Logit-Analyse mit ordinalen und nominalen abhängigen Variablen

Regressionsmodelle für Politikwissenschaftler

Übersicht

Wiederholung

Modelle für ordinale Daten
- Alternative Herleitung des binären Logit-Modells
- Proportional Odds / Parallel Lines
- Verallgemeinerung des ordinalen Modells

Modelle für nominale Daten
- Das multinomiale Logit-Modell
- Das konditionale Logit-Modell

Darstellen und bewerten
- R^2
- Standardisierte Koeffizienten
- Standardfehler und Tests
- Erwartete Wahrscheinlichkeiten

Wie war das mit der Wahlbeteiligung?

- Für einen Befragten mit mittlerer Bildung und mittlerem Interesse beträgt der *Logit* der Wahlbeteiligung $0.08 + 2 \times 0.4 + 1 \times 0.5 = 1.38$

Wie war das mit der Wahlbeteiligung?

- Für einen Befragten mit mittlerer Bildung und mittlerem Interesse beträgt der *Logit* der Wahlbeteiligung $0.08 + 2 \times 0.4 + 1 \times 0.5 = 1.38$
- Die zugehörige (Wahl-)Wahrscheinlichkeit ist $\frac{\exp(1.38)}{1+\exp(1.38)} = 0.799$

Wie war das mit der Wahlbeteiligung?

- Für einen Befragten mit mittlerer Bildung und mittlerem Interesse beträgt der *Logit* der Wahlbeteiligung $0.08 + 2 \times 0.4 + 1 \times 0.5 = 1.38$
- Die zugehörige (Wahl-)Wahrscheinlichkeit ist $\frac{\exp(1.38)}{1+\exp(1.38)} = 0.799$
- Wenn das Interesse um zwei Punkte steigt, ist der Logit $= 0.08 + 4 \times 0.4 + 1 \times 0.5 = 2.18$; die entsprechende Wahrscheinlichkeit $= 0.898$, also rund zehn Prozentpunkte höher

Wie war das mit der Wahlbeteiligung?

- Für einen Befragten mit mittlerer Bildung und mittlerem Interesse beträgt der *Logit* der Wahlbeteiligung $0.08 + 2 \times 0.4 + 1 \times 0.5 = 1.38$
- Die zugehörige (Wahl-)Wahrscheinlichkeit ist $\frac{\exp(1.38)}{1+\exp(1.38)} = 0.799$
- Wenn das Interesse um zwei Punkte steigt, ist der Logit $= 0.08 + 4 \times 0.4 + 1 \times 0.5 = 2.18$; die entsprechende Wahrscheinlichkeit $= 0.898$, also rund zehn Prozentpunkte höher
- Für einen Befragten mit hoher Bildung und mittlerem Interesse ist der Logit gleich 1.88, was einer Wahrscheinlichkeit von 86.8 Prozent entspricht

Wie war das mit der Wahlbeteiligung?

- Für einen Befragten mit mittlerer Bildung und mittlerem Interesse beträgt der *Logit* der Wahlbeteiligung $0.08 + 2 \times 0.4 + 1 \times 0.5 = 1.38$
- Die zugehörige (Wahl-)Wahrscheinlichkeit ist $\frac{\exp(1.38)}{1+\exp(1.38)} = 0.799$
- Wenn das Interesse um zwei Punkte steigt, ist der Logit $= 0.08 + 4 \times 0.4 + 1 \times 0.5 = 2.18$; die entsprechende Wahrscheinlichkeit $= 0.898$, also rund zehn Prozentpunkte höher
- Für einen Befragten mit hoher Bildung und mittlerem Interesse ist der Logit gleich 1.88, was einer Wahrscheinlichkeit von 86.8 Prozent entspricht
- Wenn bei diesem Befragten das Interesse um zwei Punkte steigt, liegt der Logit bei 2.68, was einer Wahrscheinlichkeit von 93.6 Prozent entspricht d. h. die Wahrscheinlichkeit steigt jetzt um weniger als sieben Prozentpunkte

Wie war das mit der Wahlbeteiligung?

- Für einen Befragten mit mittlerer Bildung und mittlerem Interesse beträgt der *Logit* der Wahlbeteiligung $0.08 + 2 \times 0.4 + 1 \times 0.5 = 1.38$
- Die zugehörige (Wahl-)Wahrscheinlichkeit ist $\frac{\exp(1.38)}{1+\exp(1.38)} = 0.799$
- Wenn das Interesse um zwei Punkte steigt, ist der Logit $= 0.08 + 4 \times 0.4 + 1 \times 0.5 = 2.18$; die entsprechende Wahrscheinlichkeit $= 0.898$, also rund zehn Prozentpunkte höher
- Für einen Befragten mit hoher Bildung und mittlerem Interesse ist der Logit gleich 1.88, was einer Wahrscheinlichkeit von 86.8 Prozent entspricht
- Wenn bei diesem Befragten das Interesse um zwei Punkte steigt, liegt der Logit bei 2.68, was einer Wahrscheinlichkeit von 93.6 Prozent entspricht d. h. die Wahrscheinlichkeit steigt jetzt um weniger als sieben Prozentpunkte
- Auch ohne Interaktionsterme sind die Veränderungen auf der Ebene der Wahrscheinlichkeiten nicht wie im linearen Modell konstant, sondern abhängig vom Niveau der x-Variablen

Wie sieht die alternative Herleitung des Logit-Modells aus?

- Das GLM besteht aus (1) einer systematischen Komponente $\mathbf{X}\beta$, (2) einer Verteilungsannahme für y und (3) einem „Link" zwischen dem systematischen Teil und y

Wie sieht die alternative Herleitung des Logit-Modells aus?

- Das GLM besteht aus (1) einer systematischen Komponente $\mathbf{X}\beta$, (2) einer Verteilungsannahme für y und (3) einem „Link“ zwischen dem systematischen Teil und y
- Für eine dichotome abhängige Variable ist (2) die Bernoulli-Verteilung und der „kanonische“ Link die Logit-Funktion

Wie sieht die alternative Herleitung des Logit-Modells aus?

- Das GLM besteht aus (1) einer systematischen Komponente $\mathbf{X}\beta$, (2) einer Verteilungsannahme für y und (3) einem „Link" zwischen dem systematischen Teil und y
- Für eine dichotome abhängige Variable ist (2) die Bernoulli-Verteilung und der „kanonische" Link die Logit-Funktion
- Das Logit-Modell läßt sich aber auch etwas anders herleiten

Wie sieht die alternative Herleitung des Logit-Modells aus?

- Hinter der beobachteten dichotomen Variable y (z. B. Wahlbeteiligung) steht eine latente kontinuierliche Variable y^* (z. B. „Tendenz“ zu wählen)

Wie sieht die alternative Herleitung des Logit-Modells aus?

- Hinter der beobachteten dichotomen Variable y (z. B. Wahlbeteiligung) steht eine latente kontinuierliche Variable y^* (z. B. „Tendenz“ zu wählen)
- Der konditionale Mittelwert von y^* hängt direkt von $\mathbf{X}\boldsymbol{\beta}$ ab

Wie sieht die alternative Herleitung des Logit-Modells aus?

- Hinter der beobachteten dichotomen Variable y (z. B. Wahlbeteiligung) steht eine latente kontinuierliche Variable y^* (z. B. „Tendenz“ zu wählen)
- Der konditionale Mittelwert von y^* hängt direkt von $\mathbf{X}\beta$ ab
- Die Werte von y^* folgen einer bestimmten Verteilung um den konditionalen Mittelwert, d. h. $y^* = \mu + \epsilon$

Wie sieht die alternative Herleitung des Logit-Modells aus?

- Hinter der beobachteten dichotomen Variable y (z. B. Wahlbeteiligung) steht eine latente kontinuierliche Variable y^* (z. B. „Tendenz" zu wählen)
- Der konditionale Mittelwert von y^* hängt direkt von $\mathbf{X}\beta$ ab
- Die Werte von y^* folgen einer bestimmten Verteilung um den konditionalen Mittelwert, d. h. $y^* = \mu + \epsilon$
- Für Werte von y^*, die einen Schwellenwert τ überschreiten, wird $y = 1$ beobachtet, ansonsten ist $y = 0$

Welche zusätzlichen Annahmen müssen getroffen werden?

- In dieser Form ist das Modell nicht identifiziert, d. h. β kann nicht geschätzt werden (da y^* nicht beobachtet werden kann)

Welche zusätzlichen Annahmen müssen getroffen werden?

- In dieser Form ist das Modell nicht identifiziert, d. h. β kann nicht geschätzt werden (da y^* nicht beobachtet werden kann)
- Zusätzliche Annahmen

Welche zusätzlichen Annahmen müssen getroffen werden?

- In dieser Form ist das Modell nicht identifiziert, d. h. β kann nicht geschätzt werden (da y^* nicht beobachtet werden kann)
- Zusätzliche Annahmen
  - Mittelwert von $\epsilon = 0$

Welche zusätzlichen Annahmen müssen getroffen werden?

- In dieser Form ist das Modell nicht identifiziert, d. h. β kann nicht geschätzt werden (da y^* nicht beobachtet werden kann)
- Zusätzliche Annahmen
  - Mittelwert von $\epsilon = 0$
  - $\tau = 0$

Welche zusätzlichen Annahmen müssen getroffen werden?

- In dieser Form ist das Modell nicht identifiziert, d. h. β kann nicht geschätzt werden (da y^* nicht beobachtet werden kann)
- Zusätzliche Annahmen
  - Mittelwert von $\epsilon = 0$
  - $\tau = 0$
  - Für ϵ wird entweder (1) eine Normalverteilung mit einer Varianz von 1 (Standardnormalverteilung) oder (2) eine logistische Verteilung mit einer Varianz von $\pi^2/3 \approx 3.28$ (Dichtefunktion: $\frac{\exp(x)}{(1+\exp(x))^2}$) angenommen

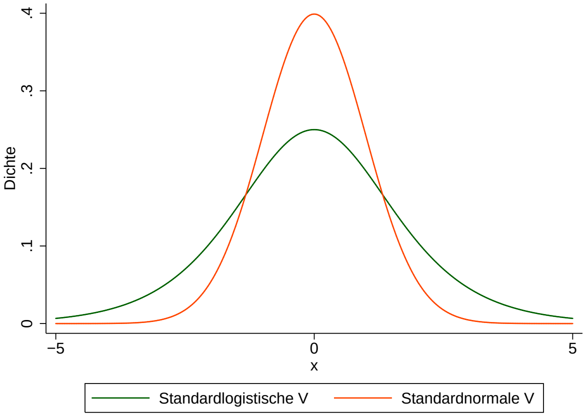
Welche zusätzlichen Annahmen müssen getroffen werden?

- In dieser Form ist das Modell nicht identifiziert, d. h. β kann nicht geschätzt werden (da y^* nicht beobachtet werden kann)
- Zusätzliche Annahmen
  - Mittelwert von $\epsilon = 0$
  - $\tau = 0$
  - Für ϵ wird entweder (1) eine Normalverteilung mit einer Varianz von 1 (Standardnormalverteilung) oder (2) eine logistische Verteilung mit einer Varianz von $\pi^2/3 \approx 3.28$ (Dichtefunktion: $\frac{\exp(x)}{(1+\exp(x))^2}$) angenommen
- Bei Fall (1) erhält man das Probit-, bei Fall (2) das Logit-Modell

Welche zusätzlichen Annahmen müssen getroffen werden?

- In dieser Form ist das Modell nicht identifiziert, d. h. β kann nicht geschätzt werden (da y^* nicht beobachtet werden kann)
- Zusätzliche Annahmen
  - Mittelwert von $\epsilon = 0$
  - $\tau = 0$
  - Für ϵ wird entweder (1) eine Normalverteilung mit einer Varianz von 1 (Standardnormalverteilung) oder (2) eine logistische Verteilung mit einer Varianz von $\pi^2/3 \approx 3.28$ (Dichtefunktion: $\frac{\exp(x)}{(1+\exp(x))^2}$) angenommen
- Bei Fall (1) erhält man das Probit-, bei Fall (2) das Logit-Modell
- Beide Herleitungen des logistischen Modells sind vollständig äquivalent

Wie sieht die Standardlogistische Verteilung aus?



Wie sieht die Standardlogistische Verteilung aus?

- Die Standardlogistische Verteilung ähnelt der Normalverteilung, hat aber etwas größere Wahrscheinlichkeiten für niedrige / höhere Werte

Wie sieht die Standardlogistische Verteilung aus?

- Die Standardlogistische Verteilung ähnelt der Normalverteilung, hat aber etwas größere Wahrscheinlichkeiten für niedrige / höhere Werte
- Die kumulative Dichtefunktion ist $\frac{\exp(x)}{1+\exp(x)}$, was der Umkehrfunktion der Logit-Transformation entspricht

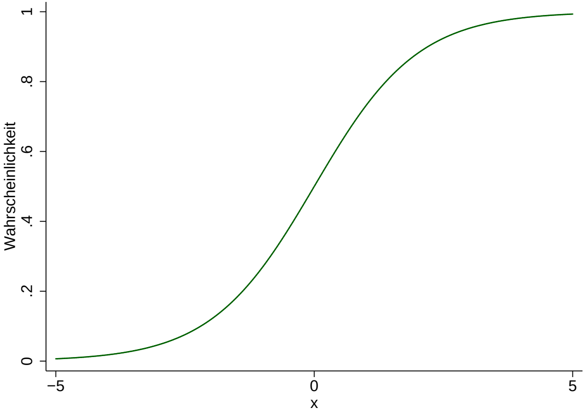
Wie sieht die Standardlogistische Verteilung aus?

- Die Standardlogistische Verteilung ähnelt der Normalverteilung, hat aber etwas größere Wahrscheinlichkeiten für niedrige / höhere Werte
- Die kumulative Dichtefunktion ist $\frac{\exp(x)}{1+\exp(x)}$, was der Umkehrfunktion der Logit-Transformation entspricht
- Mit anderen Worten: Auf diese Weise erhält man wieder den Logit-Link

Wie sieht die Standardlogistische Verteilung aus?

- Die Standardlogistische Verteilung ähnelt der Normalverteilung, hat aber etwas größere Wahrscheinlichkeiten für niedrige / höhere Werte
- Die kumulative Dichtefunktion ist $\frac{\exp(x)}{1+\exp(x)}$, was der Umkehrfunktion der Logit-Transformation entspricht
- Mit anderen Worten: Auf diese Weise erhält man wieder den Logit-Link
- Latente Variable mit standardlogistischer Verteilung und Schwellenwert $\Leftrightarrow$ beobachtete Variable mit Bernoulli-Verteilung und logistischem Link

Wie sieht die kumulative Dichtefunktion der standardlogistischen Verteilung aus?



Wie läßt sich das Modell für ordinale Daten erweitern?

- Ordinale abhängige Variable

Wie läßt sich das Modell für ordinale Daten erweitern?

- Ordinale abhängige Variable
  - Mehrere Ausprägungen

Wie läßt sich das Modell für ordinale Daten erweitern?

- Ordinale abhängige Variable
  - Mehrere Ausprägungen
  - Zwischen den Größer-Kleiner-Relation besteht

Wie läßt sich das Modell für ordinale Daten erweitern?

- Ordinale abhängige Variable
  - Mehrere Ausprägungen
  - Zwischen den Größer-Kleiner-Relation besteht
  - Abstände zwischen den Ausprägungen aber nicht konstant

Wie läßt sich das Modell für ordinale Daten erweitern?

- Ordinale abhängige Variable
  - Mehrere Ausprägungen
  - Zwischen den Größer-Kleiner-Relation besteht
  - Abstände zwischen den Ausprägungen aber nicht konstant
- Läßt sich sehr gut als Schwellenwert-Modell formulieren

Wie war das mit dem Beispiel im Text?

- Im dichotomen Fall haben wir einen Schwellenwert, der auf null gesetzt wird

Wie war das mit dem Beispiel im Text?

- Im dichotomen Fall haben wir einen Schwellenwert, der auf null gesetzt wird
- Bei vier Ausprägungen SD (1), D (2), A (3), SA (4) braucht man fünf Schwellenwerte $\tau_0 \ldots \tau_4$, von denen der erste und der letzte $+\infty$ beziehungsweise $-\infty$ entspricht, d. h. eigentlich nur drei (vgl. Long: 116

Wie war das mit dem Beispiel im Text?

- Im dichotomen Fall haben wir einen Schwellenwert, der auf null gesetzt wird
- Bei vier Ausprägungen SD (1), D (2), A (3), SA (4) braucht man fünf Schwellenwerte $\tau_0 \ldots \tau_4$, von denen der erste und der letzte $+\infty$ beziehungsweise $-\infty$ entspricht, d. h. eigentlich nur drei (vgl. Long: 116
- Diese drei legen fest, welcher Wert von y^* welcher Antwort entspricht (Abstände nicht gleich, Long: 117)

Wie war das mit dem Beispiel im Text?

- Im dichotomen Fall haben wir einen Schwellenwert, der auf null gesetzt wird
- Bei vier Ausprägungen SD (1), D (2), A (3), SA (4) braucht man fünf Schwellenwerte $\tau_0 \ldots \tau_4$, von denen der erste und der letzte $+\infty$ beziehungsweise $-\infty$ entspricht, d. h. eigentlich nur drei (vgl. Long: 116
- Diese drei legen fest, welcher Wert von y^* welcher Antwort entspricht (Abstände nicht gleich, Long: 117)
- Für ϵ kann wieder eine standardlogistische Verteilung (oder eine Standardnormalverteilung) angenommen werden

Wie war das mit dem Beispiel im Text?

- Im dichotomen Fall haben wir einen Schwellenwert, der auf null gesetzt wird
- Bei vier Ausprägungen SD (1), D (2), A (3), SA (4) braucht man fünf Schwellenwerte $\tau_0 \ldots \tau_4$, von denen der erste und der letzte $+\infty$ beziehungsweise $-\infty$ entspricht, d. h. eigentlich nur drei (vgl. Long: 116
- Diese drei legen fest, welcher Wert von y^* welcher Antwort entspricht (Abstände nicht gleich, Long: 117)
- Für ϵ kann wieder eine standardlogistische Verteilung (oder eine Standardnormalverteilung) angenommen werden
- Die Wahrscheinlichkeit, daß y einen bestimmten Wert annimmt, entspricht der Wahrscheinlichkeit, daß y^* zwischen zwei Schwellenwerten liegt

Wie war das mit dem Beispiel im Text?

- Im dichotomen Fall haben wir einen Schwellenwert, der auf null gesetzt wird
- Bei vier Ausprägungen SD (1), D (2), A (3), SA (4) braucht man fünf Schwellenwerte $\tau_0 \ldots \tau_4$, von denen der erste und der letzte $+\infty$ beziehungsweise $-\infty$ entspricht, d. h. eigentlich nur drei (vgl. Long: 116
- Diese drei legen fest, welcher Wert von y^* welcher Antwort entspricht (Abstände nicht gleich, Long: 117)
- Für ϵ kann wieder eine standardlogistische Verteilung (oder eine Standardnormalverteilung) angenommen werden
- Die Wahrscheinlichkeit, daß y einen bestimmten Wert annimmt, entspricht der Wahrscheinlichkeit, daß y^* zwischen zwei Schwellenwerten liegt
- Diese wiederum entspricht der Differenz zwischen den beiden Werten der zugehörigen kumulativen Dichtefunktion (im Text: F) an diesen beiden Stellen

Wie können die Modellparameter geschätzt werden?

- Um das Modell zu identifizieren, muß wieder eine Annahme über die Varianz getroffen und entweder der Achsenabschnitt (STATA) oder τ_1 auf null gesetzt werde (kein Einfluß auf geschätzte Wahrscheinlichkeiten)

Wie können die Modellparameter geschätzt werden?

- Um das Modell zu identifizieren, muß wieder eine Annahme über die Varianz getroffen und entweder der Achsenabschnitt (Stata) oder τ_1 auf null gesetzt werde (kein Einfluß auf geschätzte Wahrscheinlichkeiten)
- Die Wahrscheinlichkeit, daß $y = m$ ist $F(\tau_m - \mathbf{x}_i\boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i\boldsymbol{\beta})$

Wie können die Modellparameter geschätzt werden?

- Um das Modell zu identifizieren, muß wieder eine Annahme über die Varianz getroffen und entweder der Achsenabschnitt (STATA) oder τ_1 auf null gesetzt werde (kein Einfluß auf geschätzte Wahrscheinlichkeiten)
- Die Wahrscheinlichkeit, daß $y = m$ ist $F(\tau_m - \mathbf{x}_i \boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i \boldsymbol{\beta})$
- D. h. für jeden Fall i gibt es eine Wahrscheinlichkeit, daß $y = m$, die von den x-Werten, dem Koeffizientenvektor β und dem Vektor mit den Schwellenwerten τ abhängt

Wie können die Modellparameter geschätzt werden?

- Um das Modell zu identifizieren, muß wieder eine Annahme über die Varianz getroffen und entweder der Achsenabschnitt (STATA) oder τ_1 auf null gesetzt werde (kein Einfluß auf geschätzte Wahrscheinlichkeiten)
- Die Wahrscheinlichkeit, daß $y = m$ ist $F(\tau_m - \mathbf{x}_i\boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i\boldsymbol{\beta})$
- D. h. für jeden Fall i gibt es eine Wahrscheinlichkeit, daß $y = m$, die von den x-Werten, dem Koeffizientenvektor β und dem Vektor mit den Schwellenwerten τ abhängt
- Die (Log-)Likelikhood-Funktion ist etwas komplizierter, weil für jeden Fall mehrere Wahrscheinlichkeiten berücksichtigt werden müssen

Wie können die Modellparameter geschätzt werden?

- Um das Modell zu identifizieren, muß wieder eine Annahme über die Varianz getroffen und entweder der Achsenabschnitt (STATA) oder τ_1 auf null gesetzt werde (kein Einfluß auf geschätzte Wahrscheinlichkeiten)
- Die Wahrscheinlichkeit, daß $y = m$ ist $F(\tau_m - \mathbf{x}_i\boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}_i\boldsymbol{\beta})$
- D. h. für jeden Fall i gibt es eine Wahrscheinlichkeit, daß $y = m$, die von den x-Werten, dem Koeffizientenvektor β und dem Vektor mit den Schwellenwerten τ abhängt
- Die (Log-)Likelikhood-Funktion ist etwas komplizierter, weil für jeden Fall mehrere Wahrscheinlichkeiten berücksichtigt werden müssen
- Ansonsten funktioniert die ML-Schätzung wie im binären Fall

Wie sieht das praktisch aus?

```
. tab warm

    Mom can |
  have warm |
  relations |
 with child |      Freq.     Percent        Cum.
------------+-----------------------------------
        1SD |        297       12.95       12.95
         2D |        723       31.53       44.48
         3A |        856       37.33       81.81
        4SA |        417       18.19      100.00
------------+-----------------------------------
      Total |      2,293      100.00
```

Wie sieht das praktisch aus?

```
. ologit warm age , table nolog

Ordered logistic regression                       Number of obs   =       2293
                                                  LR chi2(1)      =     102.14
                                                  Prob > chi2     =     0.0000
Log likelihood = -2944.6999                       Pseudo R2       =     0.0170

------------------------------------------------------------------------------
        warm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0229234   .0022893   -10.01   0.000    -.0274103   -.0184366
-------------+----------------------------------------------------------------
       /cut1 |  -2.993187   .1275942                     -3.243267   -2.743107
       /cut2 |   -1.25572   .1120547                     -1.475344   -1.036097
       /cut3 |   .5243538   .1104667                       .307843    .7408646
------------------------------------------------------------------------------
```

Wie war das mit den Wahrscheinlichkeiten?

- Die Wahrscheinlichkeit der Antwort „1" ist gleich $F(\tau_1 - \mathbf{x}\boldsymbol{\beta}) = \frac{\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}{1+\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}$

Wie war das mit den Wahrscheinlichkeiten?

- Die Wahrscheinlichkeit der Antwort „1“ ist gleich $F(\tau_1 - \mathbf{x\beta}) = \frac{\exp(\tau_1 - \mathbf{x\beta})}{1+\exp(\tau_1 - \mathbf{x\beta})}$
- Die Wahrscheinlichkeit der Antwort „2“ ist gleich $F(\tau_2 - \mathbf{x\beta}) - F(\tau_1 - \mathbf{x\beta})$

Wie war das mit den Wahrscheinlichkeiten?

- Die Wahrscheinlichkeit der Antwort „1“ ist gleich $F(\tau_1 - \mathbf{x}\boldsymbol{\beta}) = \frac{\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}{1+\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}$
- Die Wahrscheinlichkeit der Antwort „2“ ist gleich $F(\tau_2 - \mathbf{x}\boldsymbol{\beta}) - F(\tau_1 - \mathbf{x}\boldsymbol{\beta})$
- Die Wahrscheinlichkeit, daß $y \leqslant 2$ ist gleich $F(\tau_2 - \mathbf{x}\boldsymbol{\beta})$

Wie war das mit den Wahrscheinlichkeiten?

- Die Wahrscheinlichkeit der Antwort „1" ist gleich $F(\tau_1 - \mathbf{x}\boldsymbol{\beta}) = \frac{\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}{1+\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}$
- Die Wahrscheinlichkeit der Antwort „2" ist gleich $F(\tau_2 - \mathbf{x}\boldsymbol{\beta}) - F(\tau_1 - \mathbf{x}\boldsymbol{\beta})$
- Die Wahrscheinlichkeit, daß $y \leqslant 2$ ist gleich $F(\tau_2 - \mathbf{x}\boldsymbol{\beta})$
- Die Wahrscheinlichkeit, daß $y \leqslant 1$ ist gleich $F(\tau_1 - \mathbf{x}\boldsymbol{\beta})$

Wie war das mit den Wahrscheinlichkeiten?

- Die Wahrscheinlichkeit der Antwort „1“ ist gleich $F(\tau_1 - \mathbf{x}\boldsymbol{\beta}) = \frac{\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}{1+\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}$
- Die Wahrscheinlichkeit der Antwort „2“ ist gleich $F(\tau_2 - \mathbf{x}\boldsymbol{\beta}) - F(\tau_1 - \mathbf{x}\boldsymbol{\beta})$
- Die Wahrscheinlichkeit, daß $y \leqslant 2$ ist gleich $F(\tau_2 - \mathbf{x}\boldsymbol{\beta})$
- Die Wahrscheinlichkeit, daß $y \leqslant 1$ ist gleich $F(\tau_1 - \mathbf{x}\boldsymbol{\beta})$
- Auf diesen letzten Punkt bezieht sich die Sache mit der „parallelen Regression“

Wie war das mit den Wahrscheinlichkeiten?

- Die Wahrscheinlichkeit der Antwort „1" ist gleich $F(\tau_1 - \mathbf{x}\boldsymbol{\beta}) = \frac{\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}{1+\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}$
- Die Wahrscheinlichkeit der Antwort „2" ist gleich $F(\tau_2 - \mathbf{x}\boldsymbol{\beta}) - F(\tau_1 - \mathbf{x}\boldsymbol{\beta})$
- Die Wahrscheinlichkeit, daß $y \leqslant 2$ ist gleich $F(\tau_2 - \mathbf{x}\boldsymbol{\beta})$
- Die Wahrscheinlichkeit, daß $y \leqslant 1$ ist gleich $F(\tau_1 - \mathbf{x}\boldsymbol{\beta})$
- Auf diesen letzten Punkt bezieht sich die Sache mit der „parallelen Regression"
- Die Form der Kurve ist für alle Ausprägungen gleich, die Kurven sind aber jeweils um τ verschoben

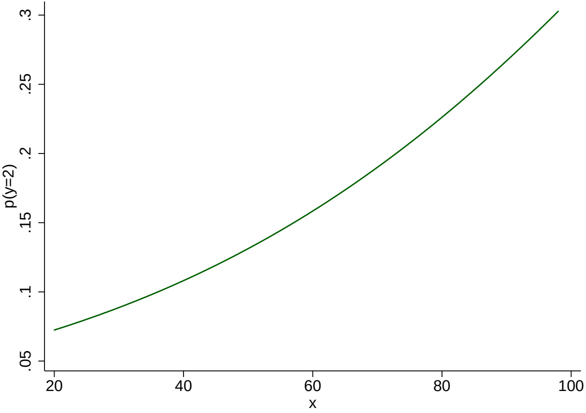
Wie war das mit den Wahrscheinlichkeiten?

- Die Wahrscheinlichkeit der Antwort „1“ ist gleich $F(\tau_1 - \mathbf{x}\boldsymbol{\beta}) = \frac{\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}{1+\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}$
- Die Wahrscheinlichkeit der Antwort „2“ ist gleich $F(\tau_2 - \mathbf{x}\boldsymbol{\beta}) - F(\tau_1 - \mathbf{x}\boldsymbol{\beta})$
- Die Wahrscheinlichkeit, daß $y \leqslant 2$ ist gleich $F(\tau_2 - \mathbf{x}\boldsymbol{\beta})$
- Die Wahrscheinlichkeit, daß $y \leqslant 1$ ist gleich $F(\tau_1 - \mathbf{x}\boldsymbol{\beta})$
- Auf diesen letzten Punkt bezieht sich die Sache mit der „parallelen Regression“
- Die Form der Kurve ist für alle Ausprägungen gleich, die Kurven sind aber jeweils um τ verschoben
- Dabei handelt es eigentlich um Serie von drei bivariaten Regressionen: 1 vs. 2,3,4; 1,2 vs. 3,4, 1,2,3 vs. 4

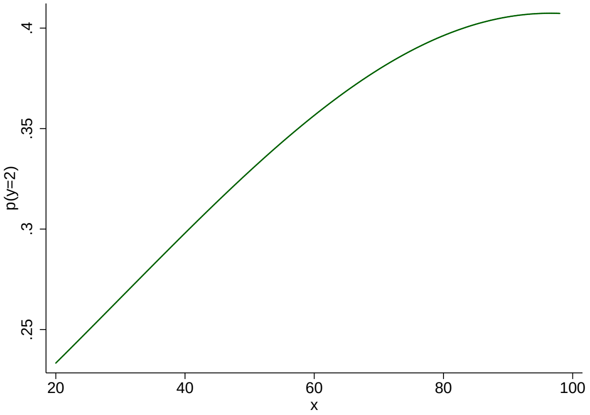
Wie war das mit den Wahrscheinlichkeiten?

- Die Wahrscheinlichkeit der Antwort „1" ist gleich $F(\tau_1 - \mathbf{x}\boldsymbol{\beta}) = \frac{\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}{1+\exp(\tau_1 - \mathbf{x}\boldsymbol{\beta})}$
- Die Wahrscheinlichkeit der Antwort „2" ist gleich $F(\tau_2 - \mathbf{x}\boldsymbol{\beta}) - F(\tau_1 - \mathbf{x}\boldsymbol{\beta})$
- Die Wahrscheinlichkeit, daß $y \leqslant 2$ ist gleich $F(\tau_2 - \mathbf{x}\boldsymbol{\beta})$
- Die Wahrscheinlichkeit, daß $y \leqslant 1$ ist gleich $F(\tau_1 - \mathbf{x}\boldsymbol{\beta})$
- Auf diesen letzten Punkt bezieht sich die Sache mit der „parallelen Regression"
- Die Form der Kurve ist für alle Ausprägungen gleich, die Kurven sind aber jeweils um τ verschoben
- Dabei handelt es eigentlich um Serie von drei bivariaten Regressionen: 1 vs. 2,3,4; 1,2 vs. 3,4, 1,2,3 vs. 4 (Implizite Annahme, daß sich Kategorien zusammenfassen lassen)

Wahrscheinlichkeit von Antwort 1 (SD)?



Wahrscheinlichkeit von Antwort 2 (D)?



Warum und wie soll das Modell verallgemeinert werden?

- Annahme der parallelen Regression ist häufig unplausibel/verletzt (Brant-Test)

Warum und wie soll das Modell verallgemeinert werden?

- Annahme der parallelen Regression ist häufig unplausibel/verletzt (Brant-Test)
- Im generalisierten ordinalen Modell wird diese Annahme aufgegeben

Warum und wie soll das Modell verallgemeinert werden?

- Annahme der parallelen Regression ist häufig unplausibel/verletzt (Brant-Test)
- Im generalisierten ordinalen Modell wird diese Annahme aufgegeben
- Für j Ausprägungen $j - 1$ Koeffizienten pro Variable

Warum und wie soll das Modell verallgemeinert werden?

- Annahme der parallelen Regression ist häufig unplausibel/verletzt (Brant-Test)
- Im generalisierten ordinalen Modell wird diese Annahme aufgegeben
- Für j Ausprägungen $j-1$ Koeffizienten pro Variable
- Ausprägung j und alle niedrigeren Werte auf 0 rekodiert, $j+1$ und alle höheren Werte auf 1 rekodiert

Warum und wie soll das Modell verallgemeinert werden?

- Annahme der parallelen Regression ist häufig unplausibel/verletzt (Brant-Test)
- Im generalisierten ordinalen Modell wird diese Annahme aufgegeben
- Für j Ausprägungen $j-1$ Koeffizienten pro Variable
- Ausprägung j und alle niedrigeren Werte auf 0 rekodiert, $j+1$ und alle höheren Werte auf 1 rekodiert
- Simultane Schätzung

Ein Beispiel?

	b
1SD	
age	-.0214454
cons	2.918156
2D	
age	-.0250633
cons	1.355112
3A	
age	-.018903
cons	-.68643

- Der Logit von 2/3/4 vs. 1 nimmt mit jedem Lebensjahr um 0.02 ab. . .

Ein Beispiel?

	b
1SD	
age	-.0214454
cons	2.918156
2D	
age	-.0250633
cons	1.355112
3A	
age	-.018903
cons	-.68643

- Der Logit von 2/3/4 vs. 1 nimmt mit jedem Lebensjahr um 0.02 ab...
- Parallele Regression Sonderfall, mit identischen Schätzungen für alle Ausprägungen

Ein Beispiel?

	b
1SD	
age	-.0214454
cons	2.918156
2D	
age	-.0250633
cons	1.355112
3A	
age	-.018903
cons	-.68643

- Der Logit von 2/3/4 vs. 1 nimmt mit jedem Lebensjahr um 0.02 ab. . .
- Parallele Regression Sonderfall, mit identischen Schätzungen für alle Ausprägungen
- Annahme kann für einzelne unabhängige Variablen aufgegeben werden

Probleme?

- Im (normalen) ordinalen Modell Interpretation scheinbar einfach

Probleme?

- Im (normalen) ordinalen Modell Interpretation scheinbar einfach (Ist sie es tatsächlich? Eventuell doch besser gleich dichotomisieren?)

Probleme?

- Im (normalen) ordinalen Modell Interpretation scheinbar einfach (Ist sie es tatsächlich? Eventuell doch besser gleich dichotomisieren?)
- Generalisiertes ordinales Modell wegen Vielzahl von Koeffizienten recht unübersichtlich

Probleme?

- Im (normalen) ordinalen Modell Interpretation scheinbar einfach (Ist sie es tatsächlich? Eventuell doch besser gleich dichotomisieren?)
- Generalisiertes ordinales Modell wegen Vielzahl von Koeffizienten recht unübersichtlich
- Zahl der zu schätzenden Parameter jedoch immer noch durch Ordinalitätsannahme begrenzt

Probleme?

- Im (normalen) ordinalen Modell Interpretation scheinbar einfach (Ist sie es tatsächlich? Eventuell doch besser gleich dichotomisieren?)
- Generalisiertes ordinales Modell wegen Vielzahl von Koeffizienten recht unübersichtlich
- Zahl der zu schätzenden Parameter jedoch immer noch durch Ordinalitätsannahme begrenzt
- Dies Beschränkung fällt im multinomialen Modell leider weg

Warum noch ein Modell?

- Eine Vielzahl von interessanten Variablen ist polytom, aber ungeordnet (z. B. Wahlabsicht)

Warum noch ein Modell?

- Eine Vielzahl von interessanten Variablen ist polytom, aber ungeordnet (z. B. Wahlabsicht)
- In diesem Fall ist das multinomiale Modell eine einfach, aber unübersichtliche Erweiterung

Was beinhaltet das multinomiale Modell?

- Prinzipiell als eine Serie von dichotomen Vergleichen vorstellbar: 1 vs. 2, 1 vs. 3, 1 vs. 4; 2 vs. 3, 2 vs. 4, 3 vs. 4

Was beinhaltet das multinomiale Modell?

- Prinzipiell als eine Serie von dichotomen Vergleichen vorstellbar: 1 vs. 2, 1 vs. 3, 1 vs. 4; 2 vs. 3, 2 vs. 4, 3 vs. 4
- Beispiel im Text: A vs. B, A vs. C, B vs. C

Was beinhaltet das multinomiale Modell?

- Prinzipiell als eine Serie von dichotomen Vergleichen vorstellbar: 1 vs. 2, 1 vs. 3, 1 vs. 4; 2 vs. 3, 2 vs. 4, 3 vs. 4
- Beispiel im Text: A vs. B, A vs. C, B vs. C
- Von diesen drei Vergleichen ist einer redundant: $\ln(\Pr(A|\mathbf{x}/\Pr(B|\mathbf{x}))) + \ln(\Pr(B|\mathbf{x}/\Pr(C|\mathbf{x}))) = \ln(\Pr(A|\mathbf{x}/\Pr(C|\mathbf{x})))$ (weil $\ln(a) + \ln(b) = \ln(a \times b)$)

Was beinhaltet das multinomiale Modell?

- Prinzipiell als eine Serie von dichotomen Vergleichen vorstellbar: 1 vs. 2, 1 vs. 3, 1 vs. 4; 2 vs. 3, 2 vs. 4, 3 vs. 4
- Beispiel im Text: A vs. B, A vs. C, B vs. C
- Von diesen drei Vergleichen ist einer redundant: $\ln(\Pr(A|\mathbf{x}/\Pr(B|\mathbf{x}))) + \ln(\Pr(B|\mathbf{x}/\Pr(C|\mathbf{x}))) = \ln(\Pr(A|\mathbf{x}/\Pr(C|\mathbf{x})))$ (weil $\ln(a) + \ln(b) = \ln(a \times b)$)
- Daraus folgt z. B. daß $\beta_{0,A|B} + \beta_{0,B|C} = \beta_{0,A|C}$ ist

Was beinhaltet das multinomiale Modell?

- Prinzipiell als eine Serie von dichotomen Vergleichen vorstellbar: 1 vs. 2, 1 vs. 3, 1 vs. 4; 2 vs. 3, 2 vs. 4, 3 vs. 4
- Beispiel im Text: A vs. B, A vs. C, B vs. C
- Von diesen drei Vergleichen ist einer redundant: $\ln(\Pr(A|\mathbf{x}/\Pr(B|\mathbf{x}))) + \ln(\Pr(B|\mathbf{x}/\Pr(C|\mathbf{x}))) = \ln(\Pr(A|\mathbf{x}/\Pr(C|\mathbf{x})))$ (weil $\ln(a) + \ln(b) = \ln(a \times b)$)
- Daraus folgt z. B. daß $\beta_{0,A|B} + \beta_{0,B|C} = \beta_{0,A|C}$ ist
- Achtung: Das sind notwendige Beziehungen für Parameter bei simultaner Betrachtung

Was beinhaltet das multinomiale Modell?

- Prinzipiell als eine Serie von dichotomen Vergleichen vorstellbar: 1 vs. 2, 1 vs. 3, 1 vs. 4; 2 vs. 3, 2 vs. 4, 3 vs. 4
- Beispiel im Text: A vs. B, A vs. C, B vs. C
- Von diesen drei Vergleichen ist einer redundant: $\ln(\Pr(A|\mathbf{x}/\Pr(B|\mathbf{x}))) + \ln(\Pr(B|\mathbf{x}/\Pr(C|\mathbf{x}))) = \ln(\Pr(A|\mathbf{x}/\Pr(C|\mathbf{x})))$ (weil $\ln(a) + \ln(b) = \ln(a \times b)$)
- Daraus folgt z. B. daß $\beta_{0,A|B} + \beta_{0,B|C} = \beta_{0,A|C}$ ist
- Achtung: Das sind notwendige Beziehungen für Parameter bei simultaner Betrachtung
- Wenn man tatsächliche separate Analysen rechnet, erhält man andere Ergebnisse, weil nicht die ganze Stichprobe verwendet wird

Was beinhaltet das multinomiale Modell?

- Prinzipiell als eine Serie von dichotomen Vergleichen vorstellbar: 1 vs. 2, 1 vs. 3, 1 vs. 4; 2 vs. 3, 2 vs. 4, 3 vs. 4
- Beispiel im Text: A vs. B, A vs. C, B vs. C
- Von diesen drei Vergleichen ist einer redundant: $\ln(\Pr(A|\mathbf{x}/\Pr(B|\mathbf{x}))) + \ln(\Pr(B|\mathbf{x}/\Pr(C|\mathbf{x}))) = \ln(\Pr(A|\mathbf{x}/\Pr(C|\mathbf{x})))$ (weil $\ln(a) + \ln(b) = \ln(a \times b)$)
- Daraus folgt z. B. daß $\beta_{0,A|B} + \beta_{0,B|C} = \beta_{0,A|C}$ ist
- Achtung: Das sind notwendige Beziehungen für Parameter bei simultaner Betrachtung
- Wenn man tatsächliche separate Analysen rechnet, erhält man andere Ergebnisse, weil nicht die ganze Stichprobe verwendet wird
- Inhaltlich bedeutet das: $\Pr(A|\mathbf{x})$ und $\Pr(B|\mathbf{x})$ addieren sich *nicht* wie im binären Modell zu eins, da es noch Möglichkeit $C \ldots$ gibt

Welche Wahrscheinlichkeiten werden hier betrachtet?

- Die abhängige Variable y hat J Ausprägungen

Welche Wahrscheinlichkeiten werden hier betrachtet?

- Die abhängige Variable y hat J Ausprägungen
- Dann ist die Wahrscheinlichkeit, daß:

Welche Wahrscheinlichkeiten werden hier betrachtet?

- Die abhängige Variable y hat J Ausprägungen
- Dann ist die Wahrscheinlichkeit, daß:

$$\Pr(y_i = m | \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_m)}{\sum_{j=1}^{J} \exp(\mathbf{x}_i \boldsymbol{\beta}_j)}$$

Welche Wahrscheinlichkeiten werden hier betrachtet?

- Die abhängige Variable y hat J Ausprägungen
- Dann ist die Wahrscheinlichkeit, daß:

$$\Pr(y_i = m|\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i\boldsymbol{\beta}_m)}{\sum_{j=1}^{J}\exp(\mathbf{x}_i\boldsymbol{\beta}_j)}$$

- Um das Modell zu identifizieren, wird außerdem angenommen, daß $\boldsymbol{\beta}_1 = 0$

Welche Wahrscheinlichkeiten werden hier betrachtet?

- Die abhängige Variable y hat J Ausprägungen
- Dann ist die Wahrscheinlichkeit, daß:

$$\Pr(y_i = m|\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_m)}{\sum_{j=1}^{J} \exp(\mathbf{x}_i \boldsymbol{\beta}_j)}$$

- Um das Modell zu identifizieren, wird außerdem angenommen, daß $\boldsymbol{\beta}_1 = 0$
- D. h. eine (beliebige) Ausprägung wird zur Referenzkategorie erklärt (die Schätzungen für diese Kategorie können rekonstruiert werden)

Welche Wahrscheinlichkeiten werden hier betrachtet?

- Die abhängige Variable y hat J Ausprägungen
- Dann ist die Wahrscheinlichkeit, daß:

$$\Pr(y_i = m|\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_m)}{\sum_{j=1}^{J} \exp(\mathbf{x}_i \boldsymbol{\beta}_j)}$$

- Um das Modell zu identifizieren, wird außerdem angenommen, daß $\boldsymbol{\beta}_1 = 0$
- D. h. eine (beliebige) Ausprägung wird zur Referenzkategorie erklärt (die Schätzungen für diese Kategorie können rekonstruiert werden)
- Schätzungen wiederum durch numerische Maximierung der Log-Likelihood-Funktion

Was ist das konditionale Logit-Modell?

- Multinomiales Logit-Modell kann als Entscheidungsmodell interpretiert werden

Was ist das konditionale Logit-Modell?

- Multinomiales Logit-Modell kann als Entscheidungsmodell interpretiert werden
- Entscheidung hängt ab von Eigenschaften des Entscheiders (x-Variablen)

Was ist das konditionale Logit-Modell?

- Multinomiales Logit-Modell kann als Entscheidungsmodell interpretiert werden
- Entscheidung hängt ab von Eigenschaften des Entscheiders (x-Variablen)
- Im konditionalen Logit-Modell ähnliche Struktur, aber hier hängt Entscheidung ab von alternativen-spezifischen Eigenschaften, die (1) mit der Entscheidung und (2) mit Eigenschaften des Entscheiders variieren

Was ist das konditionale Logit-Modell?

- Multinomiales Logit-Modell kann als Entscheidungsmodell interpretiert werden
- Entscheidung hängt ab von Eigenschaften des Entscheiders (x-Variablen)
- Im konditionalen Logit-Modell ähnliche Struktur, aber hier hängt Entscheidung ab von alternativen-spezifischen Eigenschaften, die (1) mit der Entscheidung und (2) mit Eigenschaften des Entscheiders variieren
- Politikwissenschaftliches Beispiel: Wahlentscheidung in Mehrparteiensystemen mit wahrgenommener Distanz zwischen eigenen Issue-Positionen und Positionen der Parteien als erklärender Variablen

Was ist das konditionale Logit-Modell?

- Multinomiales Logit-Modell kann als Entscheidungsmodell interpretiert werden
- Entscheidung hängt ab von Eigenschaften des Entscheiders (x-Variablen)
- Im konditionalen Logit-Modell ähnliche Struktur, aber hier hängt Entscheidung ab von alternativen-spezifischen Eigenschaften, die (1) mit der Entscheidung und (2) mit Eigenschaften des Entscheiders variieren
- Politikwissenschaftliches Beispiel: Wahlentscheidung in Mehrparteiensystemen mit wahrgenommener Distanz zwischen eigenen Issue-Positionen und Positionen der Parteien als erklärender Variablen
- Sehr viele komplexe Möglichkeiten

Warum ist R^2 in Verruf gekommen?

- Viele Wissenschaftler glauben, daß ein hohes R^2 ein „gutes“ Modell signalisiert

Warum ist R^2 in Verruf gekommen?

- Viele Wissenschaftler glauben, daß ein hohes R^2 ein „gutes“ Modell signalisiert
- R^2 (und ähnlich Pseudo-R^2) vermitteln einen Eindruck davon, wie stark zufällige und systematische Effekte das y beeinflussen

Warum ist R^2 in Verruf gekommen?

- Viele Wissenschaftler glauben, daß ein hohes R^2 ein „gutes“ Modell signalisiert
- R^2 (und ähnlich Pseudo-R^2) vermitteln einen Eindruck davon, wie stark zufällige und systematische Effekte das y beeinflussen
- Hohes R^2 nur dann ein Zeichen für „Güte“, wenn in ϵ tatsächlich noch systematische Effekte enthalten, die bisher nicht modelliert wurden

Warum ist R^2 in Verruf gekommen?

- Viele Wissenschaftler glauben, daß ein hohes R^2 ein „gutes“ Modell signalisiert
- R^2 (und ähnlich Pseudo-R^2) vermitteln einen Eindruck davon, wie stark zufällige und systematische Effekte das y beeinflussen
- Hohes R^2 nur dann ein Zeichen für „Güte“, wenn in ϵ tatsächlich noch systematische Effekte enthalten, die bisher nicht modelliert wurden
- R^2 (und der Korrelationskoeffizient!) hängen ab

Warum ist R^2 in Verruf gekommen?

- Viele Wissenschaftler glauben, daß ein hohes R^2 ein „gutes“ Modell signalisiert
- R^2 (und ähnlich Pseudo-R^2) vermitteln einen Eindruck davon, wie stark zufällige und systematische Effekte das y beeinflussen
- Hohes R^2 nur dann ein Zeichen für „Güte“, wenn in ϵ tatsächlich noch systematische Effekte enthalten, die bisher nicht modelliert wurden
- R^2 (und der Korrelationskoeffizient!) hängen ab
  - Von der Stärke des Zusammenhangs

Warum ist R^2 in Verruf gekommen?

- Viele Wissenschaftler glauben, daß ein hohes R^2 ein „gutes“ Modell signalisiert
- R^2 (und ähnlich Pseudo-R^2) vermitteln einen Eindruck davon, wie stark zufällige und systematische Effekte das y beeinflussen
- Hohes R^2 nur dann ein Zeichen für „Güte“, wenn in ϵ tatsächlich noch systematische Effekte enthalten, die bisher nicht modelliert wurden
- R^2 (und der Korrelationskoeffizient!) hängen ab
  - Von der Stärke des Zusammenhangs
  - Von der Varianz von x *in der betreffenden Stichprobe*

Warum ist R^2 in Verruf gekommen?

- Viele Wissenschaftler glauben, daß ein hohes R^2 ein „gutes“ Modell signalisiert
- R^2 (und ähnlich Pseudo-R^2) vermitteln einen Eindruck davon, wie stark zufällige und systematische Effekte das y beeinflussen
- Hohes R^2 nur dann ein Zeichen für „Güte“, wenn in ϵ tatsächlich noch systematische Effekte enthalten, die bisher nicht modelliert wurden
- R^2 (und der Korrelationskoeffizient!) hängen ab
  - Von der Stärke des Zusammenhangs
  - Von der Varianz von x *in der betreffenden Stichprobe*
- R^2, Korrelationskoeffizienten und standardisierte Regressionskoeffizienten deshalb nicht über Stichproben hinweg vergleichen

Warum ist R^2 in Verruf gekommen?

- Viele Wissenschaftler glauben, daß ein hohes R^2 ein „gutes“ Modell signalisiert
- R^2 (und ähnlich Pseudo-R^2) vermitteln einen Eindruck davon, wie stark zufällige und systematische Effekte das y beeinflussen
- Hohes R^2 nur dann ein Zeichen für „Güte“, wenn in ϵ tatsächlich noch systematische Effekte enthalten, die bisher nicht modelliert wurden
- R^2 (und der Korrelationskoeffizient!) hängen ab
  - Von der Stärke des Zusammenhangs
  - Von der Varianz von x *in der betreffenden Stichprobe*
- R^2, Korrelationskoeffizienten und standardisierte Regressionskoeffizienten deshalb nicht über Stichproben hinweg vergleichen
- Diese Maße sind nicht in einem absoluten Sinn hoch oder niedrig; bestenfalls für interne Vergleiche nützlich (Aggregatdaten)

Warum standardisiert man Koeffizienten?

- Vollstandardisierte Koeffizienten:

Warum standardisiert man Koeffizienten?

- Vollstandardisierte Koeffizienten:(„Betas“): Um wieviele Standardabweichungen verändert sich y, wenn sich x um eine Standardabweichung verändert?

Warum standardisiert man Koeffizienten?

- Vollstandardisierte Koeffizienten:(„Betas“): Um wieviele Standardabweichungen verändert sich y, wenn sich x um eine Standardabweichung verändert?
- y und x vorab durch jeweilige Standardabweichung teilen oder ex post Koeffizienten durch Standardabweichung von y teilen und mit Standardabweichung von x multiplizieren

Warum standardisiert man Koeffizienten?

- Vollstandardisierte Koeffizienten:(„Betas“): Um wieviele Standardabweichungen verändert sich y, wenn sich x um eine Standardabweichung verändert?
- y und x vorab durch jeweilige Standardabweichung teilen oder ex post Koeffizienten durch Standardabweichung von y teilen und mit Standardabweichung von x multiplizieren
- Warum macht man das?

Warum standardisiert man Koeffizienten?

- Vollstandardisierte Koeffizienten:(„Betas“): Um wieviele Standardabweichungen verändert sich y, wenn sich x um eine Standardabweichung verändert?
- y und x vorab durch jeweilige Standardabweichung teilen oder ex post Koeffizienten durch Standardabweichung von y teilen und mit Standardabweichung von x multiplizieren
- Warum macht man das? Welcher Effekt ist der „wichtigste“ (Skalierung)

Warum standardisiert man Koeffizienten?

- Vollstandardisierte Koeffizienten:(„Betas“): Um wieviele Standardabweichungen verändert sich y, wenn sich x um eine Standardabweichung verändert?
- y und x vorab durch jeweilige Standardabweichung teilen oder ex post Koeffizienten durch Standardabweichung von y teilen und mit Standardabweichung von x multiplizieren
- Warum macht man das? Welcher Effekt ist der „wichtigste“ (Skalierung)
- Ist die Standardabweichung von dichotomen Variablen in irgendeiner Form interpretierbar?

Warum standardisiert man Koeffizienten?

- Vollstandardisierte Koeffizienten:(„Betas"): Um wieviele Standardabweichungen verändert sich y, wenn sich x um eine Standardabweichung verändert?
- y und x vorab durch jeweilige Standardabweichung teilen oder ex post Koeffizienten durch Standardabweichung von y teilen und mit Standardabweichung von x multiplizieren
- Warum macht man das? Welcher Effekt ist der „wichtigste" (Skalierung)
- Ist die Standardabweichung von dichotomen Variablen in irgendeiner Form interpretierbar?
- Nicht-standardisierte Koeffizienten

Warum standardisiert man Koeffizienten?

- Vollstandardisierte Koeffizienten:(„Betas“): Um wieviele Standardabweichungen verändert sich y, wenn sich x um eine Standardabweichung verändert?
- y und x vorab durch jeweilige Standardabweichung teilen oder ex post Koeffizienten durch Standardabweichung von y teilen und mit Standardabweichung von x multiplizieren
- Warum macht man das? Welcher Effekt ist der „wichtigste“ (Skalierung)
- Ist die Standardabweichung von dichotomen Variablen in irgendeiner Form interpretierbar?
- Nicht-standardisierte Koeffizienten
  - Sind über Stichproben hinweg vergleichbar

Warum standardisiert man Koeffizienten?

- Vollstandardisierte Koeffizienten:(„Betas“): Um wieviele Standardabweichungen verändert sich y, wenn sich x um eine Standardabweichung verändert?
- y und x vorab durch jeweilige Standardabweichung teilen oder ex post Koeffizienten durch Standardabweichung von y teilen und mit Standardabweichung von x multiplizieren
- Warum macht man das? Welcher Effekt ist der „wichtigste“ (Skalierung)
- Ist die Standardabweichung von dichotomen Variablen in irgendeiner Form interpretierbar?
- Nicht-standardisierte Koeffizienten
  - Sind über Stichproben hinweg vergleichbar
  - Beziehen sich auf ursprüngliche Skalen ➔ leicht zu interpretieren

Warum standardisiert man Koeffizienten?

- Vollstandardisierte Koeffizienten:(„Betas"): Um wieviele Standardabweichungen verändert sich y, wenn sich x um eine Standardabweichung verändert?
- y und x vorab durch jeweilige Standardabweichung teilen oder ex post Koeffizienten durch Standardabweichung von y teilen und mit Standardabweichung von x multiplizieren
- Warum macht man das? Welcher Effekt ist der „wichtigste" (Skalierung)
- Ist die Standardabweichung von dichotomen Variablen in irgendeiner Form interpretierbar?
- Nicht-standardisierte Koeffizienten
  - Sind über Stichproben hinweg vergleichbar
  - Beziehen sich auf ursprüngliche Skalen → leicht zu interpretieren
  - Erfordern und erleichtern inhaltliche Auseinandersetzung mit Bedeutsamkeit

Und wenn es doch Probleme mit der Vergleichbarkeit gibt?

- Variablen multiplizieren/dividieren, um Streuung vergleichbar zu machen (Euros vs. K-Euros)

Und wenn es doch Probleme mit der Vergleichbarkeit gibt?

- Variablen multiplizieren/dividieren, um Streuung vergleichbar zu machen (Euros vs. K-Euros)
- Maximale Effekte vergleichen

Und wenn es doch Probleme mit der Vergleichbarkeit gibt?

- Variablen multiplizieren/dividieren, um Streuung vergleichbar zu machen (Euros vs. K-Euros)
- Maximale Effekte vergleichen
- Interquartilsabstand

Wie kommt man eigentlich an Standardfehler?

- Beim ML-Verfahren wird die Log-Likelihood-Funktion maximiert ...

Wie kommt man eigentlich an Standardfehler?

- Beim ML-Verfahren wird die Log-Likelihood-Funktion maximiert ...
- indem ihre erste Ableitung (Score-Funktion) auf null gesetzt wird

Wie kommt man eigentlich an Standardfehler?

- Beim ML-Verfahren wird die Log-Likelihood-Funktion maximiert ...
- indem ihre erste Ableitung (Score-Funktion) auf null gesetzt wird
- Diese Nullstelle wird iterativ gesucht; dazu braucht man die erste Ableitung der Score-Funktion (zweite Ableitung der Log-Likelihood-Funktion)

Wie kommt man eigentlich an Standardfehler?

- Beim ML-Verfahren wird die Log-Likelihood-Funktion maximiert ...
- indem ihre erste Ableitung (Score-Funktion) auf null gesetzt wird
- Diese Nullstelle wird iterativ gesucht; dazu braucht man die erste Ableitung der Score-Funktion (zweite Ableitung der Log-Likelihood-Funktion)
- Hesse-Matrix: Die Matrix, die die zweiten partiellen Ableitungen der Log-Likelihood-Funktion enthält

Wie kommt man eigentlich an Standardfehler?

- Beim ML-Verfahren wird die Log-Likelihood-Funktion maximiert ...
- indem ihre erste Ableitung (Score-Funktion) auf null gesetzt wird
- Diese Nullstelle wird iterativ gesucht; dazu braucht man die erste Ableitung der Score-Funktion (zweite Ableitung der Log-Likelihood-Funktion)
- Hesse-Matrix: Die Matrix, die die zweiten partiellen Ableitungen der Log-Likelihood-Funktion enthält
- Information-Matrix: Der negative Erwartungswert der Hesse-Matrix (bzw. die Hesse-Matrix evaluiert an der ML-Schätzung)

Wie kommt man eigentlich an Standardfehler?

- Beim ML-Verfahren wird die Log-Likelihood-Funktion maximiert ...
- indem ihre erste Ableitung (Score-Funktion) auf null gesetzt wird
- Diese Nullstelle wird iterativ gesucht; dazu braucht man die erste Ableitung der Score-Funktion (zweite Ableitung der Log-Likelihood-Funktion)
- Hesse-Matrix: Die Matrix, die die zweiten partiellen Ableitungen der Log-Likelihood-Funktion enthält
- Information-Matrix: Der negative Erwartungswert der Hesse-Matrix (bzw. die Hesse-Matrix evaluiert an der ML-Schätzung)
- Die Inverse zur Information-Matrix ist die Varianz-Kovarianz-Matrix

Worum geht es bei Signifikanztests?

- Wie wahrscheinlich ist es, diese Parameterschätzungen zu erhalten, wenn die entsprechenden Parameter in der Grundgesamtheit gleich null sind (anderer Werte als null sind ebenfalls möglich)

Worum geht es bei Signifikanztests?

- Wie wahrscheinlich ist es, diese Parameterschätzungen zu erhalten, wenn die entsprechenden Parameter in der Grundgesamtheit gleich null sind (anderer Werte als null sind ebenfalls möglich)
- Logik:

Worum geht es bei Signifikanztests?

- Wie wahrscheinlich ist es, diese Parameterschätzungen zu erhalten, wenn die entsprechenden Parameter in der Grundgesamtheit gleich null sind (anderer Werte als null sind ebenfalls möglich)
- Logik:
  - In der Grundgesamtheit gilt H_0, d. h. in Wahrheit ist Parameter $=0$

Worum geht es bei Signifikanztests?

- Wie wahrscheinlich ist es, diese Parameterschätzungen zu erhalten, wenn die entsprechenden Parameter in der Grundgesamtheit gleich null sind (anderer Werte als null sind ebenfalls möglich)
- Logik:
  - In der Grundgesamtheit gilt H_0, d. h. in Wahrheit ist Parameter $=0$
  - Unter dieser Bedingung hat Prüfgröße eine bestimmte Verteilung (z. B. χ^2)

Worum geht es bei Signifikanztests?

- Wie wahrscheinlich ist es, diese Parameterschätzungen zu erhalten, wenn die entsprechenden Parameter in der Grundgesamtheit gleich null sind (anderer Werte als null sind ebenfalls möglich)
- Logik:
  - In der Grundgesamtheit gilt H_0, d. h. in Wahrheit ist Parameter $=0$
  - Unter dieser Bedingung hat Prüfgröße eine bestimmte Verteilung (z. B. χ^2)
  - Verbindung Parameter – Prüfgröße

Worum geht es bei Signifikanztests?

- Wie wahrscheinlich ist es, diese Parameterschätzungen zu erhalten, wenn die entsprechenden Parameter in der Grundgesamtheit gleich null sind (anderer Werte als null sind ebenfalls möglich)
- Logik:
  - In der Grundgesamtheit gilt H_0, d. h. in Wahrheit ist Parameter $=0$
  - Unter dieser Bedingung hat Prüfgröße eine bestimmte Verteilung (z. B. χ^2)
  - Verbindung Parameter – Prüfgröße
  - Wenn Parameter in Wirklichkeit gleich null (und Verteilungsannahme für Prüfgröße korrekt), kann man ausrechnen, wie wahrscheinlich es ist, empirisch beobachtete Prüfgröße zu erhalten, wenn Parameter in Grundgesamtheit gleich null

Worum geht es bei Signifikanztests?

- Wie wahrscheinlich ist es, diese Parameterschätzungen zu erhalten, wenn die entsprechenden Parameter in der Grundgesamtheit gleich null sind (anderer Werte als null sind ebenfalls möglich)
- Logik:
  - In der Grundgesamtheit gilt H_0, d. h. in Wahrheit ist Parameter $=0$
  - Unter dieser Bedingung hat Prüfgröße eine bestimmte Verteilung (z. B. χ^2)
  - Verbindung Parameter – Prüfgröße
  - Wenn Parameter in Wirklichkeit gleich null (und Verteilungsannahme für Prüfgröße korrekt), kann man ausrechnen, wie wahrscheinlich es ist, empirisch beobachtete Prüfgröße zu erhalten, wenn Parameter in Grundgesamtheit gleich null
- Wald-Test, Likelihood-Ratio-Test, Lagrange-Multiplier-Test

Wie funktioniert der Wald-Test?

- Test der Standardhypothese, daß ein Koeffizient gleich null ist: Vergleich mit Standardfehler (Normalverteilungsannahme)

Wie funktioniert der Wald-Test?

- Test der Standardhypothese, daß ein Koeffizient gleich null ist: Vergleich mit Standardfehler (Normalverteilungsannahme)
- Wald-Test: Verallgemeinerung, Vergleich von Schätzungen für Modell, das Bedingungen (constraints) enthält (z. B. $\beta_0 = 0$ oder $\beta_1 = \beta_2$), mit Modell, das keine Bedingungen enthält ➔ Wenn in der Grundgesamtheit diese Bedingungen gelten, wie wahrscheinlich ist es, solche Stichprobenwerte zu erhalten

Wie funktioniert der Wald-Test?

- Test der Standardhypothese, daß ein Koeffizient gleich null ist: Vergleich mit Standardfehler (Normalverteilungsannahme)
- Wald-Test: Verallgemeinerung, Vergleich von Schätzungen für Modell, das Bedingungen (constraints) enthält (z. B. $\beta_0 = 0$ oder $\beta_1 = \beta_2$), mit Modell, das keine Bedingungen enthält ➔ Wenn in der Grundgesamtheit diese Bedingungen gelten, wie wahrscheinlich ist es, solche Stichprobenwerte zu erhalten
- Distanz zwischen beschränkter und unbeschränkter Schätzung wird zur Wölbung der Likelihood-Funktion beziehungsweise (äquivalent) zur Streuung der Parameterschätzungen ins Verhältnis gesetzt (flache Wölbung – große Streuung, starke Wölbung – geringe Streuung)

Wie lassen sich lineare Bedingungen testen?

- Allgemeine Form der linearen Bedingungen

Wie lassen sich lineare Bedingungen testen?

- Allgemeine Form der linearen Bedingungen

$$\mathbf{Q}\boldsymbol{\beta} = \mathbf{r}$$

Wie lassen sich lineare Bedingungen testen?

- Allgemeine Form der linearen Bedingungen

$$\mathbf{Q}\beta = \mathbf{r}$$

- **Q** und **r** sind Matrix/Vektor mit Konstanten

Wie lassen sich lineare Bedingungen testen?

- Allgemeine Form der linearen Bedingungen

$$\mathbf{Q}\beta = \mathbf{r}$$

- **Q** und **r** sind Matrix/Vektor mit Konstanten
- Z. B. die typische Nullhypothese $\beta_1 = 0$

Wie lassen sich lineare Bedingungen testen?

- Allgemeine Form der linearen Bedingungen

$$\mathbf{Q}\beta = \mathbf{r}$$

- **Q** und **r** sind Matrix/Vektor mit Konstanten
- Z. B. die typische Nullhypothese $\beta_1 = 0$

$$\begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}$$

Wie lassen sich lineare Bedingungen testen?

- Allgemeine Form der linearen Bedingungen

$$\mathbf{Q}\beta = \mathbf{r}$$

- $\mathbf{Q}$ und $\mathbf{r}$ sind Matrix/Vektor mit Konstanten
- Z. B. die typische Nullhypothese $\beta_1 = 0$

$$\begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}$$

- Oder die komplexere Nullhypothese $\beta_1 = \beta_2 = 0$

Wie lassen sich lineare Bedingungen testen?

- Allgemeine Form der linearen Bedingungen

$$\mathbf{Q}\beta = \mathbf{r}$$

- **Q** und **r** sind Matrix/Vektor mit Konstanten
- Z. B. die typische Nullhypothese $\beta_1 = 0$

$$\begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}$$

- Oder die komplexere Nullhypothese $\beta_1 = \beta_2 = 0$

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Wie sieht der eigentliche Test aus?

- Zu diesem Ausdruck $\mathbf{Q}\boldsymbol{\beta} = \mathbf{r}$ gehört die Prüfstatistik W, die abhängt (1) von der Distanz zwischen den in der Stichprobe geschätzten und den in der Nullhypothese unterstellten Werten und (2) von der Streuung der Schätzwerte

Wie sieht der eigentliche Test aus?

- Zu diesem Ausdruck $\mathbf{Q}\boldsymbol{\beta} = \mathbf{r}$ gehört die Prüfstatistik W, die abhängt (1) von der Distanz zwischen den in der Stichprobe geschätzten und den in der Nullhypothese unterstellten Werten und (2) von der Streuung der Schätzwerte

$$W = [\mathbf{Q}\boldsymbol{\beta} - \mathbf{r}]'[\mathbf{Q}\hat{\mathrm{V}}(\hat{\boldsymbol{\beta}})\mathbf{Q}']^{-1}[\mathbf{Q}\boldsymbol{\beta} - \mathbf{r}]$$

Wie sieht der eigentliche Test aus?

- Zu diesem Ausdruck $\mathbf{Q}\boldsymbol{\beta} = \mathbf{r}$ gehört die Prüfstatistik W, die abhängt (1) von der Distanz zwischen den in der Stichprobe geschätzten und den in der Nullhypothese unterstellten Werten und (2) von der Streuung der Schätzwerte

$$W = [\mathbf{Q}\boldsymbol{\beta} - \mathbf{r}]'[\mathbf{Q}\hat{\mathrm{V}}(\hat{\boldsymbol{\beta}})\mathbf{Q}']^{-1}[\mathbf{Q}\boldsymbol{\beta} - \mathbf{r}]$$

- Im einfachen Fall $\beta_1 = 0$ reduziert sich der erste und der dritte Teil der Formel zu $(\hat{\beta}_1 - 0)^2$

Wie sieht der eigentliche Test aus?

- Zu diesem Ausdruck $\mathbf{Q}\beta = \mathbf{r}$ gehört die Prüfstatistik W, die abhängt (1) von der Distanz zwischen den in der Stichprobe geschätzten und den in der Nullhypothese unterstellten Werten und (2) von der Streuung der Schätzwerte

$$W = [\mathbf{Q}\beta - \mathbf{r}]'[\mathbf{Q}\hat{\mathrm{V}}(\hat{\beta})\mathbf{Q}']^{-1}[\mathbf{Q}\beta - \mathbf{r}]$$

- Im einfachen Fall $\beta_1 = 0$ reduziert sich der erste und der dritte Teil der Formel zu $(\hat{\beta}_1 - 0)^2$
- Und der mittlere Teil zu $\frac{1}{\hat{\sigma}^2_{\hat{\beta}_1}}$

Wie sieht der eigentliche Test aus?

- Zu diesem Ausdruck $\mathbf{Q}\boldsymbol{\beta} = \mathbf{r}$ gehört die Prüfstatistik W, die abhängt (1) von der Distanz zwischen den in der Stichprobe geschätzten und den in der Nullhypothese unterstellten Werten und (2) von der Streuung der Schätzwerte

$$W = [\mathbf{Q}\boldsymbol{\beta} - \mathbf{r}]'[\mathbf{Q}\hat{\mathrm{V}}(\hat{\boldsymbol{\beta}})\mathbf{Q}']^{-1}[\mathbf{Q}\boldsymbol{\beta} - \mathbf{r}]$$

- Im einfachen Fall $\beta_1 = 0$ reduziert sich der erste und der dritte Teil der Formel zu $(\hat{\beta}_1 - 0)^2$
- Und der mittlere Teil zu $\frac{1}{\hat{\sigma}^2_{\hat{\beta}_1}}$
- D. h. $W = \frac{(\hat{\beta}_1 - 0)^2}{\hat{\sigma}^2_{\hat{\beta}_1}} = \left(\frac{\hat{\beta}_1 - 0}{\hat{\sigma}_{\hat{\beta}_1}}\right)^2$

Wie sieht der eigentliche Test aus?

- Zu diesem Ausdruck $\mathbf{Q}\beta = \mathbf{r}$ gehört die Prüfstatistik W, die abhängt (1) von der Distanz zwischen den in der Stichprobe geschätzten und den in der Nullhypothese unterstellten Werten und (2) von der Streuung der Schätzwerte

$$W = [\mathbf{Q}\beta - \mathbf{r}]'[\mathbf{Q}\hat{\mathrm{V}}(\hat{\beta})\mathbf{Q}']^{-1}[\mathbf{Q}\beta - \mathbf{r}]$$

- Im einfachen Fall $\beta_1 = 0$ reduziert sich der erste und der dritte Teil der Formel zu $(\hat{\beta}_1 - 0)^2$
- Und der mittlere Teil zu $\frac{1}{\hat{\sigma}^2_{\hat{\beta}_1}}$
- D. h. $W = \frac{(\hat{\beta}_1 - 0)^2}{\hat{\sigma}^2_{\hat{\beta}_1}} = \left(\frac{\hat{\beta}_1 - 0}{\hat{\sigma}_{\hat{\beta}_1}}\right)^2$
- W folgt einer χ^2-Verteilung; die Zahl der Freiheitsgrade entspricht der Zahl der Beschränkungen ➔ Hier äquivalent zum „normalen“ Signifikanztest

Wie funktioniert der LR-Test?

- Der Likelihood-Ratio-Test ist im Grunde einfacher

Wie funktioniert der LR-Test?

- Der Likelihood-Ratio-Test ist im Grunde einfacher
- Vergleich der Likelihood für unbedingtes Modell und der Likelihood für Modell, das Bedingungen enthält (z. B. $\beta_1 = 0$), indem das Modell einfach ohne entsprechenden Koeffizienten geschätzt wird

Wie funktioniert der LR-Test?

- Der Likelihood-Ratio-Test ist im Grunde einfacher
- Vergleich der Likelihood für unbedingtes Modell und der Likelihood für Modell, das Bedingungen enthält (z. B. $\beta_1 = 0$), indem das Modell einfach ohne entsprechenden Koeffizienten geschätzt wird
- In der Praxis eigentlich nur eine Art von Bedingungen: Ein oder mehrere Koeffizienten gleich null

Wie funktioniert der LR-Test?

- Der Likelihood-Ratio-Test ist im Grunde einfacher
- Vergleich der Likelihood für unbedingtes Modell und der Likelihood für Modell, das Bedingungen enthält (z. B. $\beta_1 = 0$), indem das Modell einfach ohne entsprechenden Koeffizienten geschätzt wird
- In der Praxis eigentlich nur eine Art von Bedingungen: Ein oder mehrere Koeffizienten gleich null
- Das Modell mit den Bedingungen ist im nicht-restringierten Modell enthalten (nested)

Wie funktioniert der LR-Test?

- Der Likelihood-Ratio-Test ist im Grunde einfacher
- Vergleich der Likelihood für unbedingtes Modell und der Likelihood für Modell, das Bedingungen enthält (z. B. $\beta_1 = 0$), indem das Modell einfach ohne entsprechenden Koeffizienten geschätzt wird
- In der Praxis eigentlich nur eine Art von Bedingungen: Ein oder mehrere Koeffizienten gleich null
- Das Modell mit den Bedingungen ist im nicht-restringierten Modell enthalten (nested)
- Die Likelihood Ratio Statistic entspricht der Differenz zwischen zweimal der Log-Likelihood für das unbeschränkte Modell - zweimal der Log-Likelihood für das restringierte Modell

Wie funktioniert der LR-Test?

- Der Likelihood-Ratio-Test ist im Grunde einfacher
- Vergleich der Likelihood für unbedingtes Modell und der Likelihood für Modell, das Bedingungen enthält (z. B. $\beta_1 = 0$), indem das Modell einfach ohne entsprechenden Koeffizienten geschätzt wird
- In der Praxis eigentlich nur eine Art von Bedingungen: Ein oder mehrere Koeffizienten gleich null
- Das Modell mit den Bedingungen ist im nicht-restringierten Modell enthalten (nested)
- Die Likelihood Ratio Statistic entspricht der Differenz zwischen zweimal der Log-Likelihood für das unbeschränkte Modell - zweimal der Log-Likelihood für das restringierte Modell
- Diese Größe ist asymptotisch χ^2-verteilt

Wie funktioniert der Lagrange Multiplier Test?

- Hier wird nur das restringierte Modell geschätzt und der Wert der Score-Funktion (Steigung der Likelihood-Funktion) berechnet

Wie funktioniert der Lagrange Multiplier Test?

- Hier wird nur das restringierte Modell geschätzt und der Wert der Score-Funktion (Steigung der Likelihood-Funktion) berechnet
- Diese Funktion sollte nahe null sein (Maximum der Likelihood-Funktion)

Wie funktioniert der Lagrange Multiplier Test?

- Hier wird nur das restringierte Modell geschätzt und der Wert der Score-Funktion (Steigung der Likelihood-Funktion) berechnet
- Diese Funktion sollte nahe null sein (Maximum der Likelihood-Funktion)
- Wenn die Nullhypothese gilt, ist dieser Wert ebenfalls χ^2

Kann man das mal praktisch sehen?

- Ja

Kann man das mal praktisch sehen?

- Ja
- Preisfrage: Wie müssen $\mathbf{Q}$ und $\mathbf{r}$ für den letzten Test aussehen?

Kann man das mal praktisch sehen?

- Ja
- Preisfrage: Wie müssen **Q** und **r** für den letzten Test aussehen?

$$\begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}$$

Kann man das mal praktisch sehen?

- Ja
- Preisfrage: Wie müssen **Q** und **r** für den letzten Test aussehen?

$$\begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}$$

- Können wir auch testen, ob drei oder mehr Koeffizienten gleich sind?

Kann man das mal praktisch sehen?

- Ja
- Preisfrage: Wie müssen **Q** und **r** für den letzten Test aussehen?

$$\begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}$$

- Können wir auch testen, ob drei oder mehr Koeffizienten gleich sind?

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Was denn nun?

- Alle drei Tests sind asymptotisch äquivalent

Was denn nun?

- Alle drei Tests sind asymptotisch äquivalent
- In endlichen Stichproben können sich die Ergebnisse unterscheiden

Was denn nun?

- Alle drei Tests sind asymptotisch äquivalent
- In endlichen Stichproben können sich die Ergebnisse unterscheiden
- Viele Statistiker halten den Likelihood-Ratio-Test für den besten

Was denn nun?

- Alle drei Tests sind asymptotisch äquivalent
- In endlichen Stichproben können sich die Ergebnisse unterscheiden
- Viele Statistiker halten den Likelihood-Ratio-Test für den besten
- In der Praxis entscheidet im wesentlichen die Konvenienz, welcher Test benutzt wird

Was helfen erwartete Wahrscheinlichkeiten?

- Offengestanden kann sich kein Mensch etwas unter einem Logit oder seiner Veränderung vorstellen

Was helfen erwartete Wahrscheinlichkeiten?

- Offengestanden kann sich kein Mensch etwas unter einem Logit oder seiner Veränderung vorstellen
- Deshalb: Logits in Wahrscheinlichkeiten zurücktransformieren

Was helfen erwartete Wahrscheinlichkeiten?

- Offengestanden kann sich kein Mensch etwas unter einem Logit oder seiner Veränderung vorstellen
- Deshalb: Logits in Wahrscheinlichkeiten zurücktransformieren
- Problem: Wahrscheinlichkeiten hängen vom Wert aller unabhängiger Variablen ab

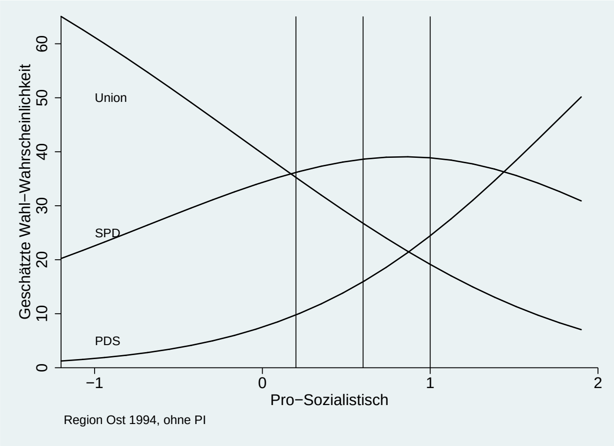
Was helfen erwartete Wahrscheinlichkeiten?

- Offengestanden kann sich kein Mensch etwas unter einem Logit oder seiner Veränderung vorstellen
- Deshalb: Logits in Wahrscheinlichkeiten zurücktransformieren
- Problem: Wahrscheinlichkeiten hängen vom Wert aller unabhängiger Variablen ab
- Tabellen oder Grafiken mit typischen Szenarien

Ein Beispiel?

	CDU/CSU	SPD	FDP	B90/Grüne	PDS
PI: Union	3.381**	-0.280	1.621**	-0.030	-0.740
	(0.143)	(0.176)	(0.203)	(0.281)	(0.411)
PI: SPD	-0.436*	2.805**	0.297	1.627**	0.866**
	(0.183)	(0.129)	(0.250)	(0.176)	(0.209)
PI: FDP	1.014*	-0.240	4.681**	0.768	-0.679
	(0.441)	(0.508)	(0.408)	(0.636)	(1.079)
PI: Grüne	-0.643	0.634**	-0.965	4.060**	1.067**
	(0.335)	(0.239)	(0.745)	(0.231)	(0.348)
PI: PDS	-2.276**	0.086	-0.307	0.178	3.527**
	(0.742)	(0.276)	(0.751)	(0.509)	(0.238)
Pro-Sozialismus	-0.600**	-0.190**	-0.628**	-0.219*	0.525**
	(0.077)	(0.070)	(0.119)	(0.099)	(0.111)

Ein Beispiel?



Übung für heute

- Versuchen Sie, in zwei bis drei Sätzen zu erklären, warum die SPD bei den nicht-parteigebundenen Ostdeutschen 1994 relativ gut abgeschnitten hat. Gehen Sie dabei auf den Median und die Quartile ein.