



# Allgemeines Lineares Modell

BSc Psychologie, SoSe 2025

Joram Soch

## (8) F-Statistiken

## Modellformulierung

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

## Modellschätzung

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

## Modellevaluation

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

---

## 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.

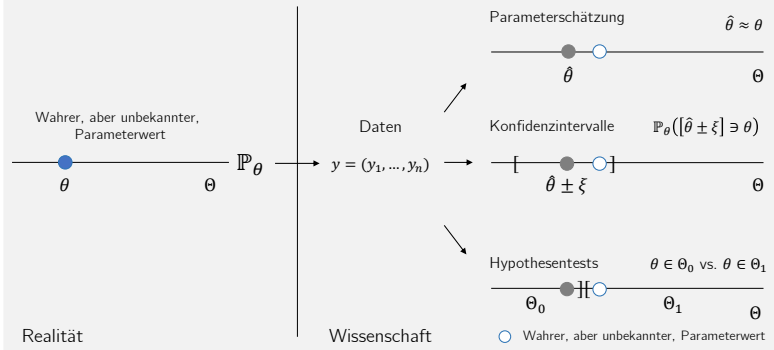
### (2) Konfidenzintervalle

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



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

## 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 Betaparameterkomponenten probabilistisch zu evaluieren, sondern die Modellanpassung an einen Datensatz insgesamt zu evaluieren.
- Die Modellvergleichskapazität von F-Statistiken ist allerdings etwas begrenzt, da sich die F-Statistik nur auf ALMs und insbesondere geschachtelte ALMs bezieht, d.h. auf ALMs, in denen ein Modell Bestandteil eines anderen Modells ist.
- F-Statistiken bilden üblicherweise die Grundlage für Hypothesentests im Rahmen varianzanalytischer Verfahren (vgl. Einheiten (10) Einfaktorielle Varianzanalyse, (11) Zweifaktorielle Varianzanalyse und (14) Kovarianzanalyse in *Allgemeines Lineares Modell*). Der Einsatz von F-Statistiken ist aber *per se* nicht auf Varianzanalysen beschränkt, sondern kann auch bei parametrischen ALM-Designs angebracht sein.

---

F-Zufallsvariablen

Likelihood-Quotienten

F-Statistiken

Selbstkontrollfragen

---

## **F-Zufallsvariablen**

Likelihood-Quotienten

F-Statistiken

Selbstkontrollfragen

## Definition ( $f$ -Zufallsvariable)

$U_1 \sim \chi^2(n_1)$  und  $U_2 \sim \chi^2(n_2)$  seien zwei unabhängige  $\chi^2$ -Zufallsvariablen mit Freiheitsgradparametern  $n_1$  und  $n_2$ . Dann nennen wir die Zufallsvariable

$$F := \frac{U_1/n_1}{U_2/n_2} \quad (4)$$

eine  $f$ -verteilte Zufallsvariable mit Freiheitsgradparametern  $n_1$  und  $n_2$ . Wir schreiben  $F \sim f(n_1, n_2)$  ab. Die Wahrscheinlichkeitsdichtefunktion (WDF) einer  $f$ -Zufallsvariable bezeichnen wir mit  $f(x; n_1, n_2)$ , die kumulative Verteilungsfunktion (KVF) und inverse KVF einer  $f$ -Zufallsvariable bezeichnen wir mit  $\varphi(x; n_1, n_2)$  bzw.  $\varphi^{-1}(x; n_1, n_2)$ .

### 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.

## Theorem (WDF einer $f$ -Zufallsvariable)

$F$  sei eine  $f$ -Zufallsvariable mit Ergebnisraum  $\mathbb{R}_{>0}$  und Freiheitsgradparametern  $n_1$  und  $n_2$ . Dann ist die Wahrscheinlichkeitsdichtefunktion von  $F$  gegeben durch

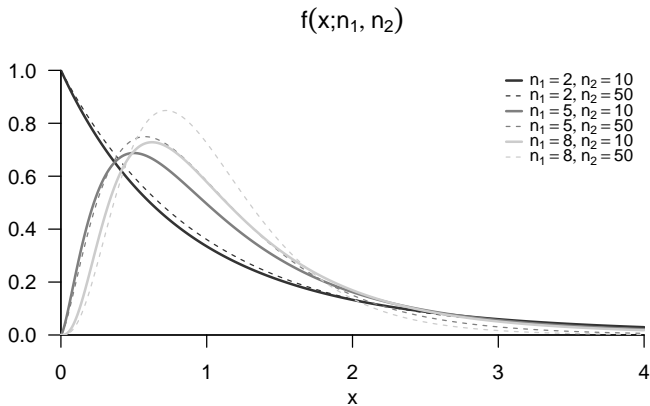
$$f(\cdot; n_1, n_2) : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, x \mapsto f(x; n_1, n_2) := n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \frac{x^{\frac{n_1}{2}-1}}{(n_1x + n_2)^{\frac{n_1+n_2}{2}}}, \quad (5)$$

wobei  $\Gamma$  die Gammafunktion bezeichne.

### Bemerkungen

- Wir verzichten auf einen Beweis, der [sich hier findet](#) (englisch).
- 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  $\chi^2$ -verteilten Zufallsvariablen anwendet. Dabei ist die Regel zur Integration durch Substitution von zentraler Bedeutung.
- Die  $f$ -Verteilung ist nicht um 0 symmetrisch, da sie auf die positiven reellen Zahlen beschränkt ist. Steigendes  $n_2$  bei gleichbleibendem  $n_1$  verschiebt die Wahrscheinlichkeitsmasse in den höheren positiven Bereich.

## Wahrscheinlichkeitsdichtefunktionen von $f$ -Verteilungen



## Definition (Nichtzentrale $f$ -Zufallsvariable)

$U_1 \sim \chi^2(\delta, n_1)$  und  $U_2 \sim \chi^2(n_2)$  seien eine nichtzentrale  $\chi^2$ -Zufallsvariable und eine  $\chi^2$ -Zufallsvariable mit Freiheitsgradparametern  $n_1$  und  $n_2$ . Weiterhin seien  $U_1$  und  $U_2$  unabhängige Zufallsvariablen. Dann nennen wir die Zufallsvariable

$$F := \frac{U_1/n_1}{U_2/n_2} \quad (6)$$

eine nichtzentral  $f$ -verteilte Zufallsvariable mit Nichtzentralitätsparameter  $\delta$  und Freiheitsgradparametern  $n_1$  und  $n_2$ . Wir schreiben  $F \sim f(\delta, n_1, n_2)$ . Die WDF einer nichtzentralen  $f$ -Zufallsvariable bezeichnen wir mit  $f(x; \delta, n_1, n_2)$ , die KVF und inverse KVF einer nichtzentralen  $f$ -Zufallsvariable bezeichnen wir mit  $\varphi(x; \delta, n_1, n_2)$  bzw.  $\varphi^{-1}(x; \delta, n_1, n_2)$ .

### Bemerkungen

- Eine nichtzentrale  $f$ -Zufallsvariable mit  $\delta = 0$  ist eine  $f$ -Zufallsvariable. Es gilt also  $f(x; 0, n_1, n_2) = f(x; n_1, n_2)$ . Dies ist der Fall, weil eine nichtzentrale  $\chi^2$ -Zufallsvariable mit  $\delta = 0$  eine  $\chi^2$ -Zufallsvariable ist, also  $\chi^2(x; 0, n) = \chi^2(x; n)$  gilt.

## Theorem (WDF einer nichtzentralen $f$ -Zufallsvariable)

$F$  sei eine nichtzentrale  $f$ -Zufallsvariable mit Ergebnisraum  $\mathbb{R}_{>0}$ , Nichtzentralitätsparameter  $\delta$  und Freiheitsgradparametern  $n_1$  und  $n_2$ . Dann ist die Wahrscheinlichkeitsdichtefunktion von  $F$  gegeben durch

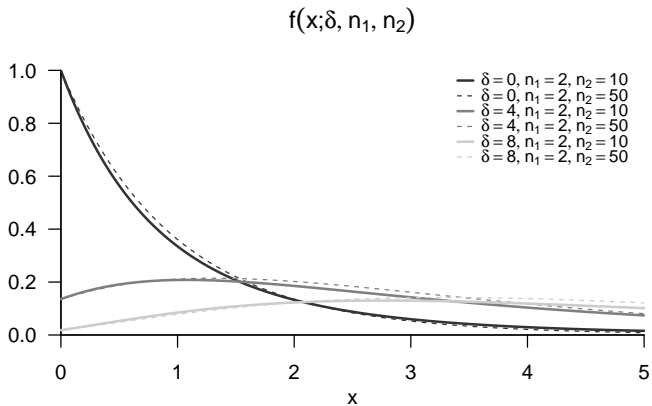
$$f(\cdot; \delta, n_1, n_2) : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, x \mapsto$$
$$f(x; \delta, n_1, n_2) := \sum_{k=0}^{\infty} \frac{e^{-\delta/2} (\delta/2)^k}{\frac{\Gamma(n_2/2)\Gamma(n_1/2+k)}{\Gamma(n_2/2+n_1/2+k)} k!} \left(\frac{n_1}{n_2}\right)^{n_1/2+k} \left(\frac{n_2}{n_2 + n_1 x}\right)^{(n_1+n_2)/2+k} x^{n_1/2-1+k} \quad (7)$$

wobei  $k!$  die Fakultät von  $k$  und  $\Gamma$  die Gammafunktion bezeichne.

### Bemerkungen

- Wir verzichten auf einen Beweis.

## Wahrscheinlichkeitsdichtefunktionen nichtzentraler $f$ -Verteilungen



---

F-Zufallsvariablen

**Likelihood-Quotienten**

F-Statistiken

Selbstkontrollfragen

## Definition (Likelihood-Quotienten-Statistik)

Gegeben seien zwei parametrische statistische Modelle

$$\mathcal{M}_0 := (Y, \mathcal{A}, \{\mathbb{P}_{\theta_0} | \theta_0 \in \Theta_0\}) \quad \text{und} \quad \mathcal{M}_1 := (Y, \mathcal{A}, \{\mathbb{P}_{\theta_1} | \theta_1 \in \Theta_1\}) \quad (8)$$

mit identischem Datenraum  $Y$ , identischer  $\sigma$ -Algebra  $\mathcal{A}$ , aber potentiell distinkten Wahrscheinlichkeitsmaßmengen  $\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1}$  und Parameterräumen  $\Theta_1, \Theta_2$ . Es seien weiterhin  $y$  ein Zufallsvektor mit Ergebnisraum  $Y$  und schließlich  $L_0(y, \theta_0)$  und  $L_1(y, \theta_1)$  die Likelihood-Funktionen von  $\mathcal{M}_0$  bzw.  $\mathcal{M}_1$ , wobei das Argument  $y$  jeweils die Datenabhängigkeit der Likelihood-Funktion anzeigt. Dann wird

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

*Likelihood-Quotienten-Statistik* genannt.

## Bemerkungen

- Eine Likelihood-Quotienten-Statistik setzt die Wahrscheinlichkeitsdichten (oder -massen) eines beobachteten Datensatzes  $y \in \mathcal{Y}$  unter zwei statistischen Modellen *nach Optimierung der jeweiligen Modellparameter* ins Verhältnis. Ein hoher Wert der Likelihood-Quotienten-Statistik entspricht einer höheren Wahrscheinlichkeitsdichte des beobachteten Datensatzes  $y$  unter  $\mathcal{M}_0$  als unter  $\mathcal{M}_1$  und umgekehrt.
- Die Wahrscheinlichkeitsdichten 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 Entstehung 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-Faktoren oder des Bayesianischen Informationskriteriums (BIC) auf allgemeine probabilistische Modelle generalisiert. Allerdings sind 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  $\mathcal{M}_0$  bzw.  $\mathcal{M}_1$  im Kontext des ALMs.

## Definition (Vollständiges und reduziertes Modell)

Für  $p > 1$  mit  $p = p_0 + p_1$  seien

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

sowie

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

Partitionierungen einer  $n \times p$  Designmatrix und eines  $p$ -dimensionalen Betaparametervektors. Dann sprechen wir von einer *Partitionierung eines Allgemeinen Linearen Modells* und nennen

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

das *vollständige Modell* und

$$y = X_0\beta_0 + \varepsilon_0 \quad \text{mit} \quad \varepsilon_0 \sim N(0_n, \sigma_0^2 I_n) \quad (13)$$

das *reduzierte Modell*.

### Bemerkungen

- Man sagt auch, dass das reduzierte Modell im vollständigen Modell *geschachtelt* ist.
- Im der englischsprachigen Literatur ist der Begriff *nested models* verbreitet.

## Theorem (Likelihood-Quotient von vollständigem & reduziertem Modell)

Für  $p = p_0 + p_1$  mit  $p > 1$  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 bzw. reduziertem Modell. 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

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

### Bemerkungen

- Informell gilt hier

$$\mathcal{M}_0 : y = X_0 \beta_0 + \varepsilon_0, \quad \varepsilon_0 \sim N(0_n, \sigma_0^2 I_n) \quad (15)$$

und

$$\mathcal{M}_1 : y = X \beta + \varepsilon, \quad \varepsilon \sim N(0_n, \sigma^2 I_n) . \quad (16)$$

# Likelihood-Quotienten

## Beweis

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

$$\hat{\sigma}^2 = \frac{1}{n} (y - X\hat{\beta})^T (y - X\hat{\beta}) \quad \text{und} \quad \hat{\sigma}_0^2 = \frac{1}{n} (y - X_0\hat{\beta}_0)^T (y - X_0\hat{\beta}_0) \quad (17)$$

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

$$\begin{aligned} L_1(y, \hat{\beta}, \hat{\sigma}^2) &= \frac{1}{\sqrt{(2\pi)^n (\hat{\sigma}^2)^n}} \cdot \exp\left(-\frac{1}{2\hat{\sigma}^2} (y - X\hat{\beta})^T (y - X\hat{\beta})\right) \\ &= \frac{1}{\sqrt{(2\pi)^n (\hat{\sigma}^2)^n}} \cdot \exp\left(-\frac{n}{2} \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{(y - X\hat{\beta})^T (y - X\hat{\beta})}\right) \\ &= \frac{1}{\sqrt{(2\pi)^n (\hat{\sigma}^2)^n}} \cdot e^{-\frac{n}{2}}, \end{aligned} \quad (18)$$

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

$$L_0(y, \hat{\beta}_0, \hat{\sigma}_0^2) = \frac{1}{\sqrt{(2\pi)^n (\hat{\sigma}_0^2)^n}} \cdot e^{-\frac{n}{2}}. \quad (19)$$

Damit ergibt sich dann aber

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

□

---

F-Zufallsvariablen

Likelihood-Quotienten

**F-Statistiken**

Selbstkontrollfragen

## Definition (F-Statistik für Partitionierung eines ALM)

Gegeben seien das ALM

$$y = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0_n, \sigma^2 I_n), \quad (20)$$

wobei  $X \in \mathbb{R}^{n \times p}$ , und mit  $p = p_0 + p_1$  eine Partitionierung

$$X = \begin{pmatrix} X_0 & X_1 \end{pmatrix}, \quad X_0 \in \mathbb{R}^{n \times p_0}, \quad X_1 \in \mathbb{R}^{n \times p_1} \quad \text{und} \\ \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \beta_0 \in \mathbb{R}^{p_0}, \quad \beta_1 \in \mathbb{R}^{p_1}. \quad (21)$$

Weiterhin seien  $\hat{\beta}$  und  $\hat{\beta}_0$  die Betaparameterschätzer des vollständigen bzw. reduzierten Modells sowie  $\hat{\varepsilon}$  und  $\hat{\varepsilon}_0$  die Residuenvektoren des vollständigen bzw. reduzierten Modells:

$$\hat{\varepsilon} := y - X\hat{\beta} \quad \text{mit} \quad \hat{\beta} := (X^T X)^{-1} X^T y \\ \hat{\varepsilon}_0 := y - X_0 \hat{\beta}_0 \quad \text{mit} \quad \hat{\beta}_0 := (X_0^T X_0)^{-1} X_0^T y. \quad (22)$$

Dann ist die F-Statistik definiert als

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

## Bemerkungen

- Der Zähler der F-Statistik

$$\frac{\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon}}{p_1} \quad (24)$$

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 Residualquadratsummenreduktion (und gleichem Nenner) ein größerer  $F$ -Wert resultiert, wenn diese durch weniger zusätzliche Regressoren erreicht wird, also  $p_1$  klein ist (und umgekehrt). Im Sinne der Anzahl der Spalten von  $X$  und der entsprechenden Komponenten von  $\beta$  favorisiert die  $F$ -Statistik also weniger "komplexe" Modelle.

- Für den Nenner der F-Statistik gilt

$$\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n - p} = \hat{\sigma}^2, \quad (25)$$

wobei  $\hat{\sigma}^2$  hier der aufgrund des vollständigen Modells geschätzte Schätzer von  $\sigma^2$  ist. Wenn die Daten tatsächlich unter dem reduzierten Modell generiert werden, so kann das vollständige Modell dies durch  $\hat{\beta}_1 \approx 0_{p_1}$  abbilden und erreicht eine ähnliche Varianzschätzung wie das reduzierte Modell. Wenn die Daten tatsächlich unter dem vollständigem Modell generiert werden, so ist  $\hat{\varepsilon}^T \hat{\varepsilon} / (n - p)$  ein besserer Schätzer von  $\sigma^2$  als  $\hat{\varepsilon}_0^T \hat{\varepsilon}_0 / (n - p_0)$ , da sich für ersteren Datenvariabilität, die nicht durch die  $p_0$  Regressoren in  $X_0$  erklärt wird, in der Schätzung von  $\sigma^2$  widerspiegeln würde. Der Nenner der F-Statistik ist also in beiden Fällen der sinnvollere Schätzer von  $\sigma^2$ .

- Zusammengefasst 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 ( $\hat{\sigma}^2$ )- und Regressor ( $p_1$ )-Einheit.

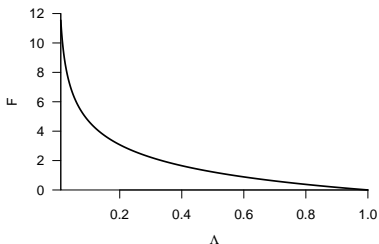
## Theorem (F-Statistik und Likelihood-Quotienten-Statistik)

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

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

### Bemerkungen

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



# F-Statistiken

## Beweis

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

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

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

$$\Lambda = \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \right)^{\frac{n}{2}} \quad (28)$$

ergibt sich dann

$$\begin{aligned} F &= \frac{(\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon})/p_1}{\hat{\varepsilon}^T \hat{\varepsilon}/(n-p)} \\ &= \frac{n(\hat{\sigma}_0^2 - \hat{\sigma}^2)/p_1}{n\hat{\sigma}^2/(n-p)} \\ &= \frac{n-p}{p_1} \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \\ &= \frac{n-p}{p_1} \left( \frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} - 1 \right) \\ &= \frac{n-p}{p_1} \left( \Lambda^{-\frac{2}{n}} - 1 \right). \end{aligned} \quad (29)$$

□

# F-Statistiken

## Beispiel: Einfache lineare Regression

$$X = \begin{pmatrix} X_0 & X_1 \end{pmatrix}, \quad X_0 := \mathbf{1}_n, \quad X_1 := (x_1, \dots, x_n)^T, \quad \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \beta_0 = \text{Offset}, \quad \beta_1 = \text{Steigung} \quad (30)$$

```
# Modellformulierung
library(MASS)
nmod = 2
n = 10
p_0 = 1
p_1 = 1
p = p_0 + p_1
x = 1:n
X = matrix(c(rep(1,n),x), nrow = n)
X_0 = X[,1]
I_n = diag(n)
beta = matrix(c(1,0,
                1,0.5), nrow = 2)
nscn = ncol(beta)
sigsqr = 1

# multivariate Normalverteilung
# Anzahl Modelle
# Anzahl Datenpunkte
# Anzahl Betaparameter reduziertes Modell
# Anzahl zusätzlicher Betaparameter vollständiges Modell
# Anzahl Betaparameter im vollständigem Modell
# Prädiktorwerte
# Designmatrix des vollständigen Modells
# Designmatrix des reduzierten Modells
# Einheitsmatrix

# wahre, aber unbekannte Betaparameter
# Anzahl wahrer, aber unbekannter Hypothesenszenarien
# wahrer, aber unbekannter Varianzparameter

# Simulation und Evaluierung
Eff = matrix(rep(NA, nscn), nrow = nscn)
for(s in 1:nscn){
  y = mvrnorm(1, X %>% beta[,s], sigsqr*I_n)
  beta_0_hat = solve(t(X_0)%*%X_0) %*% t(X_0)%*%y
  beta_hat = solve(t(X) %*% X) %*% t(X) %*%y
  eps_0_hat = y - X_0 %*% beta_0_hat
  eps_hat = y - X %*% beta_hat
  eps_0_eps_0_hat = t(eps_0_hat) %*% eps_0_hat
  eps_eps_hat = t(eps_hat) %*% eps_hat
  Eff[s] = (((eps_0_eps_0_hat-eps_eps_hat)/p_1)/
            (eps_eps_hat/(n-p)))
}

# F-Statistik-Realisierungsarray
# Szenarieniterationen
# Datenrealisierung
# Betaparameterschätzer reduziertes Modell
# Betaparameterschätzer vollständiges Modell
# Residuenvektor reduziertes Modell
# Residuenvektor vollständiges Modell
# RQS reduziertes Modell
# RQS vollständiges Modell
# F-Statistik
```

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

## Theorem (Verteilung der F-Statistik)

Gegeben seien das ALM

$$y = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0_n, \sigma^2 I_n), \quad (31)$$

wobei  $X \in \mathbb{R}^{n \times p}$ , und mit  $p = p_0 + p_1$  eine Partitionierung

$$X = \begin{pmatrix} X_0 & X_1 \end{pmatrix}, \quad X_0 \in \mathbb{R}^{n \times p_0}, \quad X_1 \in \mathbb{R}^{n \times p_1} \quad \text{und} \\ \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \beta_0 \in \mathbb{R}^{p_0}, \quad \beta_1 \in \mathbb{R}^{p_1}. \quad (32)$$

Schließlich sei ein Vektor  $c$  gegeben durch

$$c := \begin{pmatrix} 0_{p_0} \\ 1_{p_1} \end{pmatrix} \in \mathbb{R}^p. \quad (33)$$

Dann gilt

$$F \sim f(\delta, p_1, n - p) \quad \text{mit} \quad \delta := \frac{c^T \beta (c^T (X^T X)^{-1} c)^{-1} c^T \beta}{\sigma^2} \quad (34)$$

### Bemerkungen

- Wir verzichten auf einen vollständigen Beweis.
- $F$  ist eine Funktion der Parameterschätzer,  $\delta$  ist eine Funktion der wahren, aber unbekannt Parameter.
- Diese Verteilung von  $F$  kann zum Nullhypothesentesten und zur Powerfunktionsevaluation genutzt werden.
- Insbesondere gilt  $F \sim f(0, p_1, n - p) = f(p_1, n - p)$ , wenn  $\beta_1 = 0_{p_1}$  und damit  $c^T \beta = 0$ .

# F-Statistiken

## Beweisandeutung

Wir haben bereits gesehen, dass folgende Funktion des Varianzparameterschätzers einer Chi-Quadrat-Verteilung folgt (siehe Einheit (6) in *Allgemeines Lineares Modell*):

$$U_2 = \frac{n-p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n-p). \quad (35)$$

Des Weiteren lässt sich zeigen, dass folgende Funktion der Varianzparameterschätzer von vollständigem und reduziertem Modell einer nichtzentralen Chi-Quadrat-Verteilung folgt:

$$U_1 = \frac{n-p_0}{\sigma^2} \hat{\sigma}_0^2 - \frac{n-p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2 \left( \frac{c^T \beta (c^T (X^T X)^{-1} c)^{-1} c^T \beta}{\sigma^2}, p-p_0 \right). \quad (36)$$

Definitionsgemäß ist der Quotient aus der Zufallsvariable  $U_1$ , geteilt durch ihren Freiheitsgradparameter, und der Zufallsvariable  $U_2$ , geteilt durch ihren Freiheitsgradparameter, eine nichtzentrale  $f$ -Zufallsvariable

$$F = \frac{U_1/(p-p_0)}{U_2/(n-p)} \sim f(\delta, p-p_0, n-p), \quad (37)$$

wobei der Nichtzentralitätsparameter der nichtzentralen  $f$ -Verteilung durch den Nichtzentralitätsparameter der nichtzentralen Chi-Quadrat-Verteilung gegeben ist und sich die F-Statistik wie folgt ergibt:

$$\begin{aligned} F &= \frac{\left( \frac{n-p_0}{\sigma^2} \hat{\sigma}_0^2 - \frac{n-p}{\sigma^2} \hat{\sigma}^2 \right) / (p-p_0)}{\frac{n-p}{\sigma^2} \hat{\sigma}^2 / (n-p)} = \frac{\frac{1}{\sigma^2} \left( (n-p_0) \hat{\sigma}_0^2 - (n-p) \hat{\sigma}^2 \right) / (p-p_0)}{\frac{1}{\sigma^2} (n-p) \hat{\sigma}^2 / (n-p)} \\ &= \frac{\left( (n-p_0) \hat{\sigma}_0^2 - (n-p) \hat{\sigma}^2 \right) / p_1}{(n-p) \hat{\sigma}^2 / (n-p)} = \frac{(\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon}) / p_1}{\hat{\varepsilon}^T \hat{\varepsilon} / (n-p)}. \end{aligned} \quad (38)$$

# F-Statistiken

## Beispiel: Einfache lineare Regression

$$X = \begin{pmatrix} X_0 & X_1 \end{pmatrix}, \quad X_0 := \mathbf{1}_n, \quad X_1 := (x_1, \dots, x_n)^T, \quad \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \beta_0 = \text{Offset}, \quad \beta_1 = \text{Steigung} \quad (39)$$

```
# Modellformulierung
library(MASS)
nmod = 2 # multivariate Normalverteilung
n = 10 # Anzahl Modelle
p_0 = 1 # Anzahl Datenpunkte
p_1 = 1 # Anzahl Betaparameter im reduzierten Modell
p = p_0 + p_1 # Anzahl additiver Betaparameter im vollständigen Modell
x = 1:n # Anzahl Betaparameter im vollständigen Modell
X = matrix(c(rep(1,n),x), nrow = n) # Prädiktorwerte
X_0 = X[,1] # Designmatrix des vollständigen Modells
I_n = diag(n) # Designmatrix des reduzierten Modells
beta = matrix(c(1,0, # Einheitsmatrix
               1,0.5), nrow = 2) # wahre, aber unbekannte Betaparameter

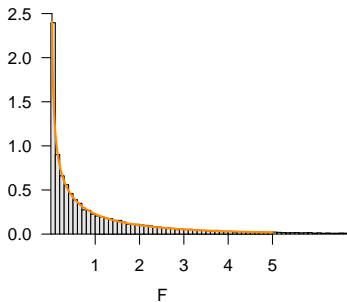
nscn = ncol(beta) # Anzahl wahrer, aber unbekannter Hypothesenszenarien
sigsqr = 1 # wahrer, aber unbekannter Varianzparameter
c = matrix(c(0,1), nrow = 2) # Vektor c

# Frequentistische Simulation
nsim = 1e4 # Anzahl Realisierungen des n-dimensionalen ZVs
delta = rep(NaN,nscn) # Nichtzentralitätsparameterarray
Eff = matrix(rep(NaN, nscn*nsim), nrow = nscn) # F-Statistik-Realisierungsarray
for(s in 1:nscn){ # Szenarieriterationen
  delta[s] = (t(t(c)%*%beta[,s])%*% # Nichtzentralitätsparameter
             solve(t(X_0)%*%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_0_hat = 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_0_hat # 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))))}
}
```

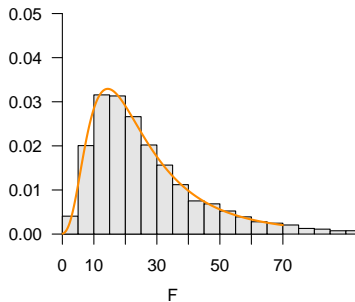
# F-Statistiken

## Beispiel: Einfache lineare Regression

$$\beta = (1, 0)^T$$



$$\beta = (1, 0.5)^T$$



## Ausblick

- Die Theorie von T- und F-Statistiken wird unter dem Begriff der *Allgemeinen Linearen Hypothese* verallgemeinert und integriert. Dabei betrachtet man allgemeine lineare Funktionen der Betaparameter der Form  $C^T\beta$ , wobei  $C \in \mathbb{R}^{p \times k}$  eine beliebige Matrix als Grundlage von Hypothesen ist. Zum Beispiel ergeben sich für  $C \in \mathbb{R}^{p \times 1}$  die hier für F-Statistiken und zuvor für T-Statistiken betrachteten 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

F-Statistiken

**Selbstkontrollfragen**

# Selbstkontrollfragen

---

1. Geben Sie die Definition einer  $f$ -Zufallsvariable wieder.
2. Geben Sie die Definition einer nichtzentralen  $f$ -Zufallsvariable wieder.
3. Skizzieren Sie die WDF einer  $f$ -Verteilung mit  $n_1 = 4, n_2 = 26$ .
4. Skizzieren Sie die WDF einer nichtzentralen  $f$ -Verteilung mit  $\delta = 4, n_1 = 2, n_2 = 26$ .
5. Geben Sie die Definition der Likelihood-Quotienten-Statistik wieder.
6. Erläutern Sie die Definition der Likelihood-Quotienten-Statistik.
7. Geben Sie die Definition von vollständigem und reduziertem ALM wieder.
8. Geben Sie das Theorem zum Likelihood-Quotienten von vollständigem und reduziertem ALM wieder.
9. Definieren Sie die F-Statistik.
10. Erläutern Sie den Zähler der F-Statistik.
11. Erläutern Sie den Nenner der F-Statistik.
12. Erläutern Sie die F-Statistik.
13. Geben Sie das Theorem zum Zusammenhang von F-Statistik und Likelihood-Quotienten-Statistik wieder.
14. Geben Sie das Theorem zur Verteilung der F-Statistik wieder.

- 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*. Wiley Classics Library. New York, NY: Wiley.