

Einführung Mikroökometrie

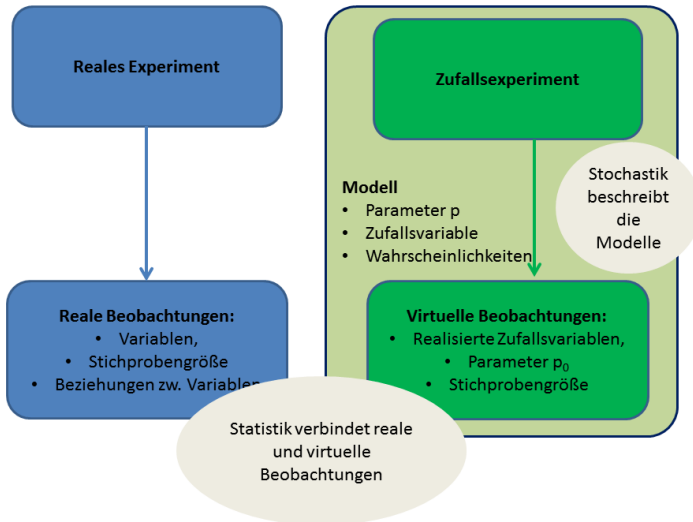
310208 (VO)

Dietmar Bauer

Wiederholung

Letztes Mal haben wir begonnen, ökonometrische Modellierung zu untersuchen:

- Modelle werden formuliert in der Sprache der Stochastik.
- Zufällige Effekte haben unterschiedliche Gründe (unbeobachtete, unbeobachtbare Effekte, Stichprobenziehung, Begrenzung der Modellkomplexität, ...) und können als Zufallsvariablen modelliert werden.
- Daher haben wir für einen beobachteten Datensatz ein Modell, das durch ein Zufallsexperiment beschrieben wird, welches als Ergebnis Zufallsvariable liefert.
- Die Zufallsvariablen werden mit Verteilungen beschrieben: PMF für diskrete, PDF für kontinuierliche Variable.
- Zur Charakterisierung von Zufallsvariablen gibt es Maßzahlen: Lage durch Erwartungswert, Streuung mit Varianz und Standardabweichung.
- Auf Ebene der Daten gibt es die selben Konzepte als Stichprobengrößen (Mittelwert, empirische Varianz).



Zusammenhang zwischen Variablen

gemeinsame Verteilung von educ und wage in der Population der CPS
Befragung: $p_{i,j}$

educ years	wage [\$]				
	0-5	5-10	10-15	15-20	20-25
0	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00
4	0.01	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00
6	0.01	0.00	0.00	0.00	0.00
7	0.01	0.00	0.00	0.00	0.00
8	0.03	0.02	0.00	0.00	0.00
9	0.03	0.00	0.00	0.00	0.00
10	0.04	0.01	0.00	0.00	0.00
11	0.04	0.01	0.00	0.00	0.00
12	0.23	0.13	0.01	0.00	0.00
13	0.04	0.03	0.01	0.00	0.00
14	0.06	0.03	0.01	0.00	0.00
15	0.02	0.02	0.01	0.00	0.00
16	0.03	0.07	0.02	0.01	0.00
17	0.00	0.01	0.01	0.00	0.00
18	0.01	0.01	0.01	0.00	0.00

Randverteilungen: $p_{wage}(i)$

wage:

[\$]	0-5	5-10	10-15	15-20	20-25
Anzahl	292	182	36	11	5
Prob.	0.56	0.35	0.07	0.02	0.01

Bedingte Verteilung: wage für educ = 12: $p_{wage|educ=12}(i)$

[\$]	0-5	5-10	10-15	15-20	20-25
Anzahl	121	68	6	1	2
gem. P	0.23	0.13	0.01	0.00	0.00
bed. P	0.61	0.34	0.03	0.01	0.01

Bedingte Verteilung: wage für educ = 16: $p_{wage|educ=16}(i)$

[\$]	0-5	5-10	10-15	15-20	20-25
Anzahl	15	39	8	5	1
gem. P	0.03	0.07	0.02	0.01	0.00
bed. P	0.22	0.57	0.12	0.07	0.02

Bedingter Erwartungswert: Erwartungswert der bedingten Verteilung:

- diskret: $\mathbb{E}(\mathbf{y}|\mathbf{x} = i) = \sum_{j=1}^J j p_{\mathbf{y}|\mathbf{x}=i}(j)$
- Der bedingte Erwartungswert hängt von der bedingenden Variablen ab:
 $\mathbb{E}(\mathbf{y}|\mathbf{x}) = g(\mathbf{x})$.

Beispiel:

wage [\$]	5	10	15	20	25	bed. Erwartungswert
bed. P (educ = 12)	0.61	0.34	0.03	0.01	0.01	7.35
bed. P (educ = 16)	0.22	0.57	0.12	0.07	0.02	10.44

Berechnung *educ* = 12:

$$5 * 0.61 + 10 * 0.34 + 15 * 0.03 + 20 * 0.01 + 25 * 0.01 = 7.35$$

Berechnung *educ* = 16:

$$5 * 0.22 + 10 * 0.57 + 15 * 0.12 + 20 * 0.07 + 25 * 0.02 = 10.44$$

- Für bedingte Erwartungswerte gelten die gleichen Rechenregeln wie für Erwartungswerte bezüglich Skalierung, Linearität, Konstanten.
- Gesetz der iterierten Erwartungswert: $\mathbb{E} g(\mathbf{x}, \mathbf{y}) = \mathbb{E}(\mathbb{E}(g(\mathbf{x}, \mathbf{y})|\mathbf{x}))$.
- Funktionen von bedingenden Variablen:
$$\mathbb{E}(h(\mathbf{x})g(\mathbf{x}, \mathbf{y})|\mathbf{x}) = h(\mathbf{x}) \mathbb{E}(g(\mathbf{x}, \mathbf{y})|\mathbf{x})$$

Alternativen

- Die bedingten Verteilungen sind oft nicht einfach zu charakterisieren.
- Eine Alternative zur Darstellung der Abhängigkeit: Kovarianzen bzw. Korrelationen.
- $\text{Cov}(\mathbf{x}, \mathbf{y}) = \mathbb{E}(\mathbf{x} - \mathbb{E} \mathbf{x})(\mathbf{y} - \mathbb{E} \mathbf{y})$.
- Korrelation: $\text{Corr}(\mathbf{x}, \mathbf{y}) = \text{Cov}(\mathbf{x}, \mathbf{y}) / \sqrt{\text{Var}(\mathbf{x}) \text{Var}(\mathbf{y})}$.
- Stichprobenäquivalent:

$$\widehat{\text{Cov}}(x_i, y_i) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})$$

empirische Korrelation: $\widehat{\text{Corr}}(x_i, y_i)$.

Rechenregeln für Zufallsvektoren

- Erwartungswerte, Varianzen und Kovarianzen werden elementweise gebildet.
- $\mathbb{E} \mathbf{A}\mathbf{X} = \mathbf{A} \mathbb{E} \mathbf{X}$.
- $\text{Var}(\mathbf{A}\mathbf{X} + \mathbf{B}) = \mathbf{A} \text{Var}(\mathbf{X}) \mathbf{A}'$.
- $\text{Cov}(\mathbf{A}_x \mathbf{X} + \mathbf{B}_x, \mathbf{A}_y \mathbf{Y} + \mathbf{B}_y) = \mathbf{A}_x \text{Cov}(\mathbf{X}, \mathbf{Y}) \mathbf{B}'$.



Das lineare Regressionsmodell in Matrixform

Grundmodell zur multiplen Regression: linearer Zusammenhang in der Population:

Assumption (MLR.1 (lineares Modell))

Die Daten entstammen dem folgenden Modell in der Population:

$$\mathbf{y}_i = \beta_0 + \sum_{k=1}^K \beta_k \mathbf{x}_{k,i} + \mathbf{u}_i = \mathbf{X}_i' \boldsymbol{\beta} + \mathbf{u}_i, \mathbf{X}_i = (1, \mathbf{x}_{1,i}, \dots, \mathbf{x}_{i,K})',$$

wobei β_k unbekannte Koeffizienten und \mathbf{u}_i ein unbeobachteter Fehlerterm ist.

Der datengenerierende Prozess soll aus diesem Modell kommen:

Assumption (MLR.2 (Stichprobe))

Wir beobachten eine zufällige Stichprobe von n Beobachtungen (y_i, \mathbf{X}_i) , $\mathbf{X}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,K}]' \in \mathbb{R}^K$ aus der Population gemäß MLR.1.

Die Regressoren müssen genug Information enthalten:

Assumption (MLR.3 (Information in den Regressoren))

In der Stichprobe (und daher auch in der Population) ist keine der Variablen $x_{i,k}$, $i = 1, \dots, n$ konstant (Standardabweichung null) und es gibt keine linearen Abhängigkeiten zwischen den Regressoren.

Die Fehler weisen keine systematischen Abhängigkeiten von den Regressoren auf:

Assumption (MLR.4 (bedingte Unkorreliertheit der Fehler))

Der Fehlerterm u_i hat bedingt auf die Regressoren Erwartung gleich null:

$$\mathbb{E}(u_i | \mathbf{X}_i) = 0$$

Die bedingte Varianz der Fehler weist keine systematische Abhängigkeit von den Regressoren auf:

Assumption (MLR.5 (Homoskedastizität))

Die Varianz des Fehlers bedingt auf die Regressoren ist konstant:

$$\text{Var}(\mathbf{u}_i | \mathbf{X}_i) = \sigma^2$$

Annahmen MLR.1, MLR.2 und MLR.4 implizieren:

- Die Variablen \mathbf{y}_i , \mathbf{X}_i und \mathbf{y}_j , \mathbf{X}_j für $i \neq j$ sind unabhängig.
- Die Fehler weisen Erwartungswert null auf, sowohl bedingt als auch unbedingt: $\mathbb{E} \mathbf{u}_i = 0$.
- Das lineare Modell ist korrekt spezifiziert.

MLR.3 impliziert, dass genug Information in den Regressoren ist, um die verschiedenen Effekte der verschiedenen Regressoren zu trennen.

Um Informationen über die Parameter zu gewinnen, wird die Residuenquadratsumme minimiert:

$$\hat{\beta} = \arg \min_{b \in \mathbb{R}^{K+1}} \sum_{i=1}^n (y_i - X_i' b)^2$$

Matrixschreibweise:

$$\arg \min Q(b) := (Y - Xb)'(Y - Xb)$$

wobei

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,K} \\ 1 & x_{2,1} & \vdots & & \vdots \\ \vdots & \vdots & & & \vdots \\ 1 & x_{n,1} & x_{n,2} & \dots & x_{n,K} \end{pmatrix},$$

$$Q(b) = (Y - Xb)'(Y - Xb) = Y'Y - b'X'Y - Y'Xb + b'X'Xb$$

Ableiten und Nullsetzen ergibt die folgenden Gleichungen:

$$X'Y = \sum_{i=1}^n X_i y_i' = \sum_{i=1}^n X_i X_i' \hat{\beta} = X'X \hat{\beta}$$

In Matrixschreibweise ergibt sich dann die Schätzung:

$$\hat{\beta} = (X'X)^{-1} X'Y = \left(\sum_{i=1}^n X_i X_i' \right)^{-1} \sum_{i=1}^n X_i y_i$$



Zwei Konzepte, die unterschieden werden müssen:

Schätzer: Funktion einer (multivariaten) Zufallsvariable der Form

$$\hat{\beta} = \hat{\beta}(\mathbf{X}), \mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^n.$$

Beispiel: $\bar{\mathbf{X}} = n^{-1} \sum_{i=1}^n \mathbf{x}_i.$

- Ist eine Zufallsvariable als Transformation einer Zufallsvariablen.
- Als Zufallsvariable hat der Schätzer eine Verteilung, einen Erwartungswert und eine Varianz.

Schätzung: gleiche Funktion angewandt auf die Stichprobe X : $\hat{\beta}(X).$

Beispiel: \bar{X}

- als Funktion der Stichprobe fix gegeben
- hat keine Verteilung, null Varianz



Eigenschaften von Schätzern: angenommen der datengenerierende Prozess stammt aus einer Zufallsvariablen mit Parametervektor β :

- **Unverzerrtheit (Englisch: unbiasedness):** $\mathbb{E} \hat{\beta}(\mathbf{X}) = \beta, \forall \beta$.
- **Schätzvarianz:** $\text{Var}(\hat{\beta}(\mathbf{X}))$: je kleiner die Varianz, desto genauer (präziser) ist der Schätzer.
- Die Varianz hängt sowohl vom datengenerierenden Prozess als auch von der Definition des Schätzers ab. Wir sind klarerweise daran interessiert, möglichst präzise Schätzer zu verwenden.
- **Konsistenz:** je mehr Daten wir zur Verfügung haben, desto präziser wird der Schätzer (genaue Definition folgt noch).

Schätzfehler

- In Matrixform ist der Schätzfehler relativ einfach auszurechnen:

$$\begin{aligned}\hat{\beta} - \beta &= (X'X)^{-1}X'Y - \beta = (X'X)^{-1}X'(X\beta + U) - \beta \\ &= (X'X)^{-1}X'(X\beta + (X'X)^{-1}X'U - \beta) = (X'X)^{-1}X'U\end{aligned}$$

- Daher $\mathbb{E}(\hat{\beta}|\mathbf{X}) = \beta + \mathbb{E}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}|\mathbf{X}) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}(\mathbf{U}|\mathbf{X})$.
- Daher ist der Schätzer unverzerrt für $\mathbb{E}(\mathbf{U}|\mathbf{X}) = 0$ (MLR.1, MLR.2, MLR.4).
- Und: $\text{Var}(\hat{\beta}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{U}|\mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$.
- Aus MLR.2 und MLR.5: $\text{Var}(\mathbf{U}|\mathbf{X}) = \sigma^2 I_n$. Dann haben wir $\text{Var}(\hat{\beta}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

Zusammenfassung aus Einführung in Ökonometrie

Theorem (Unverzerrtheit des KQ-Schätzers)

Unter den Annahmen MLR.1-MLR.4 ist der KQ-Schätzer unverzerrt. Das heißt, wenn der datengenerierende Prozess gemäß MLR.1 den Parametervektor β aufweist, dann gilt:

$$\mathbb{E}\hat{\beta} = \beta$$

Bedeutung der Unverzerrtheit:

- Wenn wir viele verschiedene Datensätze aus dem gleichen datengenerierenden Prozess nehmen würden, und für jeden dieser Datensätze getrennt die KQ-Schätzung durchführten, dann wäre das arithmetische Mittel dieser vielen KQ-Schätzungen gleich den wahren Parametern.
- Das gilt unabhängig vom Wert der wahren Parameter, von der Stichprobengröße und von der Verteilung der Fehler.

Das R^2

Maß für die Güte der Anpassung:

$$R^2 = SSE/SST = 1 - SSR/SST$$

wobei

- $SST = \sum_{i=1}^n (y_i - \bar{Y})^2 = SSE + SSR$: total sum of squares
- $SSE = \sum_{i=1}^n \hat{\beta}' (X_i - \iota \bar{X})(X_i - \iota \bar{X})' \hat{\beta}$: explained sum of squares
- $SSR = \sum_{i=1}^n \hat{u}_i^2$ residual sum of squares, $\hat{u}_i = y_i - X_i \hat{\beta}$

$$0 \leq R^2 \leq 1$$

Adjustiertes R^2 berücksichtigt Zahl der Regressoren:

$$\bar{R}^2 = 1 - [SSR/(n - K - 1)]/[SST/(n - 1)].$$

R^2 und der Satz von Pythagoras

$$\begin{aligned}\hat{U}'X\hat{\beta} &= (Y - X\hat{\beta})X\hat{\beta} = Y'X\hat{\beta} - \hat{\beta}'X'X\hat{\beta} \\ &= Y'X(X'X)^{-1}X'Y - Y'X(X'X)^{-1}X'X(X'X)^{-1}X'Y = 0.\end{aligned}$$

Daher gilt: $Y = Y - X\hat{\beta} + X\hat{\beta} = \hat{U} + X\hat{\beta}$ und $\hat{U}'\iota_n = 0$ woraus wir folgern $\bar{Y} = \bar{X}\hat{\beta}$.

Damit:

$$\begin{aligned}(Y - \iota_n\bar{Y} - (X - \iota_n\bar{X})\hat{\beta})'(Y - X\hat{\beta}) &= \hat{U}'\hat{U} + ((X - \iota_n\bar{X})\hat{\beta})'(X - \iota_n\bar{X})\hat{\beta} \\ SST &= SSR + SSE.\end{aligned}$$

Theorem (Varianz der KQ-Schätzer)

Unter den Annahmen MLR.1-MLR.5 ist die Varianz der KQ-Schätzer bedingt auf die Regressoren \mathbf{X} gegeben durch:

$$\text{Var}(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \right)^{-1}$$

Man kann zeigen, dass die Varianz der j -ten Komponente des Vektors β durch folgende Formel gegeben ist:

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

wobei R_j^2 das R^2 einer Regression von $\mathbf{x}_{i,j}$ auf alle anderen Regressoren (inklusive der Konstanten) ist und $SST_j = \sum_{i=1}^n \mathbf{x}_{i,j}^2$.

Interpretation der Varianz:

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

- Die Varianz hängt direkt proportional von der Varianz der Fehler ab.
- Die Varianz hängt indirekt proportional von SST_j ab: je mehr Variation in dem Regressor steckt, desto genauer kann der Koeffizient geschätzt werden.
- Die Varianz hängt indirekt proportional von R_j^2 ab: je mehr der Variation des Regressors bereits durch die anderen Regressoren erklärt wird, desto geringer ist die neue Information im betrachteten Regressor.
- Die Varianz hängt indirekt proportional von der Stichprobengröße ab: je größer die Stichprobe, desto kleiner die Varianz.



Theorem (Gauss-Markov)

Unter den Annahmen MRL.1-MLR.5 ist der KQ-Schätzer der beste lineare unverzerzte Schätzer (BLUS; englisch: best linear unbiased estimator BLUE).

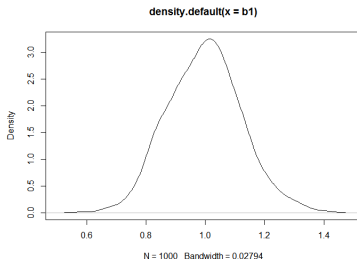
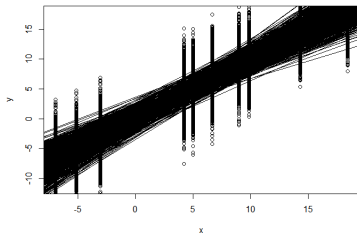
Dieses Resultat zeigt, dass der KQ-Schätzer optimal ist, wenn wir uns auf:

- unverzerzte
- lineare: im Sinne einer linearen Funktion von \mathbf{y}_i

beschränken.

Visualisierung der Ergebnisse

$$M = 1000, T = 10 : \mathbf{y}_i = \mathbf{x}_i + \mathbf{u}_i$$



Zusammenfassung aus Einführung in Ökonometrie

Schätzung der Residualvarianz:

Theorem (unverzerrte Schätzung der Fehlervarianz)

Unter den Annahmen MLR.1-MLR.5 kann die Varianz $\text{Var}(\mathbf{u})$ unverzerrt geschätzt werden durch

$$\hat{\sigma}^2 = \frac{SSR}{n - K - 1} = \frac{\sum_{i=1}^n \hat{u}_i^2}{n - K - 1}$$

das heißt $\mathbb{E} \hat{\sigma}^2 = \sigma^2$.

Das ermöglicht es uns, die Varianz der Parameterschätzer unverzerrt zu schätzen:

$$\widehat{\text{Var}}(\hat{\beta}) = \hat{\sigma}^2 (\tilde{X}' \tilde{X})^{-1}$$

Zusammenfassung aus Einführung in Ökonometrie

Für Inferenz im KQ-Setting verwenden wir noch folgende Annahme:

Assumption (MLR.6 (Normalität der Fehler))

Die Fehler \mathbf{u} sind unabhängig von den Regressoren \mathbf{X} und normalverteilt mit Mittel null und Varianz σ^2 .

MLR.6 ist stärker als MLR.4 und MLR.5.

Assumption (KLR (Klassisches lineares Regressionsmodell; Englisch: classical linear model: CLM))

Die Annahmen des klassischen linearen Regressionsmodells (KLR) sind MLR.1-MLR.6.

Zusammenfassung aus Einführung in Ökonometrie

Theorem

Unter den Annahmen KLR sind die Schätzer $\hat{\beta}_j$ bedingt auf die Regressoren normalverteilt mit Mittel null und Varianz $\text{Var}(\hat{\beta}_j) = \sigma^2 / (SST_j(1 - R_j^2))$, das heißt

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \text{Var}(\hat{\beta}_j))$$

Weiters ist (mit $\widehat{\text{Var}}(\hat{\beta}_j) = \hat{\sigma}^2 / (SST_j(1 - R_j^2))$)

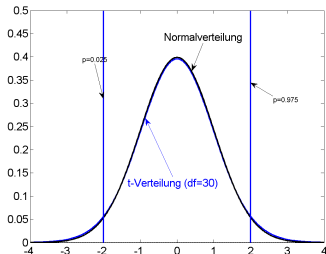
$$(\hat{\beta}_j - \beta_j) / \sqrt{\widehat{\text{Var}}(\hat{\beta}_j)} \sim t_{n-k-1}$$

wobei t_{df} die t -Verteilung mit df Freiheitsgraden bezeichnet.

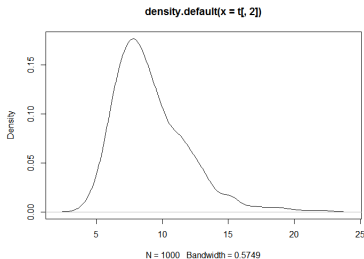
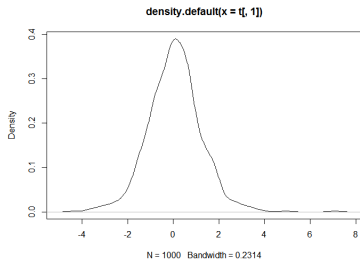
Zusammenfassung aus Einführung in Ökonometrie

Dieses Theorem ermöglicht Inferenz, genauer einen sogenannten t-Test:

- Nullhypothese: $H_0 : \beta_j = 0$
- Gegenhypothese: $H_1 : \beta_j \neq 0$
- Teststatistik: $t : \hat{\beta}_j / \widehat{\text{Var}}(\hat{\beta}_j)$
- Test mit Konfidenzniveau $1 - \alpha$ wird verworfen, wenn $|t| > t_{1-\alpha/2, n-K-1} \approx 2$



Zusammenfassung aus Einführung in Ökonometrie



Zusammenfassung aus Einführung in Ökonometrie

Simultane Tests auf mehrere Koeffizienten:

- Nullhypothese: $H_0 : \beta_1 = \beta_2 = \dots = \beta_K = 0$
- Gegenhypothese: mindestens ein Koeffizient ist nicht null.
- Teststatistik:

$$F = \hat{\beta}' \tilde{X}' \tilde{X} \hat{\beta} / (K \hat{\sigma}^2) = \frac{R^2 / K}{(1 - R^2) / (n - K - 1)}$$

- F verteilt nach F -Verteilung mit K und $n - K - 1$ Freiheitsgrade.
- Wir verwerfen H_0 , wenn F zu groß.
- Entscheidung wird anhand des p-Wertes getroffen: p-Wert größer als 0.05 (bei 95% Konfidenz): H_0 wird nicht verworfen.