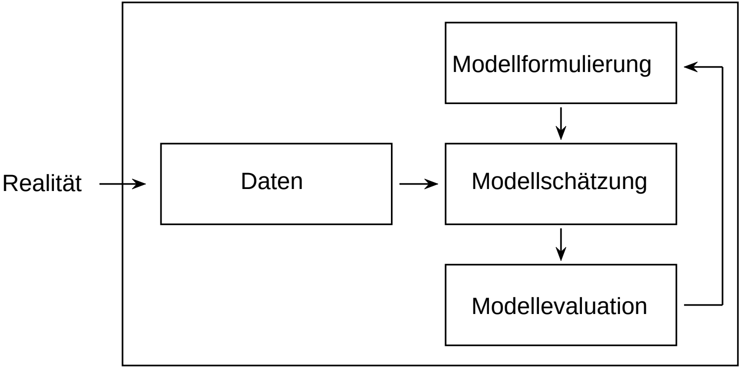
Allgemeines Lineares Modell

BSc Psychologie SoSe 2023

Prof. Dr. Dirk Ostwald

(8) F-Statistiken

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.

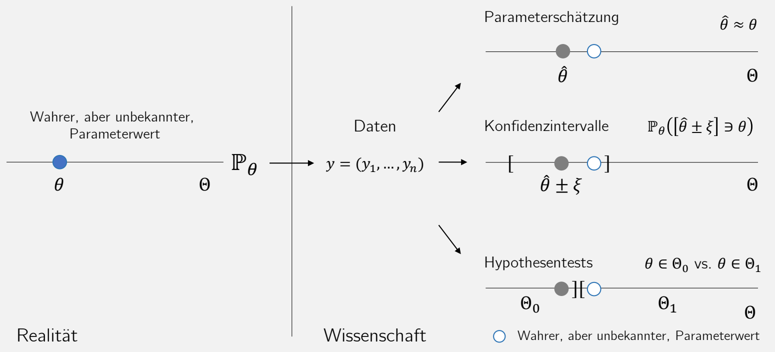
(2) Konfidenzintervalle

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

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.

Modell und Standardprobleme Frequentistischer Inferenz



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

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

$$
\begin{aligned}
&\text{Datensatz (1)} : y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, ..., y_n^{(1)}\right)^T \text{ mit } \hat{\beta}^{(1)} = (X^TX)^{-1}X^Ty^{(1)} \\
&\text{Datensatz (2)} : y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, ..., y_n^{(2)}\right)^T \text{ mit } \hat{\beta}^{(2)} = (X^TX)^{-1}X^Ty^{(2)} \\
&\text{Datensatz (3)} : y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, ..., y_n^{(3)}\right)^T \text{ mit } \hat{\beta}^{(3)} = (X^TX)^{-1}X^Ty^{(3)} \\
&\text{Datensatz (4)} : y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, ..., y_n^{(4)}\right)^T \text{ mit } \hat{\beta}^{(4)} = (X^TX)^{-1}X^Ty^{(4)} \\
&\text{Datensatz (5)} : y^{(5)} = ...
\end{aligned}
$$

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 $\hat{\beta} := (X^TX)^{-1}X^T\upsilon$? 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.

Überblick

- Wir führen F-Statistiken hier vor dem Hintergrund Likelihood-Quotienten-basierter Modellvergleiche ein. Die (maximierte oder marginale) Likelihood eines Datensatzes unter einem gegebenen probabilistischen Modell als Modellvergleichskriterium heranzuziehen ist ein weit verbreitetes Verfahren in der probabilistischen Datenanalyse.
- Im Gegensatz zu T-Statistiken kann das Ziel der Berechnung von F-Statistiken damit insbesondere sein, nicht nur Linearkombinationen von Betaparameterschätzwerten probabilistisch zu evaluieren, sondern die Modellanpassung an einen Datensatz insgesamt zu evaluieren.
- Die Modellvergleichskapazität von F-Statistiken ist allerdings etwas beschränkt, da sich die F-Statistik nur auf ALMs und insbesondere geschachtelte ALMs bezieht, in denen ein Modell Bestandteil eines anderen Modells ist.
- F-Statistiken bilden üblicherweise die Grundlage für Hypothesentests im Rahmen varianzanalytischer Verfahren (vgl. (10) Einfaktorielle Varianzanalyse, (11) Zweifaktorielle Varianzanalyse und (13) Kovarianzanalyse). Der Einsatz von F-Statistiken ist aber *perse* nicht auf Varianzanalysen beschränkt, sondern kann auch bei parametrischen ALM Designs angebracht sein.

F-Zufallsvariablen

Likelihood-Quotienten

Definition und Verteilung

Selbstkontrollfragen

F-Zufallsvariablen

Likelihood-Quotienten

Definition und Verteilung

Selbstkontrollfragen

Definition (f-Zufallsvariable)

ξ sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und Wahrscheinlichkeitsdichtefunktion

$$p_\xi : \mathbb{R} \to \mathbb{R}_{>0}, x \mapsto p_\xi(x) := \nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}} \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \frac{x^{\frac{\nu_1}{2}-1}}{(\nu_1 x + \nu_2)^{\frac{\nu_1+\nu_2}{2}}}, \tag{4}$$

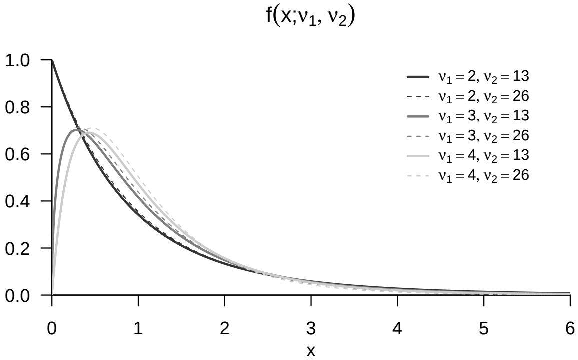
wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass ξ einer f-Verteilung mit Freiheitsgradparametern ν_1 und ν_2 unterliegt und nennen ξ eine f-Zufallsvariable mit Freiheitsgradparametern ν_1 und ν_2. Wir kürzen dies mit $\xi \sim f(\nu_1, \nu_2)$ ab. Die Wahrscheinlichkeitsdichtefunktion (WDF) einer f-Zufallsvariable bezeichnen wir mit $f(x; \nu_1, \nu_2)$, die kumulative Verteilungsfunktion (KVF) einer f-Zufallsvariable bezeichnen wir mit $\varphi(x; \nu_1, \nu_2)$, und die inverse kumulative Verteilungsfunktion einer f-Zufallsvariable bezeichnen wir mit $\varphi^{-1}(x; \nu_1, \nu_2)$.

Bemerkungen

- f-Zufallsvariablen sind nach Ronald A. Fisher benannt.
- George W. Snedecor hat die KVF der f-Verteilung wohl 1934 basierend auf Arbeiten von Fisher tabuliert.

F-Zufallsvariablen

Wahrscheinlichkeitsdichtefunktionen von f-Verteilungen



F-Zufallsvariablen

Theorem (F-Transformation)

$\zeta_1 \sim \chi^2(\nu_1)$ und $\zeta_2 \sim \chi^2(\nu_2)$ seien zwei unabhängige χ^2-Zufallfsvariablen mit ν_1 und ν_2 Freiheitsgraden, respektive. Dann ist die Zufallsvariable

$$\xi := \frac{\zeta_1/\nu_1}{\zeta_2/\nu_2} \tag{5}$$

eine f-verteilte Zufallsvariable mit ν_1, ν_2 Freiheitsgraden, es gilt also $\xi \sim f(\nu_1, \nu_2)$.

Bemerkungen

- Wir verzichten auf einen Beweis.
- Das Theorem kann bewiesen werden, in dem man zunächst ein Transformationstheorem für Quotienten von Zufallsvariablen mithilfe des multivariaten Transformationstheorems und Marginalisierung herleitet und dieses Theorem dann auf die WDF von χ^2-verteilten ZVen anwendet. Dabei ist die Regel zur Integration durch Substitution von zentraler Bedeutung.

Definition (Nichtzentrale f-Zufallsvariable)

ξ sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und Wahrscheinlichkeitsdichtefunktion

$$p_\xi : \mathbb{R} \to \mathbb{R}_{>0}, x \mapsto$$
$$p_\xi(x) := \sum_{k=0}^{\infty} \frac{e^{-\delta/2}(\delta/2)^k}{\frac{\Gamma(\nu_2/2)\Gamma(\nu_1/2+k)}{\Gamma(\nu_2/2+\nu_1/2+k)} k!} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2+k} \left(\frac{\nu_2}{\nu_2+\nu_1 x}\right)^{(\nu_1+\nu_2)/2+k} x^{\nu_1/2-1+k} \tag{6}$$

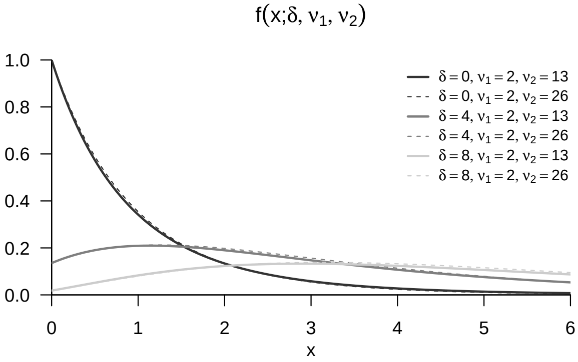
wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass ξ einer nichtzentralen f-Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparametern ν_1 und ν_2 unterliegt und nennen ξ eine nichtzentrale f-Zufallsvariable mit Nichtzentralitätsparameter δ und Freiheitsgradparametern ν_1 und ν_2. Wir kürzen dies mit $\xi \sim f(\delta, \nu_1, \nu_2)$ ab. Die Wahrscheinlichkeitsdichtefunktion (WDF) einer f-Zufallsvariable bezeichnen wir mit $f(x; \delta, \nu_1, \nu_2)$, die kumulative Verteilungsfunktion (KVF) einer nichtzentralen f-Zufallsvariable bezeichnen wir mit $\varphi(x; \delta, \nu_1, \nu_2)$, und die inverse kumulative Verteilungsfunktion einer nichtzentralen f-Zufallsvariable bezeichnen wir mit $\varphi^{-1}(x; \delta, \nu_1, \nu_2)$.

Bemerkungen

- Es gilt $f(0, \nu_1, \nu_2) = f(\nu_1, \nu_2)$.

F-Zufallsvariablen

WDFen von nichtzentralen f-Verteilungen



F-Zufallsvariablen

Theorem (Nichtzentrale F-Transformation)

$\zeta_1 \sim \chi^2(\delta, \nu_1)$ und $\zeta_2 \sim \chi^2(\nu_2)$ seien eine nichtzentrale χ^2- Zufallsvariable und eine χ^2-Zufallsvariable mit ν_1 und ν_2 Freiheitsgraden, respektive und ζ_1 und ζ_2 seien unabhängig. Dann ist die Zufallsvariable

$$\xi := \frac{\zeta_1/\nu_1}{\zeta_2/\nu_2} \tag{7}$$

eine nichtzentral f-verteilte Zufallsvariable mit Nichtzentralitätsparameter δ und ν_1, ν_2 Freiheitsgraden, es gilt also $\xi \sim f(\delta, \nu_1, \nu_2)$.

Bemerkungen

- Wir verzichten auf einen Beweis.

F-Zufallsvariablen

Likelihood-Quotienten-Statistiken

Definition und Verteilung

Selbstkontrollfragen

Definition (Likelihood-Quotienten-Statistik)

Gegeben seien zwei parametrische statistische Modelle

$$\mathcal{M}_0 := \left(\mathcal{Y}, \mathcal{A}, \left\{\mathbb{P}^0_{\theta_0} | \theta_0 \in \Theta_0\right\}\right) \text{ und } \mathcal{M}_1 := \left(\mathcal{Y}, \mathcal{A}, \left\{\mathbb{P}^1_{\theta_1} | \theta_1 \in \Theta_1\right\}\right) \tag{8}$$

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

$$\Lambda := \frac{\max_{\theta_0 \in \Theta_0} L_0^{\upsilon}(\theta_0)}{\max_{\theta_1 \in \Theta_1} L_1^{\upsilon}(\theta_1)}, \tag{9}$$

Likelihood-Quotienten-Statistik genannt.

Likelihood-Quotienten-Statistiken

Bemerkungen

- 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.
- Die Wahrscheinlichkeitsdichten/massen beobachteter Daten nach Modellschätzung unter verschiedenen Modellen zu betrachten ist ein allgemeines Vorgehen zum Vergleich von Modellen. Letztlich erlaubt dieses Vorgehen, verschiedene wissenschaftliche Theorien über die Genese beobachtbarer Daten quantitativ zu vergleichen und die damit verbundene Unsicherheit zu quantifizieren.
- Modellvergleiche sind ein zentrales Thema in der Bayesianischen Inferenz die die Logik von Likelihood-Quotienten-Statistiken zum Beispiel unter den Begriffen der Bayes Factors oder der des Bayesian Information Criterions auf allgemeine probabilistische Modelle generalisiert. Allerdings sind, wie hier gesehen, Modellvergleiche auch im Rahmen der Frequentistischen Inferenz möglich und sinnvoll, Modellvergleiche sind also kein Alleinstellungsmerkmal der Bayesianischen gegenüber der Frequentistischen Inferenz.
- Mit dem *reduzierten Modell* und dem *vollständigen Modell* betrachten wir im Folgenden zwei spezielle Formen von $\mathcal{M}_0$ und $\mathcal{M}_1$, respektive, im Kontext des ALMs.

Definition (Vollständiges und reduziertes Modell)

Für $p > 1$ mit $p = p_0 + p_1$ seien

$$X := \begin{pmatrix} X_0 & X_1 \end{pmatrix} \in \mathbb{R}^{n \times p} \text{ mit } X_0 \in \mathbb{R}^{n \times p_0} \text{ und } X_1 \in \mathbb{R}^{n \times p_1}, \quad (10)$$

sowie

$$\beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \in \mathbb{R}^{p} \text{ mit } \beta_0 \in \mathbb{R}^{p_0} \text{ und } \beta_1 \in \mathbb{R}^{p_1} \quad (11)$$

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

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (12)$$

das *vollständige Modell* und

$$\upsilon = X_0\beta_0 + \varepsilon_0 \text{ mit } \varepsilon_0 \sim N(0_n, \sigma^2 I_n) \quad (13)$$

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

Bemerkungen

- Man sagt auch, dass das reduzierte Modell im vollständigen Modell *geschachtelt (nested)* ist.

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

$$\Lambda = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{\frac{n}{2}} \tag{14}$$

Bemerkungen

- Informell gilt hier

$$\mathcal{M}_0 : \upsilon = X_0\beta_0 + \varepsilon_0 \text{ und } \mathcal{M}_1 : \upsilon = X\beta + \varepsilon \tag{15}$$

Likelihood-Quotienten-Statistiken

Beweis

Wir erinnern zunächst daran, dass die Maximum-Likelihood Schätzer des Varianzparameters durch

$$\hat{\sigma}^2 = \frac{1}{n}\left(\upsilon - X\hat{\beta}\right)^T\left(\upsilon - X\hat{\beta}\right) \text{ und } \hat{\sigma}_0^2 = \frac{1}{n}\left(\upsilon - X_0\hat{\beta}_0\right)^T\left(\upsilon - X_0\hat{\beta}_0\right) \tag{16}$$

respektive, gegeben sind, wobei $\hat{\beta}$ und $\hat{\beta}_0$ die Maximum-Likelihood Schätzer der Betaparameter unter vollständigem und reduziertem Modell, respektive, bezeichnen. Weiterhin halten wir fest, dass für die Likelihood-Funktion des vollständigem Modells an der Stelle der Maximum-Likelihood Schätzer gilt, dass

$$\begin{aligned}
L_1^y(\hat{\beta}, \hat{\sigma}^2) &= (2\pi)^{-\frac{n}{2}}(\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\hat{\sigma}^2}(y - X\hat{\beta})^T(y - X\hat{\beta})\right) \\
&= (2\pi)^{-\frac{n}{2}}(\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left(-\frac{n}{2}\frac{(y - X\hat{\beta})^T(y - X\hat{\beta})}{(y - X\hat{\beta})^T(y - X\hat{\beta})}\right) \\
&= (2\pi)^{-\frac{n}{2}}(\hat{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}
\end{aligned} \tag{17}$$

und analog, dass für die Likelihood-Funktion des reduzierten Modells an der Stelle der Maximum-Likelihood Schätzer gilt, dass

$$L_0^y(\hat{\beta}_0, \hat{\sigma}_0^2) = (2\pi)^{-\frac{n}{2}}(\hat{\sigma}_0^2)^{-\frac{n}{2}} e^{-\frac{n}{2}} \tag{18}$$

Damit ergibt sich dann aber

$$\Lambda = \frac{\max_{\theta_0 \in \Theta_0} L_0^\upsilon(\theta_0)}{\max_{\theta_1 \in \Theta_1} L_1^\upsilon(\theta_1)} = \frac{L_0^\upsilon(\hat{\beta}_0, \hat{\sigma}_0^2)}{L_1^\upsilon(\hat{\beta}, \hat{\sigma}^2)} = \frac{(2\pi)^{-\frac{n}{2}}(\hat{\sigma}_0^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}}{(2\pi)^{-\frac{n}{2}}(\hat{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}} = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\right)^{-\frac{n}{2}} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{\frac{n}{2}}$$

F-Zufallsvariablen

Likelihood-Quotienten

Definition und Verteilung

Selbstkontrollfragen

Definition und Verteilung

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

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{19}$$

mit der Partitionierung

$$X = \begin{pmatrix} X_0 & X_1 \end{pmatrix}, X_0 \in \mathbb{R}^{n \times p_0}, X_1 \in \mathbb{R}^{n \times p_1}, \text{ und } \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \beta_0 \in \mathbb{R}^{p_0}, \beta_1 \in \mathbb{R}^{p_1}, \tag{20}$$

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

$$\hat{\beta}_0 := (X_0^T X_0)^{-1} X_0^T \upsilon \text{ und } \hat{\beta} := (X^T X)^{-1} X^T \upsilon \tag{21}$$

die Residuenvektoren

$$\hat{\varepsilon}_0 := \upsilon - X_0 \hat{\beta}_0 \text{ und } \hat{\varepsilon} := \upsilon - X\hat{\beta} \tag{22}$$

definiert. Dann ist die F-Statistik definiert als

$$F := \frac{(\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon})/p_1}{\hat{\varepsilon}^T \hat{\varepsilon}/(n-p)} \tag{23}$$

Definition und Verteilung

Bemerkungen

- Der Zähler der F-Statistik
$$\frac{\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon}}{p_1} \tag{24}$$
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.
- Für den Nenner der F-Statistik gilt
$$\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-p} = \hat{\sigma}^2, \tag{25}$$
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 $\hat{\beta}_1 \approx 0_{p_1}$ 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 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.
- Zusammengenommen misst die F-Statistik also 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.

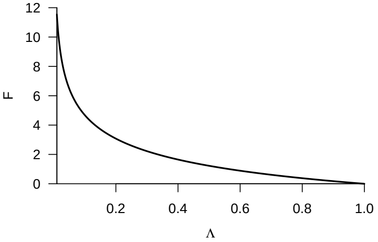
Theorem (F-Statistik und Likelihood-Quotienten-Statistik)

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

$$F = \frac{n-p}{p_1}\left(\Lambda^{-\frac{2}{n}} - 1\right). \tag{26}$$

Bemerkungen

- Zwischen der F- und der Likelihood-Quotienten-Statistik besteht ein funktionaler reziproker Zusammenhang.
- Für $\Lambda = 1$ gilt $F = 0$.
- Wir visualisieren den funktionalen Zusammenhang für $n = 12, p = 2, p_1 = 1$ untenstehend.



Definition und Verteilung

Beweis

Wir erinnern zunächst daran, dass die Maximum-Likelihood Schätzer des Varianzparameters durch

$$\hat{\sigma}^2 = \frac{1}{n}\left(\upsilon - X\hat{\beta}\right)^T\left(\upsilon - X\hat{\beta}\right) = \frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n} \text{ und } \hat{\sigma}_0^2 = \frac{1}{n}\left(\upsilon - X_0\hat{\beta}_0\right)^T\left(\upsilon - X_0\hat{\beta}_0\right) = \frac{\hat{\varepsilon}_0^T\hat{\varepsilon}_0}{n} \quad (27)$$

gegeben sind. Mit der Definition der F-Statistik und der Form der Likelihood-Quotienten-Statistik für den Vergleich von reduziertem und vollständigem Modell ergibt sich dann

$$\begin{aligned} F &= \frac{(\hat{\varepsilon}_0^T\hat{\varepsilon}_0 - \hat{\varepsilon}^T\hat{\varepsilon})/p_1}{\hat{\varepsilon}^T\hat{\varepsilon}/(n-p)} \\ &= \frac{n(\hat{\sigma}_0^2 - \hat{\sigma}^2)/p_1}{n\hat{\sigma}^2/(n-p)} \\ &= \frac{n-p}{p_1}\frac{\hat{\sigma}_0^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \\ &= \frac{n-p}{p_1}\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} - \frac{\hat{\sigma}^2}{\hat{\sigma}^2}\right) \\ &= \frac{n-p}{p_1}\left(\Lambda^{-\frac{2}{n}} - 1\right) \end{aligned} \quad (28)$$

Definition und Verteilung

Beispiel (1) Einfache lineare Regression

$$X = (X_0 \quad X_1)\,, X_0 := 1_n\,, X_1 := (x_1, ..., x_n)^T, \beta := \begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}, \beta_0 = \text{Interzept}, \beta_1 = \text{Steigung} \tag{29}$$

```
# Modellformulierung
library(MASS)                                               # Multivariate Normalverteilung
nmod   = 2                                                  # Anzahl Modelle
n      = 10                                                 # Anzahl Datenpunkte
p      = 2                                                  # Anzahl Betaparameter
p_0    = 1                                                  # Anzahl Betaparameter reduziertes Modell
p_1    = 1                                                  # Anzahl zusätzlicher Betaparameter vollständiges Modell
p      = p_0 + p_1                                          # Anzahl Betaparameter im vollständigem Modell
x      = 1:n                                                # Prädiktorwerte
X      = matrix(c(rep(1,n),x), nrow = n)                    # Designmatrix des vollständigen Modells
X_0    = X[,1]                                              # Designmatrix des reduzierten Modells
I_n    = diag(n)                                            # n x n Einheitsmatrix
beta   = matrix(c(1,0,1,.5), nrow = 2)                      # wahre , aber unbekannte , Betaparameter
nscn   = ncol(beta)                                         # Anzahl wahrer, aber unbekannter, Hypothesenszenarien
sigsqr = 1                                                  # wahrer, aber unbekannter, Varianzparameter

# Modellsimulation und Evaluierung
Eff    = matrix(rep(NaN, nscn), nrow = nscn)                # F-Statistiik Realisierungsarray
for(s in 1:nscn){                                           # Szenarieniterationen
  y               = mvrnorm(1, X %*%beta[,s], sigsqr*I_n)   # Datenrealisierung
  beta_hat_0      = solve(t(X_0)%*%X_0)%*%t(X_0)%*%y        # Betaparameterschätzer reduziertes Modell
  beta_hat        = solve(t(X)  %*%X  )%*%t(X)  %*%y        # Betaparameterschätzer vollständiges Modell
  eps_0_hat       = y-X_0%*%beta_hat_0                      # Residuenvektor reduziertes Modell
  eps_hat         = y-X%*%beta_hat                          # Residuenvektor vollständiges Modell
  eps_0_eps_0_hat = t(eps_0_hat) %*% eps_0_hat              # RQS reduziertes Modell
  eps_eps_hat     = t(eps_hat)   %*% eps_hat                # RQS vollständiges Modell
  Eff[s]          = (((eps_0_eps_0_hat-eps_eps_hat)/p_1)/   # F-Statistik
                       (eps_eps_hat/(n-p)))}
```

```
> F-Statistik für beta_1  = 0_{p_1}: 0.357
> F-Statistik für beta_1 != 0_{p_1}: 18.8
```

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

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

mit der Partitionierung

$$X = \begin{pmatrix} X_0 & X_1 \end{pmatrix}, X_0 \in \mathbb{R}^{n \times p_0}, X_1 \in \mathbb{R}^{n \times p_1}, \text{ und } \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \beta_0 \in \mathbb{R}^{p_0}, \beta_1 \in \mathbb{R}^{p_1}, \tag{31}$$

mit $p = p_0 + p_1$ gegeben. Schließlich sei

$$c := \begin{pmatrix} 0_{p_0} \\ 1_{p_1} \end{pmatrix} \in \mathbb{R}^p \tag{32}$$

ein Vektor. Dann gilt

$$F \sim f(\delta, p_1, n-p) \text{ mit } \delta := \frac{c^T\beta \left(c^T(X^TX)^{-1}c\right)^{-1} c^T\beta}{\sigma^2} \tag{33}$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- F ist eine Funktion der Parameterschätzer, δ ist eine Funktion der wahren, aber unbekannten, Parameter.
- Diese Verteilung von F kann zum Nullhypothesentesten und zur Powerfunktionsevaluation genutzt werden.

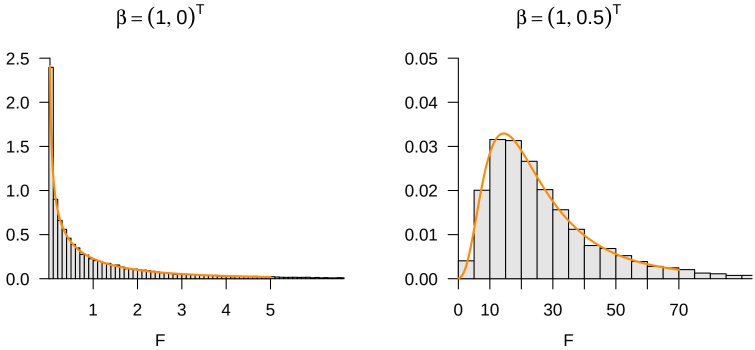
Beispiel (1) Einfache lineare Regression

$$X = \begin{pmatrix} X_0 & X_1 \end{pmatrix}, X_0 := 1_n, X_1 := (x_1, \ldots, x_n)^T, \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \beta_0 = \text{Interzept}, \beta_1 = \text{Steigung} \quad (34)$$

```
# Modellformulierung
library(MASS)                                                      # Multivariate Normalverteilung
nmod   = 2                                                         # Anzahl Modelle
n      = 10                                                        # Anzahl Datenpunkte
p_0    = 1                                                         # Anzahl Betaparameter im reduzierten Modell
p_1    = 1                                                         # Anzahl additiver Betaparameter im vollständigen Modell
p      = p_0 + p_1                                                 # Anzahl Betaparameter im vollständigem Modell
x      = 1:n                                                       # Prädiktorwerte
X      = matrix(c(rep(1,n),x), nrow = n)                           # Designmatrix des vollständigen Modells
X_0    = X[,1]                                                     # Designmatrix des reduzierten Modells
I_n    = diag(n)                                                   # n x n Einheitsmatrix
beta   = matrix(c(1,0,1,.5), nrow = 2)                             # wahre , aber unbekannte , Betaparameter
nscn   = ncol(beta)                                                # Anzahl wahrer, aber unbekannter, Hypothesenszenarien
sigsqr = 1                                                         # wahrer, aber unbekannter, Varianzparameter
c      = matrix((c(0,1)), nrow = 2)                                # Vektor

# Frequentistische Simulation
nsim   = 1e4                                                       # Anzahl Realisierungen des n-dimensionalen ZVs
delta  = rep(NaN,nscn)                                             # Nichtzentralitätsparameterarray
Eff    = matrix(rep(NaN, nscn*nsim), nrow = nscn)                  # F-Statistiik Realisierungsarray
for(s in 1:nscn){                                                  # Szenarieniterationen
  delta[s] = (t(t(c)%*%beta[,s])%*%                                # Nichtzentralitätsparameter
              solve(t(c)%*%solve(t(X)%*%X)%*%c) %*%
              (t(c)%*%beta[,s])/sigsqr)
  for(i in 1:nsim){                                                # Simulationsiterationen
    y               = mvrnorm(1, X %*%beta[,s], sigsqr*I_n)        # Datenrealisierung
    beta_hat_0      = solve(t(X_0)%*%X_0)%*%t(X_0)%*%y             # Betaparameterschätzer reduziertes Modell
    beta_hat        = solve(t(X)  %*%X  )%*%t(X)  %*%y             # Betaparameterschätzer vollständiges Modell
    eps_0_hat       = y-X_0%*%beta_hat_0                           # Residuenvektor reduziertes Modell
    eps_hat         = y-X%*%beta_hat                               # Residuenvektor vollständiges Modell
    eps_0_eps_0_hat = t(eps_0_hat) %*% eps_0_hat                   # RQS reduziertes Modell
    eps_eps_hat     = t(eps_hat)   %*% eps_hat                     # RQS vollständiges Modell
    Eff[s,i]        = (((eps_0_eps_0_hat-eps_eps_hat)/p_1)/        # F-Statistik
                        (eps_eps_hat/(n-p)))}}
```

Beispiel (1) Einfache lineare Regression



Definition und Verteilung

Ausblick

- Die Theorie von T- und F-Statistiken wird unter dem Begriff der *Allgemeinen Linearen Hypothese* verallgemeinert und integriert. Dabei betrachtet allgemeine lineare Funktionen der Betaparameter der Form $C^T\beta$, wobei $C \in \mathbb{R}^{p\times k}$ eine beliebige Matrix ist als Grundlage von Hypothesen. Zum Beispiel ergibt sich für $C \in \mathbb{R}^{p\times 1}$ die hier und in Einheit (7) T-Statistiken betrachten Kontrastgewichtsvektoren. Im Kontext der *Allgemeinen Linearen Hypothese* kann man weiterhin zeigen, dass $F = T^2$, dass also auch das Quadrat einer T-Statistik f-verteilt ist und T-Statistiken damit (nur) spezielle F-Statistiken sind.

- Dennoch wird in der Anwendung sehr stark zwischen T- und F-Statistiken unterschieden und es ist sinnvoll, sich der unterschiedlichen Anwendungsfälle von T- und F-Statistiken bewusst zu sein. Gute Einführungen in die Theorie der Allgemeinen Linearen Hypothese bieten z.B. Searle (1971), Kapitel 3, Rencher and Schaalje (2008), Kapitel 8 oder Christensen (2011), Kapitel 3.

F-Zufallsvariablen

Likelihood-Quotienten-Statistiken

Definition und Verteilung

Selbstkontrollfragen

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 zum Zusammenhang von F-Statistik und Likelihood-Quotienten-Statistik wieder.

Referenzen

Christensen, Ronald. 2011. *Plane Answers to Complex Questions*. Springer Texts in Statistics. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-9816-3.

Rencher, Alvin C., and G. Bruce Schaalje. 2008. *Linear Models in Statistics*. 2nd ed. Hoboken, N.J: Wiley-Interscience.

Searle, S. R. 1971. *Linear Models*. A Wiley Publication in Mathematical Statistics. New York: Wiley.