Tutorium Allgemeines Lineares Modell

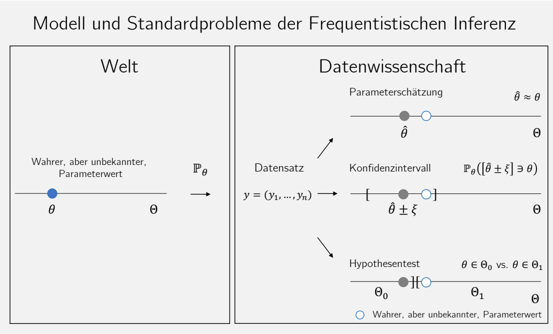
BSc Psychologie SoSe 2022

8. Termin (Teil 2): Modellevaluation

Belinda Fleischmann

Inhalte basieren auf ALM Kursmaterialien von Dirk Ostwald, lizenziert unter CC BY-NC-SA 4.0

Wiederholung - Frequentistisches Weltbild



- Wir nehmen an, dass wahre, aber unbekannte Parameter existieren.
- Wir nehmen weiterhin an, dass probabilistische Prozesse existieren, die, gegeben dieser wahren, aber unbekannten Parameter, Datensätze generieren können.
- Für diese probabilistischen Prozesse gehen wir davon aus, dass ihnen bestimmte Verteilungen bzw. Wahrscheinlichkeitsdichten zugrundeliegen (z.B. Normalverteilung der Zufallsfehler)
- Wir verwenden erhobene Daten dafür, Parameterwerte zu schätzen (Wir berechnen eine "Einschätzung", was der wahre Wert sein könnte, den wir nicht beobachten können).
- Dabei bilden die angenommenen Verteilungen bzw. Warhscheinlichkeitsdichten der probabilistischen Prozesse (die, wie wir annehmen, die Daten generiert haben), die Grundlage für Parameterschätzung und Modellevaluation.

Wiederholung - Standardannahmen frequentistischer Inferenz

- Gegeben sei ein statistisches Modell $\mathbb{P}_\theta$, in dem probabilistische Prozesse definiert sind, die Datensätze generieren können.
- Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen der Daten des Modells ist. Eine mögliche Realisierung wäre $y^{(1)}$, $y^{(2)}$ wäre eine andere, $y^{(3)}$ wäre nochmals eine andere.
- Aus frequentistischer Sicht kann man unendlich oft Datensätze basierend auf einem Modell generieren und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. den Betaparameterschätzer (z.B. $\hat{\beta}^{(1)}$ für Datensatz $y^{(1)}$)

Datensatz (1) : $y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, \ldots, y_n^{(1)}\right)^T$ mit $\hat{\beta}^{(1)} = (X^T X)^{-1} X^T y^{(1)}$

Datensatz (2) : $y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, \ldots, y_n^{(2)}\right)^T$ mit $\hat{\beta}^{(2)} = (X^T X)^{-1} X^T y^{(2)}$

Datensatz (3) : $y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, \ldots, y_n^{(3)}\right)^T$ mit $\hat{\beta}^{(3)} = (X^T X)^{-1} X^T y^{(3)}$

Datensatz (4) : $y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, \ldots, y_n^{(4)}\right)^T$ mit $\hat{\beta}^{(4)} = (X^T X)^{-1} X^T y^{(4)}$

Datensatz (5) : $y^{(5)} = \ldots$

- Um die Qualität statistischer Methoden zu beurteilen betrachtet die frequentistische Statistik die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme der Datenverteilung (z.B. im ALM die Verteilung des Datenvektors $y = X\beta + \epsilon$ mit $\epsilon \sim N(0_n, \sigma^2 I_n)$, und damit $y \sim N(x\beta, \sigma^2 I_n)$, siehe Einheit (5) Theorem zu ALM Datenverteilung).
- Was ist zum Beispiel die Verteilung (möglicher) Betaparameterschätzer ($\hat{\beta}^{(1)}$, $\hat{\beta}^{(2)}$, $\hat{\beta}^{(3)}$, $\hat{\beta}^{(4)}$), ... also die Verteilung der Zufallsvariable $\hat{\beta} := (X^T X)^{-1} X^T y$?

Selbstkontrollfragen - Modellevaluation

1. Geben Sie das Theorem zur Frequentistischen Verteilung des Betaparameterschätzers wieder.
2. Geben Sie die Verteilung des Betaparameterschätzers im Szenarion von n u.i.v. Zufallsvariablen wieder.
3. Geben Sie das Theorem zur Frequentistischen Verteilung des Varianzparameterschätzers wieder.
4. Geben Sie die Verteilung des skalierten Varianzparameterschätzers bei n u.i.v. Zufallsvariablen wieder.
5. Skizzieren Sie die WDFen von t-Zufallsvariablen mit 2, 10 und 30 Freiheitsgraden.
6. Skizzieren Sie die WDFen von nichtzentralen t-Zufallsvariablen mit Nichtzentralitätsparametern $0{,}5$ und 15.
7. Geben Sie die Definition der T-Statistik wieder.
8. Erläutern Sie die Definition der T-Statistik.
9. Warum kann die T-Statistik als Signal-zu-Rauschen Verhältnis interpretiert werden.
10. Geben Sie die Form der T-Statistik im Szenario von n u.i.v. Zufallsvariablen wieder.
11. Erläutern Sie den Zusammenhang der T-Statistik und Cohen's d.
12. Geben Sie die Definition der T-Teststatistik wieder.

Selbstkontrollfragen - Modellevaluation

13. Erläutern Sie die Definition der T-Teststatistik.
14. Geben Sie das Theorem zur Verteilung der T-Teststatistik wieder.
15. Geben Sie die Definition eines vollständigem und eines reduzierten ALMs wieder.
16. Geben Sie die Definition des Likelihood-Quotienten eines vollständigen und eines reduzierten ALMs wieder.
17. Geben Sie das Theorem zu Likelihood-Quotient und Residualquadratsummendifferenz wieder.
18. Definieren Sie die F-Statistik.
19. Erläutern Sie den Zusammenhang zwischen F-Statistik und Likelihood-Quotient.
20. Definieren Sie und erläutern Sie den Zähler der F-Statistik.
21. Definieren Sie und erläutern Sie den Nenner der F-Statistik.
22. Erläutern Sie die F-Statistik.
23. Geben Sie das Theorem zur Verteilung der F-Statistik wieder.

1. Geben Sie das Theorem zur Frequentistischen Verteilung des Betaparameterschätzers wieder.

Theorem (Frequentistische Verteilung des Betaparameterschätzers)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (1)$$

das ALM in generativer Form. Weiterhin sei

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2)$$

der Betaparameterschätzer. Dann gilt

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}). \quad (3)$$

Anmerkungen:

- Da es sich bei $\hat{\beta}$ um eine linear-affine Transformation von y handelt, leitet sich die die hier angegeben Verteilung für $\hat{\beta}$ aus dem Theorem zur linearen Transformation von multivariaten Normalverteilung aus Einheit (4) ab.
- Im Spezifischen besagt das Theorem in diesem Fall, dass wenn $y \sim N(X\beta, \sigma^2 I_n)$ und $\hat{\beta} = Ay + b$, dann gilt $\hat{\beta} \sim N(AX\beta + b, A\sigma^2 I_n A^T)$, wobei $A = (X^T X)^{-1} X^T$ der Faktor der linearen-affinen Transformation ist.

2. Geben Sie die Verteilung des Betaparameterschätzers im Szenario von n u.i.v. Zufallsvariablen wieder.

Das Szenario von n u.i.v. Zufallsvariablen ist in Matrixschreibweise gegeben durch

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := 1_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0,$$

wobei der Betaparameterschätzer gegeben ist durch $\hat{\beta} = \bar{y}$. Durch Einsetzen von $\beta := \mu$ und $X := 1_n$ in $\hat{\beta} \sim N(\beta, \sigma^2(X^T X)^{-1})$ (vlg. Theorem zur Frequentistischen Verteilung des Betaparameterschätzers), erhalten wir

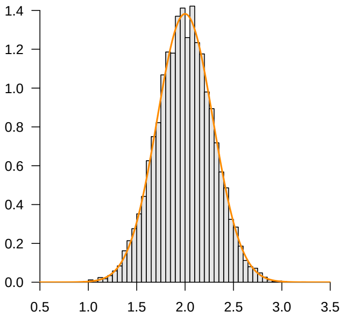
$$\bar{y} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Das Stichprobenmittel ($\bar{y}$) von n u.i.v. Zufallsvariablen mit Erwartungswertparameter μ und Varianzparameter σ^2 ist also normalverteilt mit Erwartungswertparameter μ und Varianzparameter σ^2/n.

2. Geben Sie die Verteilung des Betaparameterschätzers im Szenario von n u.i.v. Zufallsvariablen wieder.

Visualisierung von $N\left(\hat{\beta}; \beta, \frac{\sigma^2}{n}\right)$ für $y \sim N(X\beta, \sigma^2 I_n)$ mit $\beta = \mu = 2$ und $\sigma^2 = 1$. Mit anderen Worten, wie sieht die Verteilung der Zufallvariable $\hat{\beta}$ aus, wenn wir aus der von uns angenommenen "wahren" Verteilung des Zufallsvektors $y \sim N(X\beta, \sigma^2 I_n)$ ganz viele Realisierungen generieren, (z.B. 10000 Realisierungen $y^{(1)}, \ldots, y^{(10000)}$), und für jede dieser Datensätze einen Betaparameter schätzen ($\hat{\beta}^{(1)}, \ldots, \hat{\beta}^{(10000)}$).

$N\left(\hat{\beta}; \beta, \frac{\sigma^2}{n}\right)$



3. Geben Sie das Theorem zur Frequentistischen Verteilung des Varianzparameterschätzers wieder.

Theorem (Frequentistische Verteilung des Varianzparameterschätzers)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{4}$$

das ALM in generative Form. Weiterhin sei

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})^T(y - X\hat{\beta})}{n - p} \tag{5}$$

der Varianzparameterschätzer. Dann gilt

$$\frac{n-p}{\sigma^2}\hat{\sigma}^2 \sim \chi^2(n-p) \tag{6}$$

4. Geben Sie die Verteilung des skalierten Varianzparameterschätzers bei n u.i.v. Zufallsvariablen wieder.

Wie in Aufg. 2, haben wir im Szenario von n u.i.v. Zufallsvariablen

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := 1_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0,$$

wobei der Betaparameterschätzer gegeben ist durch $\hat{\beta} = \bar{y}$ und der Varianzparameterschätzer durch $\hat{\sigma}^2 = s^2$. Durch Einsetzen von $p = 1$ in die Verteilung des skalierten Varianzparameterschätzers ($\frac{n-p}{\sigma^2}\hat{\sigma}^2$), wie im Theorem zur Frequentistischen Verteilung des Varianzparamterschätzers gegeben, erhalten wir

$$\frac{n-1}{\sigma^2}\hat{\sigma}^2 \sim \chi^2(n-1)$$

Das ist identisch mit der in Einheit (11) Konfidenzintervalle von Wahrscheinlichkeitstheorie und Frequentistische Inferenz gelernten U-Statistik. Wdhl.: Für den Fall von n unabhängig und identisch normalverteilten Zufallsvariablen ist die U-Statistik definiert als

$$U := \frac{n-1}{\sigma^2} s^2,$$

wobei

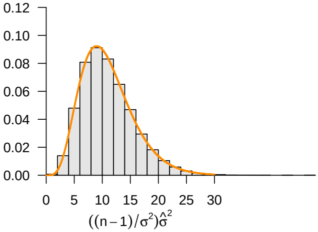
$$U \sim \chi^2(n-1).$$

Offenbar ist U für $p = 1$ mit der im obigen Theorem betrachten Zufallsvariable $\frac{n-p}{\sigma^2}\hat{\sigma}^2$ identisch.

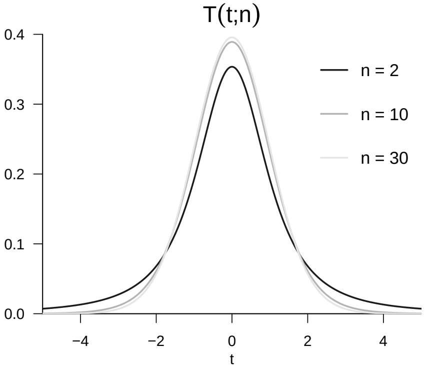
4. Geben Sie die Verteilung des skalierten Varianzparameterschätzers bei n u.i.v. Zufallsvariablen wieder.

Visualisierung von $\chi^2(\frac{n-1}{\sigma^2}\hat{\sigma}^2; n-1)$ für $y \sim N(X\beta, \sigma^2 I_n)$ mit $\beta = \mu = 2$ und $\sigma^2 = 1$. Mit anderen Worten, wie sieht die Verteilung der skalierten Zufallvariable $\hat{\sigma^2}$, also $\frac{n-p}{\sigma}^2\hat{\sigma}^2$ aus, wenn wir aus der von uns angenommenen "wahren" Verteilung des Zufallsvektors $y \sim N(X\beta, \sigma^2 I_n)$ ganz viele Realisierungen generieren, (z.B. 10000 Realisierungen $y^{(1)}, ..., y^{(10000)}$), und für jede dieser Datensätze einen Varianzparameter schätzen ($\hat{\sigma}^{2(1)}, ..., \hat{\sigma}^{2(10000)}$).

$\chi^2(\frac{n-p}{\sigma^2}\hat{\sigma}^2; n-1)$



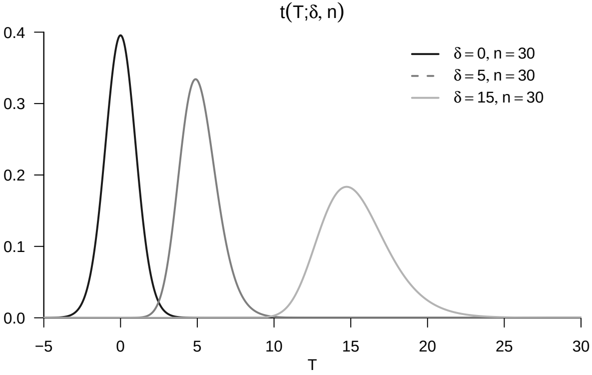
5. Skizzieren Sie die WDFen von t-Zufallsvariablen mit 2, 10 und 30 Freiheitsgraden.



Anmerkungen:

- Die Verteilung ist um 0 symmetrisch
- Steigendes n verschiebt Wahrscheinlichkeitsmasse aus den Ausläufen zum Zentrum

6. Skizzieren Sie die WDFen von nichtzentralen t-Zufallsvariablen mit Nichtzentralitätsparametern $0{,}5$ und 15.



7. Geben Sie die Definition der T-Statistik wieder.

Definition (T-Statistik)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{7}$$

das ALM in generativer Form. Weiterhin seien

$$\hat{\beta} := (X^T X)^{-1} X^T y \text{ und } \hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \tag{8}$$

die Betaparameter- und Varianzparameterschätzer, respektive. Dann ist für einen *Kontrastgewichtsvektor* $c \in \mathbb{R}^p$ die *T-Statistik* definiert als

$$T := \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}. \tag{9}$$

8. Erläutern Sie die Definition der T-Statistik.

- Die T-Statistik hängt durch die Parameterschätzer $\hat{\beta}$ und $\hat{\sigma}^2$ von den Daten y ab.
- Der Kontrastgewichtsvektor $c \in \mathbb{R}^p$ projiziert $\hat{\beta}$ auf einen Skalar $c^T\hat{\beta}$.
- Intuitiv quantifiziert die T-Statistik das Verhältnis von geschätzte Effektstärke zur geschätzten stichprobenumfangskalierten Datenvariabilität

$$T = \frac{\text{Geschätzte Effektstärke}}{\text{Geschätzte stichprobenumfangskalierte Datenvariabilität}}$$

und repräsentiert damit ein Signal-zu-Rauschen Verhältnis (signal-to-noise ratio).

9. Warum kann die T-Statistik als Signal-zu-Rauschen Verhältnis interpretiert werden.

- Im Zähler steht der Betaparameterschätzer, und im Nenner steht der Varianzparameterschätzer
- Der Term $c^T\hat{\beta}$ quantifiziert den geschätzten Effekt (wie groß ist der Beitrag eines oder mehrerer Faktoren (je nachdem wie man den Kontrast definiert) zur Gesamtdatenvariabilität und $\sqrt{\hat{\sigma}^2 c^T(X^TX)^{-1}c}$ quantifiziert die geschätzte stichprobenumfangskalierte Datenvariabilität "Restrauschen (?), Fehlervarianz (?)"
- Der Effekt ist also das ("eigentliche") Signal und die Fehlervarianz das Restrauschen".
- Die T-Statistik bestimmt das Verhältnis von Signal zu Rauschen und quantifiziert somit auch Unsicherheit.

10. Geben Sie die Form der T-Statistik im Szenario von n u.i.v. Zufallsvariablen wieder.

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Wir haben das ALM Szenario unabhängiger und identisch normalverteilter Zufallsvariablen

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := 1_n \in \mathbb{R}^{n\times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0.$$

Weiterhin sei $c := 1$. Dann gilt für die T-Statistik

$$T = \frac{c^T\hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} = \frac{1^T \bar{y}}{\sqrt{s_y^2 1^T (1_n^T 1_n)^{-1} 1}} = \sqrt{n}\frac{\bar{y}}{s_y}$$

was der Einstichproben-T-Teststatistik für den Fall $\mu_0 = 0$ entspricht (vgl. Einheit (13) Einstichproben-T-Tests in Wahrscheinlichkeitstheorie und Frequentistische Inferenz und Einheit (9) T-Tests in Allgemeines Lineares Modell).

11. Erläutern Sie den Zusammenhang der T-Statistik und Cohen's d.

Die T-Statistik ist im ALM Szenario unabhängiger und identisch normalverteilter Zufallsvariablen gegeben durch

$$T = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} = \frac{1^T \bar{y}}{\sqrt{s_y^2 1^T (1_n^T 1_n)^{-1} 1}} = \sqrt{n} \frac{\bar{y}}{s_y}$$

Die T-Statistik (in allen Fällen oder nur u.i.v. ?) nimmt hohe Werte für hohe Werte von $\bar{y}$ (Effekt), kleine Werte von s_y^2 (Datenvariabilität) und hohe Werte von n (Stichprobenumfang) an.

Cohen's d als *Effektstärkenmaß* ist gegeben als

$$d := \frac{\bar{y}}{s_y^2},$$

so dass

$$T = \sqrt{n} d \text{ bzw. } d = \frac{1}{\sqrt{n}} T$$

Cohen's d ist also ein Stichprobenumfangunabhängiges Signal-zu-Rauschen Verhältnis.

12. Geben Sie die Definition der T-Teststatistik wieder.

Theorem (T-Teststatistik, Teil 1/2)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{10}$$

das ALM in generativer Form. Weiterhin seien

$$\hat{\beta} := (X^T X)^{-1} X^T y \text{ und } \hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \tag{11}$$

die Betaparameter- und Varianzparameterschätzer, respektive. Schließlich sei für einen *Kontrastgewichtsvektor* $c \in \mathbb{R}^p$ und einen *Nullhypothesenbetaparameter* $\beta_0 \in \mathbb{R}^p$ die *T-Teststatistik* definiert als

$$T := \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}. \tag{12}$$

13. Erläutern Sie die Definition der T-Teststatistik.

- T ist eine Funktion der Parameterschätzer, δ ist eine Funktion der wahren, aber unbekannten, Parameter
- Für $c^T\beta = c^T\beta_0$, also bei Zutreffen der Nullhypothese, gilt $\delta = 0$ und damit $T \sim t(n-p)$.

14. Geben Sie das Theorem zur Verteilung der T-Teststatistik wieder.

Theorem (T-Teststatistik, Teil 2/2)

Dann gilt

$$T \sim t(\delta, n-p) \text{ mit } \delta := \frac{c^T\beta - c^T\beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}} \tag{13}$$

15. Geben Sie die Definition eines vollständigen und eines reduzierten ALMs wieder.

Definition (Vollständiges und reduziertes Modell)

Für $p > 1$ mit $p = p_1 + p_2$ seien

$$X := \begin{pmatrix} X_1 & X_2 \end{pmatrix} \in \mathbb{R}^{n \times p} \text{ mit } X_1 \in \mathbb{R}^{n \times p_1} \text{ und } X_2 \in \mathbb{R}^{n \times p_2}, \tag{14}$$

sowie

$$\beta := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^{p} \text{ mit } \beta_1 \in \mathbb{R}^{p_1} \text{ und } \beta_2 \in \mathbb{R}^{p_2} \tag{15}$$

Partitionierungen einer $n \times p$ Designmatrix und eines p-dimensionalen Betaparametervektors. Dann nennen wir

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{16}$$

das *vollständige Modell* und

$$y = X_1\beta_1 + \varepsilon_1 \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{17}$$

das *reduzierte Modell* und sprechen von einer *Partionierung eines (vollständigen) Modells*.

16. Geben Sie die Definition des Likelihood-Quotienten eines vollständigen und eines reduzierten ALMs wieder.

Definition (Likelihood-Quotient von vollständigem und reduzierten Modell)

Für $p = p_1 + p_2, p > 1$ sei eine Partitionierung eines vollständigen ALMs gegeben und $\sigma^2 > 0$ als bekannt vorausgesetzt. Weiterhin seien

$$L : \mathbb{R}^p \to \mathbb{R}_{>0}, \beta \mapsto L(\beta) := p_\beta(\upsilon) = N(\upsilon; X\beta, \sigma^2 I_n) \tag{18}$$

und

$$L_1 : \mathbb{R}^{p_1} \to \mathbb{R}_{>0}, \beta_1 \mapsto L(\beta_1) := p_{\beta_1}(\upsilon) = N(\upsilon; X_1\beta_1, \sigma^2 I_n) \tag{19}$$

die Likelihood-Funktionen von vollständigem und reduzierten Modell, respektive. Dann ist für einen Datenvektor $\upsilon \in \mathbb{R}^n$ der *Likelihood-Quotient von vollständigem und reduziertem Modell* gegeben als

$$\varrho := \frac{\max_{\beta \in \mathbb{R}^p} p_\beta(\upsilon)}{\max_{\beta_1 \in \mathbb{R}^{p_1}} p_{\beta_1}(\upsilon)} = \frac{\max_{\beta \in \mathbb{R}^p} L(\beta)}{\max_{\beta_1 \in \mathbb{R}^{p_1}} L_1(\beta_1)} = \frac{L(\hat{\beta})}{L_1(\hat{\beta}_1)} \tag{20}$$

17. Geben Sie das Theorem zu Likelihood-Quotient und Residualquadratsummendifferenz wieder.

Theorem (Likelihood-Quotient und Residualquadratsummendifferenz)

Für $p = p_1 + p_2, p > 1$ sei eine Partitionierung eines vollständigen ALMs gegeben und $\sigma^2 > 0$ als bekannt voraussgesetzt. Weiterhin seien

$$\hat{\varepsilon} := y - X\hat{\beta} \text{ und } \hat{\varepsilon}_1 := y - X_1\hat{\beta}_1 \tag{21}$$

die Residuenvektoren des vollständigen und es reduzierten Modells, respektive. Dann gilt

$$\ln \varrho = \frac{\hat{\varepsilon}_1^T \hat{\varepsilon}_1 - \hat{\varepsilon}^T \hat{\varepsilon}}{2\sigma^2} \tag{22}$$

18. Definieren Sie die F-Statistik.

Definition (F-Statistik)

Für $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^p$ und $\sigma^2 > 0$ sei ein ALM der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{23}$$

mit der Partitionierung

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}, X_1 \in \mathbb{R}^{n \times p_1}, X_2 \in \mathbb{R}^{n \times p_2}, \text{ und } \beta := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \beta_1 \in \mathbb{R}^{p_1}, \beta_2 \in \mathbb{R}^{p_2}, \tag{24}$$

mit $p = p_1 + p_2$ gegeben. Weiterhin seien mit

$$\hat{\beta}_1 := (X_1^T X_1)^{-1} X_1^T y \text{ und } \hat{\beta} := (X^T X)^{-1} X^T y \tag{25}$$

die Residuenvektoren

$$\hat{\varepsilon}_1 := y - X_1 \hat{\beta}_1 \text{ und } \hat{\varepsilon} := y - X\hat{\beta} \tag{26}$$

definiert. Dann ist die F-Statistik definiert als

$$F := \frac{\frac{\hat{\varepsilon}_1^T \hat{\varepsilon}_1 - \hat{\varepsilon}^T \hat{\varepsilon}}{p_2}}{\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-p}} \tag{27}$$

19. Erläutern Sie den Zusammenhang zwischen F-Statistik und Likelihood-Quotient.

- Die F-Statistik ist nicht das gleiche wie der Likelihood-Quotient, misst aber intuitiv das gleiche.
- Wie auch der Likelihood-Quotient, misst die F-Statistik, wie viel wahrscheinlicher die DAten sind unter dem vollständigen im Vergleich zum reduzierten Modell. Anders ausgedrückt, inwieweit das Hinzunehmen des Regressors p_2 die Erklärung der Daten verbessert.

20. Definieren Sie und erläutern Sie den Zähler der F-Statistik.

- Für den Nenner der F-Statistik gilt
$$\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-p} = \hat{\sigma}^2, \tag{28}$$
wobei $\hat{\sigma}^2$ hier der aufgrund des vollständigen Modells geschätzte Schätzer von σ^2 ist. Werden die Daten tatsächlich unter dem reduzierten Modell generiert, so kann das vollständige Modell dies durch $\hat{\beta}_2 \approx 0_{p_2}$ abbilden und erreicht eine ähnliche σ^2 Schätzung wie das reduzierte Modell. Werden die Daten de-facto unter dem vollständigem Modell generiert, so ist $\hat{\varepsilon}^T \hat{\varepsilon}/(n-p)$ ein besserer Schätzer von σ^2 als $\hat{\varepsilon}_1^T \hat{\varepsilon_1}/(n-p)$, da sich für diesen Datenvariabilität, die nicht durch die p_1 Regressoren in X_1 erklärt wird, in der Schätzung von σ^2 widerspiegeln würde. Der Nenner der F-Statistik ist also in beiden Fällen der sinnvollere Schätzer von σ^2.

21. Definieren Sie und erläutern Sie den Nenner der F-Statistik.

- Der Zähler der F-Statistik

$$\frac{\hat{\varepsilon}_1^T \hat{\varepsilon}_1 - \hat{\varepsilon}^T \hat{\varepsilon}}{p_2} \tag{29}$$

misst, inwieweit die p_2 Regressoren in X_2 die Residualquadratsumme reduzieren und zwar im Verhältnis zur Anzahl dieser Regressoren. Das heißt, dass bei gleicher Größe der Residualquadratsummenreduktion (und gleichem Nenner) ein größerer F Wert resultiert, wenn diese durch weniger zusätzliche Regressoren resultiert, also p_2 klein ist (und vice versa). Im Sinne der Anzahl der Spalten von X und der entsprechenden Komponenten von β favorisiert die F-Statistik also weniger "komplexe" Modelle.

22. Erläutern Sie die F-Statistik.

- Zusammengenommen misst die F-Statistik also die Residualquadratsummenreduktion durch die p_2 Regressoren in X_2 gegenüber den p_1 Regressoren in X_1 pro Datenvariabilitäts (σ^2)- und Regressor (p_2)-Einheit.

23. Geben Sie das Theorem zur Verteilung der F-Statistik wieder.

Theorem (F-Statistik)

Für $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^p$ und $\sigma^2 > 0$ sei ein ALM der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{30}$$

mit der Partitionierung

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}, X_1 \in \mathbb{R}^{n \times p_1}, X_2 \in \mathbb{R}^{n \times p_2}, \text{ und } \beta := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \beta_1 \in \mathbb{R}^{p_1}, \beta_2 \in \mathbb{R}^{p_2}, \tag{31}$$

mit $p = p_1 + p_2$ gegeben. Schließlich sei

$$K := \begin{pmatrix} 0_{p_1} \\ 1_{p_2} \end{pmatrix} \in \mathbb{R}^p \tag{32}$$

ein Kontrastgewichtsvektor. Dann gilt

$$F \sim f(\delta, n, n-p) \text{ mit } \delta := \frac{K^T\beta \left(K^T(X^TX)^{-1}K\right)^{-1} K^T\beta}{\sigma^2} \tag{33}$$

TODO

Beispiel F-Statistik mit 5 Spalten? Video 2:45:00