Modul 4: Tests und Inferenz

PI 6250 – Ökonometrie I

Max Heinze (mheinze@wu.ac.at)
Department für Volkswirtschaftslehre, WU
Wien

8. Mai 2025

Einführung

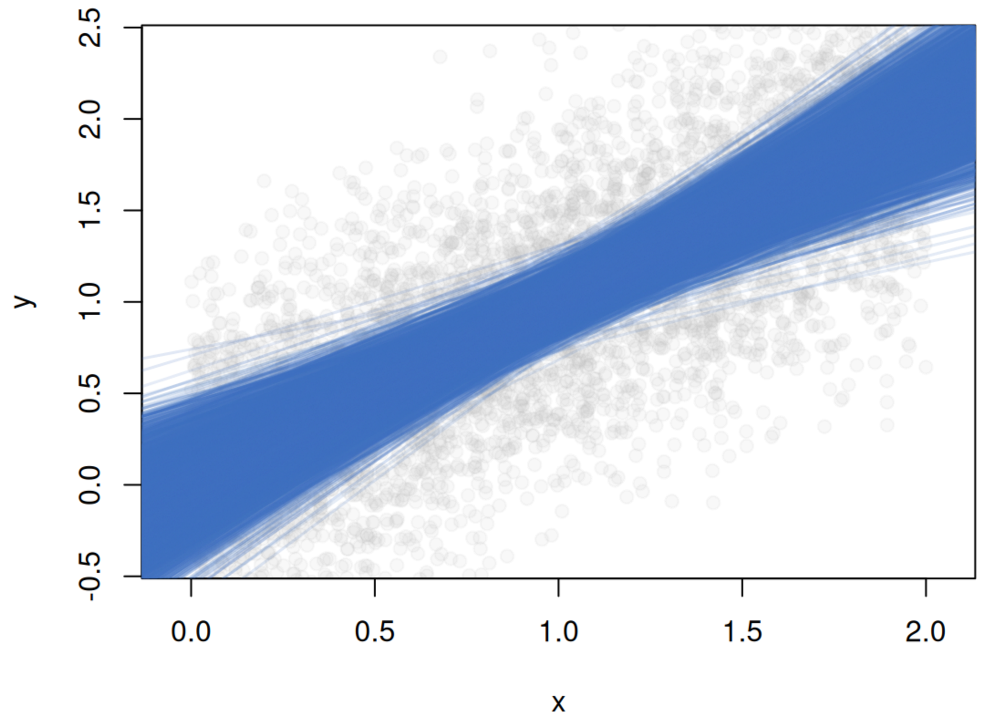
Kleine Stichproben

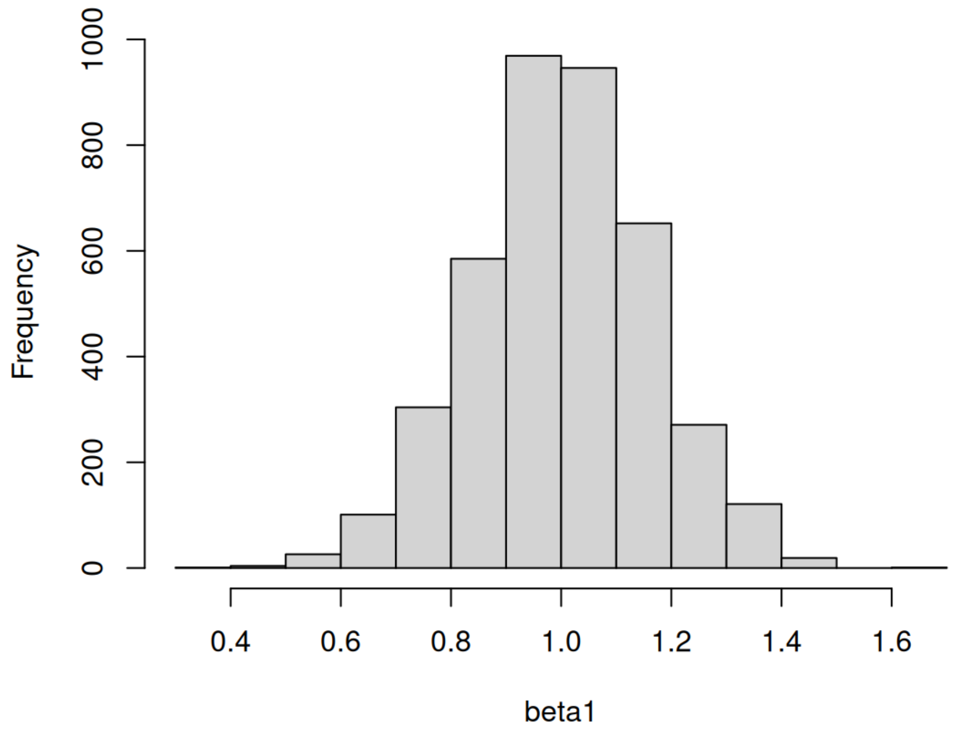
t-Test

F-Test

Motivation

In **Modul 2** (und 3) haben wir näher betrachtet, was es heißt, dass unser OLS-Schätzer eine **Zufallsvariable** ist. Wir haben **Erwartungswert** und **Varianz** des Schätzers bestimmt und auch folgende Simulation durchgeführt:





Motivation

Für alles, was wir in diesem Kapitel besprechen, brauchen wir aber mehr als nur die beiden Momente Erwartungswert und Varianz. Wir müssen uns fragen: **Was ist die Stichprobenverteilung des OLS-Schätzers?**

Wofür brauchen wir Information über diese Verteilung? In **Modul 1** haben wir gesagt:

> Um eine **Hypothese** mit **Daten** überprüfen zu können, brauchen wir **Daten** und eine **Hypothese**.

- Idealerweise haben wir eine **Theorie**, aus der wir eine **falsifizierbare Hypothese** ableiten können.
- Dann können wir versuchen, diese **Hypothese** empirisch zu **testen**.

Hypothesentests

Aber wie **testen** wir eine **Hypothese**? Angenommen, wir wollen wissen, ob der Parameter β_1 ungleich Null ist, also ob die entsprechende Variable x_1 einen Effekt auf y hat.

- **Erste Idee**: Wir schätzen unser Modell mit OLS und schauen, ob der Absolutbetrag der Schätzung $|\hat{\beta}_1| > 0$ ist.
  - Diese **Idee** ist eine **schlechte Idee**.
  - **Intuition**: Wir wissen, dass wir eine gewisse **Unsicherheit** in unserer Schätzung haben. Wenn unsere Schätzung z.B. nahe bei Null ist und/oder die Unsicherheit groß ist – wie „sicher" können wir uns da sein, dass unsere Schätzung nicht rein zufällig nicht Null ist?
- **Bessere Idee**: Wir ***nehmen an***, dass das wahre β_1 **gleich Null** ist und versuchen herauszufinden, was die **Wahrscheinlichkeit** ist, dass wir dann trotzdem **die** Schätzung bekommen, die wir erhalten haben.
  - Wenn diese Wahrscheinlichkeit klein ist, können wir sagen, dass es **unwahrscheinlich** ist, so eine Schätzung zu erhalten, **wenn** der wahre Parameter $\beta_1 = 0$ ist.

Hypothesentests

Das, was wir auf der vorherigen Folie besprochen haben, nennen wir einen **Hypothesentest** (engl. **hypothesis test**). Etwas formeller:

(1) Wir stellen eine sogenannte **Nullhypothese** (engl. **null hypothesis** oder kurz **null**) auf:

$$H_0 : \beta_1 = 0.$$

Daraus ergibt sich auch eine Alternativhypothese:

$$H_A : \beta_1 \neq 0.$$

(2) Wir nehmen an, dass die **Nullhypothese stimmt**, und berechnen, was in diesem Fall die Wahrscheinlichkeit ist, **die Schätzung $\hat{\beta}_1$ zu erhalten**.

(3) Wenn diese Wahrscheinlichkeit **ausreichend gering** ist, **verwerfen** wir die **Nullhypothese**.

Null- und Alternativhypothese

Warum ist unsere Nullhypothese $\beta_1 = 0$ und nicht $\beta_1 \neq 0$?

- Einerseits erlauben uns klassische statistische Tests nur, zu **testen, ob β_0 ein bestimmter Wert** ist, zum Beispiel 0.
  - Wir haben gesagt, wir nehmen an, die Nullhypothese stimmt, und berechnen dann die Wahrscheinlichkeit, ein bestimmtes $\hat{\beta}_1$ unter dieser Nullhypothese zu erhalten.
  - Wenn die Nullhypothese $\beta_1 = 0$ ist, dann ist das sinnvoll und intuitiv zu verstehen. Wenn $\beta_1 = 0$, dann ist es wahrscheinlicher, $\hat{\beta}_1 = 1$ zu erhalten, als $\hat{\beta}_1 = 5$.
  - Würden wir als Nullhypothese $\beta_1 \neq 0$ haben, wäre das gar nicht möglich. Die besprochene Wahrscheinlichkeit wäre völlig anders für $\beta_1 = 12$, $\beta_1 = 0.000000001$ und $\beta_1 = -10^6$.

Verwerfen heißt nicht das Gegenteil bestätigen

Warum ist unsere Nullhypothese $\beta_1 = 0$ und nicht $\beta_1 \neq 0$?

- Außerdem können wir mithilfe statistischer Tests **eine Hypothese nie bestätigen**, sondern **nur verwerfen** (engl. **reject a hypothesis**).
    - Wir wollen herausfinden, ob x_1 einen Einfluss auf y hat.
    - Wenn unsere **Nullhypothese** ist, dass sie **keinen Einfluss** hat ($\beta_1 = 0$), dann gibt uns die **Verwerfung dieser Hypothese** einen wichtigen Hinweis in die Richtung, dass die Variable Einfluss hat.
    - Wir können aber **nie bestätigen**, dass eine Variable Einfluss hat, wir können nur die Hypothese **verwerfen**, dass sie keinen Einfluss hat.
    - Das liegt daran, dass wir uns beim Verwerfen einer Hypothese mit einer **ausreichend geringen Wahrscheinlichkeit** begnügen, diese Wahrscheinlichkeit aber nie 0 ist.

Auf jeden Fall brauchen wir für dieses Testprocedere **Informationen über die Stichprobenverteilung von $\hat{\beta}_1$**, also beschäftigen wir uns erst einmal damit, bevor wir zu Hypothesentests zurückkehren.

Einführung

Kleine Stichproben

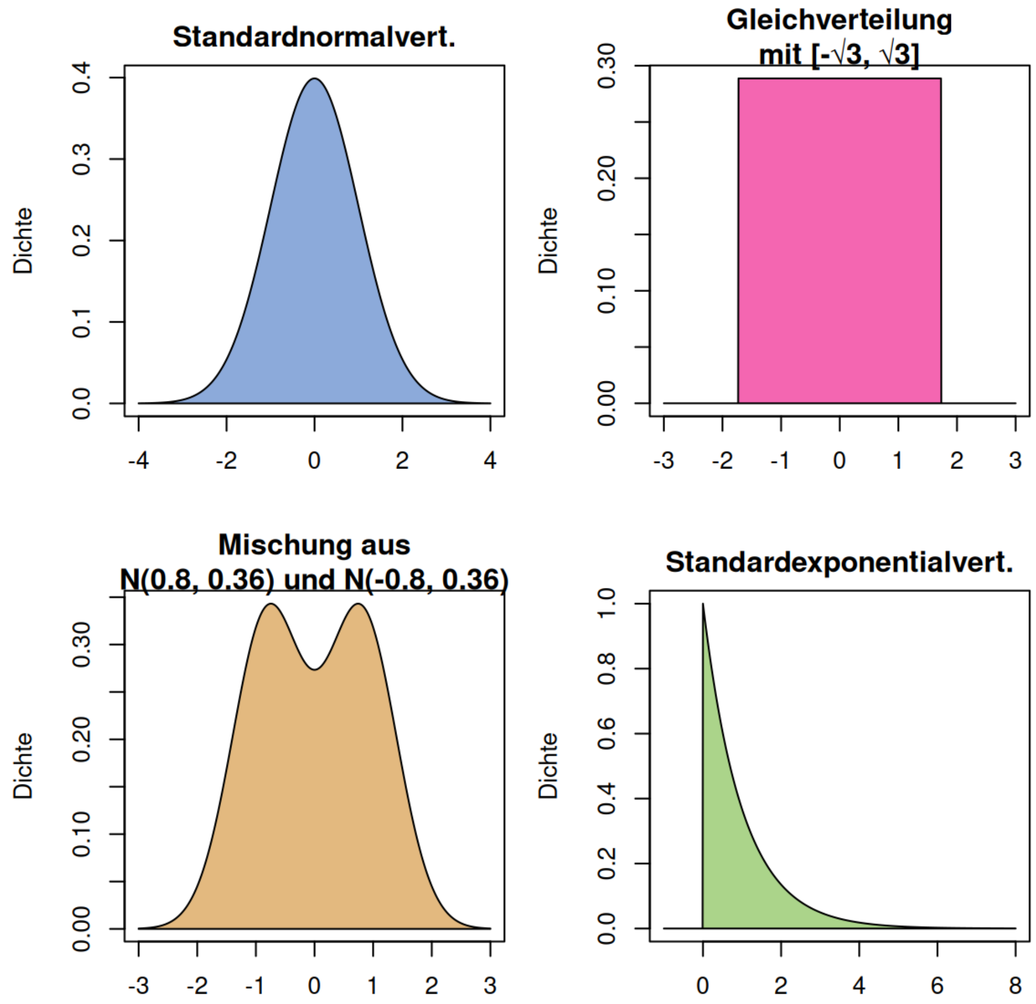
t-Test

F-Test

Interpretation von Regressionstabellen

Momente vs. Verteilung

Wir haben mithilfe der **Annahmen MLR.1 bis MLR.5** Aussagen über den Erwartungswert und die Varianz des OLS-Schätzers treffen können.



- Das reicht uns aber nicht, um Aussagen über die **Verteilung** zu treffen.
- Ein Beispiel: Die vier Verteilungen links haben alle einen Erwartungswert von 0 und eine Varianz von 1.
- Auch unter den **Gauß-Markov-Annahmen** kann die Verteilung von $\hat{\beta}_1$ sehr verschiedene Formen annehmen.
- Wir benötigen daher eine **weitere Annahme**.

(MLR.6) Normalität

Der Fehlerterm der Grundgesamtheit ist unabhängig von den erklärenden Variablen $x_1, \ldots, x_K$ und ist normalverteilt mit Erwartungswert 0 und Varianz σ^2:

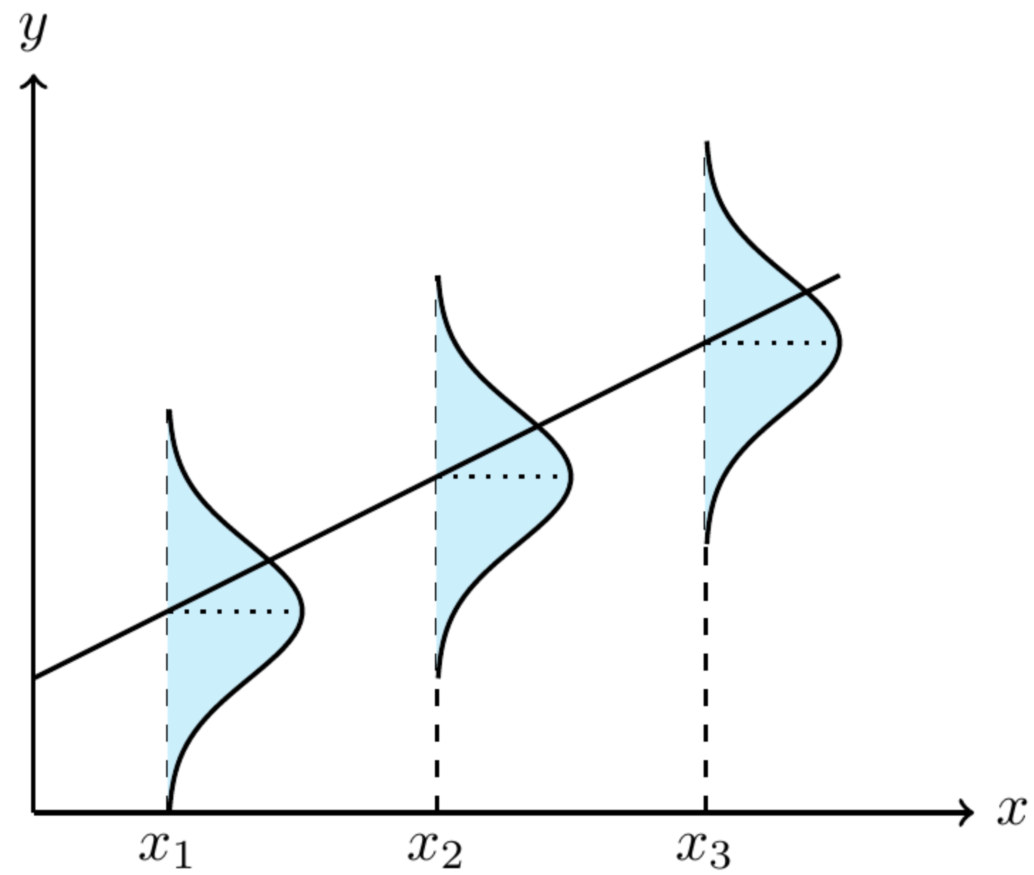
$$u \sim \mathrm{N}\left(0, \sigma^2\right)$$

- Diese Annahme impliziert die Annahmen MLR.4 und MLR.5. Wir sprechen trotzdem von MLR.1 bis MLR.6, um zu verdeutlichen, dass wir MLR.6 „zusätzlich“ annehmen.
- Diese Annahme ist eine **extrem starke Annahme**. Mehr dazu gleich.
- Wir nennen MLR.1 bis MLR.6 zusammen auch **Classical-Linear-Model-Annahmen** (**CLM**-Annahmen).
- Unter den CLM-Annahmen ist OLS nicht nur BLUE, sondern **BUE** (nicht auf lineare Schätzer beschränkt).

(MLR.6) Normalität

Wir können die CLM-Annahmen bezüglich der Grundgesamtheit so zusammenfassen:

$$y \mid \boldsymbol{x} \sim \mathrm{N}\left(\boldsymbol{x}'\boldsymbol{\beta}, \sigma^2\right).$$



- Die Grafik links illustriert diesen Fakt für den **bivariaten Fall** (die Subskripte sind also i, nicht k).
- Unter den CLM-Annahmen sind die y für eine Beobachtung i normalverteilt mit
  - Mittelwert $\boldsymbol{x}'\boldsymbol{\beta}$ (im bivariaten Fall links $\beta_1 x$) und
  - immer der gleichen Varianz σ^2.

Macht MLR.6 Sinn?

- Wie vorher erwähnt, ist $u \sim \mathrm{N}\left(0, \sigma^2\right)$ eine sehr **starke Annahme**. Können wir diese Annahme rechtfertigen?
- Ein Argument: Der Fehlerterm u ist eine **Summe aus vielen unbeobachteten Faktoren**, die y beeinflussen. Daher kann der **zentrale Grenzwertsatz** (nächste Folie) angewendet werden und u ist **annähernd normalverteilt**.
  - Allerdings können die verschiedenen Faktoren in u sehr **unterschiedliche Verteilungen** haben, was die Approximation verschlechtert.
  - Außerdem garantiert uns nichts, dass die einzelnen Faktoren **additiv** im Fehlerterm auftreten. Das ist noch ein wesentlich größeres Problem.
- Später besprechen wir, warum Nichtnormalität der Fehler **in größeren Stichproben** kein großes Problen darstellt. Bis auf weiteres nehmen wir einfach **Normalität** an.
  - Manchmal werden in kleineren Stichproben **Transformationen** (z.B. Logarithmieren) angewandt, damit die y-Werte näher an einer Normalverteilung liegen.

Zentraler Grenzwertsatz / Central Limit Theorem

Der **Zentrale Grenzwertsatz** (engl. **central limit theorem**, **CLT**) besagt:

Sei $\{X_1, X_2, \ldots, X_N\}$ eine Folge von unabhängig und identisch verteilten Zufallsvariablen mit Mittelwert μ und Varianz σ^2. Dann konvergiert die Verteilungsfunktion der standardisierten Zufallsvariable

$$Z_N = \frac{\bar{X}_N - \mu}{\sigma / \sqrt{N}},$$

wobei $\bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i$, in Verteilung gegen die Verteilungsfunktion der **Standardnormalverteilung**.

- Z_n ist eine standardisierte Version des Stichprobenmittelwerts.
- Intuitiv: Wenn N größer wird, konvergiert die Verteilung des Mittelwerts der X_i gegen eine Normalverteilung.

Verteilung des OLS-Schätzers, Teil 1

Unter den **CLM-Annahmen** MLR.1 bis MLR.6 ist der OLS-Schätzer, gegeben die Stichprobenwerte der unabhängigen Variablen normalverteilt:

$$\hat{\beta}_k \sim \mathrm{N}(\beta_k, \mathrm{Var}(\hat{\beta}_k)),$$

wobei $\mathrm{Var}(\hat{\beta}_k) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_{ik} - \bar{x}_k)^2} \times \frac{1}{1-R_k^2}$, wo wiederum R_k^2 das R^2 einer Regression von x_k auf alle anderen Regressoren $x_j, j \neq k$ ist.

- $\hat{\beta}_k$ ist **normalverteilt**, jede Linearkombination der $\hat{\beta}_k$ ist ebenso normalverteilt, und die gemeinsame Verteilung einer Teilmenge der $\hat{\beta}_j$ ist eine multivariate Normalverteilung.
- Der **standardisierte Koeffizient** $(\hat{\beta}_k - \beta_k)/\mathrm{sd}(\hat{\beta}_k)$ ist standardnormalverteilt.

Einführung

Kleine Stichproben

t-Test

F-Test

Interpretation von Regressionstabellen

Große Stichproben

Hypothesen über den OLS-Schätzer

Wir beschäftigen uns in diesem Abschnitt damit, **Hypothesen über einen der Parameter** des Populations-Modells

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u$$

zu testen.

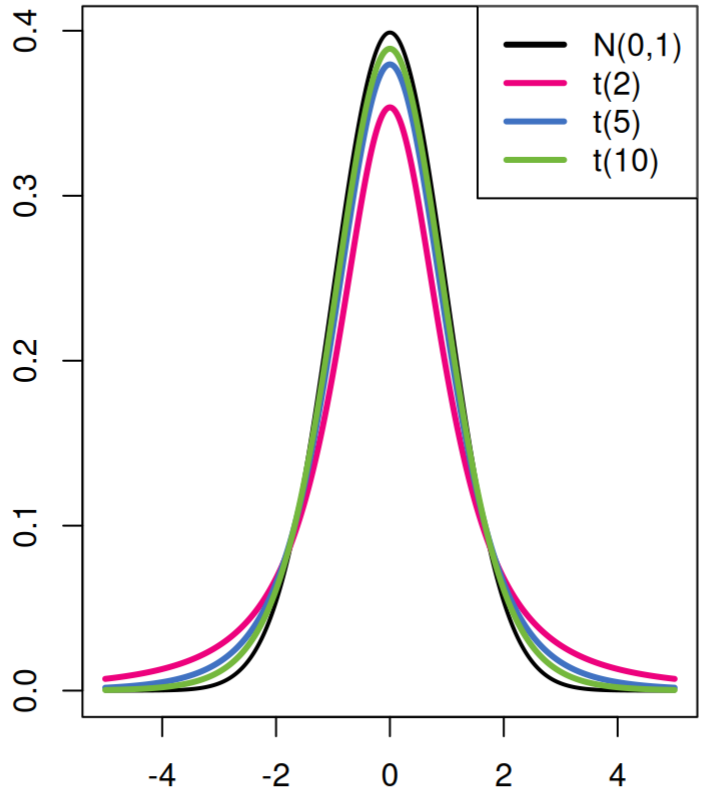
- Nach wie vor gilt: Wir kennen die β_k nicht. Wir können sie nur schätzen.
- Aber wir können **Hypothesen** über die β_k anstellen.
- Im nächsten Schritt können wir statistische Inferenz benutzen, um diese **Hypothesen zu testen**.

Verteilung des OLS-Schätzers, Teil 2

Unter den **CLM-Annahmen** MLR.1 bis MLR.6 gilt:

$$(\hat{\beta}_k - \beta_k)/\mathrm{se}(\hat{\beta}_k) \sim \mathrm{t}_{N-K-1}$$

- Wenn wir im standardisierten Koeffizienten $\mathrm{sd}(\cdot)$ durch $\mathrm{se}(\cdot)$ (also σ durch $\hat{\sigma}$) ersetzen, ist er nicht mehr standardnormalverteilt, sondern folgt einer **t-Verteilung** (engl. **t distribution**) mit $N-K-1$ **Freiheitsgraden** (engl. **degrees of freedom**).



- Die **t-Verteilung** schaut einer **Standardnormalverteilung** sehr ähnlich, sie hat aber **fettere Ränder** (engl. **fatter tails**). Je mehr Freiheitsgrade die Verteilung hat, desto eher kann sie durch eine Normalverteilung approximiert werden.

Nullhypothese und t-Statistik

Wir spezifizieren folgende **Nullhypothese**:

$$H_0 : \beta_k = 0$$

Nachdem wir alle $x_j, j \neq k$ berücksichtigt haben, hat x_k **keinen Einfluss** auf y.

Wir können diese Nullhypothese mit folgender **Teststatistik** testen:

$$t_{\hat{\beta}_k} = \frac{\hat{\beta}_k - \beta_k}{\mathrm{se}(\hat{\beta}_k)}.$$

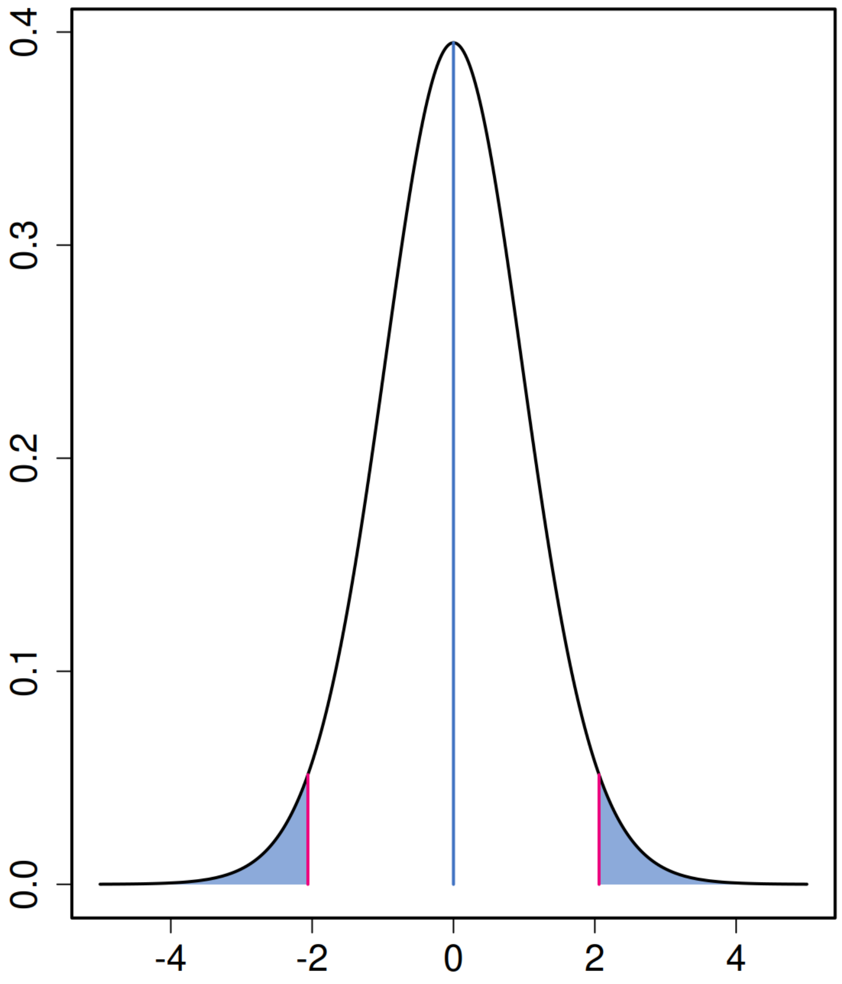
Diese bestimmte **Teststatistik** nennen wir **t-Statistik**.

Unter der Nullhypothese ist $\beta_k = 0$ und die **t-Statistik** ist

$$t_{\hat{\beta}_k} = \frac{\hat{\beta}_k}{\mathrm{se}(\hat{\beta}_k)}.$$

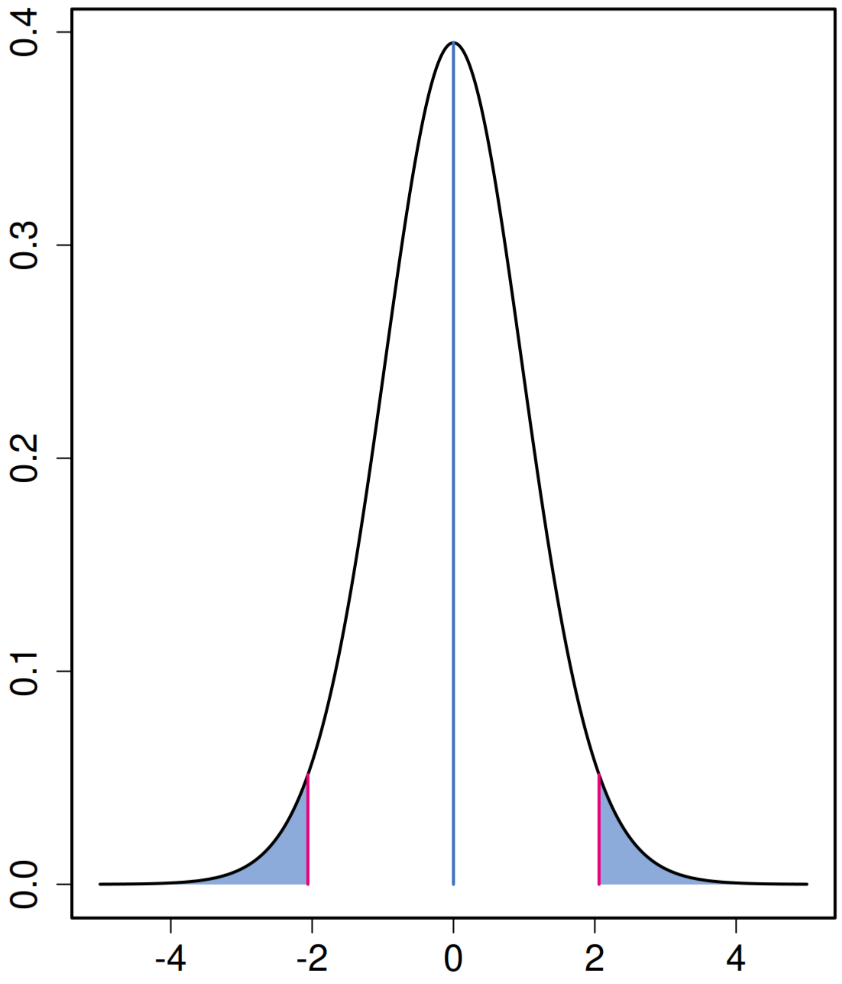
Diese **t-Statistik** ist **t-verteilt** mit Mittelwert 0 und $N - K - 1$ Freiheitsgraden.

Zweiseitige Hypothesentests



- Wir testen die Nullhypothese, $\beta_k = 0$, gegen eine **zweiseitige Alternative**, $\beta_k \neq 0$.
- In der Grafik links ist eine t-Verteilung mit 25 Freiheitsgraden gezeichnet.
- **Wenn die Nullhypothese stimmt**, dann sollten **die t-Statistiken der Schätzer, die wir bekommen** so verteilt sein wie in der Grafik links.
- Die Idee ist: Wenn unsere **tatsächliche t-Statistik** so „extrem“ (also so groß oder so klein) ist, dass sie in den blauen **Ablehnbereichen** dieser Verteilung ist, dann sehen wir es als **unwahrscheinlich an**, dass die Nullhypothese stimmt und **verwerfen sie**.

Wann verwerfen wir die Nullhypothese?

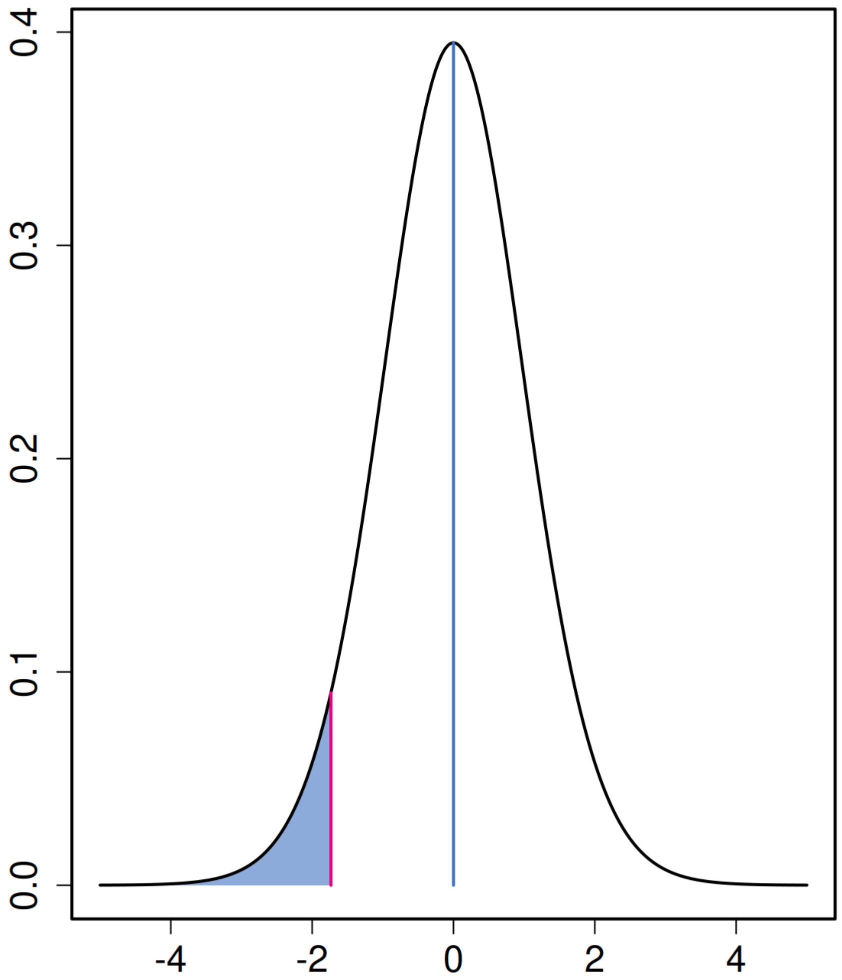


- In der Grafik links hat der **Ablehnbereich**, in denen wir die Nullhypothese verwerfen, insgesamt eine **Fläche von 0.05**. Wir nennen 0.05 das **Signifikanzniveau**.
- Bei einer t-Verteilung mit 25 Freiheitsgraden ergibt das **kritische Werte von -2.06 und 2.06**.
- Wenn der Absolutbetrag der t-Statistik also größer als 2.06 ist, verwerfen wir die Nullhypothese.
- Dass der Schwellenwert hier 2.06 ist, hängt ab von
  - der Anzahl der **Freiheitsgrade** und
  - dem **Signifikanzniveau**.

Was ist ein Signifikanzniveau?

- Wir haben ein **Signifikanzniveau von 0.05** (oder 5%) gewählt. Das bedeutet,
    - dass wir, **wenn die Nullhypothese stimmt**, sie in **5% der Fälle fälschlicherweise verwerfen**;
    - weil es unter der Nullhypothese eine Wahrscheinlichkeit von 5% gibt, dass die t-Statistik betragsmäßig größer als 2.06 ist; wir die Nullhypothese in diesen Fällen aber immer verwerfen.
    - Das nennen wir einen **Typ-1-Fehler** (engl. **type 1 error**), oder auch falsch-positives Ergebnis. Die Wahrscheinlichkeit für so einen Fehler setzen wir selber mit dem Signifikanzniveau.
    - Die Wahrscheinlichkeit für einen **Typ-2-Fehler** (engl. **type 2 error**), ein falsch-negatives Ergebnis (wir verwerfen die Nullhypothese nicht, obwohl sie falsch ist), ist schwieriger zu bestimmen.
- **0.05** ist das am häufigsten verwendete **Signifikanzniveau**; andere häufig verwendete Niveaus sind **0.10**, **0.025**, **0.01**, **0.001**, ...
    - Die **kritischen Werte** für ein bestimmtes Niveau und eine bestimmte Anzahl Freiheitsgrade können wir einer Tabelle oder einem Statistikprogramm entnehmen.

Spezielle Hypothesen testen

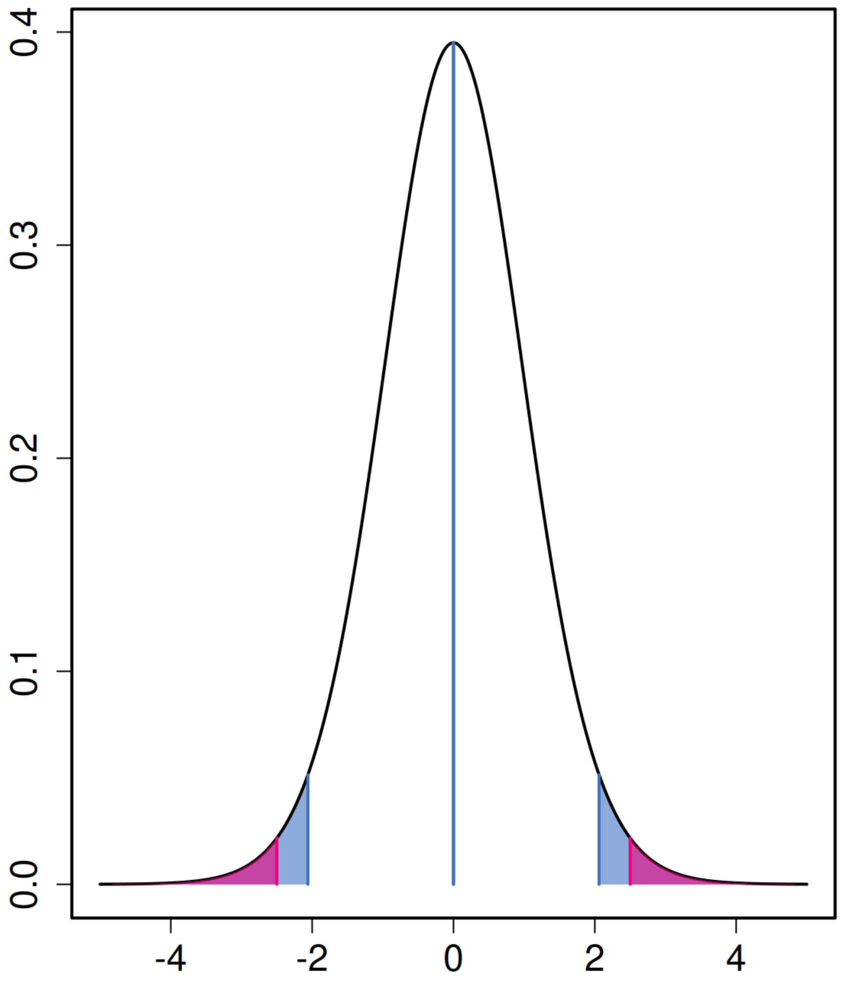


- Wir können auch einen **einseitigen t-Test** durchführen, z.B. mit

$$H_0 : \beta_k \geq 0, \qquad H_A : \beta_k < 0.$$

- Bei einem einseitigen Test ist der gesamte Ablehnbereich auf einer Seite, und der kritische Wert für dasselbe Signifikanzniveau und dieselben Freiheitsgrade ist anders.
- Wir können auch sowohl **ein- als auch zweiseitige Tests** mit **anderen Nullhypothesen** durchführen, z.B. $H_0 : \beta_k = 1$. Die Verteilung der t-Statistik ändert sich nicht.

p-Werte



- Angenommen, wir führen einen Test wie vorher $(\alpha = 0.05, \mathrm{df} = 25)$ durch und erhalten eine t-Statistik von $t = 2.5$.
- In der Grafik links ist der Bereich **mit „extremeren" t-Statistiken als 2.5** (also $|t| > 2.5$) **pink markiert**.
- Die **Wahrscheinlichkeit, eine „extremere" t-Statistik** als 2.5 zu erhalten, ist **0.019**. Das ist also auch die **Fläche beider pinken Bereiche**.
- Wir nennen diese Größe **p-Wert**. Der p-Wert erleichtert uns die Interpretation: Wir müssen keinen kritischen Wert wissen, sondern nur vergleichen, ob der p-Wert kleiner als das Signifikanzniveau ist. Dann verwerfen wir die Nullhypothese.

p-Werte und Signifikanz: Interpretation

- Wenn der **p-Wert** kleiner als das von uns gesetzte **Signifikanzniveau** ist, dann **verwerfen wir die Nullhypothese.**
- Angenommen, der zu β_2 gehörige p-Wert ist 0.03 und das Signifikanzniveau ist 0.05. Dann können wir **sagen**:
    - x_2 ist **statistisch signifikant** (engl. **statistically significant**) bei einem Signifikanzniveau von 0.05 (bzw. 5%).
    - β_2 ist *statistisch signifikant verschieden von Null* bei einem Signifikanzniveau von 0.05.
    - Wir **verwerfen die Nullhypothese** bei einem Signifikanzniveau von 0.05.
    - Bei einem Signifikanzniveau von 3% wäre der Test indifferent zwischen Verwerfen und Nicht-Verwerfen der Nullhypothese.
- Folgende Dinge sind **falsch** und wir können sie **nicht sagen**:
    - Wir nehmen die Alternativhypothese an.
    - Die Wahrscheinlichkeit, dass die Nullhypothese stimmt, ist 3%.
    - [...] bei einem Signifikanzniveau von 0.95.
    - Wir sind uns zu 97% sicher, dass x_2 einen Einfluss hat.

Statistische vs. „Ökonomische“ Signifikanz

- Wir haben uns bisher damit beschäftigt, wann eine Variable **statistisch signifikant** ist.
    - Statistische Signifikanz hängt alleine von der zu einem Koeffizienten gehörigen **t-Statistik** ab.
- Ein weiteres, bei der Interpretation wichtiges Konzept ist **ökonomische** bzw. **praktische Signifikanz**.
    - Die Idee ist: Nicht jede Variable, die statistisch signifikant ist, ist auch ein wichtiger Einflussfaktor auf y.
    - Wir beginnen damit, statistische Signifikanz zu überprüfen.
    - Wenn eine Variable statistisch signifikant ist, können wir als nächstes die (absolute) Größe des Koeffizienten überprüfen.
    - Wenn der Koeffizient sehr nahe bei Null ist, dann übt die Variable wenig Effekt auf y aus, auch wenn sie statistisch signifikant ist.
    - Eine Variable, die statistisch signifikant ist und einen großen Effekt hat, können wir als „statistisch und ökonomisch signifikant“ interpretieren.
- Bei der **Interpretation** ist es also immer wichtig, auch die **Größe** des Koeffizienten zu beachten.

Konfidenzintervalle

Unter den **CLM-Annahmen** können wir auch ein **Konfidenzintervall** für einen Parameter der Grundgesamtheit β_k berechnen. Wir besprechen das am Beispiel eines 95 %-Konfidenzintervalls.

- Wir können ein 95 %-Konfidenzintervall so **interpretieren**: Wenn wir wieder und wieder eine Stichprobe ziehen und das Konfidenzintervall berechnen, wird dieses Konfidenzintervall in 95% der Fälle den wahren Parameter abdecken.
- Wir können **nicht** sagen, dass der Parameter (der Grundgesamtheit) in 95% der Fälle in das Konfidenzintervall fällt, da sich das Konfidenzintervall ändert, und **nicht** der Parameter.
- Das 95-Prozent-Konfidenzintervall für einen Parameter β_k ist:

$$\left[\hat{\beta}_k - c \times \mathrm{se}(\hat{\beta}_k), \quad \hat{\beta}_k + c \times \mathrm{se}(\hat{\beta}_k)\right],$$

wobei c das 97.5-te Perzentil einer t_{N-K-1}-Verteilung ist.

Einführung

Kleine Stichproben

t-Test

F-Test

Interpretation von Regressionstabellen

Große Stichproben

Wie viele Restriktionen wollen wir testen?

Mit dem **t-Test** konnten wir unserem Modell eine einzige **Restriktion** auferlegen, z.B.

$$\beta_1 = 0,$$

und diese Restriktion testen.

Was ist aber, wenn wir **mehrere Restriktionen gemeinsam** testen wollen? Wir können zum Beispiel daran interessiert sein, ob eine bestimmte Menge von unabhängigen Variablen vielleicht als ganzes keinen Effekt auf y hat:

$$\beta_1 = 0, \beta_2 = 0, \beta_3 = 0.$$

Um solche Restriktionen testen zu können, benötigen wir einen anderen Test, den **F-Test**.

Mehrere Restriktionen: Hypothesen

$$\beta_1 = 0, \beta_2 = 0, \beta_3 = 0.$$

Die Nullhypothese und Alternativhypothese in diesem Fall sind:

$$H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0; \qquad H_A : H_0 \text{ ist nicht wahr.}$$

- In diesem Fall testen wir **drei Ausschluss-Restriktionen** (engl. **exclusion restrictions**), wir testen also **mehrere Hypothesen** (engl. **multiple hypotheses**) gleichzeitig.
- Da wir die Hypothesen **gleichzeitig** testen, können wir nichts mit den separaten t-Statistiken für die einzelnen Parameter anfangen.
- Wir benötigen also eine **andere Teststatistik**, deren Verteilung wir kennen, um so einen Test durchzuführen.

Vollständiges und restringiertes Modell

Wir beginnen damit, unser **vollständiges** (unrestringiertes, engl. **unrestricted**) Modell anzuschreiben, zum Beispiel:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + u.$$

Dann wenden wir alle **Restriktionen** an und erhalten das **restringierte Modell** (engl. **restricted model**):

$$y = \beta_0 + \beta_4 x_4 + \beta_5 x_5 + u.$$

Wie können wir diese Modelle **vergleichen**?

- Ein Ansatz: Wir betrachten die **Residuenquadratsummen** (SSR) der beiden Modelle.
- Da die SSR *immer* größer wird, wenn wir Variablen aus dem Modell entfernen, suchen wir eine Teststatistik, die evaluiert, **wie stark der relative Anstieg** der SSR ist, wenn wir unsere Restriktionen anwenden.

F-Statistik

Eine solche Teststatistik ist

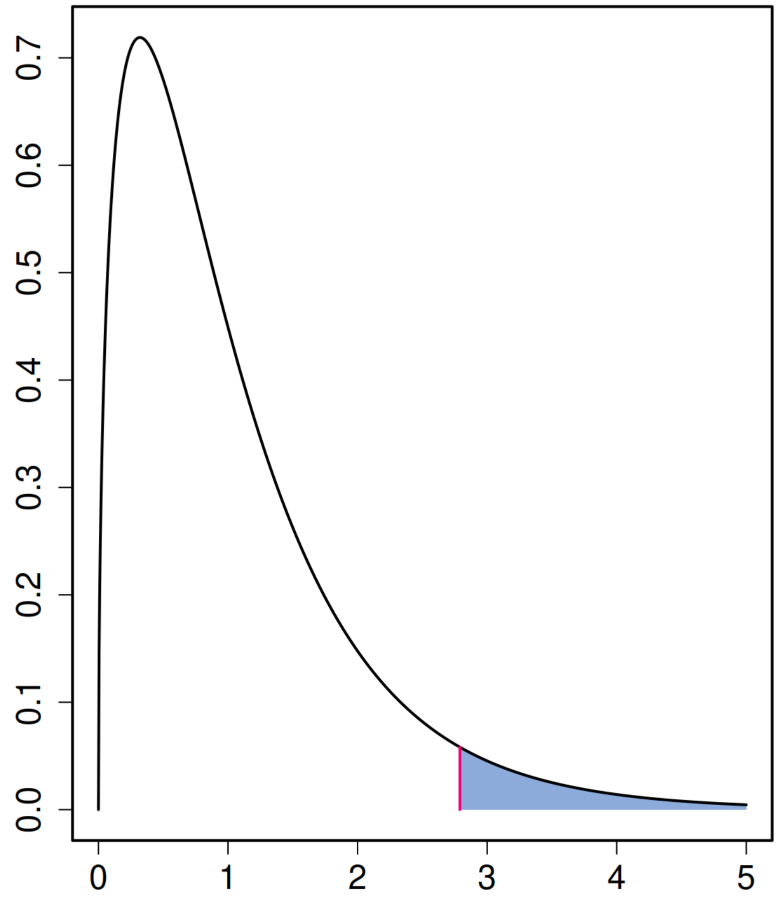
$$F = \frac{(\mathrm{SSR}_r - \mathrm{SSR}_{ur})/q}{\mathrm{SSR}_{ur}/(N-K-1)},$$

wobei q die Anzahl der Restriktionen ist, die wir auferlegen.

- Unter den CLM-Annahmen folgt diese Teststatistik, die **F-Statistik**, einer F-Verteilung mit q Freiheitsgraden im Zähler und $(N-K-1)$ Freiheitsgraden im Nenner.
- Die Teststatistik ist **immer größer als Null**.
- Eine alternative Schreibweise der F-Statistik mit R^2 statt der SSR ist

$$F = \frac{(R^2_{ur} - R^2_r)/q}{(1-R^2_{ur})/(N-K-1)}.$$

F-Verteilung



- Die Grafik zeigt eine F-Verteilung mit 3 und 50 Freiheitsgraden. Der kritische Wert ist 2.798; wir verwerfen die Nullhypothese, wenn wir eine F-Statistik erhalten, die **größer** als dieser Wert ist (**einseitiger** Test).
- Wenn wir die Nullhypothese nicht verwerfen können, sagen wir, die Variablen sind **gemeinsam nicht signifikant** (engl. **jointly insignificant**).
- Wenn wir die Nullhypothese verwerfen, sagen wir, die Variablen sind **gemeinsam signifikant**.
- Ein F-Test für nur eine Restriktion führt zum gleichen Ergebnis wie ein entsprechender t-Test. Allerdings können mehrere einzeln nicht signifikante Variablen gemeinsam signifikant sein.

Globale F-Statistik

Wenn wir eine Regression rechnen, berechnet das Statistikprogramm für gewöhnlich eine bestimmte Menge von Restriktionen:

$$H_0 : \beta_1 = 0, \beta_2 = 0, \ldots, \beta_K = 0,$$

also, dass **alle unabhängigen Variablen gemeinsam keinen Beitrag leisten**, um y **zu erklären**.

Die **F-Statistik** für diesen Fall kann geschrieben werden als

$$F = \frac{R^2/K}{(1 - R^2)/(N - K - 1)}.$$

Sowohl bei dieser „globalen" als auch bei allen anderen F-Statistiken wird von Statistikprogrammen ein **p-Wert** ausgegeben, der wie bei t-Statistiken die Interpretation erleichtert.

Kleine Stichproben

t-Test

F-Test

Interpretation von Regressionstabellen

Große Stichproben

Huh, ein Baseball-Datensatz ⚾

Wir verwenden den Baseball-Datensatz aus dem Wooldridge-Buch, um uns das Ganze an einem praktischen Beispiel anzusehen.

R Code   Start Over   Run Code

```r
# Pakete laden
library(wooldridge) # Enthält den Datensatz
library(dplyr)
library(car) # Später für F-Test

# Daten laden
data("mlb1") # Baseball-Daten

# Nur die Variablen behalten, die uns interessieren
mlb1 <- mlb1 |>
  select(salary, years, gamesyr, bavg, hrunsyr, rbisyr)
```

Wir schätzen fleißig Regressionen

R Code Start Over Run Code

```r
lm(log(salary) ~ years + gamesyr + bavg + hru
  summary()
```

R Code Start Over Run Code

```r
lm(log(salary) ~ years + gamesyr, data = mlb1
  summary()
```

Signifikant oder nicht signifikant, das ist hier die Frage

R Code Start Over Run Code

```
model <- lm(log(salary) ~ years + gamesyr + bavg + hruns

linearHypothesis(model, c("bavg = 0", "hrunsyr = 0", "rb
```

- **Keine** der drei Variablen `bavg`, `hrunsyr`, und `rbisyr` war für sich selbst genommen **signifikant**.
- Wenn wir aber testen, ob die drei Variablen **gemeinsam signifikant** sind, können wir die Nullhypothese verwerfen.
- Der p-Wert der F-Statistik für das gesamte Regressionsmodell war bei beiden Modellen **sehr klein**.

t-Test

F-Test

Interpretation von Regressionstabellen

Große Stichproben

Was ist eine große Stichprobe?

- Alle Eigenschaften des OLS-Schätzers, die wir bisher besprochen haben, treffen in allen **endlichen Stichproben** zu (engl. **finite sample properties**), egal, wie groß oder klein N ist.
    - Dazu gehören die **Unverzerrtheit** des OLS-Schätzers, das **Gauß-Markov-Theorem**, etc.
    - Dazu gehört auch alles, was wir über die Stichprobenverteilung der OLS-Schätzer, t- und F-Tests besprochen haben ... **solange wir MLR.6 annehmen**.
- Zusätzlich zu diesen Eigenschaften hat OLS bestimmte **Eigenschaften in großen Stichproben** (engl. **large sample properties**).
    - Damit meinen wir Eigenschaften, die auftreten, wenn N **gegen unendlich geht**, also nicht Eigenschaften, die für eine bestimmte Stichprobengröße N (oder gar alle möglichen N) zutreffen.
    - Gewisse Eigenschaften treffen in großen Stichproben auch zu, wenn bestimmte Annahmen nicht erfüllt sind.

Was passiert ohne Annahme MLR.6?

- **Ohne Annahme MLR.6** muss die t-Statistik keiner t-Verteilung folgen, und die F-Statistik muss keiner F-Verteilung folgen. Wir können also Hypothesen über die Parameter nicht wie bisher testen.
- Wir haben aber auch darüber gesprochen, dass MLR.6 **unrealistisch** ist. MLR.6 impliziert auch, dass $y \mid \boldsymbol{x}$ normalverteilt ist. Für bestimmte y-Variablen macht das offensichtlich keinen Sinn.
- Praktischerweise kann mithilfe des Zentralen Grenzwertsatzes für **große Stichproben** folgendes gezeigt werden:

Unter Annahmen **MLR.1 bis MLR.5** ist die t-Statistik **asymptotisch normalverteilt**:

$$\frac{\hat{\beta}_k - \beta_k}{\mathrm{se}(\hat{\beta}_k)} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1) \qquad \text{bzw.} \qquad \frac{\hat{\beta}_k - \beta_k}{\mathrm{se}(\hat{\beta}_k)} \xrightarrow{\mathrm{d}} \mathrm{t}_{N-K-1}.$$

Was passiert ohne Annahme MLR.6?

Unter Annahmen **MLR.1 bis MLR.5** ist die t-Statistik **asymptotisch normalverteilt**:

$$\frac{\hat{\beta}_k - \beta_k}{\mathrm{se}(\hat{\beta}_k)} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1) \qquad \text{bzw.} \qquad \frac{\hat{\beta}_k - \beta_k}{\mathrm{se}(\hat{\beta}_k)} \xrightarrow{\mathrm{d}} \mathrm{t}_{N-K-1}.$$

- Da die t-Verteilung für große Freiheitsgrade in Verteilung gegen eine Standardnormalverteilung konvergiert, können wir sowohl den linken als auch den rechten Ausdruck verwenden.
- Das bedeutet, dass wir die **t-Statistik** genau so verwenden können wie mit MLR.6, sofern unsere Stichprobe ausreichend groß ist.
- Die asymptotische Normalität der OLS-Schätzer impliziert auch, dass die **F-Statistik** in großen Stichproben asymptotisch F-verteilt ist.

LM-Statistik

Der **Lagrange-Multiplier-Test** (**LM**-Test) ist eine Alternative zum F-Test in großen Stichproben.

- Die LM-Statistik ist unter den Annahmen MLR.1 bis MLR.5 asymptotisch χ^2_q-verteilt (die Verteilung in kleinen Stichproben ist unbekannt).
- Wir erhalten die LM-Statistik wie folgt:
  (1) Wir schätzen nur das restringierte Modell.
  (2) Wir nehmen die Residuen aus dieser Regression und regressieren sie wiederum auf alle K unabhängigen Variablen des *vollständigen* Modells.
  (3) Wir berechnen die LM-Statistik als $LM = NR^2$, wobei R^2 das R^2 aus der Regression aus Schritt (2) ist.
- Die Idee ist also: Können die zusätzlichen erklärenden Variablen die **Residuen** aus dem restringierten Modell **erklären**?
- **LM-Test** und **F-Test** führen selten zu unterschiedlichen Resultaten.

Literatur

Wooldridge, J. M. (2020). *Introductory econometrics : a modern approach* (Seventh edition, S. xxii, 826 Seiten). Cengage. https://permalink.obvsg.at/wuw/AC15200792