Probleme im linearen Modell oder was schiefgehen kann geht schief

Regressionsmodelle für Politikwissenschaftler

Übersicht

Was haben wir letzte Woche gelernt?

- Regression betrachtet den konditionalen Mittelwert einer Variablen

Was haben wir letzte Woche gelernt?

- Regression betrachtet den konditionalen Mittelwert einer Variablen
- In Abhängigkeit vom Niveau der unabhängigen Variablen folgt dieser Mittelwert einem Pfad

Was haben wir letzte Woche gelernt?

- Regression betrachtet den konditionalen Mittelwert einer Variablen
- In Abhängigkeit vom Niveau der unabhängigen Variablen folgt dieser Mittelwert einem Pfad
- Im klassischen linearen Modell entspricht dieser Pfad der Linie / Fläche / Hyperfläche, die die SAQ minimieren → partielle Ableitungen auf null setzen

Was haben wir letzte Woche gelernt?

- Regression betrachtet den konditionalen Mittelwert einer Variablen
- In Abhängigkeit vom Niveau der unabhängigen Variablen folgt dieser Mittelwert einem Pfad
- Im klassischen linearen Modell entspricht dieser Pfad der Linie / Fläche / Hyperfläche, die die SAQ minimieren ➔ partielle Ableitungen auf null setzen
- Das Gleichungssystem, mit dessen Hilfe $b_0, b_1, \ldots$ gefunden werden, läßt sich mit Hilfe von etwas Matrix-Algebra sehr effizient analytisch lösen

Was haben wir letzte Woche gelernt?

- Regression betrachtet den konditionalen Mittelwert einer Variablen
- In Abhängigkeit vom Niveau der unabhängigen Variablen folgt dieser Mittelwert einem Pfad
- Im klassischen linearen Modell entspricht dieser Pfad der Linie / Fläche / Hyperfläche, die die SAQ minimieren ➔ partielle Ableitungen auf null setzen
- Das Gleichungssystem, mit dessen Hilfe $b_0, b_1, \ldots$ gefunden werden, läßt sich mit Hilfe von etwas Matrix-Algebra sehr effizient analytisch lösen
- Datenmatrix muß genug unabhängige Informationen enthalten ➔ keiner der Spaltenvektoren darf Linearkombination anderer Vektoren darstellen (perfekte Kollinearität)

Was haben wir letzte Woche gelernt?

- Regression betrachtet den konditionalen Mittelwert einer Variablen
- In Abhängigkeit vom Niveau der unabhängigen Variablen folgt dieser Mittelwert einem Pfad
- Im klassischen linearen Modell entspricht dieser Pfad der Linie / Fläche / Hyperfläche, die die SAQ minimieren → partielle Ableitungen auf null setzen
- Das Gleichungssystem, mit dessen Hilfe $b_0, b_1, \ldots$ gefunden werden, läßt sich mit Hilfe von etwas Matrix-Algebra sehr effizient analytisch lösen
- Datenmatrix muß genug unabhängige Informationen enthalten → keiner der Spaltenvektoren darf Linearkombination anderer Vektoren darstellen (perfekte Kollinearität)
- Mittel zur Datenverdichtung – ist OLS aber auch ein guter Schätzer für die unbekannten Parameter der Grundgesamtheit?

Vorab: Komplikationen

- Auf welche Komplikationen sind wir letzte Woche gestoßen?

Vorab: Komplikationen

- Auf welche Komplikationen sind wir letzte Woche gestoßen?

1. Kategoriale unabhängige Variablen

Vorab: Komplikationen

- Auf welche Komplikationen sind wir letzte Woche gestoßen?

1. Kategoriale unabhängige Variablen
2. Interaktionen

Vorab: Komplikationen

- Auf welche Komplikationen sind wir letzte Woche gestoßen?

1. Kategoriale unabhängige Variablen
2. Interaktionen
3. Nicht-lineare Zusammenhänge

Was tun mit kategorialen Unabhängigen?

- Keine Verteilungsannahmen für die unabhängigen Variablen?

Was tun mit kategorialen Unabhängigen?

- Keine Verteilungsannahmen für die unabhängigen Variablen?
- Unabhängige Variablen können kategorial sein

Was tun mit kategorialen Unabhängigen?

- Keine Verteilungsannahmen für die unabhängigen Variablen?
- Unabhängige Variablen können kategorial sein
- Gar kein Problem bei dichotomen Variablen → 0/1 kodieren (Dummies)

Was tun mit kategorialen Unabhängigen?

- Keine Verteilungsannahmen für die unabhängigen Variablen?
- Unabhängige Variablen können kategorial sein
- Gar kein Problem bei dichotomen Variablen → 0/1 kodieren (Dummies)
- Effekt entspricht der Differenz zwischen den Mittelwerten von Gruppe 0/1

Was tun mit kategorialen Unabhängigen?

- Keine Verteilungsannahmen für die unabhängigen Variablen?
- Unabhängige Variablen können kategorial sein
- Gar kein Problem bei dichotomen Variablen → 0/1 kodieren (Dummies)
- Effekt entspricht der Differenz zwischen den Mittelwerten von Gruppe 0/1
- Identisch mit t-Test für unabhängige Stichproben

Was tun mit kategorialen Unabhängigen?

- Nominale Variablen mit $k > 2$ Kategorien durch $k - 1$ Dummies repräsentieren

Was tun mit kategorialen Unabhängigen?

- Nominale Variablen mit $k > 2$ Kategorien durch $k - 1$ Dummies repräsentieren
- Beispiel Konfession mit Kategorien katholisch / protestantisch / andere durch kath/prot oder kath/andere oder prot/andere

Was tun mit kategorialen Unabhängigen?

- Nominale Variablen mit $k > 2$ Kategorien durch $k - 1$ Dummies repräsentieren
- Beispiel Konfession mit Kategorien katholisch / protestantisch / andere durch kath/prot oder kath/andere oder prot/andere
- Warum nicht drei Dummies?

y	p	k	a
12	1	0	0
13	0	1	0
10	0	0	1
11	0	1	0
14	1	0	0

- Jeweils einer der drei Vektoren **p**, **k**, **a** perfekte Linearkombination der anderen beiden z. B. $\mathbf{a} = 1 - \mathbf{p} - \mathbf{k}$

y	p	k	a
12	1	0	0
13	0	1	0
10	0	0	1
11	0	1	0
14	1	0	0

- Jeweils einer der drei Vektoren **p**, **k**, **a** perfekte Linearkombination der anderen beiden z. B. $\mathbf{a} = 1 - \mathbf{p} - \mathbf{k}$
- (Perfekte) *Kollinearität*

y	p	k	a
12	1	0	0
13	0	1	0
10	0	0	1
11	0	1	0
14	1	0	0

- Jeweils einer der drei Vektoren **p**, **k**, **a** perfekte Linearkombination der anderen beiden z. B. $\mathbf{a} = 1 - \mathbf{p} - \mathbf{k}$
- (Perfekte) *Kollinearität*
- Matrix $\mathbf{X}'\mathbf{X}$ nicht invertierbar

y	p	k	a
12	1	0	0
13	0	1	0
10	0	0	1
11	0	1	0
14	1	0	0

- Jeweils einer der drei Vektoren **p**, **k**, **a** perfekte Linearkombination der anderen beiden z. B. $\mathbf{a} = 1 - \mathbf{p} - \mathbf{k}$
- (Perfekte) *Kollinearität*
- Matrix $\mathbf{X}'\mathbf{X}$ nicht invertierbar
- In Normalgleichungen mehr Unbekannte als voneinander unabhängige Normalgleichungen

y	p	k	a
12	1	0	0
13	0	1	0
10	0	0	1
11	0	1	0
14	1	0	0

- Jeweils einer der drei Vektoren **p**, **k**, **a** perfekte Linearkombination der anderen beiden z. B. $\mathbf{a} = 1 - \mathbf{p} - \mathbf{k}$
- (Perfekte) *Kollinearität*
- Matrix $\mathbf{X}'\mathbf{X}$ nicht invertierbar
- In Normalgleichungen mehr Unbekannte als voneinander unabhängige Normalgleichungen
- Information redundant; einen beliebigen Dummy weglassen

y	p	k	a
12	1	0	0
13	0	1	0
10	0	0	1
11	0	1	0
14	1	0	0

- Die weggelassene Kategorie heißt „Referenzgruppe“

y	p	k	a
12	1	0	0
13	0	1	0
10	0	0	1
11	0	1	0
14	1	0	0

- Die weggelassene Kategorie heißt „Referenzgruppe"
- Konstante plus gegebenenfalls Effekte anderer Variablen = erwarteter Wert für Referenzgruppe

y	p	k	a
12	1	0	0
13	0	1	0
10	0	0	1
11	0	1	0
14	1	0	0

- Die weggelassene Kategorie heißt „Referenzgruppe“
- Konstante plus gegebenenfalls Effekte anderer Variablen = erwarteter Wert für Referenzgruppe
- Effekte der übrigen Kategorie entsprechen der Abweichung dieser Gruppen vom Wert der Referenzgruppe

y	p	k	a
12	1	0	0
13	0	1	0
10	0	0	1
11	0	1	0
14	1	0	0

- Die weggelassene Kategorie heißt „Referenzgruppe“
- Konstante plus gegebenenfalls Effekte anderer Variablen = erwarteter Wert für Referenzgruppe
- Effekte der übrigen Kategorie entsprechen der Abweichung dieser Gruppen vom Wert der Referenzgruppe
- $y = 10 + 3p + 2k$

y	p	k	a
12	1	0	0
13	0	1	0
10	0	0	1
11	0	1	0
14	1	0	0

- Die weggelassene Kategorie heißt „Referenzgruppe“
- Konstante plus gegebenenfalls Effekte anderer Variablen = erwarteter Wert für Referenzgruppe
- Effekte der übrigen Kategorie entsprechen der Abweichung dieser Gruppen vom Wert der Referenzgruppe
- $y = 10 + 3p + 2k$
- $y = 12 + 1p - 2a$

y	p	k	a
12	1	0	0
13	0	1	0
10	0	0	1
11	0	1	0
14	1	0	0

- Die weggelassene Kategorie heißt „Referenzgruppe“
- Konstante plus gegebenenfalls Effekte anderer Variablen = erwarteter Wert für Referenzgruppe
- Effekte der übrigen Kategorie entsprechen der Abweichung dieser Gruppen vom Wert der Referenzgruppe
- $y = 10 + 3p + 2k$
- $y = 12 + 1p - 2a$
- $y = 13 - 1k - 3a$

y	p	k	a
12	1	0	0
13	0	1	0
10	0	0	1
11	0	1	0
14	1	0	0

- Die weggelassene Kategorie heißt „Referenzgruppe“
- Konstante plus gegebenenfalls Effekte anderer Variablen = erwarteter Wert für Referenzgruppe
- Effekte der übrigen Kategorie entsprechen der Abweichung dieser Gruppen vom Wert der Referenzgruppe
- $y = 10 + 3p + 2k$
- $y = 12 + 1p - 2a$
- $y = 13 - 1k - 3a$
- Verfahren entspricht Varianzanalyse (auch andere Kodierungen möglich)

Wie kommt es sonst zu perfekter Kolllinearität?

- Kodierungsfehler

Wie kommt es sonst zu perfekter Kolllinearität?

- Kodierungsfehler
- Kohortenanalyse (APK-Ansatz)

Wie kommt es sonst zu perfekter Kolllinearität?

- Kodierungsfehler
- Kohortenanalyse (APK-Ansatz)
- Abhängige Variable wird beeinflußt von

Wie kommt es sonst zu perfekter Kolllinearität?

- Kodierungsfehler
- Kohortenanalyse (APK-Ansatz)
- Abhängige Variable wird beeinflußt von
    - Alter des Befragten

Wie kommt es sonst zu perfekter Kolllinearität?

- Kodierungsfehler
- Kohortenanalyse (APK-Ansatz)
- Abhängige Variable wird beeinflußt von
  - Alter des Befragten
  - Periode (Zeitpunkt der Untersuchung)

Wie kommt es sonst zu perfekter Kolllinearität?

- Kodierungsfehler
- Kohortenanalyse (APK-Ansatz)
- Abhängige Variable wird beeinflußt von
  - Alter des Befragten
  - Periode (Zeitpunkt der Untersuchung)
  - Kohorte (Generation)

Wie kommt es sonst zu perfekter Kolllinearität?

- Kodierungsfehler
- Kohortenanalyse (APK-Ansatz)
- Abhängige Variable wird beeinflußt von
  - Alter des Befragten
  - Periode (Zeitpunkt der Untersuchung)
  - Kohorte (Generation)
- Z. B. in Nichtwählerstudien

Wie kommt es sonst zu perfekter Kolllinearität?

- Kodierungsfehler
- Kohortenanalyse (APK-Ansatz)
- Abhängige Variable wird beeinflußt von
  - Alter des Befragten
  - Periode (Zeitpunkt der Untersuchung)
  - Kohorte (Generation)
- Z. B. in Nichtwählerstudien
- Aber: $A = P - K$, $P = K + A$, $K = P - A$

Wie kommt es sonst zu perfekter Kolllinearität?

- Kodierungsfehler
- Kohortenanalyse (APK-Ansatz)
- Abhängige Variable wird beeinflußt von
  - Alter des Befragten
  - Periode (Zeitpunkt der Untersuchung)
  - Kohorte (Generation)
- Z. B. in Nichtwählerstudien
- Aber: $A = P - K$, $P = K + A$, $K = P - A$
- Lösungen

Wie kommt es sonst zu perfekter Kolllinearität?

- Kodierungsfehler
- Kohortenanalyse (APK-Ansatz)
- Abhängige Variable wird beeinflußt von
  - Alter des Befragten
  - Periode (Zeitpunkt der Untersuchung)
  - Kohorte (Generation)
- Z. B. in Nichtwählerstudien
- Aber: $A = P - K$, $P = K + A$, $K = P - A$
- Lösungen
  - (Willkürliche) Restriktionen

Wie kommt es sonst zu perfekter Kolllinearität?

- Kodierungsfehler
- Kohortenanalyse (APK-Ansatz)
- Abhängige Variable wird beeinflußt von
  - Alter des Befragten
  - Periode (Zeitpunkt der Untersuchung)
  - Kohorte (Generation)
- Z. B. in Nichtwählerstudien
- Aber: $A = P - K$, $P = K + A$, $K = P - A$
- Lösungen
  - (Willkürliche) Restriktionen
  - Eine oder mehrere Variablen durch Inhaltliches ersetzen

Was ist ein Interaktionseffekt?

- Standardmodell geht davon aus, daß unabhängige Variablen additiv zusammenwirken

Was ist ein Interaktionseffekt?

- Standardmodell geht davon aus, daß unabhängige Variablen additiv zusammenwirken
- Effekt von x_1 vom Niveau von $x_2 \ldots$ unabhängig

Was ist ein Interaktionseffekt?

- Standardmodell geht davon aus, daß unabhängige Variablen additiv zusammenwirken
- Effekt von x_1 vom Niveau von $x_2 \ldots$ unabhängig
- Diese Annahme wird bei Interaktionen aufgegeben:

Was ist ein Interaktionseffekt?

- Standardmodell geht davon aus, daß unabhängige Variablen additiv zusammenwirken
- Effekt von x_1 vom Niveau von $x_2 \ldots$ unabhängig
- Diese Annahme wird bei Interaktionen aufgegeben:
  - Effekt von x_1 hängt vom Niveau von x_2 ab und umgekehrt

Was ist ein Interaktionseffekt?

- Standardmodell geht davon aus, daß unabhängige Variablen additiv zusammenwirken
- Effekt von x_1 vom Niveau von $x_2 \ldots$ unabhängig
- Diese Annahme wird bei Interaktionen aufgegeben:
  - Effekt von x_1 hängt vom Niveau von x_2 ab und umgekehrt
  - Durch Produktterm ($x_1 \times x_2$) modelliert

Was ist ein Interaktionseffekt?

- Standardmodell geht davon aus, daß unabhängige Variablen additiv zusammenwirken
- Effekt von x_1 vom Niveau von $x_2 \ldots$ unabhängig
- Diese Annahme wird bei Interaktionen aufgegeben:
  - Effekt von x_1 hängt vom Niveau von x_2 ab und umgekehrt
  - Durch Produktterm $(x_1 \times x_2)$ modelliert
  - b_1 und b_2 müssen jetzt konditional interpretiert werden

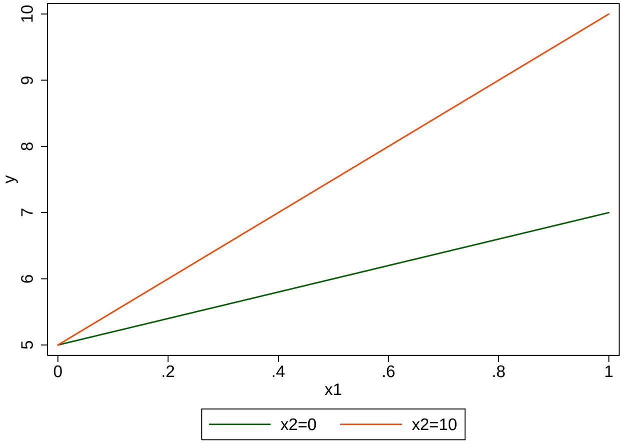
Was ist ein Interaktionseffekt?

- Standardmodell geht davon aus, daß unabhängige Variablen additiv zusammenwirken
- Effekt von x_1 vom Niveau von $x_2 \ldots$ unabhängig
- Diese Annahme wird bei Interaktionen aufgegeben:
  - Effekt von x_1 hängt vom Niveau von x_2 ab und umgekehrt
  - Durch Produktterm ($x_1 \times x_2$) modelliert
  - b_1 und b_2 müssen jetzt konditional interpretiert werden
  - b_1 entspricht der Wirkung von x_1 wenn $x_2 = 0$ und umgekehrt

Ein Beispiel?

- $y = 5 + 2x_1 + 3x_2 + 0,3x_1x_2 + \epsilon$

Ein Beispiel?



Signifikanz von Interaktionseffekten

- Koeffizient für multiplikativen Interaktionsterm = 0 → keine Interaktion in GG?
- Koeffizient entspricht Differenz zwischen Effekt von x_1 für verschiedene Niveaus von x_2 (oder umgekehrt)
- Beliebige lineare Transformationen von x_1 und x_2 zulässig → Signifikanz des Koeffizienten kann beliebig manipuliert werden
- Lösung: Stärke und Signifikanz des Effekte von x_1 für verschiedene Niveaus von x_2 berechnen/plotten und umgekehrt

Was sind nicht-lineare Effekte?

- In manchen (wenigen) Fällen ist die Linearitätsannahme offensichtlich unplausibel

Was sind nicht-lineare Effekte?

- In manchen (wenigen) Fällen ist die Linearitätsannahme offensichtlich unplausibel
- Z. B. kurvilinearer Zusammenhang zwischen Alter und Rechtsextremismus

Was sind nicht-lineare Effekte?

- In manchen (wenigen) Fällen ist die Linearitätsannahme offensichtlich unplausibel
- Z. B. kurvilinearer Zusammenhang zwischen Alter und Rechtsextremismus
- Wenn gute theoretische Begründung vorhanden, können Transformationen von y und/oder x sinnvoll sein, die den Zusammenhang zwischen beiden linearisieren

Was sind nicht-lineare Effekte?

- In manchen (wenigen) Fällen ist die Linearitätsannahme offensichtlich unplausibel
- Z. B. kurvilinearer Zusammenhang zwischen Alter und Rechtsextremismus
- Wenn gute theoretische Begründung vorhanden, können Transformationen von y und/oder x sinnvoll sein, die den Zusammenhang zwischen beiden linearisieren
- In diesem Fall ist OLS unproblematisch

Was sind nicht-lineare Effekte?

- In manchen (wenigen) Fällen ist die Linearitätsannahme offensichtlich unplausibel
- Z. B. kurvilinearer Zusammenhang zwischen Alter und Rechtsextremismus
- Wenn gute theoretische Begründung vorhanden, können Transformationen von y und/oder x sinnvoll sein, die den Zusammenhang zwischen beiden linearisieren
- In diesem Fall ist OLS unproblematisch
- Verwendet werden normalerweise das Quadrat, die Quadratwurzel, deren Kehrwerte und der natürliche Logarithmus („ladder of powers“)

Was sind nicht-lineare Effekte?

- In manchen (wenigen) Fällen ist die Linearitätsannahme offensichtlich unplausibel
- Z. B. kurvilinearer Zusammenhang zwischen Alter und Rechtsextremismus
- Wenn gute theoretische Begründung vorhanden, können Transformationen von y und/oder x sinnvoll sein, die den Zusammenhang zwischen beiden linearisieren
- In diesem Fall ist OLS unproblematisch
- Verwendet werden normalerweise das Quadrat, die Quadratwurzel, deren Kehrwerte und der natürliche Logarithmus („ladder of powers")
- Tendenziell: Vorsicht

Was wird wie geschätzt?

- OLS ist zunächst ein Verfahren, um eine Linie / Fläche / Hyperfläche durch eine Punktewolke zu legen

Was wird wie geschätzt?

- OLS ist zunächst ein Verfahren, um eine Linie / Fläche / Hyperfläche durch eine Punktewolke zu legen
- Wenn Voraussetzungen erfüllt ist, ist OLS darüber hinaus ein gutes *Schätzverfahren*

Was wird wie geschätzt?

- OLS ist zunächst ein Verfahren, um eine Linie / Fläche / Hyperfläche durch eine Punktewolke zu legen
- Wenn Voraussetzungen erfüllt ist, ist OLS darüber hinaus ein gutes *Schätzverfahren*
- Schluß von Stichprobe auf Grundgesamtheit

Was wird wie geschätzt?

- OLS ist zunächst ein Verfahren, um eine Linie / Fläche / Hyperfläche durch eine Punktewolke zu legen
- Wenn Voraussetzungen erfüllt ist, ist OLS darüber hinaus ein gutes *Schätzverfahren*
- Schluß von Stichprobe auf Grundgesamtheit
- Das Stichprobenwerte als Schätzung für Paramter der Grundgesamtheit dienen können, ist nicht selbstverstländlich

Was wird wie geschätzt?

- OLS ist zunächst ein Verfahren, um eine Linie / Fläche / Hyperfläche durch eine Punktewolke zu legen
- Wenn Voraussetzungen erfüllt ist, ist OLS darüber hinaus ein gutes *Schätzverfahren*
- Schluß von Stichprobe auf Grundgesamtheit
- Das Stichprobenwerte als Schätzung für Paramter der Grundgesamtheit dienen können, ist nicht selbstverstländlich
- Z. B. unterschätzt Stichprobenvarianz Varianz in der Grundgesamtheit

Was sind nochmal Standardfehler?

- Gedankenexperiment: Aus einer großen Grundgesamtheit immer wieder unter essentiell identischen Bedingungen Stichproben gleicher Größe ziehen

Was sind nochmal Standardfehler?

- Gedankenexperiment: Aus einer großen Grundgesamtheit immer wieder unter essentiell identischen Bedingungen Stichproben gleicher Größe ziehen
- Mit OLS Koeffizienten des Modell berechnen

Was sind nochmal Standardfehler?

- Gedankenexperiment: Aus einer großen Grundgesamtheit immer wieder unter essentiell identischen Bedingungen Stichproben gleicher Größe ziehen
- Mit OLS Koeffizienten des Modell berechnen
- Über eine unendliche Zahl von Wiederholungen hinweg *Verteilung* (mit Mittelwert, Varianz) für jeden Parameter

Was sind nochmal Standardfehler?

- Gedankenexperiment: Aus einer großen Grundgesamtheit immer wieder unter essentiell identischen Bedingungen Stichproben gleicher Größe ziehen
- Mit OLS Koeffizienten des Modell berechnen
- Über eine unendliche Zahl von Wiederholungen hinweg *Verteilung* (mit Mittelwert, Varianz) für jeden Parameter
- Außerdem Kovarianzen zwischen den Schätzungen, wenn diese nicht völlig unabhängig voneinander sind

Was sind nochmal Standardfehler?

- Gedankenexperiment: Aus einer großen Grundgesamtheit immer wieder unter essentiell identischen Bedingungen Stichproben gleicher Größe ziehen
- Mit OLS Koeffizienten des Modell berechnen
- Über eine unendliche Zahl von Wiederholungen hinweg *Verteilung* (mit Mittelwert, Varianz) für jeden Parameter
- Außerdem Kovarianzen zwischen den Schätzungen, wenn diese nicht völlig unabhängig voneinander sind
- Zu jeder Modellschätzung gehört Varianz-Kovarianz-Matrix

Was sind nochmal Standardfehler?

- Gedankenexperiment: Aus einer großen Grundgesamtheit immer wieder unter essentiell identischen Bedingungen Stichproben gleicher Größe ziehen
- Mit OLS Koeffizienten des Modell berechnen
- Über eine unendliche Zahl von Wiederholungen hinweg *Verteilung* (mit Mittelwert, Varianz) für jeden Parameter
- Außerdem Kovarianzen zwischen den Schätzungen, wenn diese nicht völlig unabhängig voneinander sind
- Zu jeder Modellschätzung gehört Varianz-Kovarianz-Matrix
- Standardfehler: Quadratwurzel aus Varianz des Parameters (über unendlich viele Stichproben hinweg)

Von was hängt Standardfehler ab?

- Formel für Varianz der Schätzungen im bivariaten Fall:

Von was hängt Standardfehler ab?

- Formel für Varianz der Schätzungen im bivariaten Fall:
  $V(b_1) = \frac{\sigma^2_\epsilon}{\sum(x_i - \bar{x})^2}$

Von was hängt Standardfehler ab?

- Formel für Varianz der Schätzungen im bivariaten Fall: $V(b_1) = \frac{\sigma_\epsilon^2}{\sum(x_i - \bar{x})^2}$
- Was bedeutet das in Worten?

Von was hängt Standardfehler ab?

- Formel für Varianz der Schätzungen im bivariaten Fall:
  $V(b_1) = \frac{\sigma_\epsilon^2}{\sum(x_i - \bar{x})^2}$
- Was bedeutet das in Worten?
- Varianz / Standardfehler umso größer, je größer Varianz von ϵ

Von was hängt Standardfehler ab?

- Formel für Varianz der Schätzungen im bivariaten Fall:
  $V(b_1) = \frac{\sigma_\epsilon^2}{\sum(x_i - \bar{x})^2}$
- Was bedeutet das in Worten?
- Varianz / Standardfehler umso größer, je größer Varianz von ϵ
- Wenn $V(\epsilon) = 0$ liegen in der Grundgesamtheit alle Punkte exakt auf der Geraden → kein Stichprobenfehler möglich

Von was hängt Standardfehler ab?

- Formel für Varianz der Schätzungen im bivariaten Fall: $V(b_1) = \frac{\sigma_\epsilon^2}{\sum(x_i - \bar{x})^2}$
- Was bedeutet das in Worten?
- Varianz / Standardfehler umso größer, je größer Varianz von ϵ
- Wenn $V(\epsilon) = 0$ liegen in der Grundgesamtheit alle Punkte exakt auf der Geraden → kein Stichprobenfehler möglich
- Varianz / Standardfehler umso kleiner, je größer die SAQ_x

Von was hängt Standardfehler ab?

- Formel für Varianz der Schätzungen im bivariaten Fall: $V(b_1) = \frac{\sigma_\epsilon^2}{\sum(x_i - \bar{x})^2}$
- Was bedeutet das in Worten?
- Varianz / Standardfehler umso größer, je größer Varianz von ϵ
- Wenn $V(\epsilon) = 0$ liegen in der Grundgesamtheit alle Punkte exakt auf der Geraden → kein Stichprobenfehler möglich
- Varianz / Standardfehler umso kleiner, je größer die SAQ_x
  - Präzisere Schätzungen mit größeren Stichproben

Von was hängt Standardfehler ab?

- Formel für Varianz der Schätzungen im bivariaten Fall: $\mathrm{V}(b_1) = \frac{\sigma_\epsilon^2}{\sum(x_i - \bar{x})^2}$
- Was bedeutet das in Worten?
- Varianz / Standardfehler umso größer, je größer Varianz von ϵ
- Wenn $\mathrm{V}(\epsilon) = 0$ liegen in der Grundgesamtheit alle Punkte exakt auf der Geraden ➔ kein Stichprobenfehler möglich
- Varianz / Standardfehler umso kleiner, je größer die SAQ_x
  - Präzisere Schätzungen mit größeren Stichproben
  - Präzisere Schätzungen, wenn mehr Varianz von x (mehr Information)

Von was hängt Standardfehler ab?

- Formel für Varianz der Schätzungen im bivariaten Fall: $V(b_1) = \frac{\sigma_\epsilon^2}{\sum(x_i - \bar{x})^2}$
- Was bedeutet das in Worten?
- Varianz / Standardfehler umso größer, je größer Varianz von ϵ
- Wenn $V(\epsilon) = 0$ liegen in der Grundgesamtheit alle Punkte exakt auf der Geraden → kein Stichprobenfehler möglich
- Varianz / Standardfehler umso kleiner, je größer die SAQ_x
  - Präzisere Schätzungen mit größeren Stichproben
  - Präzisere Schätzungen, wenn mehr Varianz von x (mehr Information)
  - Keine Schätzung möglich, wenn x nicht variiert

Von was hängt Standardfehler ab?

- Formel für die Varianz der Schätzungen im multivariaten Fall:

Von was hängt Standardfehler ab?

- Formel für die Varianz der Schätzungen im multivariaten Fall: $\mathrm{V}(b_j) = \frac{1}{1-R_j^2} \times \frac{\sigma_\epsilon^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$ mit $j \neq 0$

Von was hängt Standardfehler ab?

- Formel für die Varianz der Schätzungen im multivariaten Fall: $\mathrm{V}(b_j) = \frac{1}{1-R_j^2} \times \frac{\sigma_\epsilon^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$ mit $j \neq 0$
- beziehungsweise in Matrix-Schreibweise $\mathbf{V} = \sigma_\epsilon^2 \times (\mathbf{X}'\mathbf{X})^{-1}$

Von was hängt Standardfehler ab?

- Formel für die Varianz der Schätzungen im multivariaten Fall: $\mathrm{V}(b_j) = \frac{1}{1-R_j^2} \times \frac{\sigma_\epsilon^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$ mit $j \neq 0$
- beziehungsweise in Matrix-Schreibweise $\mathbf{V} = \sigma_\epsilon^2 \times (\mathbf{X}'\mathbf{X})^{-1}$
- **V** ist die Varianz-Kovarianz-Matrix, Quadrate der Standardfehler auf der Hauptdiagonalen

Von was hängt Standardfehler ab?

- Formel für die Varianz der Schätzungen im multivariaten Fall: $V(b_j) = \frac{1}{1-R_j^2} \times \frac{\sigma_\epsilon^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$ mit $j \neq 0$
- beziehungsweise in Matrix-Schreibweise $\mathbf{V} = \sigma_\epsilon^2 \times (\mathbf{X}'\mathbf{X})^{-1}$
- **V** ist die Varianz-Kovarianz-Matrix, Quadrate der Standardfehler auf der Hauptdiagonalen
- Was bedeutet das in Worten?

Von was hängt Standardfehler ab?

- Formel für die Varianz der Schätzungen im multivariaten Fall: $V(b_j) = \frac{1}{1-R_j^2} \times \frac{\sigma_\epsilon^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$ mit $j \neq 0$
- beziehungsweise in Matrix-Schreibweise $\mathbf{V} = \sigma_\epsilon^2 \times (\mathbf{X}'\mathbf{X})^{-1}$
- **V** ist die Varianz-Kovarianz-Matrix, Quadrate der Standardfehler auf der Hauptdiagonalen
- Was bedeutet das in Worten?
- Wenn ein x mit allen anderen x unkorreliert ist, bleibt alles wie zuvor

Von was hängt Standardfehler ab?

- Formel für die Varianz der Schätzungen im multivariaten Fall: $V(b_j) = \frac{1}{1-R_j^2} \times \frac{\sigma_\epsilon^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$ mit $j \neq 0$
- beziehungsweise in Matrix-Schreibweise $\mathbf{V} = \sigma_\epsilon^2 \times (\mathbf{X}'\mathbf{X})^{-1}$
- **V** ist die Varianz-Kovarianz-Matrix, Quadrate der Standardfehler auf der Hauptdiagonalen
- Was bedeutet das in Worten?
- Wenn ein x mit allen anderen x unkorreliert ist, bleibt alles wie zuvor
- Ansonsten Standardfehler umso größer, je enger lineare Zusammenhänge zwischen den x

Von was hängt Standardfehler ab?

- Formel für die Varianz der Schätzungen im multivariaten Fall: $\mathrm{V}(b_j) = \frac{1}{1-R_j^2} \times \frac{\sigma_\epsilon^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$ mit $j \neq 0$
- beziehungsweise in Matrix-Schreibweise $\mathbf{V} = \sigma_\epsilon^2 \times (\mathbf{X}'\mathbf{X})^{-1}$
- $\mathbf{V}$ ist die Varianz-Kovarianz-Matrix, Quadrate der Standardfehler auf der Hauptdiagonalen
- Was bedeutet das in Worten?
- Wenn ein x mit allen anderen x unkorreliert ist, bleibt alles wie zuvor
- Ansonsten Standardfehler umso größer, je enger lineare Zusammenhänge zwischen den x
- Lineare Abhängigkeiten machen Schätzungen unpräzise, im Extremfall sogar unmöglich

Welche Eigenschaften sind wichtig?

- Häufig werden drei Eigenschaften von Schätzverfahren betrachtet:

Welche Eigenschaften sind wichtig?

- Häufig werden drei Eigenschaften von Schätzverfahren betrachtet:
  1. (Asymptotische) Verzerrung (Bias)

Welche Eigenschaften sind wichtig?

- Häufig werden drei Eigenschaften von Schätzverfahren betrachtet:
  1. (Asymptotische) Verzerrung (Bias)
  2. (Asymptotische) Effizienz

Welche Eigenschaften sind wichtig?

- Häufig werden drei Eigenschaften von Schätzverfahren betrachtet:
  1. (Asymptotische) Verzerrung (Bias)
  2. (Asymptotische) Effizienz
  3. Konsistenz

Welche Eigenschaften sind wichtig?

- Häufig werden drei Eigenschaften von Schätzverfahren betrachtet:
  1. (Asymptotische) Verzerrung (Bias)
  2. (Asymptotische) Effizienz
  3. Konsistenz
  4. (Suffizienz)

Was bedeutet Verzerrung (Bias)?

- Der Mittelwert der Stichprobenkennwertverteilung (Verteilung der b ...

Was bedeutet Verzerrung (Bias)?

- Der Mittelwert der Stichprobenkennwertverteilung (Verteilung der b ...
- soll mit dem wahren Parameter β zusammenfallen

Was bedeutet Verzerrung (Bias)?

- Der Mittelwert der Stichprobenkennwertverteilung (Verteilung der b ...
- soll mit dem wahren Parameter β zusammenfallen
- Wichtig, aber nicht um jeden Preis: Was nützt geringer bias, wenn Varianz der Schätzungen sehr hoch ist?

Was bedeutet Verzerrung (Bias)?

- Der Mittelwert der Stichprobenkennwertverteilung (Verteilung der b ...
- soll mit dem wahren Parameter β zusammenfallen
- Wichtig, aber nicht um jeden Preis: Was nützt geringer bias, wenn Varianz der Schätzungen sehr hoch ist?
- Eventuell ist ein geringer bias ein akzeptabler Preis für kleine Varianz

Was bedeutet Verzerrung (Bias)?

- Der Mittelwert der Stichprobenkennwertverteilung (Verteilung der b ...
- soll mit dem wahren Parameter β zusammenfallen
- Wichtig, aber nicht um jeden Preis: Was nützt geringer bias, wenn Varianz der Schätzungen sehr hoch ist?
- Eventuell ist ein geringer bias ein akzeptabler Preis für kleine Varianz
- Mean Squared Error (MSE) $= (\text{bias})^2 + \text{V}$

Was bedeutet Verzerrung (Bias)?

- Der Mittelwert der Stichprobenkennwertverteilung (Verteilung der b ...
- soll mit dem wahren Parameter β zusammenfallen
- Wichtig, aber nicht um jeden Preis: Was nützt geringer bias, wenn Varianz der Schätzungen sehr hoch ist?
- Eventuell ist ein geringer bias ein akzeptabler Preis für kleine Varianz
- Mean Squared Error (MSE) $= (\text{bias})^2 + \text{V}$
- Asymptotisch = auf große Stichproben bezogen; asymptotisch unverzerrt = bias geht gegen null, wenn Stichprobenumfang gegen unendlich geht

Was bedeutet Verzerrung (Bias)?

- Der Mittelwert der Stichprobenkennwertverteilung (Verteilung der b ...
- soll mit dem wahren Parameter β zusammenfallen
- Wichtig, aber nicht um jeden Preis: Was nützt geringer bias, wenn Varianz der Schätzungen sehr hoch ist?
- Eventuell ist ein geringer bias ein akzeptabler Preis für kleine Varianz
- Mean Squared Error (MSE) $= (\text{bias})^2 + V$
- Asymptotisch = auf große Stichproben bezogen; asymptotisch unverzerrt = bias geht gegen null, wenn Stichprobenumfang gegen unendlich geht
- Generelles Problem: σ^2 in der Regel unbekannt, muß aus Residuen ($\mathbf{e}$) geschätzt werden

Was bedeutet Effizienz?

- Relatives Konzept

Was bedeutet Effizienz?

- Relatives Konzept
- Bezieht sich auf Varianz der Schätzungen

Was bedeutet Effizienz?

- Relatives Konzept
- Bezieht sich auf Varianz der Schätzungen
- Unter allen unverzerrten Schätzern ist der mit der geringsten Varianz der effizienteste

Was bedeutet Effizienz?

- Relatives Konzept
- Bezieht sich auf Varianz der Schätzungen
- Unter allen unverzerrten Schätzern ist der mit der geringsten Varianz der effizienteste
- Effizienz ist der Kehrwert des MSE

Was bedeutet Konsistenz?

- $\hat{\beta}_n$ ist ein Schätzer für β bei einem Stichprobenumfang von n

Was bedeutet Konsistenz?

- $\hat{\beta}_n$ ist ein Schätzer für β bei einem Stichprobenumfang von n
- Konsistenz heißt: Wenn ich für n immer größere Werte wähle …

Was bedeutet Konsistenz?

- $\hat{\beta}_n$ ist ein Schätzer für β bei einem Stichprobenumfang von n
- Konsistenz heißt: Wenn ich für n immer größere Werte wähle ...
- kann ich die Wahrscheinlichkeit, daß $\hat{\beta}_n$ um mehr als einen trivialen Betrag von β abweicht ...

Was bedeutet Konsistenz?

- $\hat{\beta}_n$ ist ein Schätzer für β bei einem Stichprobenumfang von n
- Konsistenz heißt: Wenn ich für n immer größere Werte wähle ...
- kann ich die Wahrscheinlichkeit, daß $\hat{\beta}_n$ um mehr als einen trivialen Betrag von β abweicht ...
- beliebig nahe an null heranbringen

Was bedeutet Konsistenz?

- $\hat{\beta}_n$ ist ein Schätzer für β bei einem Stichprobenumfang von n
- Konsistenz heißt: Wenn ich für n immer größere Werte wähle …
- kann ich die Wahrscheinlichkeit, daß $\hat{\beta}_n$ um mehr als einen trivialen Betrag von β abweicht …
- beliebig nahe an null heranbringen
- Wenn bias und Varianz der Schätzung bei steigender Fallzahl gegen null streben, ist das eine hinreichende Bedingung für Konsistenz

Welche Annahmen treffen wir?

- OLS ist unverzerrter und effizienter Schätzer, wenn eine Reihe von Annahmen erfüllt ist:

Welche Annahmen treffen wir?

- OLS ist unverzerrter und effizienter Schätzer, wenn eine Reihe von Annahmen erfüllt ist:

1. Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen

Welche Annahmen treffen wir?

- OLS ist unverzerrter und effizienter Schätzer, wenn eine Reihe von Annahmen erfüllt ist:

1. Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen
2. Alle unabhängigen Variablen haben Varianz

Welche Annahmen treffen wir?

- OLS ist unverzerrter und effizienter Schätzer, wenn eine Reihe von Annahmen erfüllt ist:

1. Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen
2. Alle unabhängigen Variablen haben Varianz
3. Keine perfekte Multikollineariät

Welche Annahmen treffen wir?

- OLS ist unverzerrter und effizienter Schätzer, wenn eine Reihe von Annahmen erfüllt ist:

1. Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen
2. Alle unabhängigen Variablen haben Varianz
3. Keine perfekte Multikollineariät
4. Für jede mögliche Kombination der unabhängigen Variablen ist der (konditionale) Mittelwert von $\epsilon = 0$

Welche Annahmen treffen wir?

- OLS ist unverzerrter und effizienter Schätzer, wenn eine Reihe von Annahmen erfüllt ist:

1. Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen
2. Alle unabhängigen Variablen haben Varianz
3. Keine perfekte Multikollineariät
4. Für jede mögliche Kombination der unabhängigen Variablen ist der (konditionale) Mittelwert von $\epsilon = 0$
5. Keine Kovarianz zwischen x_i und ϵ

Welche Annahmen treffen wir?

- OLS ist unverzerrter und effizienter Schätzer, wenn eine Reihe von Annahmen erfüllt ist:

1. Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen
2. Alle unabhängigen Variablen haben Varianz
3. Keine perfekte Multikollineariät
4. Für jede mögliche Kombination der unabhängigen Variablen ist der (konditionale) Mittelwert von $\epsilon = 0$
5. Keine Kovarianz zwischen x_i und ϵ
6. Für jedes beliebige Paar von Beobachtungen i und h sind ϵ_i und ϵ_h unkorreliert (keine Autokorrelation)

Welche Annahmen treffen wir?

- OLS ist unverzerrter und effizienter Schätzer, wenn eine Reihe von Annahmen erfüllt ist:

1. Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen
2. Alle unabhängigen Variablen haben Varianz
3. Keine perfekte Multikollineariät
4. Für jede mögliche Kombination der unabhängigen Variablen ist der (konditionale) Mittelwert von $\epsilon = 0$
5. Keine Kovarianz zwischen x_i und ϵ
6. Für jedes beliebige Paar von Beobachtungen i und h sind ϵ_i und ϵ_h unkorreliert (keine Autokorrelation)
7. Für jede mögliche Kombination der unabhängigen Variablen ist die (konditionale) Varianz von ϵ gleich σ^2 und damit konstant (Homoskedaszität)

Welche Annahmen treffen wir?

- OLS ist unverzerrter und effizienter Schätzer, wenn eine Reihe von Annahmen erfüllt ist:

1. Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen
2. Alle unabhängigen Variablen haben Varianz
3. Keine perfekte Multikollineariät
4. Für jede mögliche Kombination der unabhängigen Variablen ist der (konditionale) Mittelwert von $\epsilon = 0$
5. Keine Kovarianz zwischen x_i und ϵ
6. Für jedes beliebige Paar von Beobachtungen i und h sind ϵ_i und ϵ_h unkorreliert (keine Autokorrelation)
7. Für jede mögliche Kombination der unabhängigen Variablen ist die (konditionale) Varianz von ϵ gleich σ^2 und damit konstant (Homoskedaszität)
8. Für jede mögliche Kombination der unabhängigen Variablen ist ϵ normalverteilt

Was bedeutet „i.i.d.“?

- Bedingungen 6 und 7 werden auch als „independently and indentically distributed“ (i.i.d.) bezeichnet

Was bedeutet „i.i.d.“?

- Bedingungen 6 und 7 werden auch als „independently and indentically distributed“ (i.i.d.) bezeichnet
- Wie können zwei einzelne Werte eine (bzw. zwei) Verteilungen haben?
- Wie können zwei einzelne Werte („Fehler“, ϵ) eine Korrelation/Kovarianz haben?

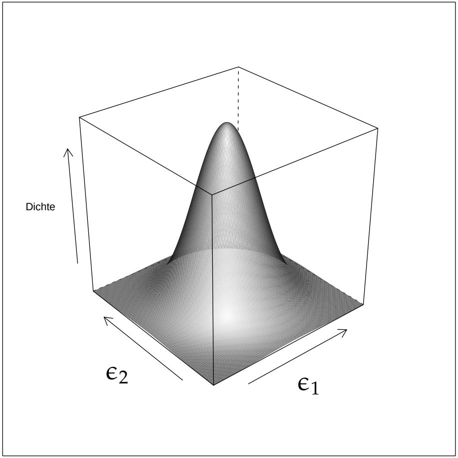
Was bedeutet „i.i.d.“?

- Bedingungen 6 und 7 werden auch als „independently and indentically distributed“ (i.i.d.) bezeichnet
- Wie können zwei einzelne Werte eine (bzw. zwei) Verteilungen haben?
- Wie können zwei einzelne Werte („Fehler“, ϵ) eine Korrelation/Kovarianz haben?
- Z.B. Fall 1 und Fall 2
  - Stichprobenziehung wiederholt → andere Fälle bzw.
  - Sozialer Prozeß läuft weiter → neue Fälle
  - Jeweils mit einem zufälligen Einfluß, der aus einer separaten Standardverteilung gezogen wird

Was bedeutet „i.i.d.“?

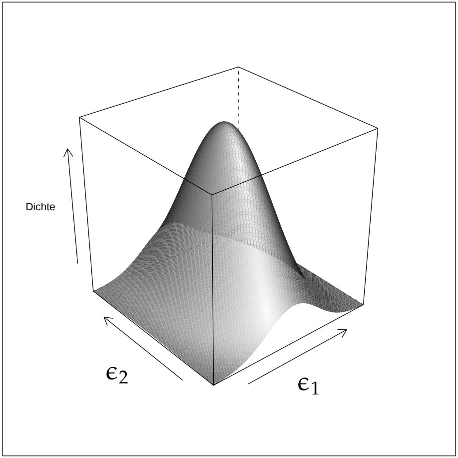
- Bedingungen 6 und 7 werden auch als „independently and indentically distributed“ (i.i.d.) bezeichnet
- Wie können zwei einzelne Werte eine (bzw. zwei) Verteilungen haben?
- Wie können zwei einzelne Werte („Fehler“, ϵ) eine Korrelation/Kovarianz haben?
- Z.B. Fall 1 und Fall 2
  - Stichprobenziehung wiederholt → andere Fälle bzw.
  - Sozialer Prozeß läuft weiter → neue Fälle
  - Jeweils mit einem zufälligen Einfluß, der aus einer separaten Standardverteilung gezogen wird
  - Diese *Verteilungen* sind unabhängig (keine Kovarianz) und haben identische Varianz

i.i.d.



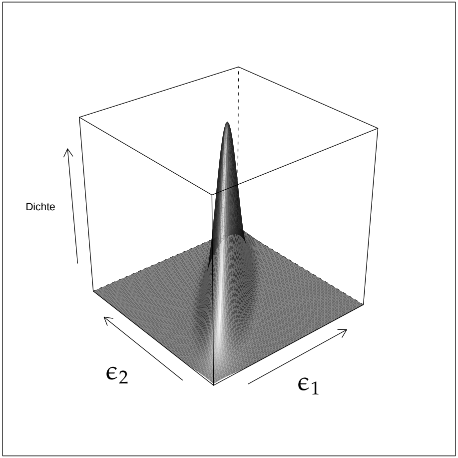
$$\hat{\epsilon}_1 = \hat{\epsilon}_2 = 0, V(\epsilon_1) = V(\epsilon_2) = 1, \mathrm{cov}(\epsilon_1, \epsilon_2) = 0$$

Heteroskedastizität



$$\hat{\epsilon}_1 = \hat{\epsilon}_2 = 0, V(\epsilon_1) = 1, V(\epsilon_2) = 3, \mathrm{cov}(\epsilon_1, \epsilon_2) = 0$$

„Fehler“ (ϵ) nicht unabhängig



$$\hat{\epsilon}_1 = \hat{\epsilon}_2 = 0, V(\epsilon_1) = V(\epsilon_2) = 1, \text{cov}(\epsilon_1, \epsilon_2) = 0.9$$

Was passiert, wenn Annahme 1 nicht erfüllt ist?

„Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen"

- Abhängige Variable hat häufig wenig diskrete Ausprägungen (Ratingskalen)

Was passiert, wenn Annahme 1 nicht erfüllt ist?

„Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen“

- Abhängige Variable hat häufig wenig diskrete Ausprägungen (Ratingskalen)
  - Erwartete Werte außerhalb des gültigen Wertebereichs

Was passiert, wenn Annahme 1 nicht erfüllt ist?

„Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen"

- Abhängige Variable hat häufig wenig diskrete Ausprägungen (Ratingskalen)
  - Erwartete Werte außerhalb des gültigen Wertebereichs
  - Annahme eines linearen Zusammenhangs, obwohl tatsächlich nicht-linear ➔ Modelle für ordinale Daten

Was passiert, wenn Annahme 1 nicht erfüllt ist?

„Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen"

- Abhängige Variable hat häufig wenig diskrete Ausprägungen (Ratingskalen)
  - Erwartete Werte außerhalb des gültigen Wertebereichs
  - Annahme eines linearen Zusammenhangs, obwohl tatsächlich nicht-linear → Modelle für ordinale Daten
  - In der Literatur wenig diskutiert, häufig wird angenommen, daß Modell relativ robust ist

Was passiert, wenn Annahme 1 nicht erfüllt ist?

„Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen“

- Abhängige Variable hat häufig wenig diskrete Ausprägungen (Ratingskalen)
  - Erwartete Werte außerhalb des gültigen Wertebereichs
  - Annahme eines linearen Zusammenhangs, obwohl tatsächlich nicht-linear ➔ Modelle für ordinale Daten
  - In der Literatur wenig diskutiert, häufig wird angenommen, daß Modell relativ robust ist
- Alle sozialwissenschaftlichen Variablen fehlerbehaftet

Was passiert, wenn Annahme 1 nicht erfüllt ist?

„Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen“

- Abhängige Variable hat häufig wenig diskrete Ausprägungen (Ratingskalen)
  - Erwartete Werte außerhalb des gültigen Wertebereichs
  - Annahme eines linearen Zusammenhangs, obwohl tatsächlich nicht-linear → Modelle für ordinale Daten
  - In der Literatur wenig diskutiert, häufig wird angenommen, daß Modell relativ robust ist
- Alle sozialwissenschaftlichen Variablen fehlerbehaftet
  - Relativ unproblematisch, wenn Fehler voneinander unabhängig

Was passiert, wenn Annahme 1 nicht erfüllt ist?

„Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen“

- Abhängige Variable hat häufig wenig diskrete Ausprägungen (Ratingskalen)
  - Erwartete Werte außerhalb des gültigen Wertebereichs
  - Annahme eines linearen Zusammenhangs, obwohl tatsächlich nicht-linear ➔ Modelle für ordinale Daten
  - In der Literatur wenig diskutiert, häufig wird angenommen, daß Modell relativ robust ist
- Alle sozialwissenschaftlichen Variablen fehlerbehaftet
  - Relativ unproblematisch, wenn Fehler voneinander unabhängig
  - Zusätzliche Unsicherheit, bei großen Stichproben kein großes Problem

Was passiert, wenn Annahme 1 nicht erfüllt ist?

„Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen“

- Abhängige Variable hat häufig wenig diskrete Ausprägungen (Ratingskalen)
  - Erwartete Werte außerhalb des gültigen Wertebereichs
  - Annahme eines linearen Zusammenhangs, obwohl tatsächlich nicht-linear → Modelle für ordinale Daten
  - In der Literatur wenig diskutiert, häufig wird angenommen, daß Modell relativ robust ist
- Alle sozialwissenschaftlichen Variablen fehlerbehaftet
  - Relativ unproblematisch, wenn Fehler voneinander unabhängig
  - Zusätzliche Unsicherheit, bei großen Stichproben kein großes Problem
  - Fehler bei y wird von ϵ absorbiert, OLS weniger effizient

Was passiert, wenn Annahme 1 nicht erfüllt ist?

„Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen"

- Abhängige Variable hat häufig wenig diskrete Ausprägungen (Ratingskalen)
  - Erwartete Werte außerhalb des gültigen Wertebereichs
  - Annahme eines linearen Zusammenhangs, obwohl tatsächlich nicht-linear → Modelle für ordinale Daten
  - In der Literatur wenig diskutiert, häufig wird angenommen, daß Modell relativ robust ist
- Alle sozialwissenschaftlichen Variablen fehlerbehaftet
  - Relativ unproblematisch, wenn Fehler voneinander unabhängig
  - Zusätzliche Unsicherheit, bei großen Stichproben kein großes Problem
  - Fehler bei y wird von ϵ absorbiert, OLS weniger effizient
  - Fehler bei x schwächt im bivariaten Fall Zusammenhang ab, multivariat auf jeden Fall bias

Was passiert, wenn Annahme 1 nicht erfüllt ist?

„Die unabhängige Variable ist intervallskaliert und unbeschränkt. Variablen werden ohne Fehler gemessen"

- Abhängige Variable hat häufig wenig diskrete Ausprägungen (Ratingskalen)
  - Erwartete Werte außerhalb des gültigen Wertebereichs
  - Annahme eines linearen Zusammenhangs, obwohl tatsächlich nicht-linear → Modelle für ordinale Daten
  - In der Literatur wenig diskutiert, häufig wird angenommen, daß Modell relativ robust ist
- Alle sozialwissenschaftlichen Variablen fehlerbehaftet
  - Relativ unproblematisch, wenn Fehler voneinander unabhängig
  - Zusätzliche Unsicherheit, bei großen Stichproben kein großes Problem
  - Fehler bei y wird von ϵ absorbiert, OLS weniger effizient
  - Fehler bei x schwächt im bivariaten Fall Zusammenhang ab, multivariat auf jeden Fall bias
  - Kann z. B. mit Strukturgleichungsmodellen explizit modelliert werden

Was passiert, wenn Annahme 2 nicht erfüllt ist?

„ Alle unabhängigen Variablen haben Varianz"

- Inhaltlich: Falls die betreffende Variable einen Effekt hat, wird dieser nicht sichtbar

Was passiert, wenn Annahme 2 nicht erfüllt ist?

„Alle unabhängigen Variablen haben Varianz"

- Inhaltlich: Falls die betreffende Variable einen Effekt hat, wird dieser nicht sichtbar
- Mathematisch: Division durch 0, Standardfehler nicht definiert

Was passiert, wenn Annahme 3 nicht erfüllt ist?

„Keine perfekte Multikollineariät“

- Oben diskutiert, keine eindeutige Lösung für Gleichungssystem

Was passiert, wenn Annahme 3 nicht erfüllt ist?

„Keine perfekte Multikollineariät"

- Oben diskutiert, keine eindeutige Lösung für Gleichungssystem
- Bei nicht perfekter Kollinearität große Standardfehler

Was passiert, wenn Annahme 3 nicht erfüllt ist?

„Keine perfekte Multikollineariät"

- Oben diskutiert, keine eindeutige Lösung für Gleichungssystem
- Bei nicht perfekter Kollinearität große Standardfehler
- Eventuell numerische Instabilität

Was passiert, wenn Annahme 3 nicht erfüllt ist?

„Keine perfekte Multikollineariät"

- Oben diskutiert, keine eindeutige Lösung für Gleichungssystem
- Bei nicht perfekter Kollinearität große Standardfehler
- Eventuell numerische Instabilität
- Eng korrelierte Variablen durch Index oder Faktor ersetzen

Was passiert, wenn Annahme 3 nicht erfüllt ist?

„Keine perfekte Multikolleariät"

- Oben diskutiert, keine eindeutige Lösung für Gleichungssystem
- Bei nicht perfekter Kollinearität große Standardfehler
- Eventuell numerische Instabilität
- Eng korrelierte Variablen durch Index oder Faktor ersetzen
- Möglichst keine überflüssigen Variablen im Modell, vor allem wenn Fallzahl klein, da OLS ansonsten ineffizient

Was passiert, wenn Annahme 3 nicht erfüllt ist?

„Keine perfekte Multikollineariät“

- Oben diskutiert, keine eindeutige Lösung für Gleichungssystem
- Bei nicht perfekter Kollinearität große Standardfehler
- Eventuell numerische Instabilität
- Eng korrelierte Variablen durch Index oder Faktor ersetzen
- Möglichst keine überflüssigen Variablen im Modell, vor allem wenn Fallzahl klein, da OLS ansonsten ineffizient
- Problem?

Was passiert, wenn Annahme 3 nicht erfüllt ist?

„Keine perfekte Multikolleariät"

- Oben diskutiert, keine eindeutige Lösung für Gleichungssystem
- Bei nicht perfekter Kollinearität große Standardfehler
- Eventuell numerische Instabilität
- Eng korrelierte Variablen durch Index oder Faktor ersetzen
- Möglichst keine überflüssigen Variablen im Modell, vor allem wenn Fallzahl klein, da OLS ansonsten ineffizient
- Problem? Möglichst alle relevanten Variablen müssen in das Modell

Was passiert, wenn Annahme 3 nicht erfüllt ist?

„Keine perfekte Multikollineariät“

- Oben diskutiert, keine eindeutige Lösung für Gleichungssystem
- Bei nicht perfekter Kollinearität große Standardfehler
- Eventuell numerische Instabilität
- Eng korrelierte Variablen durch Index oder Faktor ersetzen
- Möglichst keine überflüssigen Variablen im Modell, vor allem wenn Fallzahl klein, da OLS ansonsten ineffizient
- Problem? Möglichst alle relevanten Variablen müssen in das Modell
- Das sind Variablen, die Einfluß haben, Einfluß haben könnten, in Beziehung mit anderen relevanten Variablen stehen

Warum müssen alle relevanten Variablen ins Modell?

- Wenn x_1 und x_2 unkorreliert sind, ist Schätzung für β_1 unverzerrt, auch wenn x_2 nicht berücksichtigt wird

Warum müssen alle relevanten Variablen ins Modell?

- Wenn x_1 und x_2 unkorreliert sind, ist Schätzung für β_1 unverzerrt, auch wenn x_2 nicht berücksichtigt wird
- Aber: Wenn Korrelation zwischen x_1 und x_2 ungleich null, sind Schätzungen verzerrt (Drittvariablenproblem)

Warum müssen alle relevanten Variablen ins Modell?

- Wenn x_1 und x_2 unkorreliert sind, ist Schätzung für β_1 unverzerrt, auch wenn x_2 nicht berücksichtigt wird
- Aber: Wenn Korrelation zwischen x_1 und x_2 ungleich null, sind Schätzungen verzerrt (Drittvariablenproblem)
- Kein Probem bei echten Experimenten mit Randomisierung (warum?)

Warum müssen alle relevanten Variablen ins Modell?

- Wenn x_1 und x_2 unkorreliert sind, ist Schätzung für β_1 unverzerrt, auch wenn x_2 nicht berücksichtigt wird
- Aber: Wenn Korrelation zwischen x_1 und x_2 ungleich null, sind Schätzungen verzerrt (Drittvariablenproblem)
- Kein Probem bei echten Experimenten mit Randomisierung (warum?)
- Ansonsten: Möglichst viele mögliche unabhängige Variablen berücksichtigen?!?

Was passiert, wenn Annahme 4 nicht erfüllt ist?

„Für jede mögliche Kombination der unabhängigen Variablen ist der (konditionale) Mittelwert von $\epsilon = 0$“

- Wird u. a. verletzt, wenn Stichprobe bei Auswahl verzerrt wird (Beispiel im Text)

Was passiert, wenn Annahme 4 nicht erfüllt ist?

„Für jede mögliche Kombination der unabhängigen Variablen ist der (konditionale) Mittelwert von $\epsilon = 0$"

- Wird u. a. verletzt, wenn Stichprobe bei Auswahl verzerrt wird (Beispiel im Text)
- Falls Mittelwert von ϵ für alle Kombinationen um einen konstanten Betrag größer oder kleiner als null, wird Differenz von Konstante absorbiert → verzerrte Schätzung für Konstante

Was passiert, wenn Annahme 4 nicht erfüllt ist?

„Für jede mögliche Kombination der unabhängigen Variablen ist der (konditionale) Mittelwert von $\epsilon = 0$"

- Wird u. a. verletzt, wenn Stichprobe bei Auswahl verzerrt wird (Beispiel im Text)
- Falls Mittelwert von ϵ für alle Kombinationen um einen konstanten Betrag größer oder kleiner als null, wird Differenz von Konstante absorbiert ➔ verzerrte Schätzung für Konstante
- Ansonsten Linearitätsannahme verletzt, relevante Variable ausgelassen beziehungsweise Korrelation zwischen ϵ und x

Was passiert, wenn Annahme 4 nicht erfüllt ist?

„Für jede mögliche Kombination der unabhängigen Variablen ist der (konditionale) Mittelwert von $\epsilon = 0$“

- Wird u. a. verletzt, wenn Stichprobe bei Auswahl verzerrt wird (Beispiel im Text)
- Falls Mittelwert von ϵ für alle Kombinationen um einen konstanten Betrag größer oder kleiner als null, wird Differenz von Konstante absorbiert → verzerrte Schätzung für Konstante
- Ansonsten Linearitätsannahme verletzt, relevante Variable ausgelassen beziehungsweise Korrelation zwischen ϵ und x
- Verzerrte Schätzungen für alle Parameter

Was passiert, wenn Annahme 5 nicht erfüllt ist?

„Keine Kovarianz zwischen x_i und ϵ“

- Vorsicht: Residuen sind nur Hilfsmittel zur Schätzung von ϵ

Was passiert, wenn Annahme 5 nicht erfüllt ist?

„Keine Kovarianz zwischen x_i und ϵ"

- Vorsicht: Residuen sind nur Hilfsmittel zur Schätzung von ϵ
- Residuen und unabhängige Variablen sind immer unkorreliert (Eigenschaft des OLS-Verfahrens) → kein Rückschluß möglich

Was passiert, wenn Annahme 5 nicht erfüllt ist?

„Keine Kovarianz zwischen x_i und ϵ"

- Vorsicht: Residuen sind nur Hilfsmittel zur Schätzung von ϵ
- Residuen und unabhängige Variablen sind immer unkorreliert (Eigenschaft des OLS-Verfahrens) → kein Rückschluß möglich
- Annahme 5 garantiert nicht erfüllt wenn wechselseitige Kausalwirkung zwischen y und x (Endogenität)

Was passiert, wenn Annahme 5 nicht erfüllt ist?

„Keine Kovarianz zwischen x_i und ϵ"

- Vorsicht: Residuen sind nur Hilfsmittel zur Schätzung von ϵ
- Residuen und unabhängige Variablen sind immer unkorreliert (Eigenschaft des OLS-Verfahrens) → kein Rückschluß möglich
- Annahme 5 garantiert nicht erfüllt wenn wechselseitige Kausalwirkung zwischen y und x (Endogenität)
  - Wenn y von x und ϵ beeinflußt wird und

Was passiert, wenn Annahme 5 nicht erfüllt ist?

„Keine Kovarianz zwischen x_i und ϵ"

- Vorsicht: Residuen sind nur Hilfsmittel zur Schätzung von ϵ
- Residuen und unabhängige Variablen sind immer unkorreliert (Eigenschaft des OLS-Verfahrens) → kein Rückschluß möglich
- Annahme 5 garantiert nicht erfüllt wenn wechselseitige Kausalwirkung zwischen y und x (Endogenität)
  - Wenn y von x und ϵ beeinflußt wird und
  - y zugleich einen Einfluß auf x hat dann garantiert

Was passiert, wenn Annahme 5 nicht erfüllt ist?

„Keine Kovarianz zwischen x_i und ϵ"

- Vorsicht: Residuen sind nur Hilfsmittel zur Schätzung von ϵ
- Residuen und unabhängige Variablen sind immer unkorreliert (Eigenschaft des OLS-Verfahrens) → kein Rückschluß möglich
- Annahme 5 garantiert nicht erfüllt wenn wechselseitige Kausalwirkung zwischen y und x (Endogenität)
  - Wenn y von x und ϵ beeinflußt wird und
  - y zugleich einen Einfluß auf x hat dann garantiert
  - Zusammenhang zwischen x und ϵ

Was passiert, wenn Annahme 5 nicht erfüllt ist?

„Keine Kovarianz zwischen x_i und ϵ“

- Vorsicht: Residuen sind nur Hilfsmittel zur Schätzung von ϵ
- Residuen und unabhängige Variablen sind immer unkorreliert (Eigenschaft des OLS-Verfahrens) ➔ kein Rückschluß möglich
- Annahme 5 garantiert nicht erfüllt wenn wechselseitige Kausalwirkung zwischen y und x (Endogenität)
  - Wenn y von x und ϵ beeinflußt wird und
  - y zugleich einen Einfluß auf x hat dann garantiert
  - Zusammenhang zwischen x und ϵ
- In der Konsequenz identisch mit dem Auslassen relevanter Variablen (sofern diese nicht mit den übrigen x unkorreliert sind) ➔ Schätzungen sind verzerrt

Was passiert, wenn Annahme 5 nicht erfüllt ist?

„Keine Kovarianz zwischen x_i und ϵ“

- Vorsicht: Residuen sind nur Hilfsmittel zur Schätzung von ϵ
- Residuen und unabhängige Variablen sind immer unkorreliert (Eigenschaft des OLS-Verfahrens) ➔ kein Rückschluß möglich
- Annahme 5 garantiert nicht erfüllt wenn wechselseitige Kausalwirkung zwischen y und x (Endogenität)
  - Wenn y von x und ϵ beeinflußt wird und
  - y zugleich einen Einfluß auf x hat dann garantiert
  - Zusammenhang zwischen x und ϵ
- In der Konsequenz identisch mit dem Auslassen relevanter Variablen (sofern diese nicht mit den übrigen x unkorreliert sind) ➔ Schätzungen sind verzerrt
- Spezielle Modelle/Schätzverfahren, „Instrumente“

Was passiert, wenn Annahme 6 nicht erfüllt ist?

„Für jedes beliebige Paar von Beobachtungen i und h sind ϵ_i und ϵ_h unkorreliert (keine Autokorrelation)"

- ϵ beinhaltet (1) genuin zufällige Einflüsse und (2) solche Einflüsse, die so schwach sind, daß sie als zufällig betrachtet werden

Was passiert, wenn Annahme 6 nicht erfüllt ist?

„Für jedes beliebige Paar von Beobachtungen i und h sind ϵ_i und ϵ_h unkorreliert (keine Autokorrelation)"

- ϵ beinhaltet (1) genuin zufällige Einflüsse und (2) solche Einflüsse, die so schwach sind, daß sie als zufällig betrachtet werden
- Vor allem bei Zeitreihendaten ist es unwahrscheinlich, daß Einflüsse zwischen zwei Meßpunkten vollständig verschwinden

Was passiert, wenn Annahme 6 nicht erfüllt ist?

„Für jedes beliebige Paar von Beobachtungen i und h sind ϵ_i und ϵ_h unkorreliert (keine Autokorrelation)"

- ϵ beinhaltet (1) genuin zufällige Einflüsse und (2) solche Einflüsse, die so schwach sind, daß sie als zufällig betrachtet werden
- Vor allem bei Zeitreihendaten ist es unwahrscheinlich, daß Einflüsse zwischen zwei Meßpunkten vollständig verschwinden
  - auf überdurchschnittliche Werte folgen überdurchschnittliche Werte

Was passiert, wenn Annahme 6 nicht erfüllt ist?

„Für jedes beliebige Paar von Beobachtungen i und h sind ϵ_i und ϵ_h unkorreliert (keine Autokorrelation)"

- ϵ beinhaltet (1) genuin zufällige Einflüsse und (2) solche Einflüsse, die so schwach sind, daß sie als zufällig betrachtet werden
- Vor allem bei Zeitreihendaten ist es unwahrscheinlich, daß Einflüsse zwischen zwei Meßpunkten vollständig verschwinden
  - auf überdurchschnittliche Werte folgen überdurchschnittliche Werte
  - auf unterdurchschnittliche Werte folgen unterdurchschnittliche Werte

Was passiert, wenn Annahme 6 nicht erfüllt ist?

„Für jedes beliebige Paar von Beobachtungen i und h sind ϵ_i und ϵ_h unkorreliert (keine Autokorrelation)"

- ϵ beinhaltet (1) genuin zufällige Einflüsse und (2) solche Einflüsse, die so schwach sind, daß sie als zufällig betrachtet werden
- Vor allem bei Zeitreihendaten ist es unwahrscheinlich, daß Einflüsse zwischen zwei Meßpunkten vollständig verschwinden
  - auf überdurchschnittliche Werte folgen überdurchschnittliche Werte
  - auf unterdurchschnittliche Werte folgen unterdurchschnittliche Werte
- Für zwei beliebige Meßpunkte t und $t+1$ sind ϵ_t und ϵ_{t+1} zwei Zufallsvariable mit

Was passiert, wenn Annahme 6 nicht erfüllt ist?

„Für jedes beliebige Paar von Beobachtungen i und h sind ϵ_i und ϵ_h unkorreliert (keine Autokorrelation)“

- ϵ beinhaltet (1) genuin zufällige Einflüsse und (2) solche Einflüsse, die so schwach sind, daß sie als zufällig betrachtet werden
- Vor allem bei Zeitreihendaten ist es unwahrscheinlich, daß Einflüsse zwischen zwei Meßpunkten vollständig verschwinden
  - auf überdurchschnittliche Werte folgen überdurchschnittliche Werte
  - auf unterdurchschnittliche Werte folgen unterdurchschnittliche Werte
- Für zwei beliebige Meßpunkte t und $t+1$ sind ϵ_t und ϵ_{t+1} zwei Zufallsvariable mit
  - einem Mittelwert von null

Was passiert, wenn Annahme 6 nicht erfüllt ist?

„Für jedes beliebige Paar von Beobachtungen i und h sind ϵ_i und ϵ_h unkorreliert (keine Autokorrelation)“

- ϵ beinhaltet (1) genuin zufällige Einflüsse und (2) solche Einflüsse, die so schwach sind, daß sie als zufällig betrachtet werden
- Vor allem bei Zeitreihendaten ist es unwahrscheinlich, daß Einflüsse zwischen zwei Meßpunkten vollständig verschwinden
  - auf überdurchschnittliche Werte folgen überdurchschnittliche Werte
  - auf unterdurchschnittliche Werte folgen unterdurchschnittliche Werte
- Für zwei beliebige Meßpunkte t und $t+1$ sind ϵ_t und ϵ_{t+1} zwei Zufallsvariable mit
  - einem Mittelwert von null
  - einer positiven Korrelation

Wie kann es noch zu Autokorrelation kommen?

- Wiederholte Messungen an derselben Einheit (Panel)

Wie kann es noch zu Autokorrelation kommen?

- Wiederholte Messungen an derselben Einheit (Panel)
- Räumliche Korrelation (Klumpenstichproben)

Wie kann es noch zu Autokorrelation kommen?

- Wiederholte Messungen an derselben Einheit (Panel)
- Räumliche Korrelation (Klumpenstichproben)
- Autokorrelation ist immer dann zu erwarten, wenn die Daten strukturiert sind

Wie kann es noch zu Autokorrelation kommen?

- Wiederholte Messungen an derselben Einheit (Panel)
- Räumliche Korrelation (Klumpenstichproben)
- Autokorrelation ist immer dann zu erwarten, wenn die Daten strukturiert sind
- Was sind die Konsequenzen?

Wie kann es noch zu Autokorrelation kommen?

- Wiederholte Messungen an derselben Einheit (Panel)
- Räumliche Korrelation (Klumpenstichproben)
- Autokorrelation ist immer dann zu erwarten, wenn die Daten strukturiert sind
- Was sind die Konsequenzen? Noch einen Moment Geduld

Was passiert, wenn Annahme 7 nicht erfüllt ist?

„Für jede mögliche Kombination der unabhängigen Variablen ist die (konditionale) Varianz von ϵ gleich σ^2 und damit konstant (Homoskedaszität)"

- Impliziert: Konditionale Varianz von y ist ebenfalls für jede mögliche Kombination der unabhängigen Variablen konstant gleich σ^2

Was passiert, wenn Annahme 7 nicht erfüllt ist?

„Für jede mögliche Kombination der unabhängigen Variablen ist die (konditionale) Varianz von ϵ gleich σ^2 und damit konstant (Homoskedaszität)"

- Impliziert: Konditionale Varianz von y ist ebenfalls für jede mögliche Kombination der unabhängigen Variablen konstant gleich σ^2
- Beispiele aus dem Text: Meßfehler hängt vom Niveau der unabhängigen Variablen ab (Gewicht von Personen, Entwicklungsstand von Ländern) ➔ Korrelation zwischen x und $V(\epsilon)$

Was passiert, wenn Annahme 7 nicht erfüllt ist?

„Für jede mögliche Kombination der unabhängigen Variablen ist die (konditionale) Varianz von ϵ gleich σ^2 und damit konstant (Homoskedaszität)"

- Impliziert: Konditionale Varianz von y ist ebenfalls für jede mögliche Kombination der unabhängigen Variablen konstant gleich σ^2
- Beispiele aus dem Text: Meßfehler hängt vom Niveau der unabhängigen Variablen ab (Gewicht von Personen, Entwicklungsstand von Ländern) ➔ Korrelation zwischen x und $V(\epsilon)$
- Liegt manchmal aufgrund inhaltlicher Überlegungen nahe (Beispiel Familieneinkommen / Ausgaben für Urlaub)

Was passiert, wenn Annahme 7 nicht erfüllt ist?

„Für jede mögliche Kombination der unabhängigen Variablen ist die (konditionale) Varianz von ϵ gleich σ^2 und damit konstant (Homoskedaszität)"

- Impliziert: Konditionale Varianz von y ist ebenfalls für jede mögliche Kombination der unabhängigen Variablen konstant gleich σ^2
- Beispiele aus dem Text: Meßfehler hängt vom Niveau der unabhängigen Variablen ab (Gewicht von Personen, Entwicklungsstand von Ländern) ➔ Korrelation zwischen x und $V(\epsilon)$
- Liegt manchmal aufgrund inhaltlicher Überlegungen nahe (Beispiel Familieneinkommen / Ausgaben für Urlaub)
- Bivariat als „Fächer" im Plot erkennbar

Was passiert, wenn Annahme 7 nicht erfüllt ist?

„Für jede mögliche Kombination der unabhängigen Variablen ist die (konditionale) Varianz von ϵ gleich σ^2 und damit konstant (Homoskedaszität)“

- Impliziert: Konditionale Varianz von y ist ebenfalls für jede mögliche Kombination der unabhängigen Variablen konstant gleich σ^2
- Beispiele aus dem Text: Meßfehler hängt vom Niveau der unabhängigen Variablen ab (Gewicht von Personen, Entwicklungsstand von Ländern) → Korrelation zwischen x und $V(\epsilon)$
- Liegt manchmal aufgrund inhaltlicher Überlegungen nahe (Beispiel Familieneinkommen / Ausgaben für Urlaub)
- Bivariat als „Fächer“ im Plot erkennbar
- Inhaltlich oft Interaktion zwischen eingeschlossenen / ausgeschlossenen Variablen → Respezifikation des Modells

Was folgt aus Autokorrelation und Heteroskedastizität?

- OLS-Schätzungen sind unverzerrt

Was folgt aus Autokorrelation und Heteroskedastizität?

- OLS-Schätzungen sind unverzerrt
  - Heteroskedaszität: Größere Varianz von ϵ bei hohen Werten von x, aber Mittelwert immer noch null ➔ an Linie ändert sich nichts

Was folgt aus Autokorrelation und Heteroskedastizität?

- OLS-Schätzungen sind unverzerrt
  - Heteroskedaszität: Größere Varianz von ϵ bei hohen Werten von x, aber Mittelwert immer noch null ➔ an Linie ändert sich nichts
  - (zeitliche) Autokorrelation: Wenn positive ϵ auf positive und negative ϵ auf negative folgen, wird Linie nach oben oder unten gezogen, bleibt aber im Mittel wieder unverzerrt

Was folgt aus Autokorrelation und Heteroskedastizität?

- OLS-Schätzungen sind unverzerrt
  - Heteroskedaszität: Größere Varianz von ϵ bei hohen Werten von x, aber Mittelwert immer noch null ➔ an Linie ändert sich nichts
  - (zeitliche) Autokorrelation: Wenn positive ϵ auf positive und negative ϵ auf negative folgen, wird Linie nach oben oder unten gezogen, bleibt aber im Mittel wieder unverzerrt
- Aber: Die Streuung der Schätzwerte wird größer

Was folgt aus Autokorrelation und Heteroskedastizität?

- OLS-Schätzungen sind unverzerrt
  - Heteroskedaszität: Größere Varianz von ϵ bei hohen Werten von x, aber Mittelwert immer noch null ➜ an Linie ändert sich nichts
  - (zeitliche) Autokorrelation: Wenn positive ϵ auf positive und negative ϵ auf negative folgen, wird Linie nach oben oder unten gezogen, bleibt aber im Mittel wieder unverzerrt
- Aber: Die Streuung der Schätzwerte wird größer
- D. h. die normalen Formeln unterschätzen den realen Standardfehler

Was folgt aus Autokorrelation und Heteroskedastizität?

- OLS-Schätzungen sind unverzerrt
  - Heteroskedaszität: Größere Varianz von ϵ bei hohen Werten von x, aber Mittelwert immer noch null ➔ an Linie ändert sich nichts
  - (zeitliche) Autokorrelation: Wenn positive ϵ auf positive und negative ϵ auf negative folgen, wird Linie nach oben oder unten gezogen, bleibt aber im Mittel wieder unverzerrt
- Aber: Die Streuung der Schätzwerte wird größer
- D. h. die normalen Formeln unterschätzen den realen Standardfehler
- Bei großen Stichproben ist Heteroskedaszität nur in extremen Fällen ein ernstes Problem

Was folgt aus Autokorrelation und Heteroskedastizität?

- Bei strukturierten Daten unterschätzt die Varianz der Residuen die Varianz von ϵ jedoch oft dramatisch

Was folgt aus Autokorrelation und Heteroskedastizität?

- Bei strukturierten Daten unterschätzt die Varianz der Residuen die Varianz von ϵ jedoch oft dramatisch
- Dadurch, daß Regressionslinie in Richtung der (autokorrelierten) ϵ verzerrt wird, ergeben sich relativ geringe Abweichungen

Was folgt aus Autokorrelation und Heteroskedastizität?

- Bei strukturierten Daten unterschätzt die Varianz der Residuen die Varianz von ϵ jedoch oft dramatisch
- Dadurch, daß Regressionslinie in Richtung der (autokorrelierten) ϵ verzerrt wird, ergeben sich relativ geringe Abweichungen
- Linie paßt „zu gut“

Was folgt aus Autokorrelation und Heteroskedastizität?

- Bei strukturierten Daten unterschätzt die Varianz der Residuen die Varianz von ϵ jedoch oft dramatisch
- Dadurch, daß Regressionslinie in Richtung der (autokorrelierten) ϵ verzerrt wird, ergeben sich relativ geringe Abweichungen
- Linie paßt „zu gut"
- Alternativ gewendet: Der effektive Stichprobenumfang ist sehr viel kleiner als der numerische Stichprobenumfang, weil die Fälle nicht wirklich unabhängig voneinander sind

Was tun?

- Im Standardmodell ist ϵ ein Vektor der Länge n, dessen Werte aus identischen Normalverteilungen mit dem Mittelwert null und der Varianz σ^2 stammen: $\epsilon \sim N(0, \sigma^2)$

Was tun?

- Im Standardmodell ist $\boldsymbol{\epsilon}$ ein Vektor der Länge n, dessen Werte aus identischen Normalverteilungen mit dem Mittelwert null und der Varianz σ^2 stammen: $\boldsymbol{\epsilon} \sim N(0, \sigma^2)$
- Flexibleres Modell: Diese Definition kann erweitert werden, indem man σ^2 durch eine Matrix ersetzt: $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma}_{\epsilon\epsilon})$

Was tun?

- Im Standardmodell ist $\boldsymbol{\epsilon}$ ein Vektor der Länge n, dessen Werte aus identischen Normalverteilungen mit dem Mittelwert null und der Varianz σ^2 stammen: $\boldsymbol{\epsilon} \sim N(0, \sigma^2)$
- Flexibleres Modell: Diese Definition kann erweitert werden, indem man σ^2 durch eine Matrix ersetzt: $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma}_{\epsilon\epsilon})$
- $\boldsymbol{\Sigma}_{\epsilon\epsilon} = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{pmatrix}$

Was tun?

- Die ρ stehen für die Korrelation der ϵ zwischen den Meßpunkten

Was tun?

- Die ρ stehen für die Korrelation der ϵ zwischen den Meßpunkten
- Wenn die Elemente der Diagonalen ungleich 1 sind: Heteroskedastizität

Was tun?

- Die ρ stehen für die Korrelation der ϵ zwischen den Meßpunkten
- Wenn die Elemente der Diagonalen ungleich 1 sind: Heteroskedastizität
- Problem: $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ hat zuviele Elemente, um sie aus den Daten zu schätzen

Was tun?

- Die ρ stehen für die Korrelation der ϵ zwischen den Meßpunkten
- Wenn die Elemente der Diagonalen ungleich 1 sind: Heteroskedastizität
- Problem: $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ hat zuviele Elemente, um sie aus den Daten zu schätzen
- Es wird ein einfacher Prozeß angenommen, z. B. daß alle ρ außer ρ_1 gleich null sind

Was tun?

- Die ρ stehen für die Korrelation der ϵ zwischen den Meßpunkten
- Wenn die Elemente der Diagonalen ungleich 1 sind: Heteroskedastizität
- Problem: $\mathbf{\Sigma}_{\epsilon\epsilon}$ hat zuviele Elemente, um sie aus den Daten zu schätzen
- Es wird ein einfacher Prozeß angenommen, z. B. daß alle ρ außer ρ_1 gleich null sind
- $\epsilon_t = \rho_1 \epsilon_{t-1} + \nu_t$

Was tun?

- Die ρ stehen für die Korrelation der ϵ zwischen den Meßpunkten
- Wenn die Elemente der Diagonalen ungleich 1 sind: Heteroskedastizität
- Problem: $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ hat zuviele Elemente, um sie aus den Daten zu schätzen
- Es wird ein einfacher Prozeß angenommen, z. B. daß alle ρ außer ρ_1 gleich null sind
- $\epsilon_t = \rho_1 \epsilon_{t-1} + \nu_t$
- Vielzahl spezieller Modelle

Was tun?

- Die ρ stehen für die Korrelation der ϵ zwischen den Meßpunkten
- Wenn die Elemente der Diagonalen ungleich 1 sind: Heteroskedastizität
- Problem: $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ hat zuviele Elemente, um sie aus den Daten zu schätzen
- Es wird ein einfacher Prozeß angenommen, z. B. daß alle ρ außer ρ_1 gleich null sind
- $\epsilon_t = \rho_1 \epsilon_{t-1} + \nu_t$
- Vielzahl spezieller Modelle
- Häufig sinnvoll: „Normal" schätzen, Standardfehler korrigieren (PCSE, robuste VCE)

Warum soll ϵ normalverteil sein?

- Nur notwendig, um in kleinen Stichproben Konfidenzintervalle / Signifikanztests berechnen zu können

Warum soll ϵ normalverteil sein?

- Nur notwendig, um in kleinen Stichproben Konfidenzintervalle / Signifikanztests berechnen zu können
- In großen Stichproben Verteilung der Schätzungen auch ohne Annahme 10 näherungsweise normal

Warum soll ϵ normalverteil sein?

- Nur notwendig, um in kleinen Stichproben Konfidenzintervalle / Signifikanztests berechnen zu können
- In großen Stichproben Verteilung der Schätzungen auch ohne Annahme 10 näherungsweise normal
- Normalverteilung oft ein sinnvolles Modell für die Verteilung von ϵ

Was ist das Fazit für heute?

- In der Regel verwenden wir OLS, um von einer Stichprobe auf Parameter einer Grundgesamtheit zu schließen

Was ist das Fazit für heute?

- In der Regel verwenden wir OLS, um von einer Stichprobe auf Parameter einer Grundgesamtheit zu schließen
- Als Schätzverfahren hat OLS erfreuliche Eigenschaften...

Was ist das Fazit für heute?

- In der Regel verwenden wir OLS, um von einer Stichprobe auf Parameter einer Grundgesamtheit zu schließen
- Als Schätzverfahren hat OLS erfreuliche Eigenschaften. . .
- Sofern eine Reihe von zumindest intuitiv nachvollziehbaren Voraussetzungen erfüllt ist

Was ist das Fazit für heute?

- In der Regel verwenden wir OLS, um von einer Stichprobe auf Parameter einer Grundgesamtheit zu schließen
- Als Schätzverfahren hat OLS erfreuliche Eigenschaften. . .
- Sofern eine Reihe von zumindest intuitiv nachvollziehbaren Voraussetzungen erfüllt ist
- Ansonsten zwei Probleme

Was ist das Fazit für heute?

- In der Regel verwenden wir OLS, um von einer Stichprobe auf Parameter einer Grundgesamtheit zu schließen
- Als Schätzverfahren hat OLS erfreuliche Eigenschaften. . .
- Sofern eine Reihe von zumindest intuitiv nachvollziehbaren Voraussetzungen erfüllt ist
- Ansonsten zwei Probleme
  - Verzerrte Schätzungen

Was ist das Fazit für heute?

- In der Regel verwenden wir OLS, um von einer Stichprobe auf Parameter einer Grundgesamtheit zu schließen
- Als Schätzverfahren hat OLS erfreuliche Eigenschaften. . .
- Sofern eine Reihe von zumindest intuitiv nachvollziehbaren Voraussetzungen erfüllt ist
- Ansonsten zwei Probleme
  - Verzerrte Schätzungen
  - Zu optimistische Standardfehler

Übung für heute

- Starten Sie Stata

Übung für heute

- Starten Sie Stata
- Geben Sie folgende Befehle ein:

Übung für heute

- Starten Sie Stata
- Geben Sie folgende Befehle ein:
- `webuse set „http://www.kai-arzheimer.com/Lehre-Regression/“` (verwenden Sie die einfachen Anführungszeichen über der 2)

Übung für heute

- Starten Sie STATA
- Geben Sie folgende Befehle ein:
- `webuse set „http://www.kai-arzheimer.com/Lehre-Regression/“` (verwenden Sie die einfachen Anführungszeichen über der 2)
- `webuse beispiel1`

Übung für heute

- Starten Sie Stata
- Geben Sie folgende Befehle ein:
- `webuse set „http://www.kai-arzheimer.com/Lehre-Regression/“` (verwenden Sie die einfachen Anführungszeichen über der 2)
- `webuse beispiel1`
- Mit `reg y x1 x2 prod` schätzen Sie eine lineare Regression von y auf x_1, x_2 und einen Interaktionsterm

Übung für heute

- Starten Sie Stata
- Geben Sie folgende Befehle ein:
- `webuse set „http://www.kai-arzheimer.com/Lehre-Regression/“` (verwenden Sie die einfachen Anführungszeichen über der 2)
- `webuse beispiel1`
- Mit `reg y x1 x2 prod` schätzen Sie eine lineare Regression von y auf x_1, x_2 und einen Interaktionsterm
- mit `vce` erhalten Sie die Varianz-Kovarianz-Matrix

Übung für heute

- Starten Sie Stata
- Geben Sie folgende Befehle ein:
- `webuse set „http://www.kai-arzheimer.com/Lehre-Regression/“` (verwenden Sie die einfachen Anführungszeichen über der 2)
- `webuse beispiel1`
- Mit `reg y x1 x2 prod` schätzen Sie eine lineare Regression von y auf x_1, x_2 und einen Interaktionsterm
- mit `vce` erhalten Sie die Varianz-Kovarianz-Matrix
- Erläutern Sie in Ihren eigenen Worten die Bedeutung der Werte in dieser Matrix (einige Sätze genügen). Wie kommen Sie von dieser Matrix zu den von `reg...` ausgegebenen Standardfehlern?

Übung für heute

- Starten Sie STATA
- Geben Sie folgende Befehle ein:
- `webuse set „http://www.kai-arzheimer.com/Lehre-Regression/“` (verwenden Sie die einfachen Anführungszeichen über der 2)
- `webuse beispiel1`
- Mit `reg y x1 x2 prod` schätzen Sie eine lineare Regression von y auf x_1, x_2 und einen Interaktionsterm
- mit `vce` erhalten Sie die Varianz-Kovarianz-Matrix
- Erläutern Sie in Ihren eigenen Worten die Bedeutung der Werte in dieser Matrix (einige Sätze genügen). Wie kommen Sie von dieser Matrix zu den von `reg...` ausgegebenen Standardfehlern?
- Schicken Sie die Antwort bis nächsten Mittwoch per Mail an mich (subject: Übung 2)