Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

CAS Datenanalyse

Kapitel 3: Modellauswahl und F-Test

Prof. Dr. Raúl Gimeno
FRM,CAIA, PRM

Inhalt

Informationskriterien:

- ✓ Akaike
- ✓ Schwarz
- ✓ Hannan-Quinn

T-Test: Beschränkung

t-Test für Linearkombination von Parametern

F-Verteilung

F-Test

F-Test für Restriktionen

Gutes Regressionsmodell

- Relevanz: Das Modell sollte der Fragestellung entsprechen
- Theoretische Konsistenz: Logisch widerspruchsfrei, geschätzte Parameter → erwartete Vorzeichen
- Einfachheit (Parsimony): So einfach wie möglich
- Anpassung (goodness of fit): Verfügbaren Daten möglichst gut abbilden
- Prognosefähigkeit: Gute Prognosen liefern

Trade-off: Anpassungsgüte und Komplexität, gemessen an der Anzahl der Parameter, ausbalancieren.

Top-down Vorgehen: Möglichst umfassendes Modell → durch Tests ein adäquates Modell wählen

Problem: Keine Zauberformel liefert automatisch das «bestes Modell».

Ziel: Kriterium zur Auswahl eines geeigneten Regressionsmodells unter konkurrierenden Modellen.

Informationskriterium: Fitmass

Kriterium: Die Anzahl der Parameter wird dabei "strafend" berücksichtigt, da sonst komplexe Modelle mit vielen Parametern bevorzugt würden.

Alternative zu $\overline{R}^2$: Abwägung des Verzerrungsrisikos und des Risikos einer zu grossen Schätzvarianz.

Selektionskriterium = Fitmass + Anzahl Parameter × Straffunktion

Fitmass

Misst, wie gut sich das geschätzte Modell an die Daten anpasst.

Problem: Zusätzliche Parameter führen im Allgemeinen zu einer Fitverbesserung, niemals jedoch zu einer Fitverschlechterung.

Auswahl: Minus zweimal die Log-Likelihood Funktion

Unterscheidet sich für eine gegebene Stichprobengrösse N nur um eine Konstante von s^2.

OLS-Likelihoodfunktion: $\ell(\mathbf{b}) = -(N/2)[1 + \ln 2\pi + \ln(RSS/N)]$

$$RSS = \sum_i (y_i - \hat{y}_i)^2 = S_{ee}$$

Informationskriterium: Strafterm

Strafterm:

Produkt aus Zahl der geschätzten Parameter k und Straffunktion:

- Der Strafterm bestraft die Anzahl der Parameter, um zu vermeiden, dass überflüssige Variablen mit ins Modell aufgenommen werden und somit das Schätzverfahren ineffizient wird.
- Der Strafterm steigt mit steigendem K (# Regressoren)
- Trade-off: Regressoren werden dann in das Modell aufgenommen, wenn die Strafe geringer ausfällt als die Verbesserung des Fit.
- Durch die Wahl der Straffunktion (und damit des Kriteriums) legt man fest, wie dieser Trade-off quantitativ vorgenommen wird. Gängig sind drei verschiedene Kriterien: AIC, SIC und HQ.

Informationskriterium: Akaike

gretl: AIC = l(b) + 2k $\quad AIC = N\ln\left(\frac{RSS}{N}\right) + 2K + N\left(\left(1+\ln(2\pi)\right)\right)$

Andere Software: $AIC^{*} = \ln\left(\frac{RSS}{N}\right) + 2K$

AIC* = AIC/N – (1+ln(2π))

Hinweis: Je nach Software-Paket unterschiedliche Implementierung der verschiedenen Informationskriterien.

	Modell 1	Modell 2	Modell 3
N	20	20	20
K	2	3	4
Variablen	1	1, 2	1,2 und 3
S_{ee} = e'e	1'260'028	957'698.10	955'691.70
gretl AIC	281.7758	278.2887	280.2467
AIC*	11.251	11.077	11.174

Y: Lohnhöhe
1: Ausbildung
2: Alter
3: Firmenzugehörigkeit

residuale Variation

Regel: Unter allen konkurrierenden Regressionsmodellen wird die Spezifikation gewählt, für die das Informationskriterium den kleinsten Wert annimmt.

Informationskriterium: SIC und HQC

Schwarz Informationskriterium / Bayessches Informationskriterium (BIC)

SIC = -2l(**b**) + Kln(N)

Strafterm: Kln(N)

Faktor des Strafterms wächst logarithmisch mit der Anzahl der Beobachtungen N.

Bereits ab acht Beobachtungen (ln 8 = 2.07944 > 2) bestraft das SIC zusätzliche Parameter schärfer als das AIC.

Hannan-Quinn Kriterium (HQC)

HQC = -2l(**b**) + 2Kln[ln(N)]

Strafterm: 2Kln[ln(N)]

In günstigen Fällen führen alle Kriterien zur gleichen Modellauswahl.

SIC bestraft für Stichprobengrössen (n > 8) zusätzliche Parameter stärker als HQ, und HQ wiederum stärker als AIC.

Informationskriterien: Beispiel

Regel: Unter allen konkurrierenden Regressionsmodellen wird die Spezifikation gewählt, für die das Informationskriterium den kleinsten Wert annimmt.

Korrektes Modell 2: $y = b_1 + b_2\text{Ausbildung} + b_3\text{Alter} + b_4\text{Dauer}$

	Modell 1	Modell 2	Modell 3	
N	20	20	20	
K	2	3	4	
Variablen	1	1, 2	1,2 und 3	Strafterm
AIC	281.775	278.288	280.246	2K
SIC	283.767	281.275	284.229	Kln(N)
HQC	282.164	278.871	281.024	2Kln(ln(N))

Ungünstiger Fall: Die verschiedenen Kriterien führen zu widersprüchlichen Ergebnissen.

Variablenauswahl: t- und F-Tests

gretl Output-Fenster

```
Modell 3: KQ, benutze die Beobachtungen 1-20
Abhängige Variable: Lohn

               Koeffizient   Std.-fehler   t-Quotient   p-Wert
  -------------------------------------------------------------
  const         1000,45       225,727        4,432      0,0004 ***
  Ausbildung      62,4284      21,8345       2,859      0,0114 **
  Alter           12,3539      10,6575       1,159      0,2634
  Dauer           -2,62034     14,2970      -0,1833     0,8569

Mittel d. abh. Var.       1685,000   Stdabw. d. abh. Var.     375,6048
Summe d. quad. Res.       955691,7   Stdfehler d. Regress.    244,3987
R-Quadrat                 0,643465   Korrigiertes R-Quadrat   0,576615
F(3, 16)                  9,625466   P-Wert(F)                0,000720
1 Log-Likelihood         -136,1234 2 Akaike-Kriterium         280,2467
3 Schwarz-Kriterium       284,2296 4 Hannan-Quinn-Kriterium   281,0242
```

1. $l(b) = -\frac{N}{2}(1 + \ln 2\pi - \ln N) - \frac{N}{2}\ln RSS = -136.1234$
2. AIC = -2l(**b**) +2k = -2(-136.1234) + 2x3 = 280.2467
3. SIC = -2l(**b**) + klnN = -2(-136.1234) + 3ln(20) = 284.2296
4. HQC = -2l(**b**) + 2kln[lnN] = -2(-136.1234) + 6ln[ln(20)] = 281.0242

Modellauswahl mittels AIC

Megamodell mit 4 Variablen: X_1, X_2, X_3, X_4 $\rightarrow$ AIC = 159.55
\# Modelle: $2^4 = 16$

	AIC		AIC	
Interzept	208.75	X_2, X_3	166.97	
X_1	169.3	X_2, X_4	166.97	
X_2	190.79	X_3, X_4	164.16	
X_3	170.20	X_1, X_2, X_3	168.40	
X_4	166.02	X_1, X_2, X_4	158.01	
X_1, X_2	169.61	X_1, X_3, X_4	157.55	
X_1, X_3	167.14	X_2, X_3, X_4	165.78	
X_1, X_4	156.01			

Hypothesentests mittels t-Test

Mit dem t-Test kann überprüft werden, ob ein geschätzter Koeffizient einem theoretisch angenommenen Wert (b_i) entspricht.

Prüfgrösse: t-Wert

$$t_i = \frac{b_i - \beta_i}{se(b_i)} \sim t_{N-k}$$

$$t(\mathbf{S}) = \frac{b_i - c}{se(b_i)} \sim t_{N-k} \qquad \text{wenn } H_0\text{: } \beta_i = c \quad \text{wahr}$$

S: Menge aller möglichen Stichproben

Teststatistik t(**S**): Funktion der Stichproben **S**.

Gilt nur wenn H_0: $\beta_i = c$ wahr

Formulierung einer Alternativhypothese.

H_{1a} : $b_i \neq c$

H_{1b} : $b_i > c$

H_{1c} : $b_i < c$

t-Test: Stichprobenkennwertverteilung

$$t_i = \frac{b_i - \beta_i}{se(b_i)} \sim t_{N-k} \qquad \text{wenn } H_0\text{: } \beta_i = c \ \text{ wahr}$$

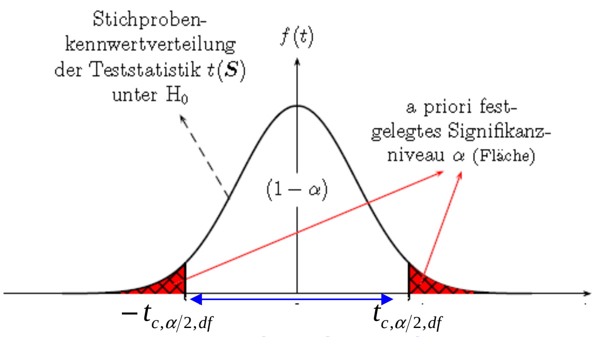
Der Wert der Teststatistik t wird desto näher bei null liegen, je geringer die Distanz zwischen dem Schätzer b_i und dem unter H_0 vermuteten Wert c ist!

Für jede Stichprobe den Wert der Teststatistik t berechnen

Bei unendlich vielen Stichproben mit Umfang N (hypothetisch) →

Histogramm all dieser t-Werte → t-Werte sind t-verteilt.

Stichprobenkennwertverteilung von t(**S**) unter H_0



Annahmebereich

Bewertung der Parameter

Regressionsfunktion (K = 3): $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$

Frage: Trägt der Regressor x_i zur Erklärung von y bei?

Test von H_0: $\beta_i = 0$ gegen H_1: $\beta_i > 0$ oder H_1: $\beta_i \neq 0$ (zweiseitig)

Interpretation: Erklärende Variable x_i hat keinen partiellen Effekt auf y

t-Verteilung für standardisierten Schätzer: $\frac{b_i - \beta_i}{se(b_i)} \sim t_{N-k}$

N: Stichprobengrösse

K: Anzahl Parameter

t-verteilte Statistik: $t_i = \frac{b_i}{se(b_i)} \sim t_{N-3}$ when H_0: $\beta_i = 0$ wahr

Interpretation t_i: Wie viele Standardabweichungen liegt b_i von null entfernt?

Nullhypothese

Ein statistisches Testproblem (statistischer Test) besteht aus einer Nullhypothese H_0 und einer Alternativhypothese H_1, die sich gegenseitig ausschliessen.

Regeln für das Aufstellen der Nullhypothese:

- ✓ Was ich zeigen oder beweisen will, gehört in die Alternativhypothese
- ✓ Das Gleichheitszeichen gehört immer in die Nullhypothese

Die Annahme der Nullhypothese führt immer zur Ablehnung der Alternativhypothese ist aber kein Beweis dafür, dass die Nullhypothese stimmt.

Die Ablehnung der Nullhypothese führt zur Annahme der Alternativ-hypothese.

Gretl-Output: Signifikanzniveau

```
Modell 3: KQ, benutze die Beobachtungen 1-20
Abhängige Variable: Lohn

               Koeffizient   Std.-fehler   t-Quotient   p-Wert
  -------------------------------------------------------------
  const         1000,45       225,727        4,432      0,0004  ***
  Ausbildung      62,4284      21,8345       2,859      0,0114  **
  Alter           12,3539      10,6575       1,159      0,2634
  Dauer           -2,62034     14,2970      -0,1833     0,8569

Mittel d. abh. Var.       1685,000   Stdabw. d. abh. Var.     375,6048
Summe d. quad. Res.       955691,7   Stdfehler d. Regress.    244,3987
R-Quadrat                 0,643465   Korrigiertes R-Quadrat   0,576615
F(3, 16)                  9,625466   P-Wert(F)                0,000720
```

1%-Niveau

5%-Niveau

$$t_i = \frac{b_i}{se(b_i)}$$

Faustregel: $|t| > 2 \quad \rightarrow \quad H_0$: $b_i = 0$ verwerfen

Mittels hochgestellten Sternen * neben dem p-Wert wird kenntlich gemacht, auf welchem Signifikanzniveau α die Nullhypothese $b_i = 0$ verworfen werden kann.

* - Signifikant auf dem 10% Niveau

** - Signifikant auf dem 5% Niveau

*** - Signifikant auf dem 1% Niveau → Koeffizient hoch signifikant

Beschränkung des t-Tests

Statistische Signifikanz vs. theoretische Validität

Einfaches Regressionsmodell mit y = UK Konsumentenpreisindex

Schätzfunktion: $\hat{y} = 10.9 - 3.2x + 0.39\,x^2 \qquad N = 21,\ \bar{R}^2 = 0.982$

$$\begin{array}{cc} (0.23) & (0.02) \\ t = -13.9 & 19.5 \end{array}$$

H_0: $b_2 = 0$ und H_0: $b_3 = 0$ zweiseitiger Test → H_0 ablehnen

Konklusion: x und x^2 statistisch signifikante Variablen für die Erklärung von y.

x: kumulierte Niederschlagsmenge in UK !

Grund der Signifikanz: Beide Variablen weisen einen Trend auf

Ungültiges Modell! → Regression ohne theoretische Fundierung →

Gefahr: Es wird nur noch geprüft, ob die Daten etwas aussagen können!

Beschränkung des t-Tests

t-Test und Wichtigkeit einer Variable

- Statistische Signifikanz gibt Aufschluss darüber, dass ein bestimmtes Stichprobenschätzergebnis nicht reiner Zufall ist.
- Keine Aussage darüber, welche Variable den grössten Teil der Variation von y erklärt
- Schlussfolgerung: Variable mit dem signifikantesten Koeffizienten (höchster t-Wert) ist nicht unbedingt die wichtigste Variable (grösster Effekt auf Y)

t-Test bei hohen Stichprobenumfängen

- Stichprobenumfang N↑: Unverzerrter Schätzwert eines Parameters liegt immer näher am wahren Parameter der Grundgesamtheit.
- $se(b_i)\downarrow \rightarrow t_i \uparrow \rightarrow H_0$ verwerfen

$$t_i = \frac{b_i - \beta_i}{se(b_i)} \sim t_{N-k}$$

- Bedeutung einer Variablen für y nicht alleine mit der Signifikanz beurteilen

t-Test für eine Linearkombination von Parametern

Regressionsfunktion: $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$

H_0: $c_1\beta_2 + c_2\beta_3 = c$ H_1: $c_1\beta_2 + c_2\beta_3 \neq c$

Beispiel 1: Parameter β_2 und β_3 weisen den gleichen Wert auf

Formulierung H_0: $\beta_2 - \beta_3 = 0$ $(c_1 = c_2 = 1)$

Beispiele

Beispiel 2: Parameter β_3 ist halb so gross wie β_2 aber mit unterschiedlichem Vorzeichen.

Formulierung H_0: $\beta_2 + 2\beta_3 = 0$ $(c_1 = 1$ und $c_2 = 2)$

Allgemeines Test-Prinzip: Wenn H_0: $c_1\beta_2 + c_2\beta_3 = c$ wahr ist, gilt in der Grundgesamtheit $c_1\beta_2 + c_2\beta_3 - c = 0$

Aufgrund von Zufallsschwankungen ist dieser Zusammenhang in der Stichprobe nicht exakt erfüllt: $c_1 b_2 + c_2 b_3 - c = v$

wobei v relativ nahe bei null liegen sollte, wenn H_0 wahr ist.

Wenn b_2 und b_3 normalverteilt sind, ist auch v normalverteilt.

t-Test für eine Linearkombination von Parametern

Wenn H_0 wahr gilt $E(v) = 0$.

Standardfehler, um eine Teststatistik zu bestimmen

$$\mathrm{var}(v) = \mathrm{var}(c_1 b_2 \pm c_2 b_3 - c) = c_1^2\,\mathrm{var}(b_2) + c_2^2\,\mathrm{var}(b_3) \pm 2 c_1 c_2\,\mathrm{cov}(b_2, b_3)$$

H_0: $c_1 b_2 \pm c_2 b_3 = c$

Unter H_0 gilt $$\frac{v}{se(v)} = \frac{c_1 b_2 \pm c_2 b_3 - c}{\sqrt{c_1^2\,\mathrm{var}(b_2) + c_2^2\,\mathrm{var}(b_3) \pm 2 c_1 c_2\,\mathrm{cov}(b_2, b_3)}} \sim t_{N-k}$$

Einschränkung des t-Tests: Kann nicht für den Test mehrerer Hypothesen verwendet werden!

Nicht möglich für die simultane Hypothese H_0: $\beta_2 = \beta_3 = 0$

t-Test für Linearkombinationen

Kann H_0 als eine einzelne Linearkombination dargestellt werden, so wird sie normalerweise auf Basis eines t-Tests überprüft.

Allgemeine Form H_0: $\mathbf{r}'\beta = c$

Zweifachregression: $y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + u$

Beispiel H_0: $\beta_2 + \beta_3 = c$ $\Leftrightarrow$ H_0: $\mathbf{r}'\beta = c$ $\Leftrightarrow$ $\underbrace{\begin{pmatrix} 0 & 1 & 1 \end{pmatrix}}_{\mathbf{r}'} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}}_{\beta} = c$

$$t = \frac{\mathbf{r}'\mathbf{b} - c}{\sqrt{\mathbf{r}'\mathrm{var}(\mathbf{b})\mathbf{r}}} \qquad t = \frac{\mathbf{r}'\mathbf{b} - c}{s_e \sqrt{\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}}}$$

$$\underbrace{\begin{pmatrix} 0 & 1 & 1 \end{pmatrix}}_{\mathbf{r}'} \underbrace{\begin{pmatrix} \mathrm{var}(b_1) & s_{12} & s_{13} \\ s_{21} & \mathrm{var}(b_2) & s_{23} \\ s_{31} & s_{32} & \mathrm{var}(b_3) \end{pmatrix}}_{\mathrm{var}(\mathbf{b})} \underbrace{\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}}_{\mathbf{r}} = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} s_{12} + s_{13} \\ \mathrm{var}(b_2) + s_{23} \\ \mathrm{var}(b_3) + s_{23} \end{pmatrix} = \underline{\mathrm{var}(b_2) + \mathrm{var}(b_3) + 2 s_{23}}$$

H_0: $\mathbf{r}'\mathbf{b} = c$

$$t = \frac{b_2 + b_3 - c}{\underline{\sqrt{\mathrm{var}(b_2) + \mathrm{var}(b_3) + 2\,\mathrm{cov}(b_2, b_3)}}} \sim t_{N-k}$$

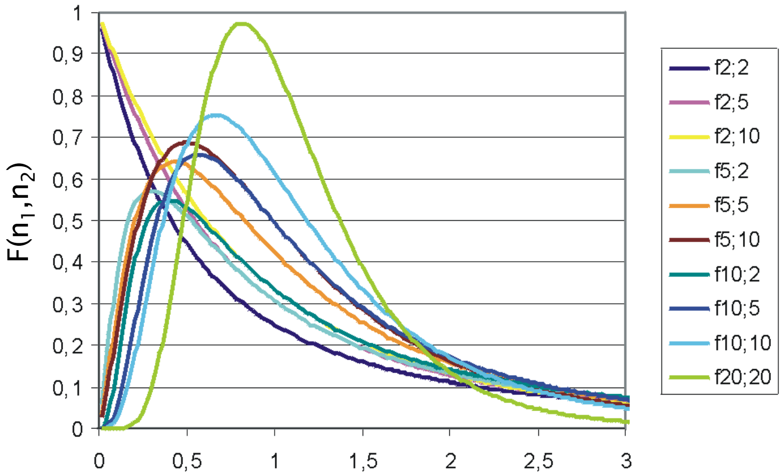
F- Verteilung

Verhältnis zweier unabhängig χ^2 verteilter Zufallsvariablen, die beide durch die entsprechenden Freiheitsgrade dividiert werden.

Es seien $X_1 \sim \chi^2_{n1}$ und $X_2 \sim \chi^2_{n2}$ Chi-Quadrat-verteilt und unabhängig.

Zufallsvariable $F = \dfrac{X_1/n_1}{X_2/n_2} \sim F(n_1, n_2)$ ist F-verteilt mit n_1 und n_2 Freiheitsgraden (df).

Die genaue Form der F-Verteilung hängt von beiden Freiheitsgraden ab.



F-Test: Interpretation

t-Test überprüft den Einfluss jeder einzelnen erklärenden Variable

F-Test testet den gemeinsamen Einfluss aller erklärender Variablen mit Ausnahme des Interzepts getestet werden.

Modell: $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \ldots + \beta_k x_{kt} + u_t \quad \Leftrightarrow \quad \mathbf{Y = Xb + u}$

Frage: Tragen alle Regressoren gesamthaft zur Erklärung von y bei?

Nullhypothese H_0: $\beta_2 = \ldots = \beta_k = 0$ (F-Test: Überwindung der Einschränkung des t-Tests → nur 1 Restriktion)

Interpretation:

Alle Koeffizienten mit Ausnahme des Interzepts (β_1) sind simultan gleich null
⇔ alle K-1 Steigungskoeffizienten sind simultan gleich null.

H_0 wahr ⇒ Regressoren x_i leisten gemeinsam keinen Erklärungsbeitrag für y

Alternativhypothese H_1: H_0 trifft nicht zu ⇔ mindestens einer der Koeffizienten ist ungleich null!

Wenn **u** normalverteilt ist und bei Zutreffen von H_0 gilt:

$$\sum(\hat{y}_i - \bar{y})^2 \Big/ \sigma^2 \sim \chi^2_{K-1} \qquad \text{und} \qquad \sum_i e_i^2 \Big/ \sigma^2 \sim \chi^2_{N-K}$$

F-Test: Formel

F-Statistik:

$$F = \frac{ESS/(K-1)}{RSS/(N-K)} \sim F_{(K-1,N-K)}$$

$$\sum (\hat{y}_i - \bar{y})^2 / \sigma^2 \sim \chi^2_{k-1}$$

$$\sum_i e_i^2 / \sigma^2 \sim \chi^2_{N-k}$$

$$F = \frac{\chi^2_{K-1}/(K-1)}{\chi^2_{N-K}/(N-K)} \sim F(K-1, N-K)$$

F ist verteilt nach F(k-1,N-k) mit k–1 Zähler- und N–k Nenner-Freiheitsgraden.

$$F = \frac{ESS/(K-1)}{RSS/(N-K)} = \frac{S_{\hat{y}\hat{y}}}{S_{ee}} \frac{N-K}{K-1} \sim F_{(K-1,N-K)}$$

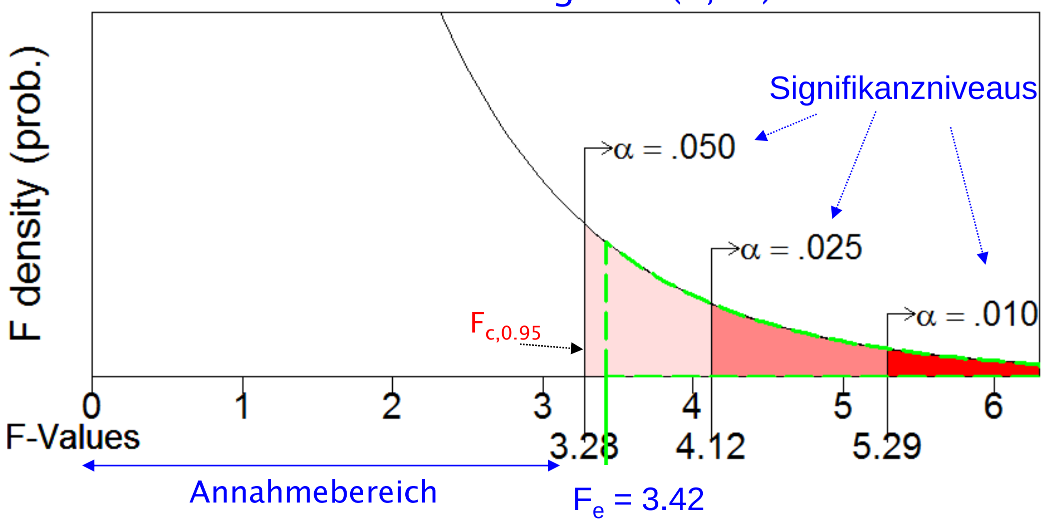
N: Anzahl Beobachtungen   k: Anzahl Regressoren

F-Test: Verhältnis erklärter (ESS) und unerklärter Streuung (RSS).

Achtung: Es ist möglich, dass kein einziger Koeffizient signifikant von null verschieden ist, aber alle Koeffizienten gemeinsam trotzdem signifikant von null verschieden sind!

F-Test: Kritischer Wert

Verteilung für F(2,34)



Für jede Verteilung F(K-1,N-K) lässt sich ein kritischer Wert F_c ermitteln (Tabelle, gretl), $F_{c,0.95}$: 5% der Wahrscheinlichkeitsmasse der Verteilung liegt rechter Hand dieses Wertes.

F_e = empirischer Wert der F-Statistik (berechneter Wert)

F_c: kritischer Wert (aus Tabelle, gretl)

$F_e > F_c \rightarrow$ Nullhypothese auf 5%-Niveau verwerfen

F-Test: Verwerfungsregel

Wenn die durch alle erklärenden X Variablen gemeinsam erklärte Streuung ESS sehr klein ist im Verhältnis zur unerklärten Streuung RSS, würde man einen sehr kleinen Wert der empirischen Statistik F_e erwarten.

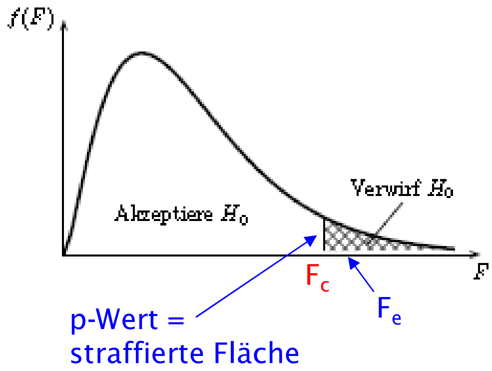
Nullhypothese H_0: $\beta_2 = \ldots = \beta_k = 0$ wird verworfen, wenn der empirische Wert (F_e) > kritischer Wert F_c.

Der zur F-Statistik gehörende p-Wert ist wieder die Fläche unter der Verteilung rechts vom berechneten F_e-Wert.

Verwerfungsregel für H_0

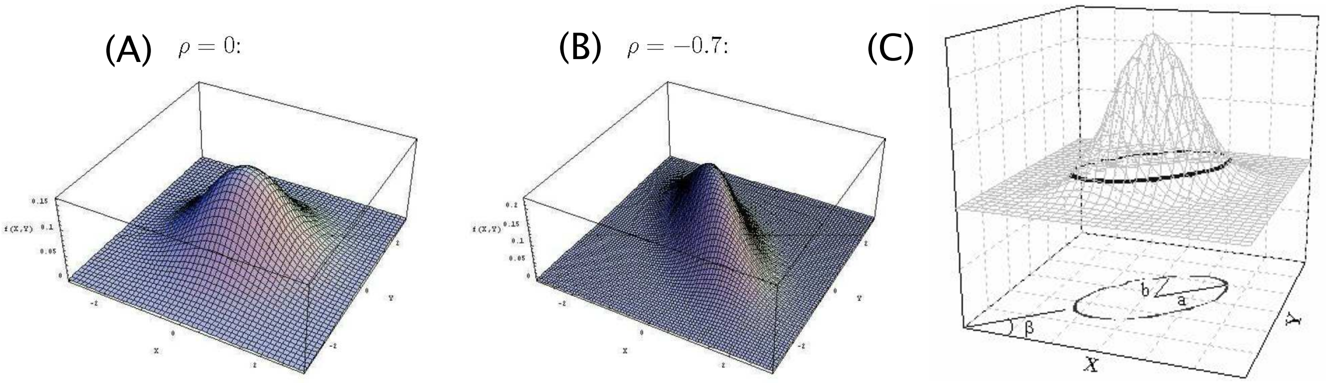
- ✓ $F_e > F_c$
- ✓ p-Wert ≤ Signifikanzniveau

$$F = \frac{ESS/(K-1)}{RSS/(N-K)} \approx F_{(K-1,N-K)}$$



Konfidenzellipse

Bivariate Normalverteilungen ohne und mit Korrelation zwischen den Zufallsvariablen:



Wenn die Variablen unkorreliert sind (A) wird die gemeinsame Signifikanz durch einen Signifikanzkreis dargestellt, bei Korrelation zwischen den Variablen (B) durch eine Konfidenzellipse, die desto schmaler wird, je höher die Korrelation ist.

Die Form der Konfidenzellipse (C) wird vom Korrelationskoeffizienten *r* bestimmt. Eine starke Korrelation bedeutet ein langes *a* (grosse Halbachse) und ein kurzes *b* (kleine Halbachse). Die Ausrichtung der Ellipse hängt ebenfalls von *r* ab.

F-Test: Konfidenzellipse

Regressionsfunktion: $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$

N = 50

Variable		Coefficient	Std. Error	t-Stat.	Prob.
const.	$\widehat{\beta}_1$	7.8878	5.0277	1.5689	0.1234
x_2	$\widehat{\beta}_2$	0.1996	0.2799	0.7131	0.4793
x_3	$\widehat{\beta}_3$	0.6339	0.3442	1.8417	0.0718

R-squared	0.5979	Log likelihood	-210.2721
Adjusted R-squared	0.5808	S.E. of regression	16.7340
Sum squared resid	13161.2380	F-statistic	34.9485
Durbin-Watson Stat.	2.1632	Prob(F-statistic)	0.0000

Einzelne t-Tests

t-Statistik < 2 → H_0: $\beta_2 = 0$ und H_0: $\beta_3 = 0$ werden für α = 5% nicht abgelehnt! Die Koeffizienten sind individuell statistisch nicht signifikant.

F-Test

Nullhypothese H_0: $\beta_2 = \beta_3 = 0$

F = 34.94 und p-Wert = 0 → H_0 verwerfen

Die Koeffizienten β_2, β_3 sind gemeinsam hochsignifikant!

F-Test

Regressionsfunktion: $Y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$

Nullhypothesen

Individuelle H_{01}: $\beta_2 = 0$ und H_{02}: $\beta_3 = 0$

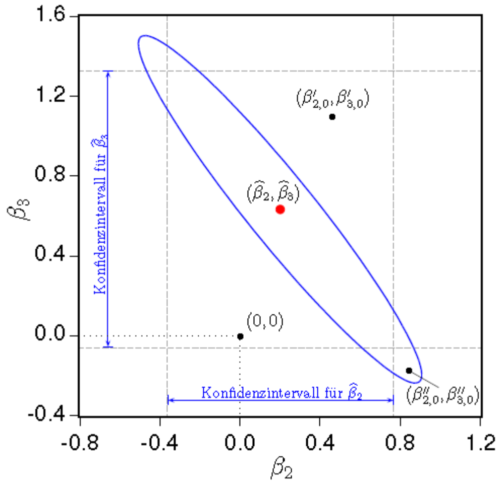
Gemeinsame H_{03}: $\beta_2 = \beta_3 = 0$

$\Rightarrow$ β_2, β_3 gemeinsam hochsignifikant

Die Punkte (0,0) und (β_2', β_3') liegen innerhalb der individuellen Konfidenzintervalle (H_{01} und H_{02} nicht ablehnen) aber ausserhalb der Konfidenzellipse (H_{03} ablehnen)

Punkt (β_2'', β_3'') liegt ausserhalb der individuellen Konfidenzintervalle (H_{01} und H_{02} ablehnen) aber innerhalb der Konfidenzellipse (H_{03} nicht ablehnen)

95%-Konfidenzellipse



Gretl Output: F-Test

Modell: Pro-Kopf-Konsum = $b_1 + b_2$Einkommen + b_3T

= 300.28 + 0.741Einkommen + 8.04T

N = 15 und K = 3 und Freiheitsgrade = N-K = 12

Nullhypothese H_0: $\beta_2 = \beta_3 = 0$ → Einkommen und Zeit liefern gemeinsam keinen Erklärungsbeitrag für Konsum (y)

F(K-1, N-k) = F(2,12)= 2'513.521

Kritischer Wert: $F_c = 3.88$ p-Wert = $P[F > F_e = 2'513] = 0 < \alpha$

Empirischer Wert: $F_e = 2'513.521 > F_c$ → H_0 verwerfen

Abhängige Variable: Konsum

	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	300,286	78,3176	3,834	0,0024	***
Einkommen	0,741981	0,0475337	15,61	2,46e-09	***
T	8,04356	2,98355	2,696	0,0195	**

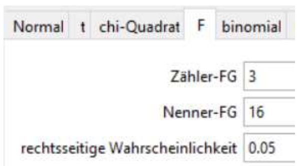
Mittel d. abh. Var.	1942,333	Stdabw. d. abh. Var.	243,5044
Summe d. quad. Res.	1976,855	Stdfehler d. Regress.	12,83503
R-Quadrat	0,997619	Korrigiertes R-Quadrat	0,997222
F(2, 12)	2513,521	P-Wert(F)	1,82e-16
Log-Likelihood	-57,89317	Akaike-Kriterium	121,7863
Schwarz-Kriterium	123,9105	Hannan-Quinn-Kriterium	121,7637
rho	-0,496597	Durbin-Watson-Stat	2,488178

Gretl-Output: F-Test

	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	1000,45	225,727	4,432	0,0004	***
Ausbildung	62,4284	21,8345	2,859	0,0114	**
Alter	12,3539	10,6575	1,159	0,2634	
Dauer	-2,62034	14,2970	-0,1833	0,8569	

Mittel d. abh. Var.	1685,000	Stdabw. d. abh. Var.	375,6048
Summe d. quad. Res.	955691,7	Stdfehler d. Regress.	244,3987
R-Quadrat	0,643465	Korrigiertes R-Quadrat	0,576615
F(3, 16)	9,625466	P-Wert(F)	0,000720

Modell 3: Lohn = $b_1 + b_2$Ausb. + b_3Alter + b_4Dauer



N = 20 und K = 4 und Freiheitsgrade (df) = N-K = 16

Nullhypothese H_0: $\beta_2 = \beta_3 = \beta_4 = 0$

F(K-1, N-K) = F(3,16)= 9.625

Kritischer Wert: $F_c = 3.238$ p-Wert = $P[F > F_e = 9.625] \cong 0 < \alpha = 0.05$

Regel: $F_e = 9.625 > F_c$ → H_0 verwerfen

Konklusion: Mindestens eine erklärende Variable ist

F-Test und t-Test

- Achtung: Es reicht nicht zu überprüfen, ob alle Koeffizienten individuell signifikant von Null verschieden sind!
- Multikollinearität: Die erklärenden Variablen sind untereinander hoch korreliert
- Folge: Kein einziger Koeffizient ist signifikant von null verschieden, aber alle Koeffizienten sind gemeinsam signifikant von null verschieden!
- Konklusion: Die mit der F-Statistik getestete gemeinsame Nullhypothese $H_0 : \beta_2 = \beta_3 = \cdots = \beta_k = 0$ darf nicht durch eine Reihe individueller t-Tests ersetzt werden!
- Grund: Die F-Statistik berücksichtigt die mögliche Korrelation zwischen den OLS-Schätzern $\beta_2, \beta_3, \ldots, \beta_k$, während diese bei individuellen t-Tests in dieser Form unberücksichtigt bleibt.

F-Test und R^2

Die Teststatistik des F-Tests als Funktion des Bestimmheitsmasses:

$$F = \frac{ESS/(K-1)}{RSS/(N-K)} = \frac{R^2}{1\text{-}R^2} \cdot \frac{N-K}{K-1}$$

Grosser Wert von $R^2 \rightarrow$ grosser Wert von F

ANOVA Analyse = Varianzanalyse →schematische Schreibweise zur Berechnung des F-Tests.

N = 20 und K = 4

	Quadratsumme	FG	quad. Mittel
Regression	1,72481e+006	3	574936
Residuum	955692	16	59730,7
Total	2,6805e+006	19	141079

R^2 = 1,72481e+006 / 2,6805e+006 = 0,643465

F(3, 16) = $\boxed{574936}$ / $\boxed{59730,7}$ = 9,62547 [p-Wert 0,0007]

ESS/(K–1) = ESS/3

RSS/(N–K) = RSS/16

TSS/(N–1) = TSS/19

$$F = \frac{R^2}{1-R^2}\frac{N-K}{K-1} = \frac{0.6434}{1-0.6434}\frac{16}{3} = 9.625$$

F-Wert entweder mittels R^2 oder mittels Varianzanalyse

F-Test in Matrixform

- Regression: $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \ldots + \beta_k x_{kt} + u_t$
- Nullhypothese H_0: $\beta_2 = \beta_3 = \ldots = \beta_k = 0$
- Allgemeine Form H_0: $\mathbf{R\beta} = \mathbf{q}$

$$\mathbf{R\beta} = \mathbf{q} \quad \Leftrightarrow \quad \underbrace{\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}}_{L \times k} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}}_{K \times 1} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

L = K-1 = # Linearkombinationen

Nullhypothese H_0: $\beta_2 + \beta_3 + \ldots + \beta_k = 1$ und $\beta_2 + 2\beta_3 = 0$

$$\mathbf{R\beta} = \mathbf{q} \Leftrightarrow \underbrace{\begin{pmatrix} 0 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & 0 & \cdots & 0 \end{pmatrix}}_{2 \times K} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} = \underbrace{\begin{pmatrix} 1 \\ 0 \end{pmatrix}}_{2 \times 1}$$

L = 2 Linearkombinationen

Vergleich t-Test und F-Test

- Nullhypothese H_0: $\mathbf{R\beta} = \mathbf{q}$ (Lx1-Vektor)

$$F = \frac{(\mathbf{Rb} - \mathbf{q})' \left[\mathbf{R}(\mathbf{X'X})^{-1}\mathbf{R'}\right]^{-1} (\mathbf{Rb} - \mathbf{q}) / L}{\mathbf{e'e} / (T - K)}$$

H_0: $\mathbf{r}'\beta = c$ (Skalar)

$$t = \frac{\mathbf{r'b} - c}{s_e \sqrt{\mathbf{r}'(\mathbf{X'X})^{-1}\mathbf{r}}}$$

- Aus dem Zähler dieses Ausdrucks wird deutlich, dass der F-Test analog wie ein t-Test verstanden werden kann, nämlich als ein Vergleich der auf Basis der beobachteten Stichprobendaten berechneten Werte **Rb** mit den in der Nullhypothese angegebenen Parameterwerten **R**β (=q).
- Ist die Differenz zwischen der Beobachtung **Rb** und der Nullhypothese **R**β zu gross, dann wird die Nullhypothese abgelehnt.

- Konklusion: Mit Hilfe der Matrix-Notation kann eine Analogie zwischen F-und t-Test gezeigt werden!

Restringiertes vs. unrestringiertes Modell

Zahlreiche Tests beruhen auf einem Vergleich zweier Modelle.

Modell ohne Restriktionen = unrestringiertes Modell: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$

Restriktion H_0: $\beta_3 = 0$

- Modell mit Restriktion = restringiertes Modell: $y_i = \beta_1' + \beta_2' x_{2i} + u_i'$
- Wenn $\beta_3 = 0$ wahr $\rightarrow$ $u_i = u_i'$

Restriktion H_0: $\beta_3 = 1$

$\beta_3 = 1$

- Restringiertes Modell: $y_i = \beta_1'' + \beta_2'' x_{2i} + x_{3i} + u_i''$
- Geschätztes Modell: $(y_i - x_{3i}) = b_1'' + b_2'' x_{2i} + e_i''$

$\beta_3 = 1 - \beta_2$

Restriktion H_0: $\beta_2 + \beta_3 = 1$

- Restringiertes Modell: $y_i = \beta_1^* + \beta_2^* x_{2i} + (1 - \beta_2^*) x_3 + u_i^*$
- Geschätztes Modell: $(y_i - x_{3i}) = b_1^* + b_2^* (x_{2i} - x_{3i}) + e_i^*$

F-Statistik: $$F = \frac{(RSS_r - RSS)/L}{RSS/(N-K)} \approx F_{(K-1,N-K)}$$

Restringiertes vs. unrestringiertes Modell

Unrestringiertes Modell: $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$ (1)

Wenn Nullhypothese H_0: $\beta_2 = \beta_3 = 0$ wahr ist:

Nullhypothesenmodell = Restringiertes Modell: $y_t = \beta_1 + u_t^0$ (2)

Schätzung: $e_t^0 = y_t - b_1$

OLS-Schätzung von (2) $\Leftrightarrow$ Minimierung

$$S_{ee}^0 = \sum_t (y_t - b_1)^2 = \sum_t (y_t^2 + b_1^2 - 2 y_t b_1) = \sum y_t^2 + N b_1^2 - 2 b_1 \sum_t y_t$$

$$dS_{ee}^0 / db_1 = 2 N b_1 - 2 \sum_t y_t = 0 \quad \Rightarrow \quad b_1 = \bar{y}$$

$$e_t^0 = y_t - \bar{y}$$

$$S_{ee}^0 = \sum e_t^{02} = \sum (y_t - \bar{y})^2 = S_{yy} = TSS$$

$$F = \frac{(S_{ee}^0 - S_{ee})/L}{S_{ee}/(N-K)} = \frac{(TSS - RSS)}{RSS/(N-K)} \frac{(N-K)}{L} \approx F_{(L,N-K)}$$

L: Anzahl der formulierten Restriktionen

N–K: Anzahl der Freiheitsgrade

F-Test: US Pro-Kopf-Konsumausgaben

Abhängige Variable: Konsum

	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	300,286	78,3176	3,834	0,0024	***
Einkommen	0,741981	0,0475337	15,61	2,46e-09	***
T	8,04356	2,98355	2,696	0,0195	**

	Mittel d. abh. Var.	1942,333	Stdabw. d. abh. Var.	243,5044
RSS	Summe d. quad. Res.	1976,855	Stdfehler d. Regress.	12,83503
	R-Quadrat	0,997619	Korrigiertes R-Quadrat	0,997222
	F(2, 12)	2513,521	P-Wert(F)	1,82e-16
	Log-Likelihood	-57,89317	Akaike-Kriterium	121,7863

N = 15
K = 3
N – K = 12 = FG (df)

H_0: $\beta_2 = \beta_3 = 0 \rightarrow L = 2$

$$F = \frac{(RSS_r - RSS)}{RSS}\frac{(N-K)}{L} =$$

$$\frac{830'121.3 - 1'976.86}{1'976.86} \cdot \frac{12}{2} = 2513.52$$

ANOVA

	Quadratsumme	FG	quad. Mittel
Regression	828144	2	414072
Residuum	1976,86	12	164,738
Total	830121	14	59294,4

R^2 = 828144 / 830121 = 0,997619

TSS F(2, 12) = 414072 / 164,738 = 2513,52 [p-Wert 1,82e-016]

$$F = \frac{(TSS_r - RSS)}{RSS}\frac{(N-K)}{L}$$

Restringiertes Modell

	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	1942,33	62,8726	30,89	2,78e-014	***

	Mittel d. abh. Var.	1942,333	Stdabw. d. abh. Var.	243,5044
RSS_r	Summe d. quad. Res.	830121,3	Stdfehler d. Regress.	243,5044
	R-Quadrat	0,000000	Korrigiertes R-Quadrat	0,000000

$$F = \frac{830'121.3 - 1'976.86}{1'976.86} \cdot \frac{12}{2} = 2513.52$$

F-Test für Gruppe von Parametern

- Nullhypothese H_0: $\beta_3 = \beta_4 = 0$
- Anzahl Restriktionen: L = 2
- Modell 3: Lohn = b_1 + b_2Ausb. + b_3Alter + b_4Dauer
- Restringiertes Modell mit $\beta_3 = \beta_4 = 0$:

	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	1354,66	94,2224	14,38	2,62e-011	***
Ausbildung	89,2817	19,8198	4,505	0,0003	***

Mittel d. abh. Var.	1685,000	Stdabw. d. abh. Var.	375,6048
Summe d. quad. Res.	1260028	Stdfehler d. Regress.	264,5781
R-Quadrat	0,529928	Korrigiertes R-Quadrat	0,503813
F(1, 18)	20,29200	P-Wert(F)	0,000274

Teststatistik: F(2, 16) = 2,54757, p-Wert 0,109522
Das Weglassen von Variablen verbesserte 1 von 3 Informationskriterien.

$$F = \frac{(RSS_r - RSS)}{RSS}\frac{(N-K)}{L} = \frac{1'260'028 - 955'691.7}{955'691.7} \cdot \frac{16}{2} = 2.547$$

- Intuition: Grosser Wert von F → Alter und Dauer leisten Erklärungsbeitrag
- $F_e < F_c(2, 16) = 3.63$ → H_0 nicht verwerfen
- p-Wert ≅ 10.9% > α = 5% → H_0 nicht verwerfen
- Beide Variablen Alter und Firmenzugehörigkeit sind gemeinsam nicht signifikant