Mathematische und statistische Methoden I

Dr. Malte Persike

persike@uni-mainz.de
lordsofthebortz.de
twitter.com/methodenlehre
tinyurl.com/gplusmethodenlehre

Sprechstunde jederzeit nach Vereinbarung und nach der Vorlesung

Wallstr. 3, 6. Stock, Raum 06-206

WiSe 2011/2012

Fachbereich Sozialwissenschaften
Psychologisches Institut
Johannes Gutenberg Universität Mainz

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Voraus-
setzungen

Residualplot

Voraussetzungen der Regression

Mathematische und statistische Betrachtung

- Mathematisch ist eine multiple Regression praktisch **immer zu rechnen**, da nur in Ausnahmefällen die Invertierung der Prädiktorinterkorrelationsmatrix fehlschlägt

- Statistisch aber sollen eine Reihe von **Voraussetzungen** erfüllt sein, damit
  - Kennwerte und inferenzstatistische Verfahren (z.B. der statistische Test der β–Gewichte) anwendbar sind
  - die Regressionsgleichung empirische Aussagekraft besitzt

Voraus-
setzungen

Residualplot

Voraussetzungen der Regression

1. Skalenniveaus

- Die **Prädiktoren** können entweder intervallskaliert oder dichotom sein
- Das **Kriterium** muss intervallskaliert sein und die Skala soll unbeschränkt sein (keine untere und obere Schranke → **Ungebundenheit**)
- Für andere Skalenniveaus des Kriteriums existieren verschiedene Regressionsvarianten:
  - **Logistische Regression** für dichotome Kriteriumsvariablen
  - **Multinomiale Regression** für nominalskalierte Kriterien
  - **Ordinale Regression** für ordinalskalierte Kriterien

Voraus-
setzungen

Residualplot

Voraussetzungen der Regression

2. Eigenschaften der Prädiktoren

- Keine zu hohen Interkorrelationen zwischen den Prädiktoren, i.e. **Vermeidung von Multikollinearität**
- Es sollen alle wesentlichen Einflussvariablen des Kriteriums erfasst werden, d.h. **hinreichend hohes R^2**
- Der Zusammenhang zwischen den Prädiktoren und dem Kriteriums soll dem Modell der Regressionsgleichung entsprechen (linear, polynomisch etc.)
- Es soll eine **hinreichend hohe Stichprobengröße** vorliegen, Daumenregeln empfehlen hier zwischen 15 und 25 Personen pro Prädiktor

Voraus-
setzungen

Residualplot

Voraussetzungen der Regression

3. Eigenschaften der Fehler bzw. Residuen

- **Hinweis:** Der Vorhersagefehler in der Regression wird auch als **Residuum** bezeichnet
- Die Residuen dürfen **nicht untereinander korreliert** sein, d.h. die Höhe des Vorhersagefehlers für Merkmalsträger 1 darf nicht den Fehler für Merkmalsträger 2 beeinflussen
- Die Residuen sollen **normalverteilt** sein
- Für die Residuen soll der **erwartete Mittelwert 0** sein
- Die Residuen sollen dem Gebot der **Homoskedastizität** genügen, d.h. ihre Varianz soll unabhängig vom Kriteriumswert sein.

Voraus-
setzungen

Residualplot

Der Residualplot

Eigenschaften der Fehler bzw. Residuen

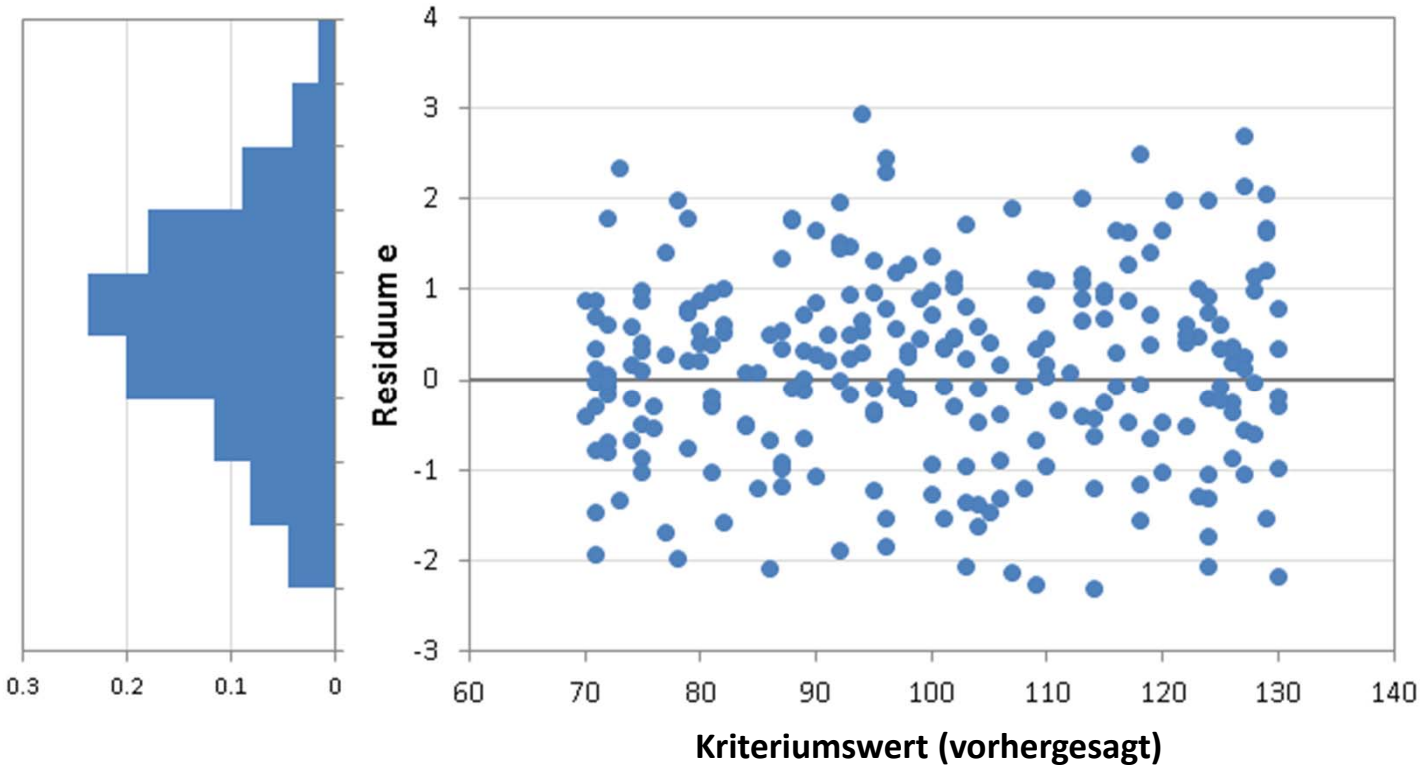
- Für die meisten der Fehlereigenschaften gibt es **statistische Tests** zur Voraussetzungsprüfung
  z.B. Variance Inflation Factor (VIF) für Multikollinearität, Durbin-Watson Test für Unkorreliertheit, Levene-Test für Homoskedastizität, Kolmogoroff-Smirnov Test für Normalverteilung
- Der **Residualplot** ist ein optisches Verfahren zur Prüfung der Voraussetzungen
- Er stellt die beobachteten Kriteriumswerte (x-Achse) und die Residuen (y-Achse) gegenüber
- An ihm kann man Homoskedastizität, Modellpassung (und auch Normalverteiltheit) optisch gut überprüfen

Voraus-setzungen

Residualplot

Der Residualplot

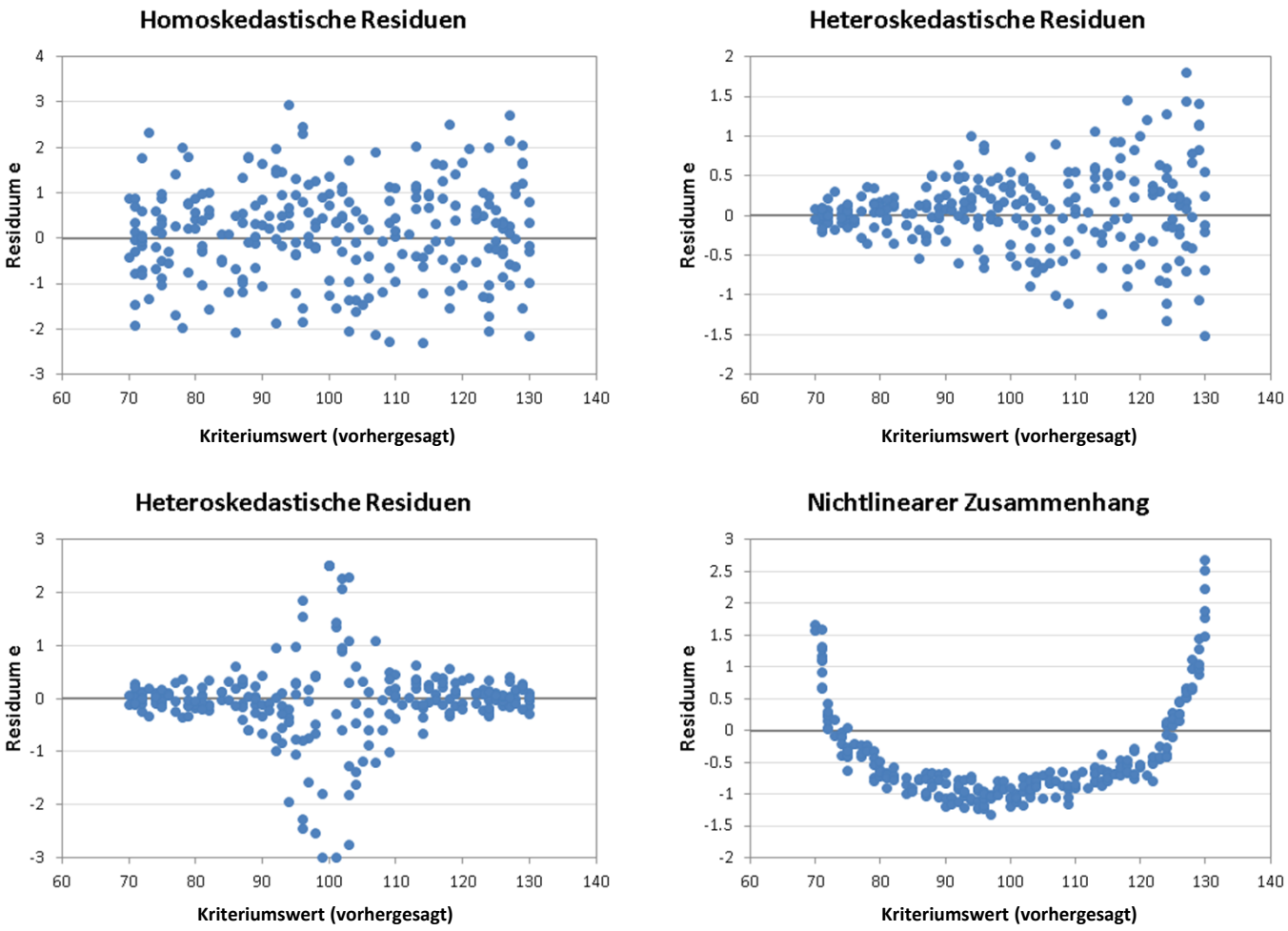
Eigenschaften der Fehler bzw. Residuen



- **Hinweis:** Für die Residuen werden zumeist die **z-standardisierten** Residuen gewählt

Voraus-
setzungen

Residualplot

Der Residualplot

Eigenschaften der Fehler bzw. Residuen