Schätzverfahren, Annahmen und ihre Verletzungen, Standardfehler. Oder: was schiefgehen kann, geht schief

Statistik II

Wiederholung

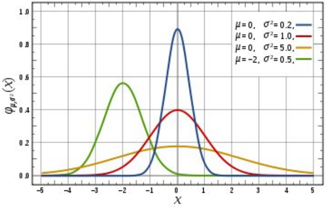
- Literatur
- Kategoriale Unabhängige,
- Interaktion, nicht-lineare Effekte

Schätzverfahren und ihre Eigenschaften

- Wiederholung: Schätzung
- Eigenschaften von Schätzverfahren

Annahmeverletzungen

Zusammenfassung



Literatur

- Zum Nachlesen für heute: Berry (1993, S. 10-45, 51, 67-83)
- Für nächste Woche: Agresti ch. 15 (Logistic Regression)

Kategoriale *unabhängige* Variablen

- Jede kategoriale Variable mit k Ausprägungen $\rightarrow k-1$ Dummy-Variablen
- z. B. Konfession (protestantisch, katholisch, andere)
  - Zwei Dummy-Variablen k und p
  - katholisch: $k = 1, p = 0$
  - protestantisch: $k = 0, p = 1$
  - andere: $k = 0, p = 0$
- Informationsverlust bei ordinalen Variablen
- Ordinale Variablen: Annahme gleicher Abstände plausibel?

Interaktion

- Definition Interaktion:

Interaktion

- Definition Interaktion: Wirkung einer Variable x_1 hängt ab vom *Niveau* anderer Variabler x_2 und umgekehrt
- Beispiel: Unterschiedlicher Effekt der Bildung für unterschiedliche ethnische Gruppen
- Modellierung:
  - Zusätzliche Produktvariable $x_1 \times x_2$ erzeugen
  - Zusätzlichen Effekt für diese Variable schätzen
- Verändert Interpretation der Haupteffekte
- Am klarsten: graphische Darstellung

Was wird wie geschätzt?

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

Was wird wie geschätzt?

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

Was wird wie geschätzt?

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

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
- Wenn $\sigma_\epsilon^2 = 0$ liegen in der Grundgesamtheit alle Punkte exakt auf der Geraden $\rightarrow$ 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 $\sigma_\epsilon^2 = 0$ liegen in der Grundgesamtheit alle Punkte exakt auf der Geraden $\rightarrow$ 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: $V(b_j) = \frac{1}{1-R_j^2} \times \frac{\sigma_\epsilon^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$ 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: $\mathrm{V}(b_j) = \frac{1}{1-R_j^2} \times \frac{\sigma_\epsilon^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$ 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
- 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 $\mathbf{V} = \sigma_\epsilon^2 \times (\mathbf{X}'\mathbf{X})^{-1}$
- $\mathbf{V}$ 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

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

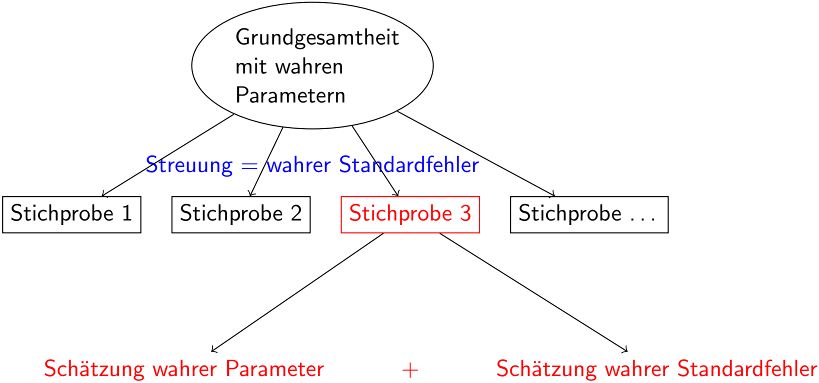
Von was hängt Standardfehler ab?

- Formel für die Varianz der Schätzungen im multivariaten Fall: $V(b_j) = \frac{1}{1-R_j^2} \times \frac{\sigma_\epsilon^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$ 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
- Lineare Abhängigkeiten machen Schätzungen unpräzise, im Extremfall sogar unmöglich (perfekte Kolllinearität)

Simulation Stichprobenziehung in Stata

- `net from http://www.kai-arzheimer.com/Statistik-II/stata/`
- `net get regsimul`

Was wird geschätzt?



Was wird geschätzt?

- Mit einer Stichprobe werden zwei Dinge geschätzt:
  1. Der/die wahren Parameter in der Grundgesamtheit (Anteilswerte, Zusammenhangsmaße, Mittelwerte, Regressionskoeffizienten etc.)
  2. Die Streuung dieser Schätzungen über wiederholte Stichprobenziehungen hinweg

Was wird geschätzt?

- Mit einer Stichprobe werden zwei Dinge geschätzt:
  1. Der/die wahren Parameter in der Grundgesamtheit (Anteilswerte, Zusammenhangsmaße, Mittelwerte, Regressionskoeffizienten etc.)
  2. Die Streuung dieser Schätzungen über wiederholte Stichprobenziehungen hinweg
- Voraussetzung: Ein „gutes“ Schätzverfahren

Welche Eigenschaften sind wichtig?

- Viele verschiedene Schätzverfahren $\rightarrow$ wie bestes wählen?
- Häufig werden drei Eigenschaften von Schätzverfahren betrachtet:
  1. (Asymptotische) Unverzerrtheit (Bias)
  2. (Asymptotische) Effizienz
  3. Konsistenz
- Idealerweise sind Schätzungen/Schätzverfahren unverzerrt, effizient und konsistent

Was bedeutet Bias?

- Der Mittelwert der Stichprobenkennwertverteilung (Verteilung der $\hat{\beta}$ ...)
- Soll mit dem wahren Parameter β zusammenfallen

Was bedeutet Bias?

- Der Mittelwert der Stichprobenkennwertverteilung (Verteilung der $\hat{\beta}$ ...)
- 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 Bias?

- Der Mittelwert der Stichprobenkennwertverteilung (Verteilung der $\hat{\beta}$ ...)
- 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
- Zussammenfassend: Mean Squared Error (MSE) $= (\text{bias}(\hat{\beta}))^2 + \text{Varianz}(\hat{\beta})$
- Asymptotisch = auf große Stichproben bezogen; asymptotisch unverzerrt = bias geht gegen null, wenn Stichprobenumfang gegen unendlich geht

Was bedeutet Effizienz?

- Relatives Konzept
- 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
- 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

Wozu das alles?

- *Wenn* bestimmte Annahmen erfüllt sind, ist OLS unverzerrt, effizient und konsistent

Wozu das alles?

- *Wenn* bestimmte Annahmen erfüllt sind, ist OLS unverzerrt, effizient und konsistent
- *Welche* Annahmen?
- *Was* passiert, wenn Annahmen verletzt?

Wozu das alles?

- *Wenn* bestimmte Annahmen erfüllt sind, ist OLS unverzerrt, effizient und konsistent
- *Welche* Annahmen?
- *Was* passiert, wenn Annahmen verletzt?
- *Welche* Alternativen?

Was sind die Standard-Annahmen?

- Zufallsstichprobe
- Wahres Modell: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \cdots + \epsilon_i$

Was sind die Standard-Annahmen?

- Zufallsstichprobe
- Wahres Modell: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \cdots + \epsilon_i$

1. Die abhängige 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 $\epsilon = 0$

Was sind die Standard-Annahmen?

- Zufallsstichprobe
- Wahres Modell: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \cdots + \epsilon_i$

5. Keine Kovarianz zwischen x_{ki} 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 (Homoskedastizität)
8. Für jede mögliche Kombination der unabhängigen Variablen ist ϵ normalverteilt

Was passiert, wenn Annahme 1 nicht erfüllt ist?

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

- Abhängige Variable hat häufig wenig diskrete Ausprägungen (Ratingskalen)
  - Erwartete Werte außerhalb des gültigen Wertebereichs
  - 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 und Stichprobe groß
  - 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 2 und 3 nicht erfüllt sind?

„Alle unabhängigen Variablen haben Varianz; keine perfekte Multikollineariät“

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

- Keine eindeutige Lösung für Gleichungssystem
- Bei nicht-perfekter Kollinearität große Standardfehler
- Möglichst keine überflüssigen Variablen im Modell, vor allem wenn Fallzahl klein, da OLS ansonsten ineffizient
- Problem: Alle relevanten Variablen müssen ins Modell (bei ex-post facto), sonst Bias

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, relevante Variable fehlt, nicht-lineare Zusammenhänge, Korrelation zwischen ϵ und x
- Falls Mittelwert von ϵ für alle Kombinationen um einen konstanten Betrag größer oder kleiner als null, wird Differenz von Konstante absorbiert $\rightarrow$ verzerrte Schätzung für Konstante
- Sonst verzerrte Schätzungen für alle Parameter

Was passiert, wenn Annahme 5 nicht erfüllt ist?

„Keine Kovarianz zwischen x_i und ϵ"

- Residuen sind nur Hilfsmittel zur Schätzung von ϵ
- Residuen und unabhängige Variablen sind immer unkorreliert (Eigenschaft des OLS-Verfahrens) $\rightarrow$ 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) $\rightarrow$ 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) schwache, nicht gemessene Einflüsse, die als zufällig betrachtet werden
- Probleme durch ...
    - Autokorrelation in Zeitreihen
    - Wiederholte Messung am selben Objekt (Panel)
    - Räumliche Korrelation
    - „Struktur" im Design $\rightarrow$ Autokorrelation
- Konsequenzen
    - Standardfehler zu optimistisch
    - Aber Schätzungen unverzerrt

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
- (Gewicht von Personen, Entwicklungsstand von Ländern, Familieneinkommen und Urlaubsausgaben) $\rightarrow$ Korrelation zwischen x und σ_ϵ^2
- Bivariat als „Fächer" im Plot erkennbar
- Inhaltlich oft Interaktion zwischen eingeschlossenen / ausgeschlossenen Variablen $\rightarrow$ Respezifikation des Modells

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 $\rightarrow$ 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
- Effektiver Stichprobenumfang sehr viel kleiner als numerischer Stichprobenumfang, weil Fälle nicht unabhängig voneinander sind
- Bei großen Stichproben ist Heteroskedaszität nur in extremen Fällen ein ernstes Problem

Was passiert, wenn Annahme 8 nicht erfüllt ist?

„Für jede mögliche Kombination der unabhängigen Variablen ist ϵ normalverteilt"

- Wenn Stichprobenumfang klein, Normalverteilung notwendig für Standardfehler und Konfidenzintervalle
- Je größer Stichprobenumfang, desto unproblematischer Abweichungen (Grenzwertsatz)
- Unabhängig von Stichprobenumfang wird Normalverteilung nur für Standardfehler benötigt – Schätzungen sind unverzerrt

Praktische Bedeutung der Annahmen?

- Typische Konstellation: relativ große Zufallsstichprobe, ex-post facto design
- Möglichst alle denkbaren Kontrollvariablen aufnehmen, sonst bias
- Heteroskedaszität nur in extremen Fällen ein Problem
- Nicht-Normalität nur in extremen Fällen ein Problem
- Autokorrelation
  - Woher?
  - Ggf. spezielle Verfahren, um korrekte Standardfehler zu bekommen, „robuste“ Standardfehler
  - Kein bias zu erwarten
- Abhängige Variable nicht kontinuierlich (z. B. Wahlabsicht) $\rightarrow$ spezielle Verfahren notwendig

Zusammenfassung

- Stichprobe $\rightarrow$ Grundgesamtheit – Schätzverfahren
- Alle Schätzverfahren setzen Zufallsstichprobe voraus
- Weitere Voraussetzungen, sonst
  - Bias
  - Ineffizienz
  - Inkonsistenz
  - Inkorrekte Standardfehler
- Problem identifizieren $\rightarrow$ geeignete Verfahren
- Nächstes Jahr: Verfahren für dichotome abhängige Variablen