Tutorium

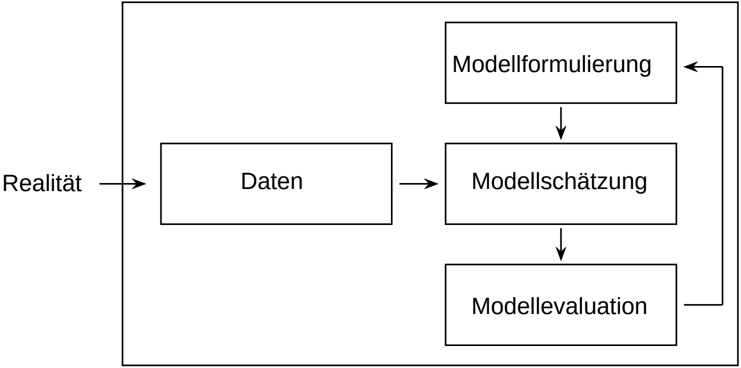
Allgemeines Lineares Modell

BSc Psychologie SoSe 2023

9. Termin: (8) F-Statistiken

Sean Mulready

Wiederholung Naturwissenschaft



Modellformulierung

$$\upsilon = X\beta + \varepsilon, \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{1}$$

Modellschätzung

$$\hat{\beta} = (X^T X)^{-1} x^T \upsilon,\ \hat{\sigma}^2 = \frac{(\upsilon - X\hat{\beta})^T(\upsilon - X\hat{\beta})}{n-p} \tag{2}$$

Modellevaluation

$$T = \frac{c^T\hat{\beta} - c^T\beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}},\ F = \frac{(\hat{\varepsilon}_0^T\hat{\varepsilon}_0 - \hat{\varepsilon}^T\hat{\varepsilon})/p_1}{\hat{\varepsilon}^T\hat{\varepsilon}/(n-p)} \tag{3}$$

Standardprobleme Frequentistischer Inferenz

(1) Parameterschätzung

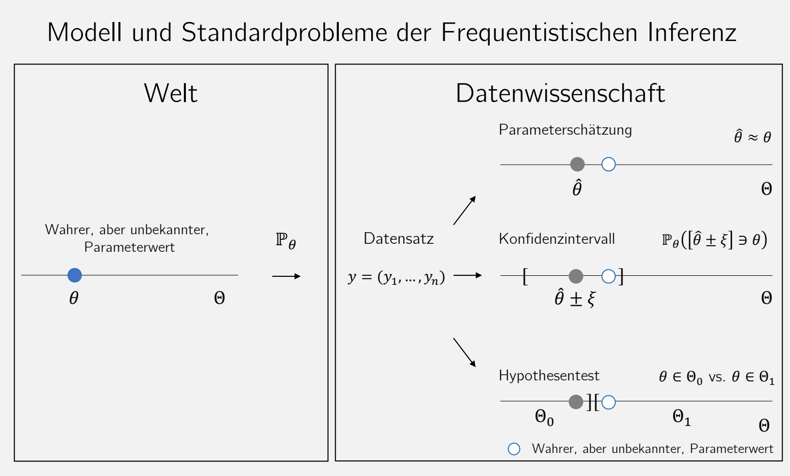
Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für wahre, aber unbekannte, Parameterwerte oder Funktionen dieser abzugeben, typischerweise mithilfe von Daten.

(2) Konfidenzintervalle

Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der angenommenen Verteilung der Daten eine quantitative Aussage über die mit Schätzwerten assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Das Ziel des Hypothesentestens ist es, basierend auf der angenommenen Verteilung der Daten in einer möglichst zuverlässigen Form zu entscheiden, ob ein wahrer, aber unbekannter Parameterwert in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes liegt.



$$\theta := (\beta, \sigma^2),\ \Theta := \mathbb{R}p \times \mathbb{R}_{>0}\ ,\ \mathbb{P}_{\beta,\sigma^2}(\upsilon) \text{ mit WDF } p_{\beta,\sigma^2}(y) := N(y; X\beta, \hat{\sigma}^2 I_n)$$

Wiederholung Standardannahmen frequentistischer Inferenz

Gegeben sei das Allgemeine Lineare Modell. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen der Daten des Modells ist. 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

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 $\hat{\beta}^{(4)} = (X^T X)^{-1} X^T y^{(4)}$

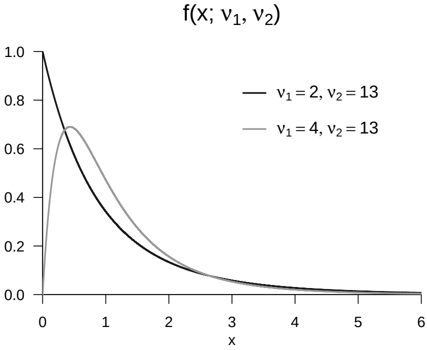
Datensatz (5) : $y^{(5)} = \ldots$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme der Datenverteilung. Was zum Beispiel ist die Verteilung von $\hat{\beta}^{(1)}$, $\hat{\beta}^{(2)}$, $\hat{\beta}^{(3)}$, $\hat{\beta}^{(4)}$, . . . also die Verteilung der Zufallsvariable $\hat{\beta} := (X^T X)^{-1} X^T \upsilon$? Wenn eine statistische Methode im Sinne der Frequentistischen Standardannahmen "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Selbstkontrollfragen

1. Skizzieren Sie die f-Verteilung für $\nu_1 = 2, \nu_2 = 13$ und $\nu_1 = 4, \nu_2 = 13$.
2. Geben Sie die Definition der Likelihood-Quotienten-Statistik wieder.
3. Erläutern Sie die Definition der Likelihood-Quotienten-Statistik.
4. Geben Sie die Definition eines vollständigem und eines reduziertem ALMs wieder.
5. Geben Sie das Theorem zum Likelihood-Quotienten von vollständigem und reduzierten ALM wieder.
6. Definieren Sie die F-Statistik.
7. Erläutern Sie den Zähler der F-Statistik.
8. Erläutern Sie den Nenner der F-Statistik.
9. Erläutern Sie die F-Statistik.
10. Geben Sie das Theorem F-Statistik und Likelihood-Quotienten-Statistik wieder.
11. Geben Sie das Theorem zur Verteilung der F-Statistik wieder.

1. Skizzieren Sie die f-Verteilung für $\nu_1 = 2, \nu_2 = 13$ und $\nu_1 = 4, \nu_2 = 13$.



2. Geben Sie die Definition der Likelihood-Quotienten-Statistik wieder.

Definition (Likelihood-Quotienten-Statistik)

Gegeben seien zwei parametrische statistische Modelle

$$\mathcal{M}_0 := \left(\mathcal{Y}, \mathcal{A}, \left\{\mathbb{P}^0_{\theta_0} | \theta_0 \in \Theta_0\right\}\right) \text{ und } \mathcal{M}_1 := \left(\mathcal{Y}, \mathcal{A}, \left\{\mathbb{P}^1_{\theta_1} | \theta_1 \in \Theta_1\right\}\right) \tag{4}$$

mit identischem Datenraum, identischer σ-Algebra und potentiell distinkten Wahrscheinlichkeitsmaßmengen und Parameterräumen. Sei weiterhin υ ein Zufallsvektor mit Datenraum $\mathcal{Y}$. Seien schließlich L_0^υ und L_1^υ die Likelihood-Funktionen von $\mathcal{M}_0$ und $\mathcal{M}_1$, respektive, wobei das Superskript $^\upsilon$ jeweils an die Datenabhängigkeit der Likelihood Funktion erinnern soll. Dann wird

$$\Lambda := \frac{\max_{\theta_0 \in \Theta_0} L_0^\upsilon(\theta_0)}{\max_{\theta_1 \in \Theta_1} L_1^\upsilon(\theta_1)}, \tag{5}$$

Likelihood-Quotienten-Statistik genannt.

3.Erläutern Sie die Definition der Likelihood-Quotienten-Statistik.

Eine Likelihood-Quotienten-Statistik setzt die Wahrscheinlichkeitsmassen/dichten eines beobachteten Datensatzes $y \in \mathcal{Y}$ unter zwei statistischen Modellen nach *Optimierung der jeweiligen Modellparameter* ins Verhältnis. Ein hoher Wert des Likelihood-Quotienten-Statistik entspricht einer höhereren Wahrscheinlichkeitsmasse/dichte des beobachteten Datensatzes $y \in \mathcal{Y}$ unter $\mathcal{M}_0$ als unter $\mathcal{M}_1$ und vice versa.

Anmerkungen:

- Grob gesagt sagt mir die Likelihoodfunktion, wie sicher ich mir sein kann, dass der beobachtete Datensatz bzw seine Komponenten so zusammen auftreten (in Abhängigkeit vom Modell und Parameter)
- Wir betrachten das Verhältnis der maximalen Funktionswerte zweier Likelihood-Funktionen
- *Optimierter Modellparameter* bedeutet, wir nehmen jeweils den ML-Schätzer der Funktion, wie wir bereits gesehen haben, maximiert dieser die Likelohoodfunktion
- Das Ganze betrachten wir immer in Abhängigkeit **eines** Datensatzes

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

Definition (Vollständiges und reduziertes Modell)

Für $p > 1$ mit $p = p_0 + p_1$ seien

$$X := \begin{pmatrix} X_0 & X_1 \end{pmatrix} \in \mathbb{R}^{n \times p} \text{ mit } X_0 \in \mathbb{R}^{n \times p_0} \text{ und } X_1 \in \mathbb{R}^{n \times p_1}, \tag{6}$$

sowie

$$\beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \in \mathbb{R}^{p} \text{ mit } \beta_0 \in \mathbb{R}^{p_0} \text{ und } \beta_1 \in \mathbb{R}^{p_1} \tag{7}$$

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

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{8}$$

das *vollständige Modell* und

$$\upsilon = X_0\beta_0 + \varepsilon_0 \text{ mit } \varepsilon_0 \sim N(0_n, \sigma^2 I_n) \tag{9}$$

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

Beispiel - vollständiges und reduziertes Modell

Wir haben das ALM

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

in Matrix- bzw. ALM-Schreibweise

$$\upsilon \sim N(X\beta, \sigma^2 I_n) \text{ mit } X \in \mathbb{R}^{n \times p}, \beta \in \mathbb{R}^p \sigma^2 > 0$$

mit der Partitionierung

$$X = \begin{pmatrix} X_0 & X_1 \end{pmatrix}, X_0 \in \mathbb{R}^{n \times p_0}, X_1 \in \mathbb{R}^{n \times p_1}, \text{ und } \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \beta_0 \in \mathbb{R}^{p_0}, \beta_1 \in \mathbb{R}^{p_1}$$

mit $p = p_1 + p_2$.

Ausgeschrieben sieht das für ein Beispiel mit $p_1 = 1$ und $p_2 = 1$ wie folgt aus:

$$X_0 = \begin{pmatrix} x_{0,1} \\ \vdots \\ x_{0,n} \end{pmatrix}, X_1 = \begin{pmatrix} x_{1,1} \\ \vdots \\ x_{1,n} \end{pmatrix}$$

$$\upsilon = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} \upsilon_1 \\ \vdots \\ \upsilon_n \end{pmatrix} = \begin{pmatrix} x_{0,1} & x_{1,1} \\ \vdots & \vdots \\ x_{0,n} & x_{1,n} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} x_{0,1}\beta_0 + x_{1,1}\beta_1 + \varepsilon_1 \\ \vdots \\ x_{0,n}\beta_0 + x_{1,n}\beta_1 + \varepsilon_n \end{pmatrix}$$

Beispiel - vollständiges und reduziertes Modell

Die Residuenvektoren

$$\hat{\varepsilon}_0 := \upsilon - X_0\hat{\beta}_0 \text{ und } \hat{\varepsilon} := \upsilon - X\hat{\beta}$$

sehen dann ausgeschrieben so aus:

$$\hat{\varepsilon}_0 := \upsilon - X_0\hat{\beta}_0 \Leftrightarrow \begin{pmatrix} \hat{\varepsilon}_{0,1} \\ \vdots \\ \hat{\varepsilon}_{0,n} \end{pmatrix} = \begin{pmatrix} \upsilon_1 \\ \vdots \\ \upsilon_n \end{pmatrix} - \begin{pmatrix} x_{0,1} \\ \vdots \\ x_{0,n} \end{pmatrix} \hat{\beta}_0 = \begin{pmatrix} \upsilon_1 - x_{0,1}\hat{\beta}_0 \\ \vdots \\ \upsilon_n - x_{0,n}\hat{\beta}_0 \end{pmatrix} \tag{10}$$

und

$$\hat{\varepsilon} := \upsilon - X\hat{\beta} \Leftrightarrow \begin{pmatrix} \hat{\varepsilon}_1 \\ \vdots \\ \hat{\varepsilon}_n \end{pmatrix} = \begin{pmatrix} \upsilon_1 \\ \vdots \\ \upsilon_n \end{pmatrix} - \begin{pmatrix} x_{0,1} & x_{1,1} \\ \vdots & \vdots \\ x_{0,n} & x_{1,n} \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \upsilon_1 - x_{0,1}\hat{\beta}_0 - x_{1,1}\hat{\beta}_1 \\ \vdots \\ \upsilon_n - x_{0,n}\hat{\beta}_0 - x_{1,n}\hat{\beta}_1 \end{pmatrix} \tag{11}$$

5.Geben Sie das Theorem zum Likelihood-Quotienten von vollständigem und reduzierten ALM wieder.

Theorem (Likelihood-Quotient von vollständigem und reduziertem Modell)

Für $p = p_0 + p_1, p > 1$ sei eine Partitionierung eines vollständigen ALMs gegeben und es seien $\hat{\sigma}^2$ und $\hat{\sigma}_0^2$ die Maximum-Likelihood-Schätzer des Varianzparameters unter vollständigem und reduziertem Modell, respektive. Weiterhin seien die zwei parametrischen statistischen Modelle $\mathcal{M}_0$ und $\mathcal{M}_1$ in der Definition der Likelihood-Quotienten-Statistik durch das reduzierte Modell und das vollständige Modell gegeben. Dann gilt

$$\Lambda = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{\frac{n}{2}} \tag{12}$$

Anmerkung: ML-Schätzer, darum n und nicht $n - p$

6.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

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{13}$$

mit der Partitionierung

$$X = \begin{pmatrix} X_0 & X_1 \end{pmatrix}, X_0 \in \mathbb{R}^{n \times p0}, X_1 \in \mathbb{R}^{n \times p1}, \text{ und } \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \beta_0 \in \mathbb{R}^{p0}, \beta_1 \in \mathbb{R}^{p1}, \tag{14}$$

mit $p = p_0 + p_1$ gegeben. Weiterhin seien mit

$$\hat{\beta}_0 := \left(X_0^T X_0\right)^{-1} X_0^T \upsilon \text{ und } \hat{\beta} := \left(X^T X\right)^{-1} X^T \upsilon \tag{15}$$

die Residuenvektoren

$$\hat{\varepsilon}_0 := \upsilon - X_0 \hat{\beta}_0 \text{ und } \hat{\varepsilon} := \upsilon - X\hat{\beta} \tag{16}$$

definiert. Dann ist die F-Statistik definiert als

$$F := \frac{(\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon})/p_1}{\hat{\varepsilon}^T \hat{\varepsilon}/(n-p)} \tag{17}$$

7.Erläutern Sie den Zähler der F-Statistik.

Der Zähler der F-Statistik

$$\frac{\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon}}{p_1} \tag{18}$$

misst, inwieweit die p_1 Regressoren in X_1 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_1 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.

8.Erläutern Sie den Nenner der F-Statistik.

Für den Nenner der F-Statistik gilt

$$\frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n-p} = \hat{\sigma}^2, \tag{19}$$

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 $\hat{\beta}_1 \approx 0_{p1}$ 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 $\hat{\varepsilon}_0^T\hat{\varepsilon}_0/(n-p)$, da sich für diesen die Datenvariabilität, die nicht durch die p_0 Regressoren in X_0 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.

9.Erläutern Sie die F-Statistik.

Die F-Statistik misst die Residualquadratsummenreduktion durch die p_1-Regressoren in X_1 gegenüber den p_0 Regressoren in X_0 pro Datenvariabilitäts(σ^2)- und Regressor(p_1)-Einheit.

10. Geben Sie das Theorem F-Statistik und Likelihood-Quotienten-Statistik wieder.

Theorem (F-Statistik und Likelihood-Quotienten-Statistik)

Es sei die Partitionierung eines ALMs in ein vollständiges und ein reduziertes Modell gegeben und F und Λ seien die entsprechenden F- und Likelihood-Quotienten-Statistiken. Dann gilt

$$F = \frac{n-p}{p_1}\left(\Lambda^{-\frac{2}{n}} - 1\right). \tag{20}$$

Anmerkung: Ich kann die F-Statistik also auch als Funktion der Likelihood-Quotienten-Statistik schreiben oder umgekehrt.