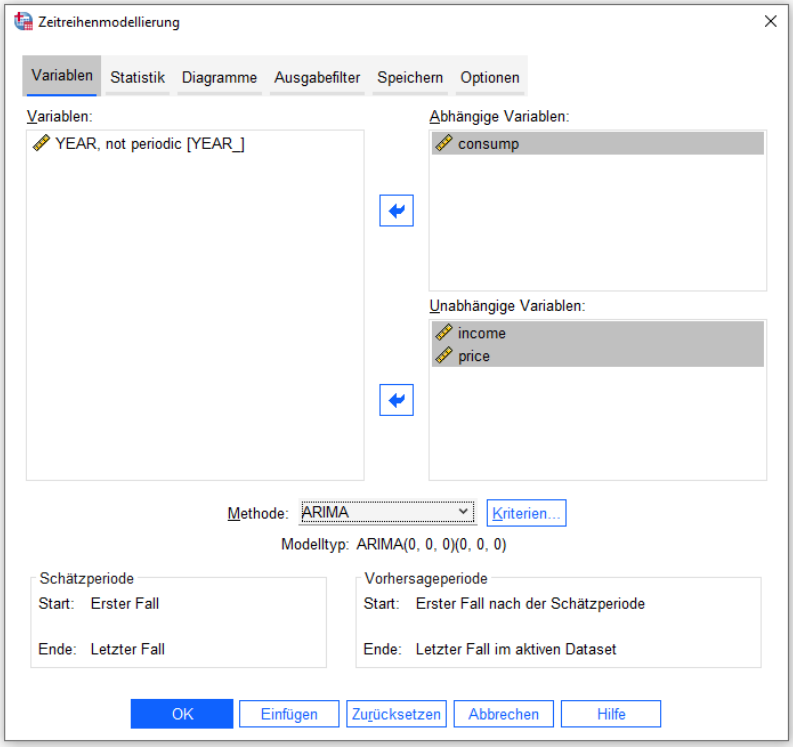
Das AR(1) - Modell ist ein einfacher Spezialfall des allgemeinen ARIMA(*p*, *d*, *q*) - Modells (*Autoregressive Integrated Moving Average*).

7.4.2 Durchführung mit SPSS

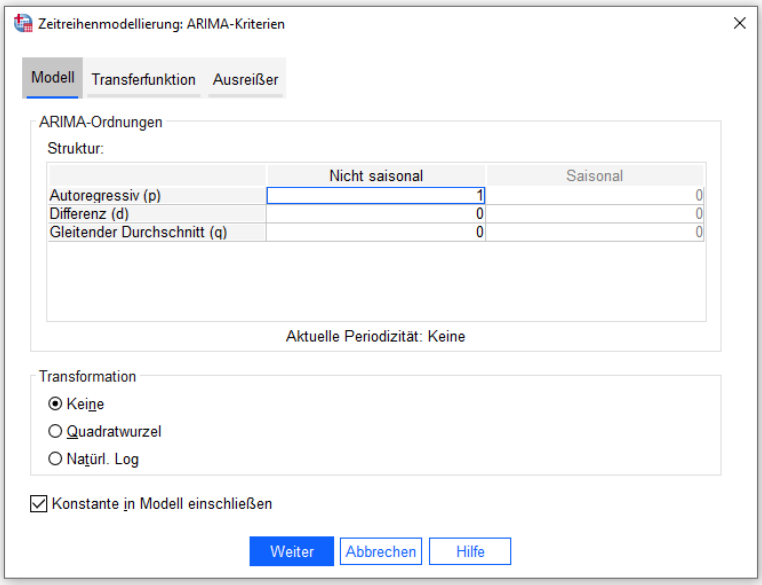
Nach dem Menübefehl:

Analysieren > Vorhersage > Traditionelle Modelle erstellen

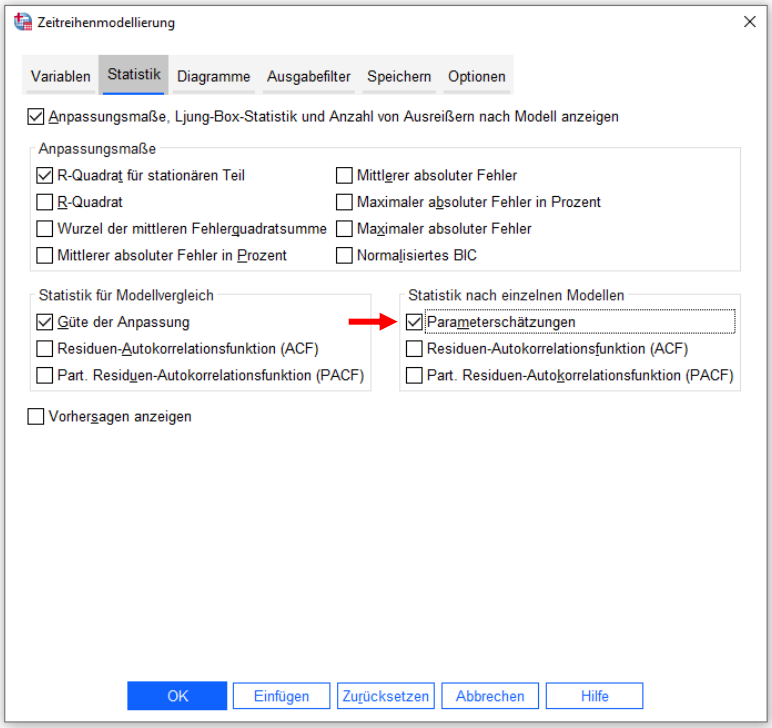
kann in der folgenden Dialogbox eine Regressionsanalyse mit AR(1) - Residualmodell angefordert werden. Im Durbin-Watson-Beispiel soll die **abhängige Variable** CONSUMP durch die **unabhängigen Variablen** INCOME und PRICE erklärt werden:



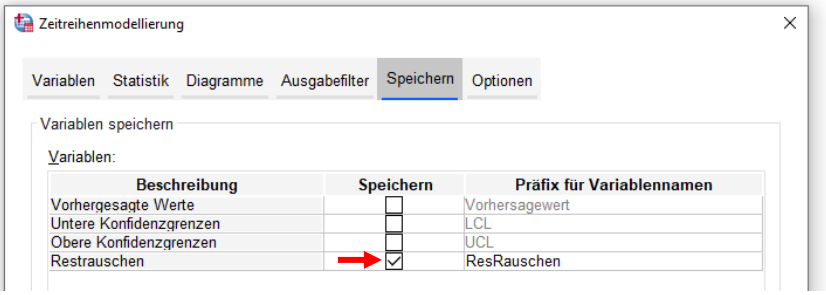
Wir wählen **ARIMA** als **Methode** und legen nach einem Klick auf den Schalter **Kriterien** die nichtsaisonale **Struktur** AR(1) für den Residualprozess fest:



Auf der Registerkarte **Statistik** müssen wir das Kontrollkästchen **Parameterschätzungen** markieren, um Schätzungen für die Modellparameter zu erhalten:



Auf der Registerkarte **Speichern** sorgen wir dafür, dass die geschätzten ε_t - Werte als zusätzliche Variable in die Arbeitsdatei geschrieben werden:



In den Ergebnissen zum Durbin-Watson - Beispiel hält der autoregressive Parameter knapp die Stationaritätsbedingung ein ($\hat{\rho} = 0{,}993 < 1$) und überwindet sehr deutlich die Signifikanzgrenze:

Parameter des ARIMA-Modells

					Schätzer	Standardfehler	t	Sig.
consump-Modell_1	consump	Keine Transformation	Konstante		2,452	,492	4,988	,000
			AR	Lag 1	,993	,021	47,153	,000
	income	Keine Transformation	Zähler	Lag 0	,621	,147	4,220	,000
	price	Keine Transformation	Zähler	Lag 0	-,929	,079	-11,712	,000

Die Residuen (geschätzten ε_t - Werte) zum Modell mit AR(1) - Residualprozess bewegen sich ohne erkennbare serielle Abhängigkeit um die Null-Lage:[1]

[1] Es ist eine Phase erhöhter Volatilität von ca. 1909 bis 1919 festzustellen, die mit dem gewählten Modellansatz nicht darstellbar ist. Abhilfe kann eventuell das speziell für Finanzzeitreihen entwickelte *GARCH-Modell* schaffen (*Generalized Autoregressive Conditional Heteroskedasticity*).

Abbildung 55: Sequenzdiagramm für die Residuen aus dem Durbin-Watson-Modell mit AR(1) - Residualprozess (ab 1871)

Das erste Beobachtungsjahr 1870 fehlt, weil es keinen Vorgänger zur Berechnung des autoregressiven Modellbestandteils hat.

In der folgenden Tabelle erfahren wir u.a. das Ergebnis des **Box-Ljung**-Tests zur Nullhypothese, dass die ersten 18 Autokorrelationen in der Residualzeitreihe alle gleich 0 sind:

Modell-Statistik

Modell	Anzahl der Prädiktoren	Statistiken zur Güte der Anpassung: R-Quadrat für stationären Teil	Ljung-Box-Q(18): Statistiken	Ljung-Box-Q(18): Freiheitsgrade	Ljung-Box-Q(18): Sig.	Anzahl der Ausreißer
consump-Modell_1	2	,986	11,761	17	,814	0

Weil ein AR(1) - bzw. ARIMA(1, 0, 0) - Modell zum Einsatz kam, hat die Prüfstatistik 18-1 = 17 Freiheitsgrade. Das p-Level von 0,814 spricht dafür, die Nullhypothese zu akzeptieren.

Im Durbin-Watson - Beispiel genügt also ein AR(1) - Prozess, um die serielle Abhängigkeit im Modell zu berücksichtigen, und wir dürfen die Parameterschätzungen interpretieren:

- Für das Einkommen erhalten wir ein hoch signifikantes, positives Regressionsgewicht, während das unangemessene OLS-Regressionsmodell einen negativen Schätzwert liefert (-0,12).
- Ein steigender Preis wirkt sich signifikant dämpfend auf den Alkoholkonsum aus.

Wenn trotz eines AR(1) - Modells Abhängigkeiten in den Residuen verbleiben, dann kann im Dialog zur Zeitreihenmodellierung ein allgemeiner ARIMA(*p*,*q*,*d*) - Fehlerprozess angenommen werden.

7.4.3 Der Box-Ljung - Test als Alternative zum Durbin-Watson - Test

Im Abschnitt 7.3 musste zur Durchführung des Durbin-Watson - Tests eine relativ umständliche Prozedur beschrieben werden, wobei insbesondere eine aus externen Quelle zu beschaffende Tabelle erforderlich war. Der im Rahmen einer ARIMA-Modellierung verfügbare Box-Ljung - Test besitzt gegenüber dem Durbin-Watson - Test zwei Vorteile:

- SPSS Statistics liefert ihn inklusive Überschreitungswahrscheinlichkeit.
- Der Box-Ljung - Test prüft nicht nur die Autokorrelation *erster* Ordnung, sondern alle Autokorrelationen bis zur Ordnung 18, sodass die serielle Abhängigkeit in einem viel breiteren Sinn berücksichtigt wird.

Wenn die lineare Regression einen verdächtig unter 2 liegenden Durbin-Watson - Wert geliefert hat, kann man folgendermaßen zu einem Box-Ljung - Testergebnis gelangen (vgl. Abschnitt 7.4.2):

- Aufruf der Zeitreihenmodellierung über den Menübefehl
  Analysieren > Vorhersage > Traditionelle Modelle erstellen
- Auf der Registerkarte **Variablen** …
  - die **abhängige Variable** und die **unabhängigen Variablen** benennen,
  - **ARIMA** als **Methode** einstellen und ein ARIMA(0, 0, 0) - Modell wählen (Verzicht auf einen ARIMA(*p*, *d*, *q*)-Prozess für die Residuen).
- Auf der Registerkarte **Statistik** die **Parameterschätzungen** aktivieren.
- Auf der Registerkarte **Diagramme** kann man optional die **Residuen-Autokorrelationsfunktion (ACF)** sowie die **Partielle Residuen-Autokorrelationsfunktion (PACF)** anfordern.

Für die Durbin-Watson - Daten wird die Box-Ljung - Nullhypothese deutlich zurückgewiesen ($p < 0{,}001$):

Modell-Statistik

Modell	Anzahl der Prädiktoren	Statistiken zur Güte der Anpassung: R-Quadrat für stationären Teil	Ljung-Box-Q(18): Statistiken	Ljung-Box-Q(18): Freiheitsgrade	Ljung-Box-Q(18): Sig.	Anzahl der Ausreißer
consump-Modell_1	2	,956	261,264	18	,000	0

Durch eine nach dem Lag 1 abbrechende PACF und eine langsam ausschwingende ACF lässt sich für die Residuen ein AR(1) - Prozess identifizieren (siehe Baltes-Götz 1997):

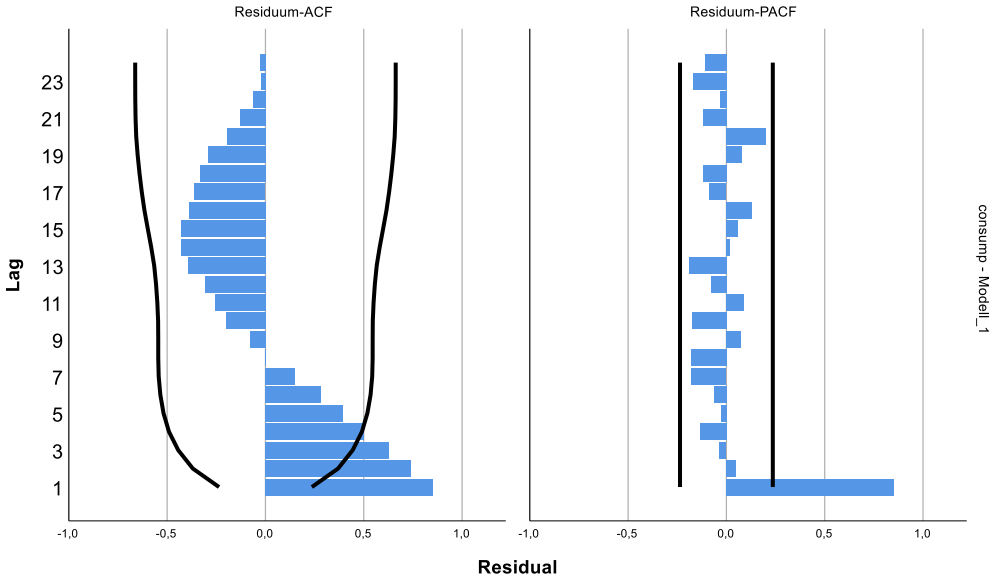


Abbildung 56: Autokorrelationsfunktion (ACF) und partielle Autokorrelationsfunktion (PACF) für die Residuen aus der Regression von CONSUMP auf INCOME und PRICE

Anhang

R 3.6.3 und die R-Integration installieren

Über das R-Integration - Plugin kann die Statistik-orientierte Programmiersprache R von SPSS Statistics aus genutzt werden. Dies ermöglicht u.a.:

- Verwendung der zahllosen Funktionen in fertigen R-Paketen
  So werden Analyseoptionen zugänglich, die in SPSS fehlen, z.B. Rasch-Itemanalyse, robuste Regression, polychorische Korrelationen. Viele R-Funktionen werden in die SPSS-Bedienoberfläche eingebunden und sind folglich mit gewohntem SPSS-Komfort zu nutzen.
- Programmieren von eigenen R-Funktionen

Die Variablen der SPSS-Arbeitsdatei werden an R übergeben und können in R-Funktionen über ihre SPSS-Namen angesprochen werden.

So installieren Sie die mit SPSS Statistics 27 kompatible R-Version 3.6.3 und die R-Integrations – Erweiterung von SPSS:[1]

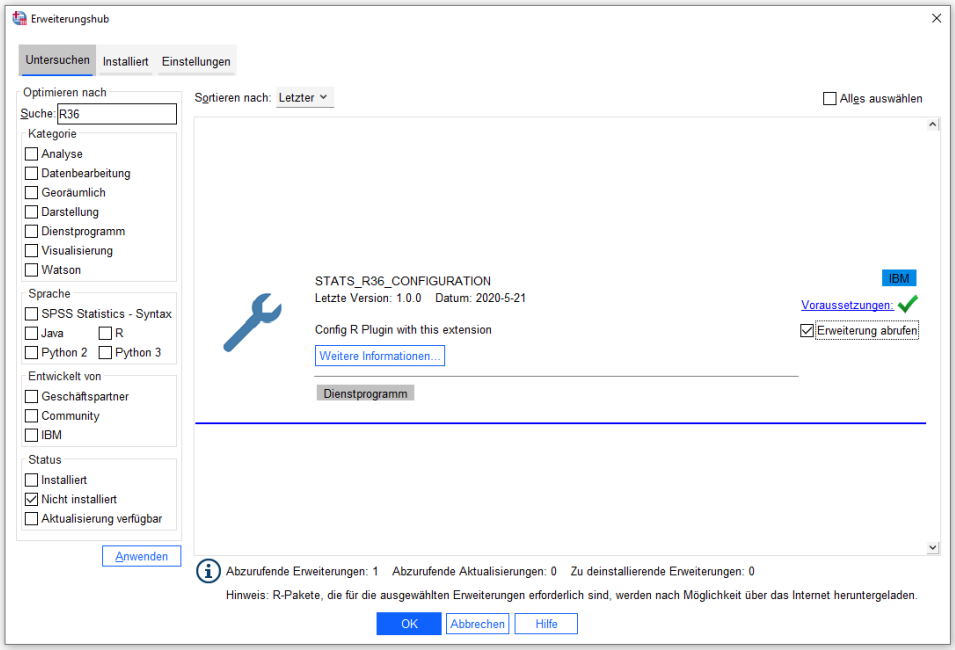
- Installieren Sie die (nicht von IBM SPSS stammende, sondern als Open Source entwickelte) Statistik-orientierte Programmiersprache R in der Version 3.6.3. Eine mögliche Bezugsquelle ist:

  https://cran.uni-muenster.de/bin/windows/base/old/3.6.3/

  Wenn Sie R auf einem Rechner mit 64-Bit - Windows installieren, müssen Sie unbedingt auch die 32-Bit-Version von R installieren (= Voreinstellung).
- Starten Sie SPSS Statistics 27, und installieren Sie über den Menübefehl

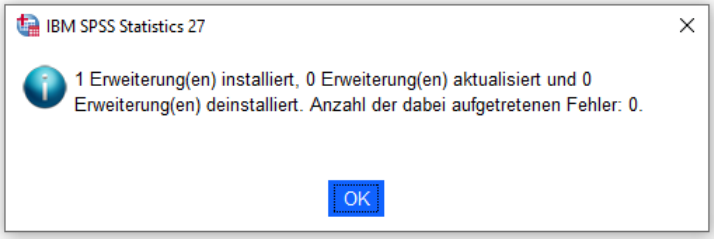
  Erweiterungen > Erweiterungsbub

  die R 3.6 - Konfigurationserweiterung **STATS_R36_CONFIGURATION**:

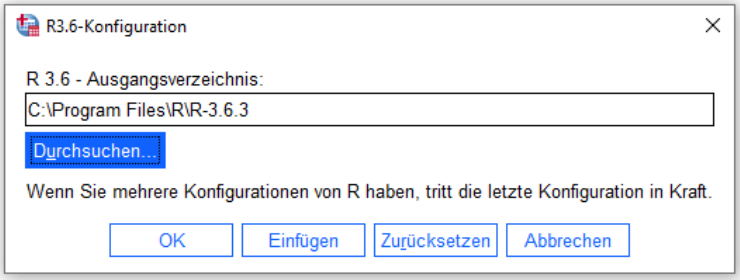


[1] Mit SPSS Statistics 28 wird automatisch die kompatible R-Version 4.0.5 installiert, sodass der im aktuellen Abschnitt beschriebene Installationsaufwand entfällt.

- Nach einer erfolgreichen Installation



befindet sich im Menü **Erweiterungen** das zusätzliche Item **R3.6-Konfiguration**. Wählen Sie es, um den Installationsordner von R 3.6 bekanntzugeben, z.B.:



Im Beispiel wird durch einen Klick auf **OK** das folgende SPSS-Kommando ausgeführt:

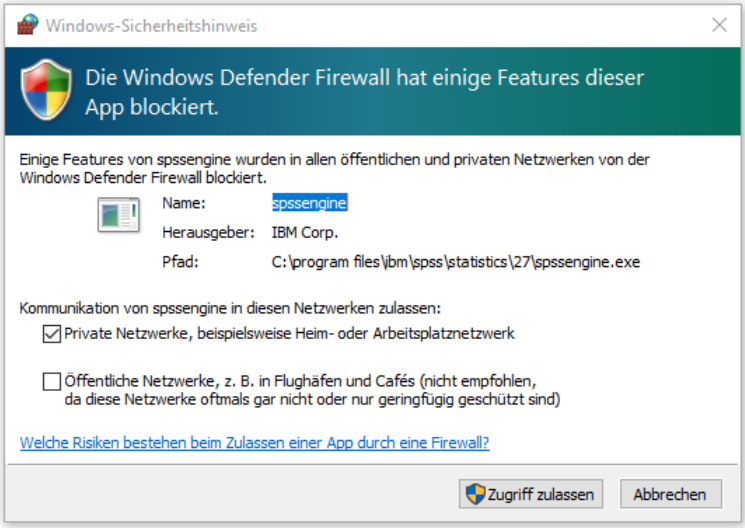
```
STATS R36 CONFIGURATION R_HOME="C:\Program Files\R\R-3.6.3".
```

Dabei werden benutzereigene Erweiterungen zur verwendeten R-Version installiert, z.B. in:

<Benutzer-Dokumente>\R\win-library\3.6\spss270

<Benutzer-Dokumente>\R\win-library\3.6\spssstatistics

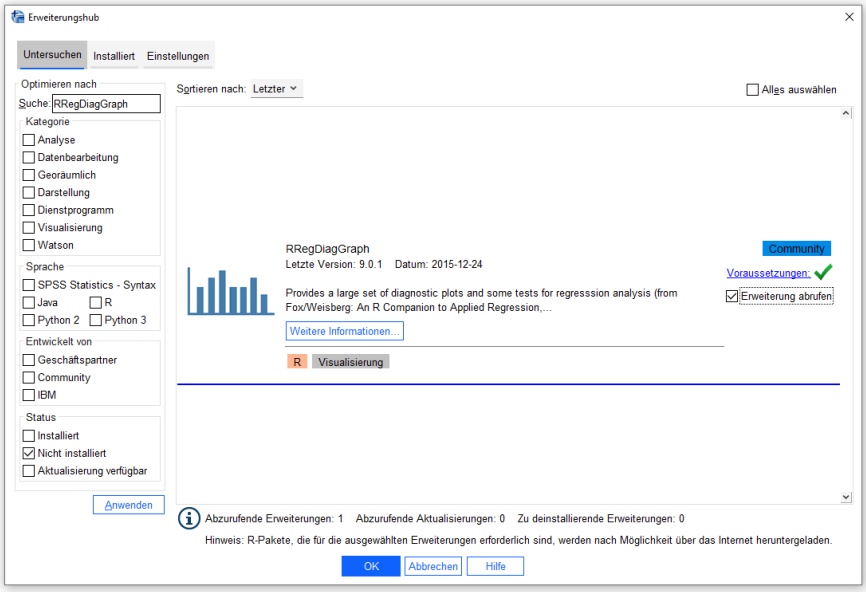
- Wenn auf Ihrem Rechner die Windows-Firewall aktiv ist, müssen Sie eventuell einer Ausnahmeregel zustimmen, welche die Kooperation von SPSS Statistics 25 und R 3.3 ermöglicht, z.B.:



- Anschließend können Sie die R-Erweiterung benutzen und z.B. R-Anweisungen über das SPSS-Syntaxfenster ausführen lassen:

```
BEGIN PROGRAM R.
x <- 1:5; sum(x)
END PROGRAM.
```

- Außerdem können Sie nun über den Menübefehl **Erweiterungen > Erweiterungsbub** die R-basierten SPSS-Erweiterungen installieren und nutzen, z.B.:


Erweiterungshub
Untersuchen
Installiert
Einstellungen
Optimieren nach
Suche: RRegDiagGraph
Kategorie
Analyse
Datenbearbeitung
Georäumlich
Darstellung
Dienstprogramm
Visualisierung
Watson
Sprache
SPSS Statistics - Syntax
Java
R
Python 2
Python 3
Entwickelt von
Geschäftspartner
Community
IBM
Status
Installiert
Nicht installiert
Aktualisierung verfügbar
Anwenden
Sortieren nach: Letzter
Alles auswählen
RRegDiagGraph
Letzte Version: 9.0.1 Datum: 2015-12-24
Provides a large set of diagnostic plots and some tests for regresssion analysis (from Fox/Weisberg: An R Companion to Applied Regression,...
Weitere Informationen...
R
Visualisierung
Community
Voraussetzungen:
Erweiterung abrufen
Abzurufende Erweiterungen: 1
Abzurufende Aktualisierungen: 0
Zu deinstallierende Erweiterungen: 0
Hinweis: R-Pakete, die für die ausgewählten Erweiterungen erforderlich sind, werden nach Möglichkeit über das Internet heruntergeladen.
OK
Abbrechen
Hilfe

Literatur

APA (2010). *Publication Manual* (6th ed.). Washington, DC: American Psychological Association.

Backhaus, K., Erichson, B, Plinke, W. & Weiber, R. (2008). *Multivariate Analysemethoden* (12. Aufl.). Berlin: Springer.

Baltes-Götz, B. (1994). *Einführung in die Analyse von Strukturgleichungsmodellen mit LISREL 7 und PRELIS unter SPSS*. Online-Dokument: http://www.uni-trier.de/index.php?id=22734

Baltes-Götz, B. (1997). *Zeitreihenanalyse mit SPSS-Trends*. Online-Dokument: http://www.uni-trier.de/index.php?id=22544

Baltes-Götz, B. (2012). *Logistische Regressionsanalyse mit SPSS*. Online-Dokument: http://www.uni-trier.de/index.php?id=22513

Baltes-Götz, B. (2013). *Behandlung fehlender Werte in SPSS und Amos*. Online-Dokument: http://www.uni-trier.de/index.php?id=23239

Baltes-Götz, B. (2015). *Analyse von Strukturgleichungsmodellen mit Amos 18*. Online-Dokument: http://www.uni-trier.de/index.php?id=22640

Baltes-Götz, B. (2016a). *Generalisierte lineare Modelle und GEE-Modelle in SPSS*. Online-Dokument: http://www.uni-trier.de/index.php?id=51455

Baltes-Götz, B. (2016b). Regressionsmodelle für Paneldaten. In A. Geissler & M. Schneider (Hrsg.). *Zwischen artes liberales und artes digitales*, S. 157-181. Marburg: Tectum-Verlag

Baltes-Götz, B. (2017). *R als Ergänzung zu SPSS*. Online-Dokument: http://www.uni-trier.de/index.php?id=44476

Baltes-Götz, B. (2020a). *Mediator- und Moderatoranalyse mit SPSS und PROCESS*. Online-Dokument: http://www.uni-trier.de/index.php?id=22528

Baltes-Götz, B. (2020b). *Analyse von hierarchischen linearen Modellen mit der SPSS-Prozedur MIXED*. Online-Dokument: http://www.uni-trier.de/index.php?id=39129

Baltes-Götz, B. (2020c). *Online-Umfragen mit Enterprise Feedback Suite Survey*. Online-Dokument: http://www.uni-trier.de/index.php?id=52985

Baltes-Götz, B. (2022). *Statistisches Praktikum mit SPSS für Windows*. Online-Dokument: http://www.uni-trier.de/index.php?id=22552

Box, G.E.P. (1979). Robustness in the Strategy of Scientific Model Building. In R. L. Launer & G. N. Wilkinson (Eds.). *Robustness in Statistics*, pp. 201-236. New York: Academic Press.

Brosius, F. (2018). *SPSS: Umfassendes Handbuch zu Statistik und Datenanalyse*. Heidelberg: mitp.

Bühner, M. & Ziegler, M (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New York: Academic Press.

Cohen, J., Cohen, P., West, S.G. & Aiken, L. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). Mahwah: Lawrence Erlbaum Associates.

Darlington, R.B & Hayes, A.F. (2017). *Regression Analysis and Linear Models*. New York: Guilford Press.

Disabato, D. (2016). *On effect sizes in multiple regression*. Online-Dokument: http://www.daviddisabato.com/blog/2016/4/8/on-effect-sizes-in-multiple-regression

Dixon, W.J. (Ed.) (1992). *BMDP Statistical software manual* (Vol 1). Berkeley: University of California Press.

Draper, N.R. & Smith, H. (1981). *Applied Regression Analysis* (2. ed). New York: Wiley.

Durbin, J, & Watson, G.S. (1951). Testing for serial correlation in least-squares regression II. *Biometrika*, 38, 159-178.

Eckstein, P.P. (2016). *Statistik für Wirtschaftswissenschaftler* (5. Aufl.). Wiesbaden: Springer.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82, 171-185.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hill.

Eicker, F. (1967). Limit theorems for regression with unequal and dependent errors. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Berkeley, CA: University of California Press.

Eid, M., Gollwitzer, M. & Schmitt, M. (2013). *Statistik und Forschungsmethoden* (3. Aufl.). Weinheim: Beltz.

Fahrmeir, L., Kneib, T. & Lang, S. (2007). *Regression. Modelle, Methoden und Anwendungen*. Berlin: Springer.

Faul, F., Erdfelder, E., Buchner, A. & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149-1160.

Fox, J. (1997). *Applied Regression Analysis, Linear Models, and related Methods*. Thousand Oaks: Sage.

Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models* (2nd ed.). Thousand Oaks, CA: Sage.

Fox, J. & Weisberg, S. (2011). *An R Companion to Applied Regression*. Thousand Oaks: Sage.

Goldberger, A.S. (1991). *A course in econometrics*. Cambridge. Cambridge, MA: Harvard University Press.

Gordon, R.A. (2010). *Regression Analysis for the Social Sciences*. New York: Routledge.

GPower (2017). *Manual 3.1*. Online-Dokument: http://www.gpower.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf

Groß, J. (2003). *Linear Regression*. Berlin: Springer.

Hartung, J. (1989). *Statistik*. München: Oldenbourg.

Hayes, A. F. (2018). *Introduction to Mediation, Moderation and Conditional Process Analysis* (2nd ed.). New York: Guilford Press.

Hayes, A. F. & Cai, L. (2007). Using heteroscedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*, 39, 709-722.

Huber, P. J. (1967). The behavior of maximum likelihood estimation under nonstandard conditions. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Berkeley, CA: University of California Press.

Hsu, J.C. (1996). *Multiple Comparisons*. London: Chapman & Hall.

Kelley, K. (2018). MBESS: The MBESS R Package. Online-Dokument: https://CRAN.R-project.org/package=MBESS

Kockläuner, G. (1988). *Angewandte Regressionsanalyse mit SPSS*. Braunschweig: Vieweg.

Long, J.S. & Erwin, L.H. (2000). Using heteroscedasticity-consistent standard errors in the linear regression model. *The American Statistician*, 54, 217-224.
Entwurf als Online-Dokument verfügbar:
http://www.indiana.edu/~jslsoc/files_research/testing_tests/hccm/99TAS.pdf

Muthén, L.K. & Muthén, B.O. (2017). *Mplus User's Guide* (8th ed.). Los Angeles, CA: Muthén & Muthén

Norušis, M.J. (2005). *SPSS 14.0. Statistical Procedures Companion*. Upper Saddle River, NJ: Prentice Hall.

Pindyck, R.S. & Rubinfeld, D.L. (1990). *Econometric models and econometric forecasts*. Auckland: McGraw-Hill.

Raudenbush, S.A. & Bryk, A. S. (2002). *Hierarchical Linear Models* (2nd ed.). Thousand Oaks: CA: Sage.

Ryan, T.S. (1997). *Modern Regression Methods*. New York: Wiley.

Savin, N.E. & White, K.J. (1977). The Durbin-Watson Test for serial correlation with extreme sample sizes or many regressors. *Econometrica*, Vol. 45, No. 8, 1989-1996.

Schlittgen, R. & Streitberg, B.H.J. (1989). *Zeitreihenanalyse*. München: Oldenbourg.

Shao, J. (1996). Bootstrap Model Selection. *Journal of the American Statistical Association*, Vol. 91 (434), 655-665.

Snedecor, G.W. & Cochran, W.C. (1980). *Statistical methods* (7th ed.). Ames, Iowa: The Iowa State University Press.

Sonnemann, E. (1982). Allgemeine Lösungen multipler Testprobleme. *EDV in Medizin und Biologie*, 13(4), 120-128.

SPSS, Inc. (2004). *SPSS Trends 13.0*. Chicago, IL.

Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245-251.

Steyer, R. (1992). *Theorie kausaler Regressionsmodelle*. Stuttgart: Fischer.

Steyer, R. (2003). *Wahrscheinlichkeit und Regression*. Berlin: Springer.

Tabachnick, B.G. & Fidell, L.S. (2013). *Using multivariate statistics* (6th ed.). Boston: Allyn & Bacon.

Urban, D. & Mayerl, J. (2018). *Angewandte Regressionsanalyse: Theorie, Technik und Anwendung* (5. Aufl.). Wiesbaden: Springer VS.

Warner, R.M. (2013). *Applied statistics: from bivariate through multivariate techniques* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Weisberg, S. (1985). *Applied linear regression* (2nd ed.). New York: Wiley.

Wentura, D. & Pospeschill, M. (2015). *Multivariate Datenanalyse. Eine kompakte Einführung*. Wiesbaden: Springer

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test of heteroskedasticity. *Econometrica*, 48, 817-838.

Wooldridge, J.M. (2013). *Introductory Econometrics: A modern approach* (5th ed.). Cengage Learning.

Stichwortverzeichnis

A

Abhängige Variable ... 8
ACF ... 170
Achsenabschnitt ... 25
Adjustierter Determinationskoeffizient ... 102
Alpha-Fehler - Kumulierung ... 99
Amos ... 15, 142
AMOS ... 158
Annahmen
  der einfachen Regression ... 12
AR(1) - Prozess ... 166
ARIMA-Modell ... 167
Ausgelassen-studentisierte Residuen ... 39
Ausgeschlossene Residuen ... 40
Ausreißer ... 41, 126
Autokorrelation ... 163
Automatische Modellsuche ... 152
Autoregressiver Prozess ... 166

B

Bedingte Verteilung ... 16
Beta-Koeffizienten ... 19, 94
Bivariate Normalverteilung ... 16
Bivariate Regression ... 8
Block von Regressoren ... 150
Blom-Anteilschätzung ... 67
BLUE-Schätzer ... 19, 93
Bonferroni-Holm ... 101
Bootstrapping ... 59
Box-Cox - Transformation ... 76
Box-Ljung - Test ... 169
Breusch-Pagan - Heteroskedastizitätstest ... 49

C

CDF.T ... 23
Cook-Distanz ... 132

D

Designmatrix ... 93
Determinationskoeffizient ... 26, 102
DFBETA-Statistik ... 133
Direkter Effekt ... 157
Durbin-Watson - Test ... 163, 169
Durbin-Watson-Test ... 163

E

Effektstärke ... 19
Einflussreiche Fälle ... 131
Einseitiger Signifikanztest ... 97
Einseitiger Test der Regressionskoeffizienten ... 22
Endogene Variable ... 8
Exogene Variable ... 8

F

Faktorenanalyse ... 141
Fallweiser Ausschluss fehlender Werte ... 89
Fehler ... 11
Fehler in den Regressoren ... 158
Fixed-X - Annahme ... 12
Fixierte Regressoren ... 10
F-Test ... 21

G

GARCH-Modelle ... 168
Gauß-Markov ... 19, 93
Geordnet-kategoriales Messniveau ... 12, 13
Gewichtete Kleinst-Quadrat-Regression ... 61
Gewichtsschätzung
  SPSS-Prozedur ... 62
GPower ... 79, 120

H

HC4 ... 138
HCCM ... 57
HCSE ... 57
Hebelwert
  zentrierter ... 38, 129
Heteroskedastizitäts-robuste Standardfehler ... 57
Hierarchische Regressionsanalyse ... 150
Homogene Regression ... 14, 28
Homogenisierung der Fehlervarianzen ... 61
Homogenitätsprüfung ... 116
Homoskedastizität ... 14

I

Indirekter Effekt ... 157
Inhomogene Regression ... 14

K

Kapitalisierung von Zufall ... 153
Kategoriale Regressoren ... 150
Kausalität ... 110, 155
Kleinst-Quadrat-Regression
  gewichtete ... 61
Kolmogorov-Smirnov-Test ... 69
Konfidenzintervall
  Determinationskoeffizient ... 27
  einseitig ... 25, 99
  zweiseitig ... 24
Konfidenzintervalle ... 21
Korrigierter Determinationskoeffizient ... 26
KQ-Schätzer ... 18
Kriterium ... 8

L

Latente Variablen ... 15
lavaan ... 158
leverage value ... 38
Linear unabhängig ... 92
Linearität ... 91
Linearitätsannahme ... 13, 18, 30
Linearitätsprüfung ... 112
Linearitätstest ... 34
LISREL ... 158
Loess-Anpassungsmethode ... 32
Logarithmus ... 52

Logistische Regression 87
Logistische Regressionsanalyse 12, 13

M

Mahalanobis-Distanzen 131
Maximalquotientenkriterium 48
Maximum-Likelihood-Schätzer 19
MBESS 27
MEANS 34
Mediation 156
Mediator 157
Mehrebenenanalyse 15
Messniveau
  Geordnet-kategorial 12, 13
  Metrisch 12
  Nominal 12
Methode der kleinsten Quadrate 18
MPlus 158
Multikollinearität 139, 147
Multiple Regression 87
Multiples Testen 99, 127

N

Nonrekursive Modelle 158
Normalgleichungen 18
Normalitätsannahme 14
Normalitätsprüfung 118
Normalverteilung 92
Normalverteilungsdiagramm 67
Normalverteilungsprüfung 66
Normalverteilungstest 68

O

OLS-Schätzung 18, 92
omitted variable error 156
Ordinatenabschnitt 13

P

Paarweise Ausschluss fehlender Werte 89
PACF 170
Partialkorrelation 97
Partielle Korrelation 110
Partieller Regressionskoeffizient 90, 113
Partielles Eta-Quadrat 111
Partielles Regressionsdiagramm 113
Pearson-Korrelation 22
Pfadmodell 158
Polynomische Regression 143
Polynomische Regressionsgleichung 143
Polynomische Regressionsgleichungen 56

Q

Quadratsummenzerlegung 21

R

Regression
  multiple 87
Regressionsgleichung
  polynomische 143
Regressionsgleichungen
  polynomische 56
Regressor 8
Replikationskrise 153
Residuen
  ausgelassen-studentisierte 39
  standardisierte 40
  studentisierte 39
  unstandardisierte 31
Residuum 11
RLM 128

S

Schiefe 70
Schrittweise Regression 152
Semipartielle Korrelation 109
Shapiro-Wilk 69
Signifikanztest
  einseitiger 97
Signifikanztests 21
Skalenqualität 12, 87
Spezifikationsfehler 33
Spread & Level - Plot 44, 52
Standardfehler des Regressionskoeffizienten 24, 98
Standardisierte Regressionskoeffizienten 19, 94
Standardisierte Residuen 40
Standardisierte Variablen 94
Stationarität 166
Stetigkeit 12
Stichprobenumfangsplanung
  bivariate Regression 78
  multiple Regression 120
Stochastische Regressoren 16
Strukturgleichungsanalyse 141
Strukturgleichungsmodelle 158
Studentisierte Residuen 39
Suppressoreffekt 107

T

Toleranz 139
Top-Down - Modellierung 154
Totaler Effekt 157

U

Unabhängige Variable 8
UNIANOVA 57
Unkorreliertheit der Residuen 14, 92, 160
Unstandardisierte Residuen 31

V

Varianzhomogenität 14, 92
Varianzhomogenitätsannahme 41
Varianzinflationierungsfaktor 140
Vertrauensintervalle 21
Vulnerabilitäts-Kompensation 106

W

Wechselwirkungen 154
Weißes Rauschen 166
WLS-Regression 61
Wölbung 70

Z

Zentraler Grenzwertsatz 66, 71
Zentrierter Hebelwert 38