universität innsbruck

Hypothesentests: Grundlagen

Grundlagen der Ökonometrie

herbert.stocker@uibk.ac.at

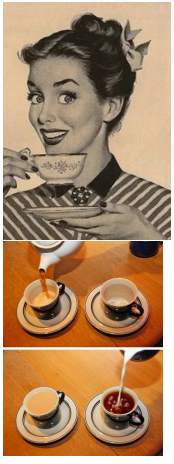
www.hsto.info/econometrics

Lady tasting tea . . .

- Tea party, Südengland, Sonntagnachmittag, vor ca. 100 Jahren . . .

Lady tasting tea . . .

- Tea party, Südengland, Sonntagnachmittag, vor ca. 100 Jahren . . .
- Kann eine Lady (d.h. die Phykologin Muriel Bristol), wie sie behauptet, unterscheiden, ob Milch in den Tee, oder Tee auf die Milch geschüttet wird??? $\rightarrow$ *Let's test her . . .*



Lady tasting tea . . .

- Tea party, Südengland, Sonntagnachmittag, vor ca. 100 Jahren . . .
- Kann eine Lady (d.h. die Phykologin Muriel Bristol), wie sie behauptet, unterscheiden, ob Milch in den Tee, oder Tee auf die Milch geschüttet wird??? $\rightarrow$ *Let's test her . . .*
- R.A. Fisher: The Design of an Experiment, 1935

Lady tasting tea . . .

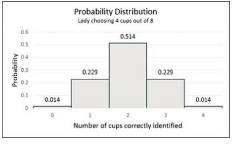
- Tea party, Südengland, Sonntagnachmittag, vor ca. 100 Jahren . . .
- Kann eine Lady (d.h. die Phykologin Muriel Bristol), wie sie behauptet, unterscheiden, ob Milch in den Tee, oder Tee auf die Milch geschüttet wird??? $\rightarrow$ *Let's test her . . .*
- R.A. Fisher: The Design of an Experiment, 1935
- Zufällige Anordnung von 8 Tassen, je vier nach einer Methode zubereitet

Lady tasting tea . . .

- Tea party, Südengland, Sonntagnachmittag, vor ca. 100 Jahren . . .
- Kann eine Lady (d.h. die Phykologin Muriel Bristol), wie sie behauptet, unterscheiden, ob Milch in den Tee, oder Tee auf die Milch geschüttet wird??? $\rightarrow$ *Let's test her . . .*
- R.A. Fisher: The Design of an Experiment, 1935
- Zufällige Anordnung von 8 Tassen, je vier nach einer Methode zubereitet
- Die Lady versucht vier Tassen . . .

Lady tasting tea . . .

- Tea party, Südengland, Sonntagnachmittag, vor ca. 100 Jahren . . .
- Kann eine Lady (d.h. die Phykologin Muriel Bristol), wie sie behauptet, unterscheiden, ob Milch in den Tee, oder Tee auf die Milch geschüttet wird??? $\rightarrow$ *Let's test her . . .*
- R.A. Fisher: The Design of an Experiment, 1935
- Zufällige Anordnung von 8 Tassen, je vier nach einer Methode zubereitet
- Die Lady versucht vier Tassen . . .
- insgesamt $\binom{8}{4} = 70$ Möglichkeiten, nur eine davon ist korrekt (hypergeometr. Verteilung!)



Hypothesentests

Nullhypothese:

- Kann man ‘beweisen’, dass die Lady den Unterschied tatsächlich schmecken kann???

Hypothesentests

Nullhypothese:

- Kann man ‘beweisen’, dass die Lady den Unterschied tatsächlich schmecken kann???
- $\Rightarrow$ es ist immer möglich, dass sie nur ‘zufällig’ die richtige Lösung erriet, aber die *‘Wahrscheinlichkeit’* ist 1/70, kleiner als 5%!

Hypothesentests

Nullhypothese:

- Kann man ‘beweisen’, dass die Lady den Unterschied tatsächlich schmecken kann???
- $\Rightarrow$ es ist immer möglich, dass sie nur ‘zufällig’ die richtige Lösung erriet, aber die *‘Wahrscheinlichkeit’* ist 1/70, kleiner als 5%!
- R.A. Fisher: ***Null hypothesis*** *is “never proved or established, but is possibly disproved, in the course of experimentation”*

Hypothesentests

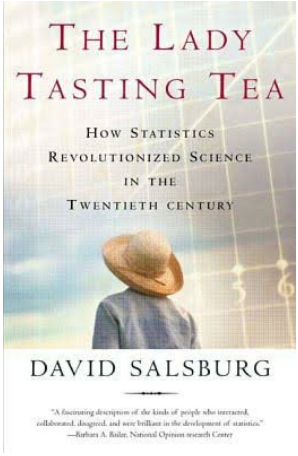
Nullhypothese:

- Kann man ‘beweisen’, dass die Lady den Unterschied tatsächlich schmecken kann???
- $\Rightarrow$ es ist immer möglich, dass sie nur ‘zufällig’ die richtige Lösung erriet, aber die *‘Wahrscheinlichkeit’* ist 1/70, kleiner als 5%!
- R.A. Fisher: ***Null hypothesis*** *is “never proved or established, but is possibly disproved, in the course of experimentation”*

- Fisher’s Exakter Test: Anteil der Erfolge *an allen möglichen Ausgängen* des Zufallsexperiments; (nur in einfachen Zusammenhängen anwendbar, da in komplexeren Fällen die Zahl der möglichen Ausgänge zu groß ist $\rightarrow$ Asymptotik)

Hypothesentests

Literaturtipp zur Geschichte der Statistik:



Exkurs 1

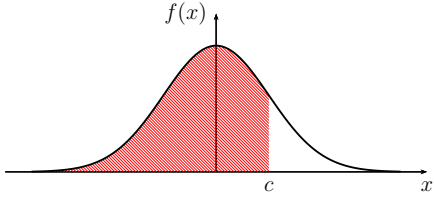
Einige Begriffe:

- **Fakten:** als Tatsache akzeptierte überprüfbare (und beweisbare) Sachverhalte
- **Thesen:** noch nicht bewiesene Behauptungen (Gedanken, Sätze)
- **Hypothesen:** Unterform einer These, die sich meist auf einen prinzipiell überprüfbaren Zusammenhang (meist zwischen Fakten) bezieht.
- **Annahmen:** spezielle Thesen, die (vorläufig) als wahr/zutreffend angenommen werden, z.B. zur Überprüfung von Hypothesen (z.B. unter welchen Annahmen ist OLS erwartungstreu?);
  Annahmen stehen nicht für sich im Zentrum des Interesses, sondern dienen eher als Mittel zum Zweck.
  Zumindest zentrale Annahmen müssen auf jeden Fall offengelegt und begründet werden. Manchmal können Annahmen auch empirisch getestet werden (Spezifikationstests).

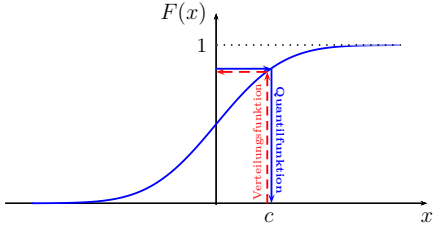
Exkurs 2: Einige theoretische Verteilungen

Allgemein: **Dichte- Verteilungs- und Quantilfunktion**

Dichtefunktion:



Verteilungsfunktion: (kumulierte Dichte)



Exkurs: Einige theoretische Verteilungen

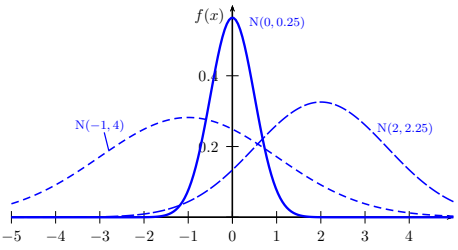
Normalverteilung: (A. de Moivre 1733, P.S: Laplace 1782, C.F. Gauss 1809)

Die Normalverteilung ist eine stetige symmetrische Verteilung mit der Dichte

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(1/2)[(x_i-\mu)/\sigma]^2}$$

mit
$\mu = \mathrm{E}(X)$ Erwartungswert,
$\sigma = \sqrt{\mathrm{var}(X)}$ Standardabweichung,
$\pi = 3.14159\ldots, \mathrm{e} = 2.71828\ldots$

$$X \sim \mathrm{N}(\mu, \sigma^2)$$



Normalverteilung ist u.a. verwandt mit χ^2 Verteilung, t Verteilung, F Verteilung.

Exkurs: Einige theoretische Verteilungen

Eigenschaften der Normalverteilung:

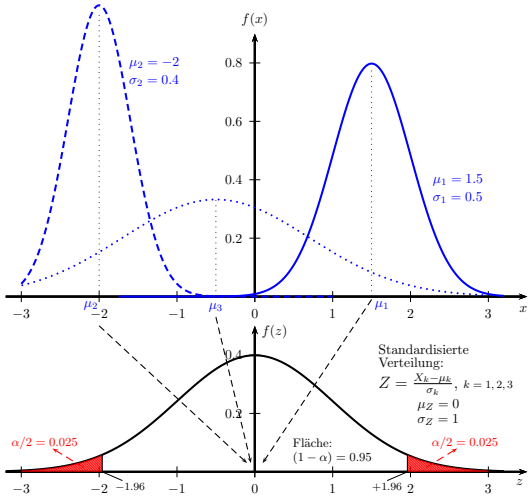
- Die stetige und symmetrische Normalverteilung ist durch die zwei Parameter μ und σ^2 vollständig spezifiziert, wir schreiben $X \sim \mathrm{N}(\mu, \sigma^2)$.
- Die Verteilung ist unbeschränkt und erstreckt sich von $-\infty$ bis $+\infty$.
- Das Maximum der Verteilung liegt beim Mittelwert μ und der Wendepunkt bei $x = \mu \pm \sigma$. Außerdem gilt:

$$\begin{aligned} \Pr(\mu - 1.96\sigma < X < \mu + 1.96\sigma) &\approx 0.95 \\ \Pr(\mu - 2.57\sigma < X < \mu + 2.57\sigma) &\approx 0.99 \end{aligned}$$

- *Reproduktivität:* jede Linearkombination von n *normalverteilten* Zufallsvariablen X_i ist selbst wieder normalverteilt.

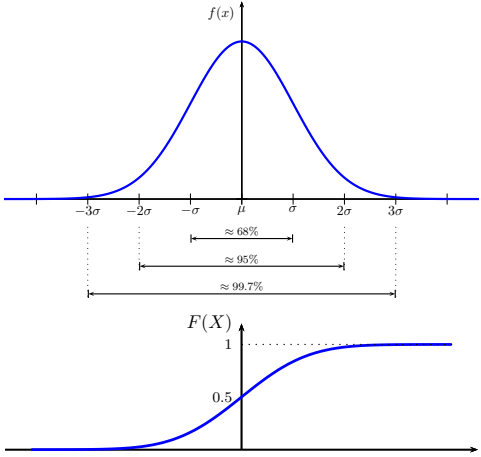
Exkurs: Einige theoretische Verteilungen

Normalverteilung: Standardisierung (z-Transformation: $Z = (X - \mu)/\sigma$)



Exkurs: Einige theoretische Verteilungen

Normalverteilung: Dichte- und Verteilungsfunktion einer standardnormalverteilten Zufallsvariable



Exkurs: Einige theoretische Verteilungen

Kumulierte Standard-Normalverteilung (Verteilungsfunktion):

z_c	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
⋮										
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

R:

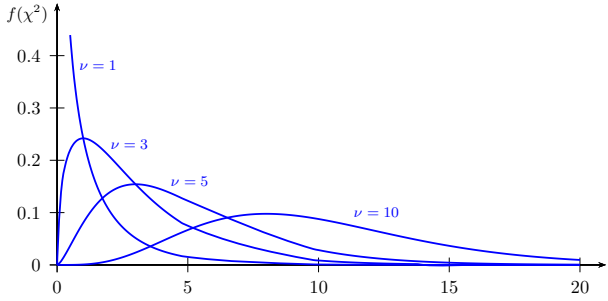
```
dnorm(x, mean = 0, sd = 1, log = FALSE) ## density
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE) ## distribution
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE) ## quantile
rnorm(n, mean = 0, sd = 1)  ## random numbers
```

Exkurs: Einige theoretische Verteilungen

χ^2 (chi–Quadrat) Verteilung: (F.R. Helmert, 1875)

Die *Quadrat*summe von ν unabhängig standardnormalverteilten Zufallsvariablen $Z_1, Z_2, \ldots, Z_\nu$ ist χ^2 verteilt mit ν Freiheitsgraden.

$$\text{Wenn } Z_i \sim \mathrm{N}(0,1) \quad \Rightarrow \quad \sum_{i=1}^{\nu} Z_i^2 \sim \chi^2_\nu \qquad \text{mit } \mathrm{E}(\chi^2) = \nu \text{ und } \mathrm{var}(\chi^2) = 2\nu$$



Exkurs: Einige theoretische Verteilungen

t Verteilung: (W.S. Gosset 1908, Pseudonym: ‘Student)

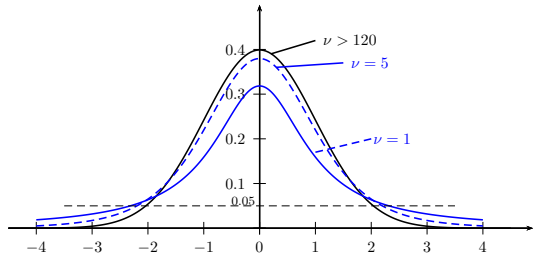
Wenn Z eine standardnormalverteilte Zufallsvariable ($Z \sim \mathrm{N}(0,1)$) und V eine χ^2-verteilte Zufallsvariable mit ν Freiheitsgraden ($V \sim \chi^2(\nu)$), und wenn Z und V stochastisch unabhängig sind, dann ist die Zufallsvariable $\hat{t}$

$$\hat{t} = \frac{Z}{\sqrt{(V/\nu)}} \sim \mathrm{t}_\nu$$

t-verteilt mit ν Freiheitsgraden.

Eigenschaften:

- symmetrisch
- ‘fat tails’
- konvergiert gegen Standardnormalverteilung



Exkurs: Einige theoretische Verteilungen

F Verteilung: (R.A. Fisher, 1924)

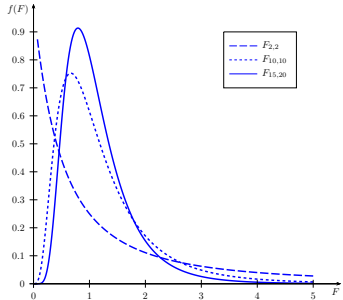
Wenn V_1 und V_2 zwei unabhängig χ^2-verteilte Zufallsvariablen mit ν_1 bzw. ν_2 Freiheitsgraden sind, dann ist

$$F = \frac{V_1/\nu_1}{V_2/\nu_2} \sim \mathrm{F}_{\nu_1,\nu_2}$$

F-verteilt mit ν_1 Zähler- und ν_2 Nennerfreiheitsgraden.

Das Quadrat einer t-verteilten Zufallsvariable mit ν Freiheitsgraden folgt einer F Verteilung mit 1 Zähler- und ν Nennerfreiheitsgraden, d.h.

$$\mathrm{t}_\nu^2 \sim \mathrm{F}_{1,\nu}$$



Hypothesentests: p-Werte nach R.A. Fisher

Grundlagen der Ökonometrie

herbert.stocker@uibk.ac.at

www.hsto.info/econometrics

Hypothesentests

- Hypothesentests sind ein zentraler Bestandteil jedes empirischen ‘Werkzeugkastens’.
- ermöglichen statistische Beurteilung, inwieweit eine theoretische Vermutung mit empirischer Evidenz kompatibel ist.
- theoretische Vermutung wird in Nullhypothese (H_0) abgebildet (meist in Form einer Gegenhypothese).
- empirische Evidenz kommt in Form einer *Teststatistik* und deren Verteilung zum Ausdruck.
- die folgenden Ausführungen gelten ziemlich allgemein, nicht nur für einfache Tests eines Regressionskoeffizienten.

Hypothesentests

Nullhypothesen:

- beziehen sich immer auf Parameter der Grundgesamtheit (DGP), nicht beobachtbar!

Hypothesentests

Nullhypothesen:

- beziehen sich immer auf Parameter der Grundgesamtheit (DGP), nicht beobachtbar!
- Einfachster Fall: $\mathrm{H}_0: \quad \theta = \theta_0$
  mit θ ... Parameter des DGP, θ_0 ... unter H_0 postulierter Wert (häufig 0)

Hypothesentests

Nullhypothesen:

- beziehen sich immer auf Parameter der Grundgesamtheit (DGP), nicht beobachtbar!
- Einfachster Fall: $\mathrm{H}_0: \quad \theta = \theta_0$
  mit θ ... Parameter des DGP, θ_0 ... unter H_0 postulierter Wert (häufig 0)
- meist *negativ* formuliert, ein Unterschied oder ein Effekt ist *nicht* vorhanden!
  (Gegenhypothese zu Anfangsvermutung)

Hypothesentests

Nullhypothesen:

- beziehen sich immer auf Parameter der Grundgesamtheit (DGP), nicht beobachtbar!
- Einfachster Fall: $\mathrm{H}_0: \quad \theta = \theta_0$
  mit θ ... Parameter des DGP, θ_0 ... unter H_0 postulierter Wert (häufig 0)
- meist *negativ* formuliert, ein Unterschied oder ein Effekt ist *nicht* vorhanden!
  (Gegenhypothese zu Anfangsvermutung)
- H_0 muss aus technischen Gründen immer '$=$' Zeichen enthalten (bzw. bei einseitigen Tests '$\leq$' oder '$\geq$').

Hypothesentests

Nullhypothesen:

- beziehen sich immer auf Parameter der Grundgesamtheit (DGP), nicht beobachtbar!
- Einfachster Fall: $\mathrm{H}_0: \quad \theta = \theta_0$
  mit $\theta \ldots$ Parameter des DGP, $\quad \theta_0 \ldots$ unter $\quad \mathrm{H}_0$ postulierter Wert (häufig 0)
- meist *negativ* formuliert, ein Unterschied oder ein Effekt ist *nicht* vorhanden!
  (Gegenhypothese zu Anfangsvermutung)
- H_0 muss aus technischen Gründen immer '$=$' Zeichen enthalten (bzw. bei einseitigen Tests '$\leq$' oder '$\geq$').
- H_0 kann als *Restriktion* begriffen werden!

Teststatistiken:

- spezielle Zufallsvariablen, deren theoretische Verteilung *unter Gültigkeit der Nullhypothese* bekannt ist.
- Teststatistiken dürfen *keine unbeobachtbaren Parameter* enthalten, damit sie berechnet werden können!
- Woher kommen Teststatistiken?
  - Arbeit theoretischer Mathematiker & Statistiker (z.B. Gosset 1908)
  - Viele (neuere) Tests sind nur asymptotisch gültig

Hypothesentests für Regressionskoeffizienten

- Reproduktivität der Normalverteilung: OLS Schätzfunktion für $\widehat{\beta}_2$ ist linear in den stochastischen ε_i

$$\widehat{\beta}_2 = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} w_i(\beta_1 + \beta_2 x_i + \varepsilon_i)$$

Hypothesentests für Regressionskoeffizienten

- Reproduktivität der Normalverteilung: OLS Schätzfunktion für $\widehat{\beta}_2$ ist linear in den stochastischen ε_i

$$\widehat{\beta}_2 = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} w_i(\beta_1 + \beta_2 x_i + \varepsilon_i)$$

- Wenn Störterme normalverteilt, $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, gilt

$$\widehat{\beta}_2 \sim \mathrm{N}(\beta_2, \sigma^2_{\widehat{\beta}_2})$$

Hypothesentests für Regressionskoeffizienten

- Reproduktivität der Normalverteilung: OLS Schätzfunktion für $\widehat{\beta}_2$ ist linear in den stochastischen ε_i

$$\widehat{\beta}_2 = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} w_i(\beta_1 + \beta_2 x_i + \varepsilon_i)$$

- Wenn Störterme normalverteilt, $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, gilt

$$\widehat{\beta}_2 \sim \mathrm{N}(\beta_2, \sigma^2_{\widehat{\beta}_2})$$

- gilt analog auch für multiple Regressionen

Hypothesentests für Regressionskoeffizienten

- Reproduktivität der Normalverteilung: OLS Schätzfunktion für $\widehat{\beta}_2$ ist linear in den stochastischen ε_i

$$\widehat{\beta}_2 = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} w_i(\beta_1 + \beta_2 x_i + \varepsilon_i)$$

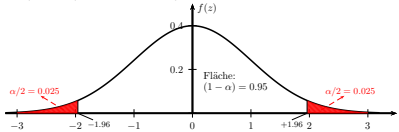
- Wenn Störterme normalverteilt, $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, gilt

$$\widehat{\beta}_2 \sim \mathrm{N}(\beta_2, \sigma^2_{\widehat{\beta}_2})$$

- gilt analog auch für multiple Regressionen
- Normalverteilte Störterme häufig unrealistisch!
  $\Rightarrow$ Zentrale Grenzwertsätze

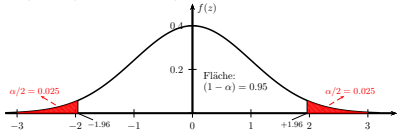
Hypothesentests für Regressionskoeffizienten

- Allgemein gilt für $X \sim \mathrm{N}(0,1)$: $\Pr(-1.96 < X < +1.96) = 0.95$



Hypothesentests für Regressionskoeffizienten

- Allgemein gilt für $X \sim \mathrm{N}(0,1)$: $\Pr(-1.96 < X < +1.96) = 0.95$

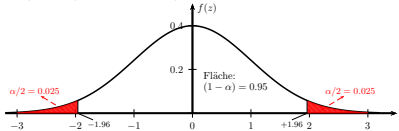


- Wenn $\widehat{\beta}_h \sim \mathrm{N}(\beta_h, \sigma^2_{\widehat{\beta}_h})$ folgt

$$\Pr\left(-1.96 \leq \frac{\widehat{\beta}_h - \beta_h}{\sigma_{\widehat{\beta}_h}} \leq +1.96\right) = 0.95$$

Hypothesentests für Regressionskoeffizienten

- Allgemein gilt für $X \sim \mathrm{N}(0,1)$: $\Pr(-1.96 < X < +1.96) = 0.95$



- Wenn $\widehat{\beta}_h \sim \mathrm{N}(\beta_h, \sigma^2_{\widehat{\beta}_h})$ folgt

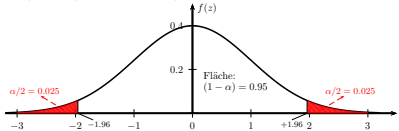
$$\Pr\left(-1.96 \leq \frac{\widehat{\beta}_h - \beta_h}{\sigma_{\widehat{\beta}_h}} \leq +1.96\right) = 0.95$$

- enthält unbeobachtbaren Parameter $\sigma_{\widehat{\beta}_h}$!
  Dieser Standardfehler muss aus Daten geschätzt werden:

$$\hat{\sigma}_{\widehat{\beta}_h} := \widehat{\mathrm{se}}(\widehat{\beta}_h) = \sqrt{\frac{\hat{\sigma}^2}{(1 - R_h^2)\sum_i (x_i - \bar{x})^2}} \qquad \text{mit } \hat{\sigma}^2 = \frac{\sum_i \hat{\varepsilon}_i^2}{n - k}$$

Hypothesentests für Regressionskoeffizienten

- Allgemein gilt für $X \sim \mathrm{N}(0,1)$: $\Pr(-1.96 < X < +1.96) = 0.95$



- Wenn $\widehat{\beta}_h \sim \mathrm{N}(\beta_h, \sigma^2_{\widehat{\beta}_h})$ folgt

$$\Pr\left(-1.96 \leq \frac{\widehat{\beta}_h - \beta_h}{\sigma_{\widehat{\beta}_h}} \leq +1.96\right) = 0.95$$

- enthält unbeobachtbaren Parameter $\sigma_{\widehat{\beta}_h}$!
  Dieser Standardfehler muss aus Daten geschätzt werden:

$$\hat{\sigma}_{\widehat{\beta}_h} := \widehat{\mathrm{se}}(\widehat{\beta}_h) = \sqrt{\frac{\hat{\sigma}^2}{(1 - R^2_h)\sum_i (x_i - \bar{x})^2}} \qquad \text{mit } \hat{\sigma}^2 = \frac{\sum_i \hat{\varepsilon}_i^2}{n - k}$$

- **aber:** Schätzung von $\hat{\sigma}_{\widehat{\beta}_h}$ bringt zusätzliche Unsicherheit!

Hypothesentests für Regressionskoeffizienten

- Wenn $\sigma_{\widehat{\beta}_2}$ durch $\hat{\sigma}_{\widehat{\beta}_2}$ ersetzt wird steht im Nenner eine Zufallsvariable, und das Resultat ist nicht länger normalverteilt

$$\frac{\widehat{\beta}_2 - \beta_2}{\hat{\sigma}_{\widehat{\beta}_2}} \nsim \mathrm{N}(0,1) \quad !$$

Hypothesentests für Regressionskoeffizienten

- Wenn $\sigma_{\widehat{\beta}_2}$ durch $\hat{\sigma}_{\widehat{\beta}_2}$ ersetzt wird steht im Nenner eine Zufallsvariable, und das Resultat ist nicht länger normalverteilt

$$\frac{\widehat{\beta}_2 - \beta_2}{\hat{\sigma}_{\widehat{\beta}_2}} \nsim \mathrm{N}(0,1) \quad !$$

- Gosset (1908) konnte zeigen, dass für normalverteilte Zufallsvariablen die durch die Schätzung von $\hat{\sigma}_{\widehat{\beta}_2}$ hinzukommende Unsicherheit durch Verwendung der t-Verteilung mit ihren *'fat tails'* berücksichtigt werden kann

$$\frac{\widehat{\beta}_2 - \beta_2}{\hat{\sigma}_{\widehat{\beta}_2}} \sim \mathrm{t}_{n-2}$$

(siehe Skript)

Hypothesentests für Regressionskoeffizienten

- Wenn $\sigma_{\widehat{\beta}_2}$ durch $\hat{\sigma}_{\widehat{\beta}_2}$ ersetzt wird steht im Nenner eine Zufallsvariable, und das Resultat ist nicht länger normalverteilt

$$\frac{\widehat{\beta}_2 - \beta_2}{\hat{\sigma}_{\widehat{\beta}_2}} \nsim \mathrm{N}(0,1) \quad !$$

- Gosset (1908) konnte zeigen, dass für normalverteilte Zufallsvariablen die durch die Schätzung von $\hat{\sigma}_{\widehat{\beta}_2}$ hinzukommende Unsicherheit durch Verwendung der t-Verteilung mit ihren *'fat tails'* berücksichtigt werden kann

$$\frac{\widehat{\beta}_2 - \beta_2}{\hat{\sigma}_{\widehat{\beta}_2}} \sim \mathrm{t}_{n-2}$$

  (siehe Skript)
- heute eher von historischer Bedeutung . . . $\rightarrow$ Asymptotik

p-Werte nach R.A. Fisher

- Gosset (1908): für PRF $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_h x_{ih} + \cdots + \beta_k x_{ik} + \varepsilon_i$

$$\frac{\widehat{\beta}_h - \beta_h}{\hat{\sigma}_{\widehat{\beta}_h}} \sim \mathrm{t}_{n-k}$$

mit $n - k$ Freiheitsgraden.

Aber: β_h ist nicht beobachtbar $\rightarrow$ keine Teststatistik!

Hypothesentests für Regressionskoeffizienten

p-Werte nach R.A. Fisher

- Gosset (1908): für PRF $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_h x_{ih} + \cdots + \beta_k x_{ik} + \varepsilon_i$

$$\frac{\widehat{\beta}_h - \beta_h}{\hat{\sigma}_{\widehat{\beta}_h}} \sim \mathrm{t}_{n-k}$$

mit $n - k$ Freiheitsgraden.

Aber: β_h ist nicht beobachtbar $\rightarrow$ keine Teststatistik!

- R.A. Fisher: ersetzt Parameter β_h durch den Wert der Nullhypothese H_0: β_0

Teststatistik:

$$\hat{t} = \frac{\widehat{\beta}_h - \beta_0}{\hat{\sigma}_{\widehat{\beta}_h}} \overset{\mathrm{H}_0}{\sim} \mathrm{t}_{n-k}$$

Gilt nur, wenn H_0 wahr ist! (*bedingte Wahrscheinlichkeit*!)

p-Werte nach R.A. Fisher

- **Theoretische Teststatistik:** Zufallsvariable beschreibt Verteilung aller möglichen Ergebnisse des Zufallsexperiments *wenn* H_0 wahr ist

$$\hat{t} = \frac{\widehat{\beta}_h - \beta_0}{\hat{\sigma}_{\widehat{\beta}_h}} \overset{\mathrm{H}_0}{\sim} \mathrm{t}_{n-k}$$

- **Empirische Teststatistik t^{emp}:** *eine* Realisation der theoret. Teststatistik $\hat{t}$ (wir ersetzen ZV durch Realisationen)

$$t^{\text{emp}} = \frac{b_h - \beta_0}{s_{\widehat{\beta}_h}}$$

Hypothesentests

Grundidee:

- Wenn Nullhypothese $\mathrm{H}_0\colon \beta_h = \beta_0$ wahr ist sollte Realisation t^{emp} aus der Verteilung der folgenden theoretischen Teststatistik stammen:

Zufallsvariable: Realisation:

$$\hat{t} = \frac{\widehat{\beta}_h - \beta_0}{\hat{\sigma}_{\widehat{\beta}_h}} \overset{\mathrm{H}_0}{\sim} \mathrm{t}_{n-k} \quad \xrightarrow{\text{Realisat.}} \quad t^{\text{emp}} = \frac{b_h - \beta_0}{s_{\widehat{\beta}_h}}$$

Hypothesentests

Grundidee:

- Wenn Nullhypothese $\mathrm{H}_0\colon \beta_h = \beta_0$ wahr ist sollte Realisation t^{emp} aus der Verteilung der folgenden theoretischen Teststatistik stammen:

Zufallsvariable:		Realisation:
$\hat{t} = \dfrac{\widehat{\beta}_h - \beta_0}{\hat{\sigma}_{\widehat{\beta}_h}} \overset{\mathrm{H}_0}{\sim} \mathrm{t}_{n-k}$	$\xrightarrow{\text{Realisat.}}$	$t^{\text{emp}} = \dfrac{b_h - \beta_0}{s_{\widehat{\beta}_h}}$

- Wenn H_0 falsch ist ($\beta_h \neq \beta_0$) stammt Realisation t^{emp} aus anderer Verteilung!

  In diesem Fall würden wir einen (in Bezug auf $\hat{t}$) extremeren Wert von t^{emp} erwarten!

Hypothesentests

Grundidee:

- Wenn Nullhypothese $\mathrm{H}_0\colon \beta_h = \beta_0$ wahr ist sollte Realisation t^{emp} aus der Verteilung der folgenden theoretischen Teststatistik stammen:

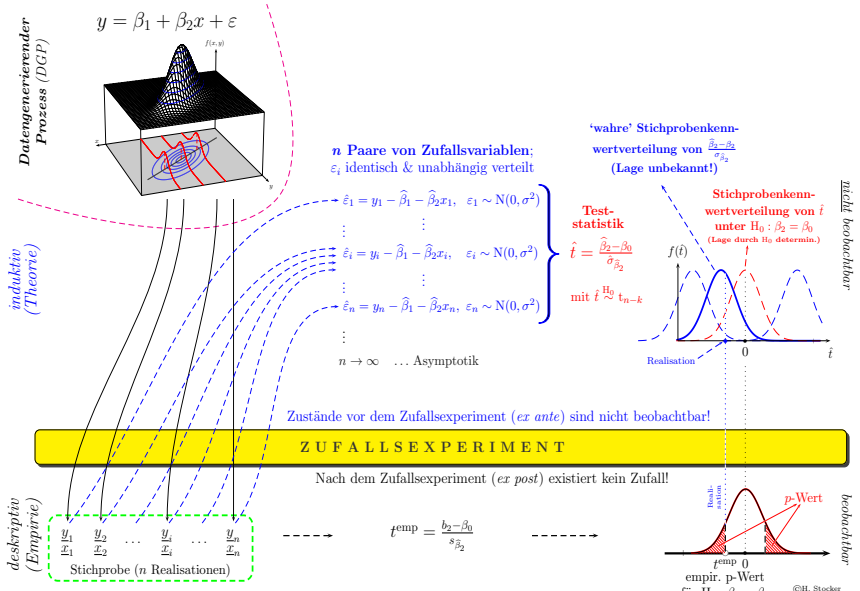
Zufallsvariable:		Realisation:
$\hat{t} = \dfrac{\widehat{\beta}_h - \beta_0}{\hat{\sigma}_{\widehat{\beta}_h}} \overset{\mathrm{H}_0}{\sim} \mathrm{t}_{n-k}$	$\xrightarrow{\text{Realisat.}}$	$t^{\text{emp}} = \dfrac{b_h - \beta_0}{s_{\widehat{\beta}_h}}$

- Wenn H_0 falsch ist ($\beta_h \neq \beta_0$) stammt Realisation t^{emp} aus anderer Verteilung!

  In diesem Fall würden wir einen (in Bezug auf $\hat{t}$) extremeren Wert von t^{emp} erwarten!

- emp. p-Wert gibt Fläche links von $-|t^{\text{emp}}|$ und rechts von $+|t^{\text{emp}}|$ an;
  $\Rightarrow$ Indikator, wie gut die H_0 den DGP beschreibt.

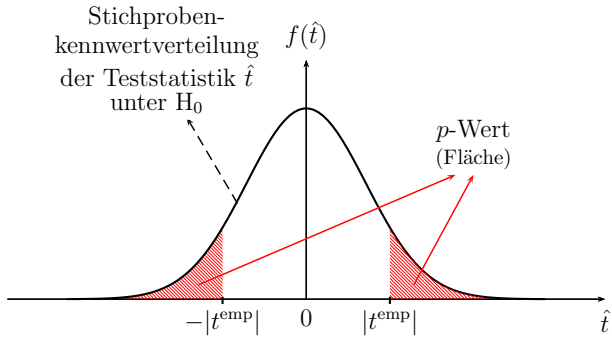
Hypothesentest für einen einzelnen Regressionskoeffizienten: p-Werte



©H. Stocker

Hypothesentests: p-Werte

Empirische Teststatistik liefert empirischen p-Wert: (zweiseitiger Test)



$\Rightarrow$ ein 'großer' p-Wert (häufig p-Wert > 0.05) wird als Unterstützung für die H_0 interpretiert, ein *kleiner* p-Wert (z.B. < 0.05) als empir. Evidenz für die Anfangsvermutung!

Hypothesentests: Berechnung der p-Werte

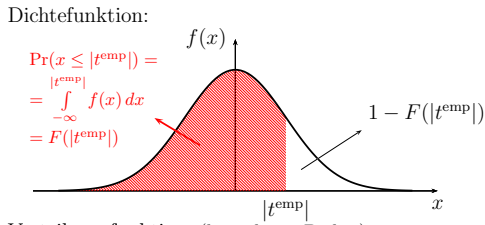
Berechnung von p-Werten:

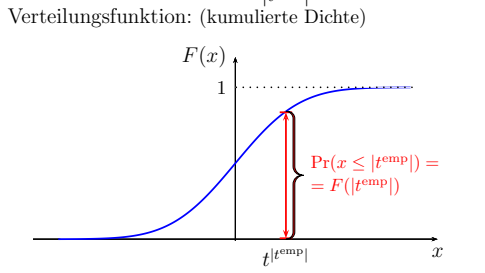
$F(t)$: Verteilungsfunktion der t-Verteilg.

$$F(t) = \int_{-\infty}^{|t^{\text{emp}}|} f(x)\,dx$$

Zweiseitiger p-Wert:

$$p^{\text{emp}} = [1 - F(|t^{\text{emp}}|, df)] \times 2$$





Hypothesentests: p-Werte

Interpretation von p-Werten

Wenn (a) die Nullhypothese wahr ist und (b) alle erforderlichen Annahmen erfüllt sind gibt der p-Wert die Wahrscheinlichkeit dafür an, dass wir bei einer *hypothetischen neuerlichen Durchführung des Zufallsexperiments* eine empirische Teststatistik erhalten würden, die noch extremer ist als die vorliegende empirische Teststatistik.

Hypothesentests: p-Werte

Interpretation von p-Werten

Wenn (a) die Nullhypothese wahr ist und (b) alle erforderlichen Annahmen erfüllt sind gibt der p-Wert die Wahrscheinlichkeit dafür an, dass wir bei einer *hypothetischen neuerlichen Durchführung des Zufallsexperiments* eine empirische Teststatistik erhalten würden, die noch extremer ist als die vorliegende empirische Teststatistik.

Achtung: Hypothesentests sind *bedingte* Wahrscheinlichkeiten!

$$\Pr(t^{\text{emp}} \mid \text{H}_0 \text{ wahr})$$

nicht umgekehrt: $\Pr(\text{H}_0 \text{ wahr} \mid t^{\text{emp}})$!!!

Hypothesentests: p-Werte

Interpretation von p-Werten:

- Da p-Werte Wahrscheinlichkeiten sind können sie nur Werte zwischen Null und Eins annehmen. Ein sehr kleiner p-Wert nahe bei Null deutet darauf hin, dass entweder die Nullhypothese falsch ist, oder dass ein sehr unwahrscheinliches Ereignis eingetreten ist (oder eine Annahme verletzt ist).

Hypothesentests: p-Werte

Interpretation von p-Werten:

- Da p-Werte Wahrscheinlichkeiten sind können sie nur Werte zwischen Null und Eins annehmen. Ein sehr kleiner p-Wert nahe bei Null deutet darauf hin, dass entweder die Nullhypothese falsch ist, oder dass ein sehr unwahrscheinliches Ereignis eingetreten ist (oder eine Annahme verletzt ist).
- Umso kleiner ein p-Wert ist, umso eher werden wir geneigt sein die Nullyhpothese zu verwerfen, was wir als empirische Evidenz für unsere Anfangsvermutung interpretieren können.

Hypothesentests: p-Werte

Interpretation von p-Werten:

- Da p-Werte Wahrscheinlichkeiten sind können sie nur Werte zwischen Null und Eins annehmen. Ein sehr kleiner p-Wert nahe bei Null deutet darauf hin, dass entweder die Nullhypothese falsch ist, oder dass ein sehr unwahrscheinliches Ereignis eingetreten ist (oder eine Annahme verletzt ist).
- Umso kleiner ein p-Wert ist, umso eher werden wir geneigt sein die Nullyhpothese zu verwerfen, was wir als empirische Evidenz für unsere Anfangsvermutung interpretieren können.
- Etwas salopp können wir den p-Wert auch einfach als Kennzahl dafür interpretieren, wie gut die Nullhypothese die Daten beschreibt.

Hypothesentests: p-Werte

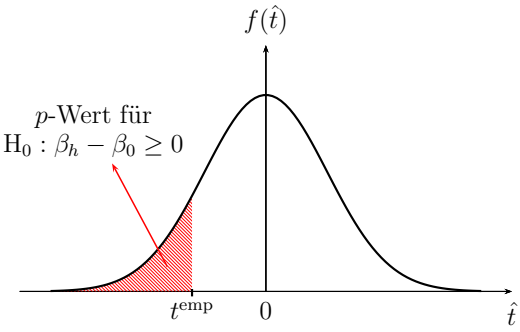
Interpretation von p-Werten:

- Da p-Werte Wahrscheinlichkeiten sind können sie nur Werte zwischen Null und Eins annehmen. Ein sehr kleiner p-Wert nahe bei Null deutet darauf hin, dass entweder die Nullhypothese falsch ist, oder dass ein sehr unwahrscheinliches Ereignis eingetreten ist (oder eine Annahme verletzt ist).
- Umso kleiner ein p-Wert ist, umso eher werden wir geneigt sein die Nullyhpothese zu verwerfen, was wir als empirische Evidenz für unsere Anfangsvermutung interpretieren können.
- Etwas salopp können wir den p-Wert auch einfach als Kennzahl dafür interpretieren, wie gut die Nullhypothese die Daten beschreibt.
- Ein sehr kleiner p-Wert wird als Evidenz *gegen* die Nullhypothese (und damit als Unterstützung für die Anfangsvermutung) interpretiert.

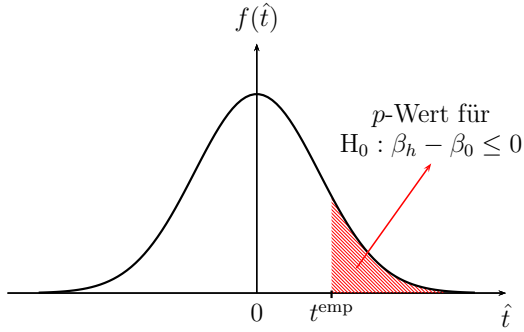
Hypothesentests: p-Werte

p–Werte für einseitige Hypothesentests:

Linksseitiger Test: $\mathrm{H}_0 : \beta_h \geq \beta_0$



Rechtsseitiger Test: $\mathrm{H}_0 : \beta_h \leq \beta_0$



$$\mathrm{H}_0 : \beta_h \geq \beta_0 \quad \text{oder} \quad \mathrm{H}_0 : \beta_h \leq \beta_0 \quad \rightarrow \quad t^{\mathrm{emp}} = \frac{b_h - \beta_0}{s_{\widehat{\beta}_h}}$$

Zum Mitnehmen . . .

Hypothesentests

- DGP enthält zufälliges Rauschen (Stochastik); können wir im Rauschen ein Muster (Gesetzmäßigkeit) entdecken?

Zum Mitnehmen . . .

Hypothesentests

- DGP enthält zufälliges Rauschen (Stochastik); können wir im Rauschen ein Muster (Gesetzmäßigkeit) entdecken?
- Intuition: signal-to-noise ratio

Teststatistik:

$$\hat{t} = \frac{\widehat{\beta}_h - \beta_0}{\hat{\sigma}_{\widehat{\beta}_h}} \quad \longleftrightarrow \quad \frac{\text{signal}}{\text{noise}}$$

Zum Mitnehmen . . .

Hypothesentests

- DGP enthält zufälliges Rauschen (Stochastik); können wir im Rauschen ein Muster (Gesetzmäßigkeit) entdecken?
- Intuition: signal-to-noise ratio

Teststatistik:

$$\hat{t} = \frac{\widehat{\beta}_h - \beta_0}{\hat{\sigma}_{\widehat{\beta}_h}} \quad \longleftrightarrow \quad \frac{\text{signal}}{\text{noise}}$$

- Zähler: stochastisches Analogon zur H_0 in impliziter Form: $\mathrm{H}_0 : \beta_h - \beta_0 = 0$

  Signal: Abweichung von der Nullhypothese

Zum Mitnehmen . . .

Hypothesentests

- DGP enthält zufälliges Rauschen (Stochastik); können wir im Rauschen ein Muster (Gesetzmäßigkeit) entdecken?
- Intuition: signal-to-noise ratio

Teststatistik:

$$\hat{t} = \frac{\widehat{\beta}_h - \beta_0}{\hat{\sigma}_{\widehat{\beta}_h}} \quad \longleftrightarrow \quad \frac{\text{signal}}{\text{noise}}$$

- Zähler: stochastisches Analogon zur H_0 in impliziter Form: $\mathrm{H}_0 : \beta_h - \beta_0 = 0$

  Signal: Abweichung von der Nullhypothese
- Nenner: $\hat{\sigma}_{\widehat{\beta}_h}$ Indikator für das *'Rauschen'* (noise) des Zufallsexperiment

Zum Mitnehmen . . .

Hypothesentests

- DGP enthält zufälliges Rauschen (Stochastik); können wir im Rauschen ein Muster (Gesetzmäßigkeit) entdecken?
- Intuition: signal-to-noise ratio

Teststatistik:

$$\hat{t} = \frac{\widehat{\beta}_h - \beta_0}{\hat{\sigma}_{\widehat{\beta}_h}} \quad \longleftrightarrow \quad \frac{\text{signal}}{\text{noise}}$$

- Zähler: stochastisches Analogon zur H_0 in impliziter Form: $\mathrm{H}_0 : \beta_h - \beta_0 = 0$

  Signal: Abweichung von der Nullhypothese
- Nenner: $\hat{\sigma}_{\widehat{\beta}_h}$ Indikator für das *'Rauschen'* (noise) des Zufallsexperiment
- Frage: hebt sich das Signal stark genug vom Hintergrundrauschen des Zufallsexperiment ab um Beachtung zu verdienen?

Zum Mitnehmen . . .

Achtung:

- Statistische Signifikanz ‘beweist’ weder die Gültigkeit noch die Ungültigkeit der H_0!
  Statistische Signifikanz ist häufig eine *notwendige Bedingung* um ein Ergebnis ernst zu nehmen, aber *nie eine hinreichende Bedingung*!!!

Zum Mitnehmen . . .

Achtung:

- Statistische Signifikanz ‘beweist’ weder die Gültigkeit noch die Ungültigkeit der H_0!
  Statistische Signifikanz ist häufig eine *notwendige Bedingung* um ein Ergebnis ernst zu nehmen, aber *nie eine hinreichende Bedingung*!!!
- Hypothesentests sind nur gültig, wenn alle zugrunde liegenden Annahmen erfüllt sind, insbesondere . . .

Zum Mitnehmen . . .

Achtung:

- Statistische Signifikanz ‘beweist’ weder die Gültigkeit noch die Ungültigkeit der H_0!
  Statistische Signifikanz ist häufig eine *notwendige Bedingung* um ein Ergebnis ernst zu nehmen, aber *nie eine hinreichende Bedingung*!!!
- Hypothesentests sind nur gültig, wenn alle zugrunde liegenden Annahmen erfüllt sind, insbesondere . . .
  - korrekte Spezifikation des DGP! (‘wahres’ Modell)
    V.a. keine Endogenität (*omitted variables, feed-backs, . . .* , $\rightarrow$ A3)

Zum Mitnehmen . . .

Achtung:

- Statistische Signifikanz ‘beweist’ weder die Gültigkeit noch die Ungültigkeit der H_0!
  Statistische Signifikanz ist häufig eine *notwendige Bedingung* um ein Ergebnis ernst zu nehmen, aber *nie eine hinreichende Bedingung*!!!
- Hypothesentests sind nur gültig, wenn alle zugrunde liegenden Annahmen erfüllt sind, insbesondere . . .
  - korrekte Spezifikation des DGP! (‘wahres’ Modell)
    V.a. keine Endogenität (*omitted variables, feed-backs, . . .* , $\rightarrow$ A3)
  - korrekt berechnete Standardfehler ($\rightarrow$ A4)

Zum Mitnehmen . . .

Achtung:

- Statistische Signifikanz ‘beweist’ weder die Gültigkeit noch die Ungültigkeit der H_0!
  Statistische Signifikanz ist häufig eine *notwendige Bedingung* um ein Ergebnis ernst zu nehmen, aber *nie eine hinreichende Bedingung*!!!
- Hypothesentests sind nur gültig, wenn alle zugrunde liegenden Annahmen erfüllt sind, insbesondere . . .
  - korrekte Spezifikation des DGP! (‘wahres’ Modell)
    V.a. keine Endogenität (*omitted variables, feed-backs, . . .* , $\rightarrow$ A3)
  - korrekt berechnete Standardfehler ($\rightarrow$ A4)
- nur manche Annahmen können mit Spezifikationstests getestet werden

Zum Mitnehmen . . .

Achtung:

- Statistische Signifikanz ‘beweist’ weder die Gültigkeit noch die Ungültigkeit der H_0!
  Statistische Signifikanz ist häufig eine *notwendige Bedingung* um ein Ergebnis ernst zu nehmen, aber *nie eine hinreichende Bedingung*!!!
- Hypothesentests sind nur gültig, wenn alle zugrunde liegenden Annahmen erfüllt sind, insbesondere . . .
  - korrekte Spezifikation des DGP! (‘wahres’ Modell)
    V.a. keine Endogenität (*omitted variables, feed-backs, . . .* , $\rightarrow$ A3)
  - korrekt berechnete Standardfehler ($\rightarrow$ A4)
- nur manche Annahmen können mit Spezifikationstests getestet werden
- Hypothesentests müssen stets kritisch hinterfragt und sollten nie überinterpretiert werden!
  Näheres folgt später . . .

Hypothesentest

Quelle: `https://xkcd.com/892/`

Thanks!

Hypothesentests: Signifikanz nach Neyman-Pearson

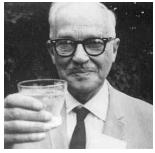
Grundlagen der Ökonometrie

herbert.stocker@uibk.ac.at

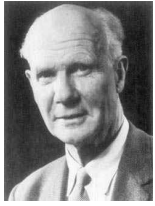
www.hsto.info/econometrics

Hypothesentests: **Neyman-Pearson**

- Auf R.A. Fisher aufbauend entwickelten Jerzy Neyman und Egon Pearson *Signifikanztests* in der heute üblichen Form.
- Ursprüngliches Ziel: Qualitätskriterien für Teststatistiken
  Typ I & Typ II Fehler, $\rightarrow$ Power
  $\rightarrow$ *'uniformly most powerful tests'*, UMP
- liefert einfache Entscheidungsregel!



Jerzy Neyman



Egon Pearson
Quelle: `https://mathshistory.st-andrews.ac.uk/Biographies/`

Hypothesentests: **Neyman-Pearson**

Neyman-Pearson: zwei wesentliche Unterschiede

1. **Signifikanznveau:** es wird *a priori* ein Signifikanznveau α festgelegt. Im Unterschied zum p–Wert von Fisher ist das Signifikanznveau α keine Zufallsvariable, sondern ein fixer Parameter, der a priori festgelegt wird.

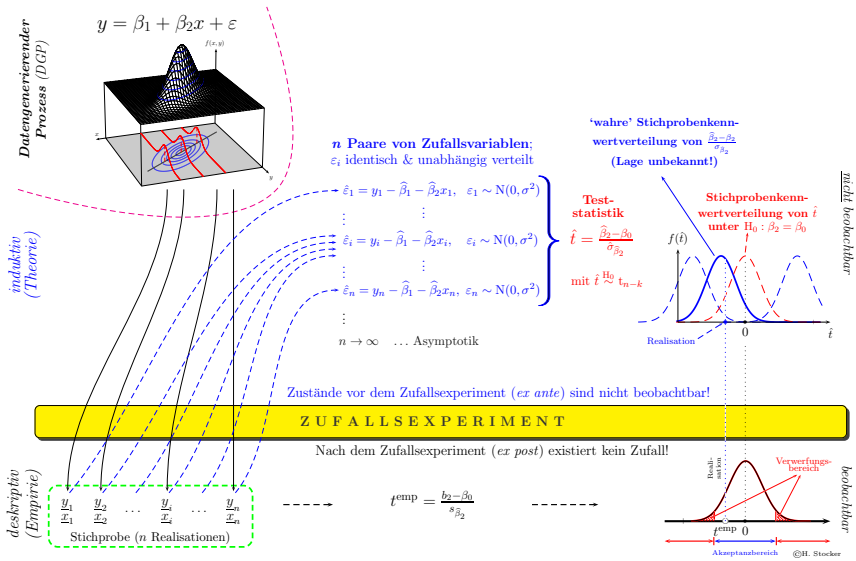
   Definiert kritischen Wert: $\alpha \rightarrow$ Quantilfunktion $\rightarrow t^{\text{crit}}_{\alpha/2}$

2. **Alternativhypothese:** der Nullhypothese wird explizit eine Alternativhypothese H_A gegenüber gestellt, die immer richtig ist, wenn die Nullhypothese falsch ist, d.h. H_0 und H_A schließen sich gegenseitig aus.

Dies erlaubt Definition von . . .

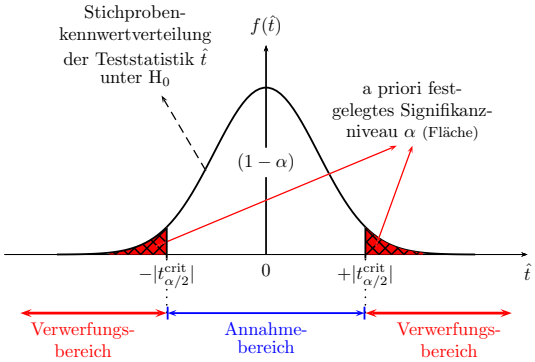
- Akzeptanz- und Verwerfungsbereich
- Typ I und Typ II Fehler sowie Trennschärfe (*Power*) von Tests

Hypothesentest für einen einzelnen Regressionskoeffizienten: Neyman-Pearson



Hypothesentests: Neyman-Pearson, zweiseitig

Signifikanzniveau α wird vorgegeben, damit wir kritischer Wert t^{crit} bestimmt (Quantilfunktion) $\rightarrow$ definiert Annahme- und Verwerfungsbereich



Hypothesentests: Neyman-Pearson

Kritische Werte: t-Tabelle (rechtsseitig)

q	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
⋮	⋮	⋮	⋮	⋮	⋮
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
⋮	⋮	⋮	⋮	⋮	⋮
30	1.310	1.697	2.042	2.457	2.750
60	1.296	1.671	2.000	2.390	2.660
∞	1.282	1.645	1.960	2.327	2.576

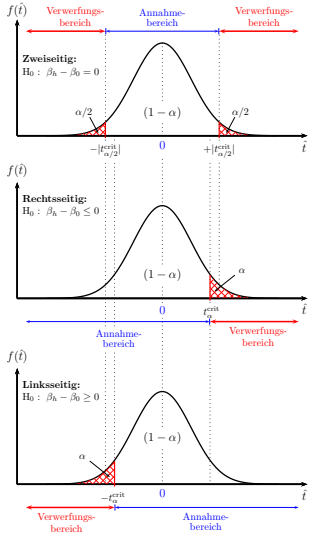
R: `qt(p = 0.975, df = 10) ## 2.228139`, (Fläche links: 0.975 $\Rightarrow$ Fl. rechts: 0.025)

Hypothesentests: Neyman-Pearson, einseitig

Einseitige Tests:

rechtsseitig:	$\mathrm{H}_0: \ \beta_h - \beta_0 \leq 0$
Akzeptanzbereich:	$[-\infty, +\mathrm{t}_\alpha^{\mathrm{crit}}]$
Verwerfungsbereich:	$(+\mathrm{t}_\alpha^{\mathrm{crit}}, +\infty]$
linksseitig:	$\mathrm{H}_0: \ \beta_h - \beta_0 \geq 0$
Akzeptanzbereich:	$[-\mathrm{t}_\alpha^{\mathrm{crit}}, +\infty]$
Verwerfungsbereich:	$[-\infty, -\mathrm{t}_\alpha^{\mathrm{crit}})$

gleiche Teststatistik (t^{emp}) wie zweiseitig!

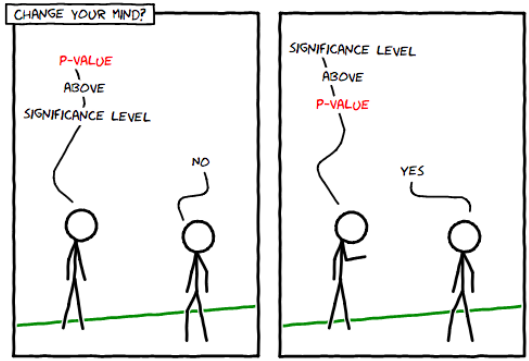


Hypothesentests: Neyman-Pearson

Hypothesentests nach Neyman-Pearson: typische Schritte

1. Problemanalyse
2. Festlegung von Null- und Alternativhypothese
3. Wahl der Teststatistik
4. Festlegung eines Signifikanzniveaus α
5. Dies definiert Annahme- und Verwerfungsbereich
6. Berechnung der empirischen Teststatistik (Stichprobe)
7. Entscheidung: wenn t^{emp} in den Verwerfungsbereich fällt verwirf H_0, sonst akzeptiere H_0.

Hypothesentests: p-Wert vs. Signifikanzniveau α



Wenn p-Wert $\geq \alpha$: akzeptiere H_0 (nichts gelernt)
Wenn p-Wert $< \alpha$: verwirf H_0, Evidenz für Anfangsvermutung :-)

Hypothesentests: Neyman-Pearson

Hypothesentests nach Neyman-Pearson sind eine Entscheidungsregel:

1. Wenn t^{emp} in den *Verwerfungsbereich* fällt:
   verwirf die Nullhypothese auf dem $\alpha \times 100$ Prozent Signifikanzniveau.

   Fehlerwahrscheinlichkeit: Wenn die Nullhypothese tatsächlich *wahr* ist (und alle Annahmen erfüllt sind) beträgt die Wahrscheinlichkeit, dass die t^{emp} in den Verwerfungsbereich fällt, $\alpha \times 100$ Prozent
   $\Rightarrow$ Typ I Fehler (α).

Hypothesentests: Neyman-Pearson

Hypothesentests nach Neyman-Pearson sind eine Entscheidungsregel:

1. Wenn t^{emp} in den *Verwerfungsbereich* fällt:
   verwirf die Nullhypothese auf dem $\alpha \times 100$ Prozent Signifikanzniveau.

   Fehlerwahrscheinlichkeit: Wenn die Nullhypothese tatsächlich *wahr* ist (und alle Annahmen erfüllt sind) beträgt die Wahrscheinlichkeit, dass die t^{emp} in den Verwerfungsbereich fällt, $\alpha \times 100$ Prozent
   $\Rightarrow$ Typ I Fehler (α).

2. Wenn t^{emp} in den *Akzeptanzbereich* fällt:
   verwirf die Nullhypothese *nicht* (auf dem $\alpha \times 100$ % Signifikanzniveau).

   Fehlerwahrscheinlichkeit: die Wahrscheinlichkeit, dass t^{emp} in den *Akzeptanzbereich* fällt, obwohl die Nullhypothese tatsächlich falsch ist
   $\Rightarrow$ Typ II Fehler (Größe hängt Alternativhypothese und α ab).

Typ I und Typ II Fehler

Analogie: Vergleich mit einem Gerichtsurteil

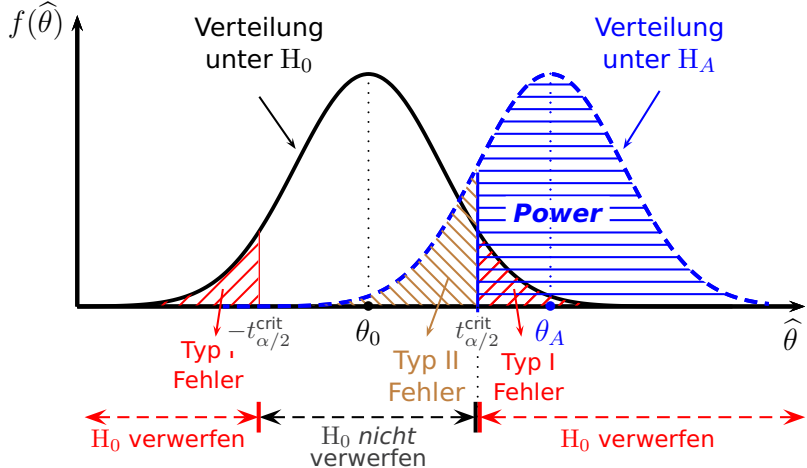
	Angeklagter ist unschuldig	**Angeklagter ist schuldig**
Gericht fällt Entscheidung “unschuldig”	richtige Entscheidung	ein Schuldiger wird freigesprochen
Gericht fällt Entscheidung “schuldig”	ein Unschuldiger wird verurteilt	richtige Entscheidung

Typ I und Typ II Fehler

Entscheidung auf Grundlage eines statistischen Tests:	**Wahrer Sachverhalt: $\mathrm{H_0}$ ist wahr**	**Wahrer Sachverhalt: $\mathrm{H_0}$ ist falsch**
Nullhypothese $\mathrm{H_0}$ wird *nicht* verworfen	korrekte Entscheidung $(1-\alpha)$	*Typ II Fehler*
Nullhypothese $\mathrm{H_0}$ wird verworfen	*Typ I Fehler*	korrekte Entscheidung $(1-\beta$: *"Power"*$)$

Typ I und Typ II Fehler, Power

Einfachster Test: $\mathrm{H}_0 : \theta = \theta_0$ vs. $\mathrm{H}_A : \theta = \theta_A$ (θ kann nur 2 Werte annehmen, θ_0 oder θ_A!)



Typ I und Typ II Fehler

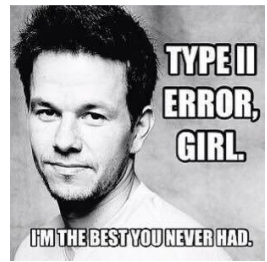
- **Typ I Fehler:** Wahrscheinlichkeit eine wahre Nullhypothese zu verwerfen (irrtümlicher Glaube an einen Zusammenhang, der tatsächlich *nicht* existiert)
  - ⇒ **Verwerfungsfehler**
    (*false positive:* Gesunder erhält Diagnose ‘krank’)

Typ I und Typ II Fehler

- **Typ I Fehler:** Wahrscheinlichkeit eine wahre Nullhypothese zu verwerfen (irrtümlicher Glaube an einen Zusammenhang, der tatsächlich *nicht* existiert)
  - ⇒ **Verwerfungsfehler**
    (*false positive:* Gesunder erhält Diagnose ‘krank’)
- **Typ II Fehler:** Wahrscheinlichkeit eine *falsche* Nullhypothese *nicht* zu verwerfen (irrtümlicher Glaube, dass *kein* Zusammenhang existiert, obwohl tatsächlich einer existiert)
  - ⇒ **Nichtverwerfungsfehler**
    (*false negative:* Kranker erhält Diagnose ‘gesund’)

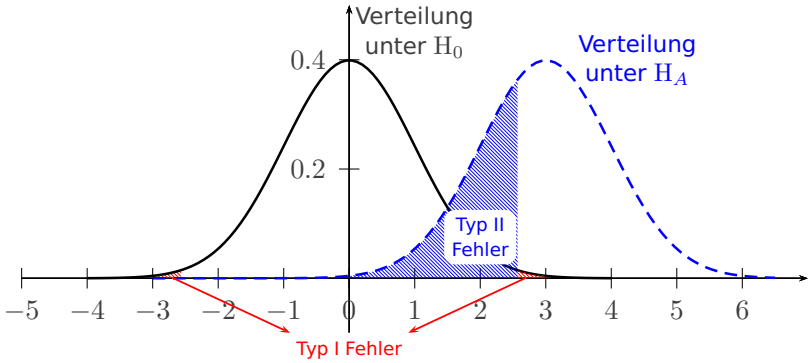
Typ I und Typ II Fehler

- **Typ I Fehler:** Wahrscheinlichkeit eine wahre Nullhypothese zu verwerfen (irrtümlicher Glaube an einen Zusammenhang, der tatsächlich *nicht* existiert)
  - ⇒ **Verwerfungsfehler**
    (*false positive:* Gesunder erhält Diagnose ‘krank’)
- **Typ II Fehler:** Wahrscheinlichkeit eine *falsche* Nullhypothese *nicht* zu verwerfen (irrtümlicher Glaube, dass *kein* Zusammenhang existiert, obwohl tatsächlich einer existiert)
  - ⇒ **Nichtverwerfungsfehler**
    (*false negative:* Kranker erhält Diagnose ‘gesund’)
  - *Verpasste Chance* ⇒
  - Größe hängt von konkreter Alternativhypothese und von Typ I Fehler ab!



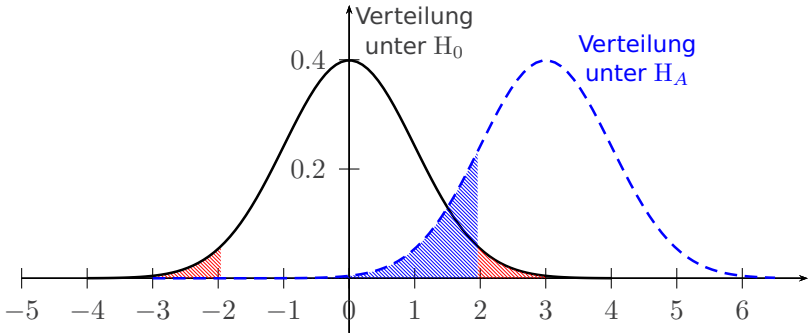
Typ I und Typ II Fehler: inverser Zusammenhang

Signifikanzniveau: $\alpha = 0.01$



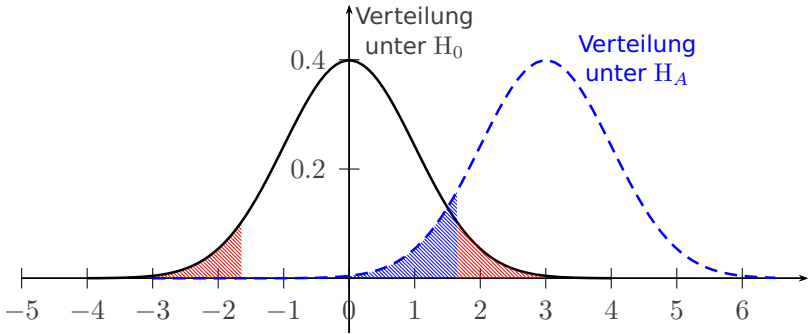
Typ I und Typ II Fehler: inverser Zusammenhang

Signifikanzniveau: $\alpha = 0.05$



Typ I und Typ II Fehler: inverser Zusammenhang

Signifikanzniveau: $\alpha = 0.1$



Power (Trennschärfe)

- Die **‘Power’** oder Trennschärfe eines Tests gibt die Wahrscheinlichkeit an, mit der ein tatsächlich vorhandener Unterschied auch aufgedeckt werden kann,

Power (Trennschärfe)

- Die **‘Power’** oder Trennschärfe eines Tests gibt die Wahrscheinlichkeit an, mit der ein tatsächlich vorhandener Unterschied auch aufgedeckt werden kann,
- in anderen Worten: die Wahrscheinlichkeit, mit der eine tatsächlich falsche Nullhypothese auch verworfen werden kann!

Power (Trennschärfe)

- Die **‘Power’** oder Trennschärfe eines Tests gibt die Wahrscheinlichkeit an, mit der ein tatsächlich vorhandener Unterschied auch aufgedeckt werden kann,
- in anderen Worten: die Wahrscheinlichkeit, mit der eine tatsächlich falsche Nullhypothese auch verworfen werden kann!
- intuitiv: ‘die Wahrscheinlichkeit, dass eine Chance *nicht* verpasst wird’.

Power (Trennschärfe)

- Die **‘Power’** oder Trennschärfe eines Tests gibt die Wahrscheinlichkeit an, mit der ein tatsächlich vorhandener Unterschied auch aufgedeckt werden kann,
- in anderen Worten: die Wahrscheinlichkeit, mit der eine tatsächlich falsche Nullhypothese auch verworfen werden kann!
- intuitiv: ‘die Wahrscheinlichkeit, dass eine Chance *nicht* verpasst wird’.
- Die Power eines Tests ist die Gegenwahrscheinlichkeit
  für das Auftreten eines Typ II Fehlers

$$\text{Power} = 1 - \Pr(\text{Typ II Fehler})$$

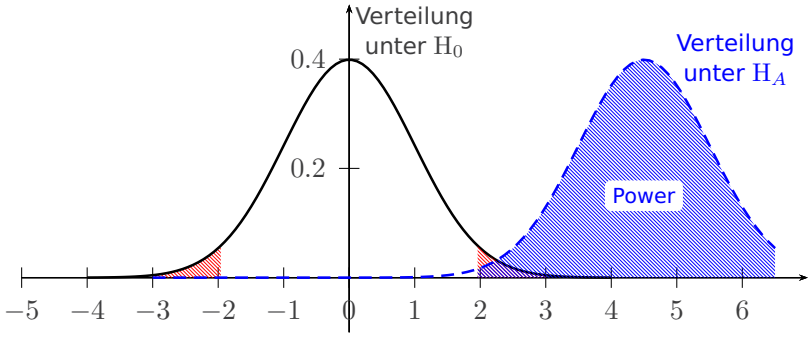
Power (Trennschärfe)

- Die **‘Power’** oder Trennschärfe eines Tests gibt die Wahrscheinlichkeit an, mit der ein tatsächlich vorhandener Unterschied auch aufgedeckt werden kann,
- in anderen Worten: die Wahrscheinlichkeit, mit der eine tatsächlich falsche Nullhypothese auch verworfen werden kann!
- intuitiv: ‘die Wahrscheinlichkeit, dass eine Chance *nicht* verpasst wird’.
- Die Power eines Tests ist die Gegenwahrscheinlichkeit
  für das Auftreten eines Typ II Fehlers

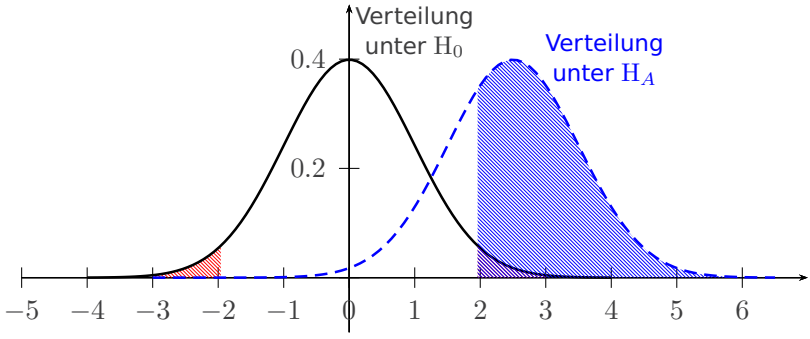
$$\text{Power} = 1 - \Pr(\text{Typ II Fehler})$$

- Verglichen wird die Verteilung unter Gültigkeit von H_0 mit der Verteilung unter Gültigkeit einer spezifischen Alternativhypothese.

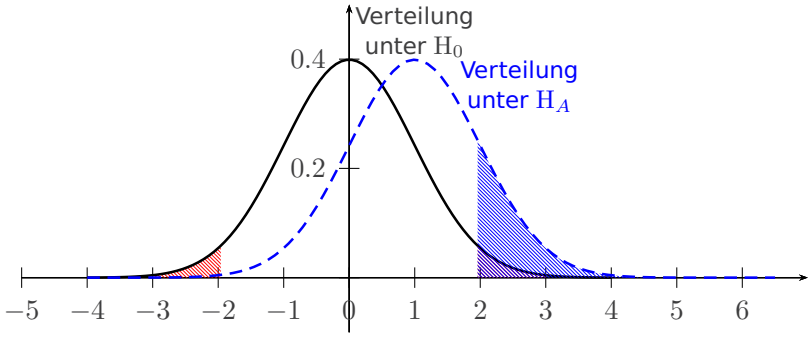
Power: verschiedene Alternativhypothesen



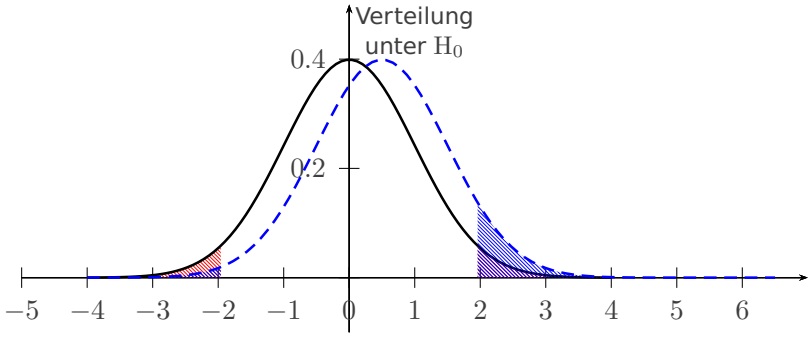
Power: verschiedene Alternativhypothesen



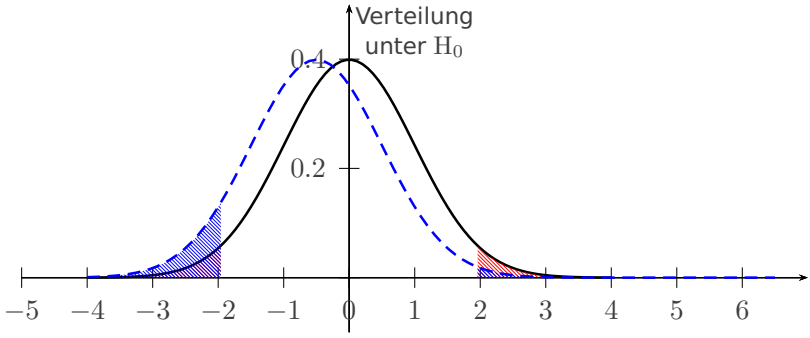
Power: verschiedene Alternativhypothesen



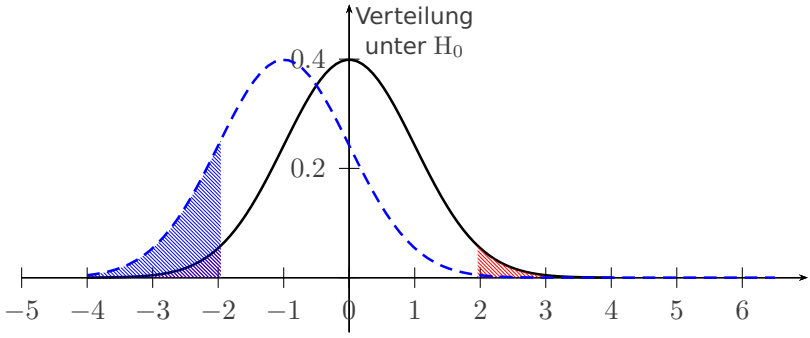
Power: verschiedene Alternativhypothesen



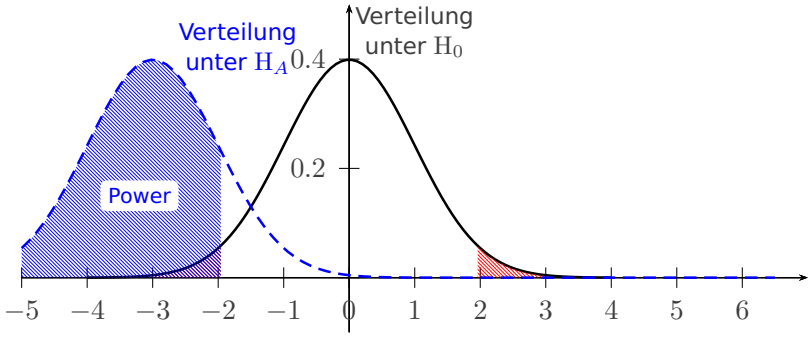
Power: verschiedene Alternativhypothesen



Power: verschiedene Alternativhypothesen



Power: verschiedene Alternativhypothesen



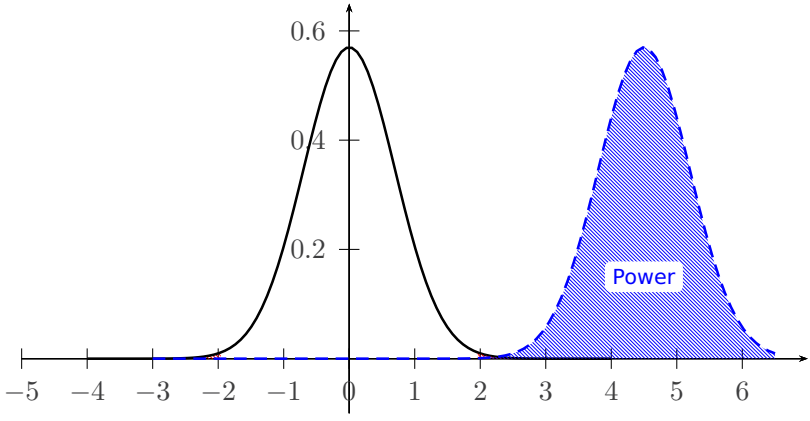
Power

- Die Power eines Tests ist *ceteris paribus* umso größer, je kleiner der Standardfehler ist

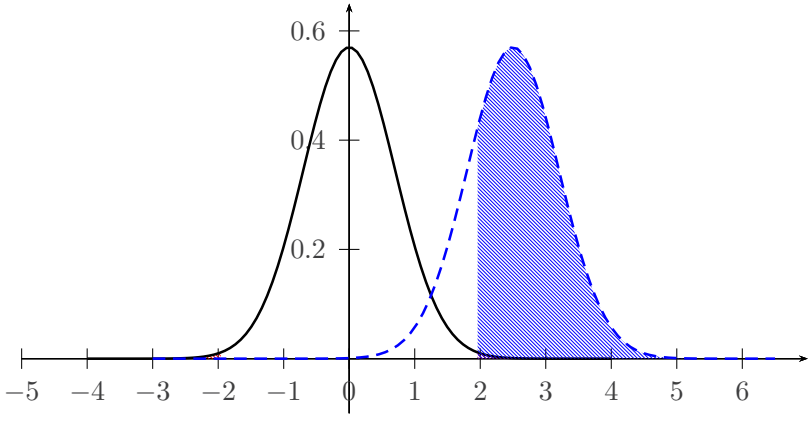
Power

- Die Power eines Tests ist *ceteris paribus* umso größer, je kleiner der Standardfehler ist
- d.h., Power eines Tests nimmt c.p. mit dem Stichprobenumfang zu!
  Umso größer eine Stichprobe ist, umso größer ist die Wahrscheinlichkeit, einen tatsächlich vorhandenen Effekt aufzuspüren.

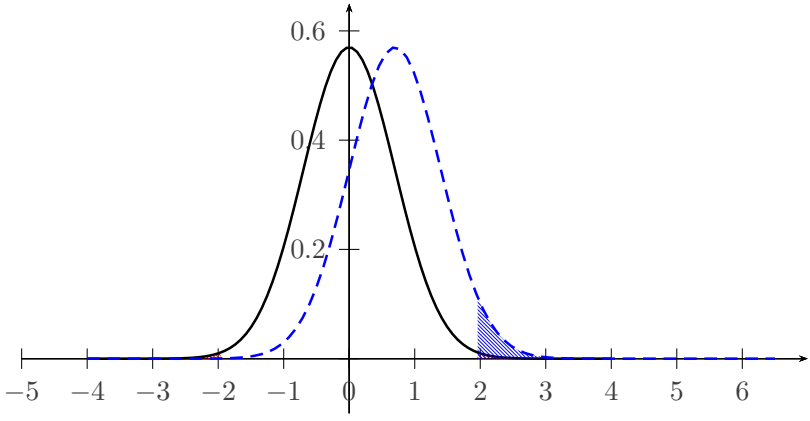
Power: alternative H_A mit größerer Stichprobe



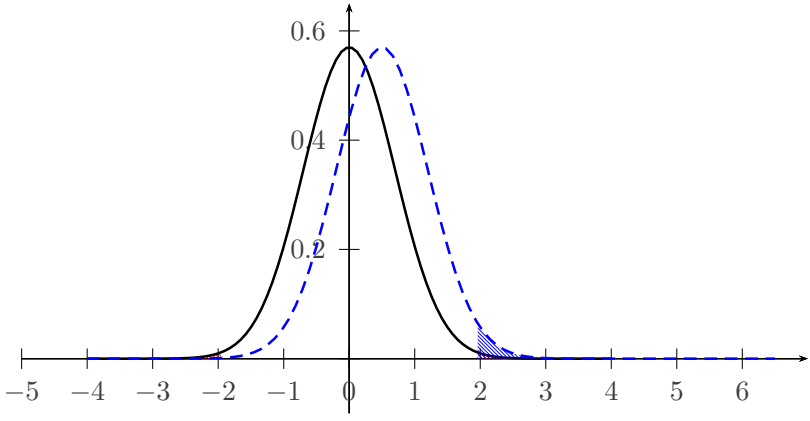
Power: alternative H_A mit größerer Stichprobe



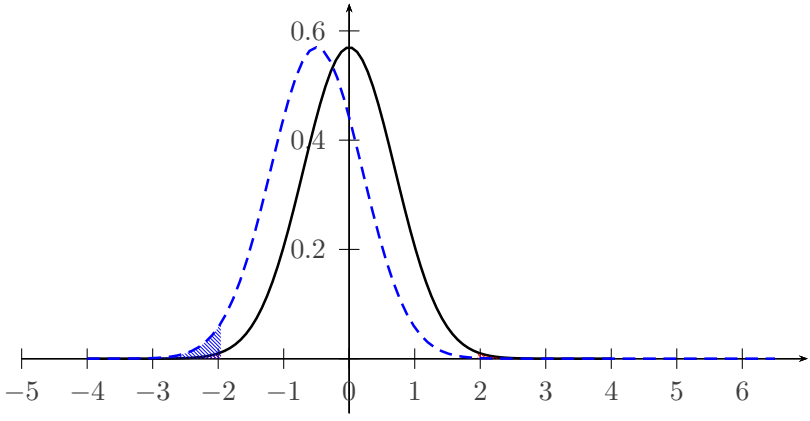
Power: alternative H_A mit größerer Stichprobe



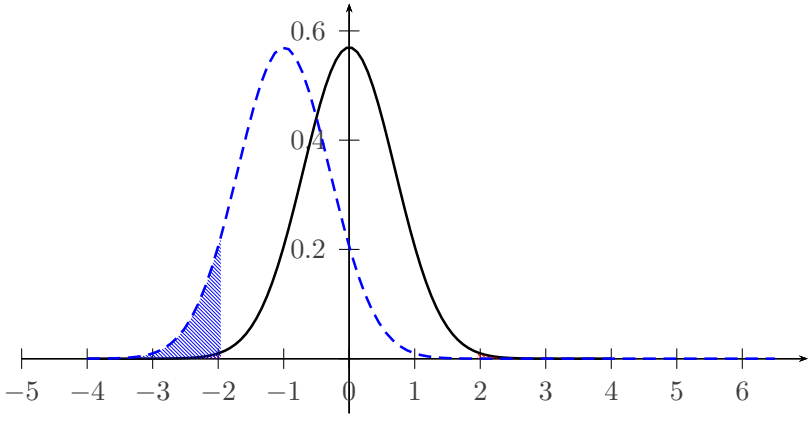
Power: alternative H_A mit größerer Stichprobe



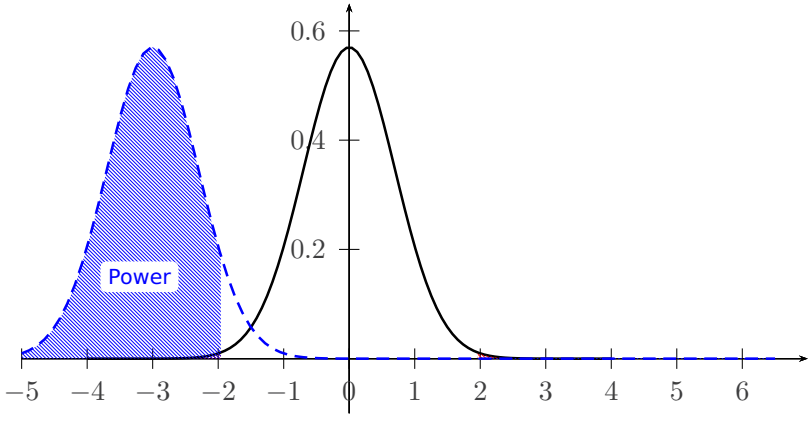
Power: alternative H_A mit größerer Stichprobe



Power: alternative H_A mit größerer Stichprobe

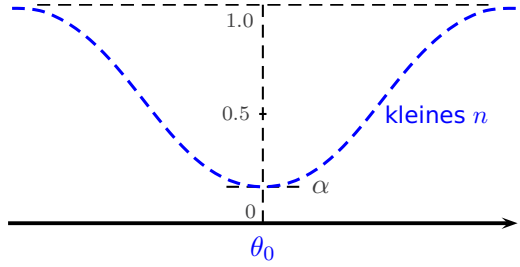


Power: alternative H_A mit größerer Stichprobe



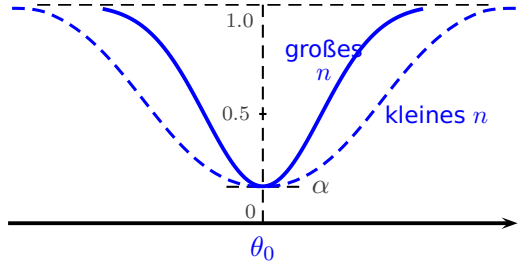
Power

Power-Funktionen eines mächtigeren und weniger mächtigen Tests (z.B. große vs. kleine Stichprobe):



Power

Power-Funktionen eines mächtigeren und weniger mächtigen Tests (z.B. große vs. kleine Stichprobe):



Zum Mitnehmen . . .

- Neyman-Pearson Signifikanztests beruhen auf der Vorgabe eines Signifikanzniveaus α und einer expliziten Alternativhypothese.

Zum Mitnehmen . . .

- Neyman-Pearson Signifikanztests beruhen auf der Vorgabe eines Signifikanzniveaus α und einer expliziten Alternativhypothese.
- Damit können Neyman-Pearson Signifikanztests eine einfache Entscheidungsregel liefern: akzeptiere oder verwirf H_0

Zum Mitnehmen . . .

- Neyman-Pearson Signifikanztests beruhen auf der Vorgabe eines Signifikanzniveaus α und einer expliziten Alternativhypothese.
- Damit können Neyman-Pearson Signifikanztests eine einfache Entscheidungsregel liefern: akzeptiere oder verwirf H_0
- Größter Vorteil: weitgehend akzeptierte Norm . . .
  (obwohl nicht unproblematisch, gilt z.B. nur in korrekt spezifizierten Modellen, siehe Gauss-Markov Annahmen)

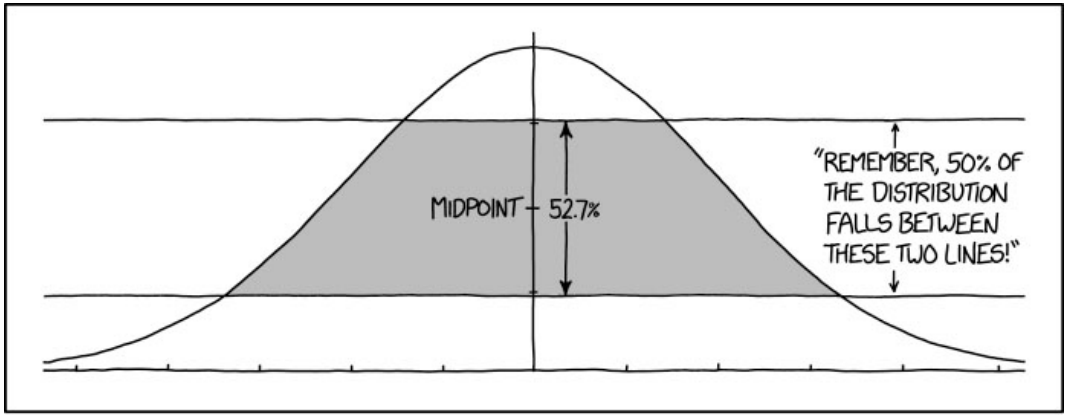
Zum Mitnehmen . . .

- Neyman-Pearson Signifikanztests beruhen auf der Vorgabe eines Signifikanzniveaus α und einer expliziten Alternativhypothese.
- Damit können Neyman-Pearson Signifikanztests eine einfache Entscheidungsregel liefern: akzeptiere oder verwirf H_0
- Größter Vorteil: weitgehend akzeptierte Norm . . .
  (obwohl nicht unproblematisch, gilt z.B. nur in korrekt spezifizierten Modellen, siehe Gauss-Markov Annahmen)
- Die Vorgabe eines Signifikanzniveaus α erlaubt die Definition von Typ I und Typ II Fehler sowie der Power

Zum Mitnehmen . . .

- Neyman-Pearson Signifikanztests beruhen auf der Vorgabe eines Signifikanzniveaus α und einer expliziten Alternativhypothese.
- Damit können Neyman-Pearson Signifikanztests eine einfache Entscheidungsregel liefern: akzeptiere oder verwirf H_0
- Größter Vorteil: weitgehend akzeptierte Norm . . .
  (obwohl nicht unproblematisch, gilt z.B. nur in korrekt spezifizierten Modellen, siehe Gauss-Markov Annahmen)
- Die Vorgabe eines Signifikanzniveaus α erlaubt die Definition von Typ I und Typ II Fehler sowie der Power
- Im übrigen gelten alle kritischen Anmerkungen wie für p-Werte nach Fisher!

XKCD: How to annoy a statistician



Quelle: XKCD, `https://xkcd.com/2118/`

Hypothesentests: Konfidenzintervalle

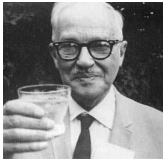
Grundlagen der Ökonometrie

herbert.stocker@uibk.ac.at

www.hsto.info/econometrics

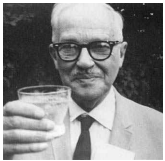
Konfidenzintervalle

- Jerzy Neyman (1934)
- Intervallschätzer: geben simultan Auskunft über ‘Genauigkeit’
- Beruhen auf den gleichen Prinzipien wie Hypothesentests:
  Für normalverteilte ZV $\widehat{\beta}_h \sim \mathrm{N}(\beta_h, \sigma^2_{\widehat{\beta}_h})$



$$\Pr\left(-1.96 < \frac{\widehat{\beta}_h - \beta_h}{\sigma_{\widehat{\beta}_h}} < +1.96\right) = 0.95$$

Konfidenzintervalle



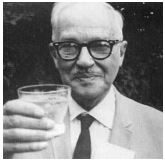
- Jerzy Neyman (1934)
- Intervallschätzer: geben simultan Auskunft über ‘Genauigkeit’
- Beruhen auf den gleichen Prinzipien wie Hypothesentests:
  Für normalverteilte ZV $\widehat{\beta}_h \sim \mathrm{N}(\beta_h, \sigma^2_{\widehat{\beta}_h})$

$$\Pr\left(-1.96 < \frac{\widehat{\beta}_h - \beta_h}{\sigma_{\widehat{\beta}_h}} < +1.96\right) = 0.95$$

- Umformen, so dass ZV $\widehat{\beta}_h$ in Unter- sowie Obergrenze und Parameter β_h im Zentrum steht

$$\Pr\left[\widehat{\beta}_h - 1.96\,\sigma_{\widehat{\beta}_h} \leq \beta_h \leq \widehat{\beta}_h + 1.96\,\sigma_{\widehat{\beta}_h}\right] = 0.95$$

Konfidenzintervalle



- Jerzy Neyman (1934)
- Intervallschätzer: geben simultan Auskunft über ‘Genauigkeit’
- Beruhen auf den gleichen Prinzipien wie Hypothesentests:
  Für normalverteilte ZV $\widehat{\beta}_h \sim \mathrm{N}(\beta_h, \sigma^2_{\widehat{\beta}_h})$

$$\Pr\left(-1.96 < \frac{\widehat{\beta}_h - \beta_h}{\sigma_{\widehat{\beta}_h}} < +1.96\right) = 0.95$$

- Umformen, so dass ZV $\widehat{\beta}_h$ in Unter- sowie Obergrenze und Parameter β_h im Zentrum steht

$$\Pr\left[\widehat{\beta}_h - 1.96\,\sigma_{\widehat{\beta}_h} \leq \beta_h \leq \widehat{\beta}_h + 1.96\,\sigma_{\widehat{\beta}_h}\right] = 0.95$$

- Wenn Parameter $\sigma_{\widehat{\beta}_h}$ durch Schätzfunktion $\hat{\sigma}_{\widehat{\beta}_h}$ ersetzt wird $\rightarrow$ t-Verteilung

Konfidenzintervalle

Beispiel: Konfidenzintervall für Regressionskoeffizenten

- $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_h x_{ih} + \cdots + \beta_k x_{ik} + \varepsilon_i$ folgt (Gosset, 1908)

$$\frac{\widehat{\beta}_h - \beta_h}{\hat{\sigma}_{\widehat{\beta}_h}} \sim \mathrm{t}_{n-k}$$

Konfidenzintervalle

Beispiel: Konfidenzintervall für Regressionskoeffizenten

- $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_h x_{ih} + \cdots + \beta_k x_{ik} + \varepsilon_i$ folgt (Gosset, 1908)

$$\frac{\widehat{\beta}_h - \beta_h}{\hat{\sigma}_{\widehat{\beta}_h}} \sim \mathrm{t}_{n-k}$$

- daraus folgt

$$\Pr\left[-t^{\text{crit}}_{\alpha/2,df} \leq \frac{\widehat{\beta}_h - \beta_h}{\hat{\sigma}_{\widehat{\beta}_h}} \leq +t^{\text{crit}}_{\alpha/2,df}\right] = 0.95$$

Konfidenzintervalle

Beispiel: Konfidenzintervall für Regressionskoeffizenten

- $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_h x_{ih} + \cdots + \beta_k x_{ik} + \varepsilon_i$ folgt (Gosset, 1908)

$$\frac{\widehat{\beta}_h - \beta_h}{\hat{\sigma}_{\widehat{\beta}_h}} \sim \mathrm{t}_{n-k}$$

- daraus folgt

$$\Pr\left[-t^{\text{crit}}_{\alpha/2,df} \leq \frac{\widehat{\beta}_h - \beta_h}{\hat{\sigma}_{\widehat{\beta}_h}} \leq +t^{\text{crit}}_{\alpha/2,df}\right] = 0.95$$

- Theoretisches Konfidenzintervall

$$\Pr\left[\widehat{\beta}_h - t^{\text{crit}}_{\alpha/2,df} \cdot \hat{\sigma}_{\widehat{\beta}_h} \;\leq\; \beta_h \;\leq\; \widehat{\beta}_h + t^{\text{crit}}_{\alpha/2,df} \cdot \hat{\sigma}_{\widehat{\beta}_h}\right] = 1 - \alpha$$

Konfidenzintervalle

Beispiel: Konfidenzintervall für Regressionskoeffizenten

- $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_h x_{ih} + \cdots + \beta_k x_{ik} + \varepsilon_i$ folgt (Gosset, 1908)

$$\frac{\widehat{\beta}_h - \beta_h}{\hat{\sigma}_{\widehat{\beta}_h}} \sim \mathrm{t}_{n-k}$$

- daraus folgt

$$\Pr\left[-t^{\text{crit}}_{\alpha/2,df} \leq \frac{\widehat{\beta}_h - \beta_h}{\hat{\sigma}_{\widehat{\beta}_h}} \leq +t^{\text{crit}}_{\alpha/2,df}\right] = 0.95$$

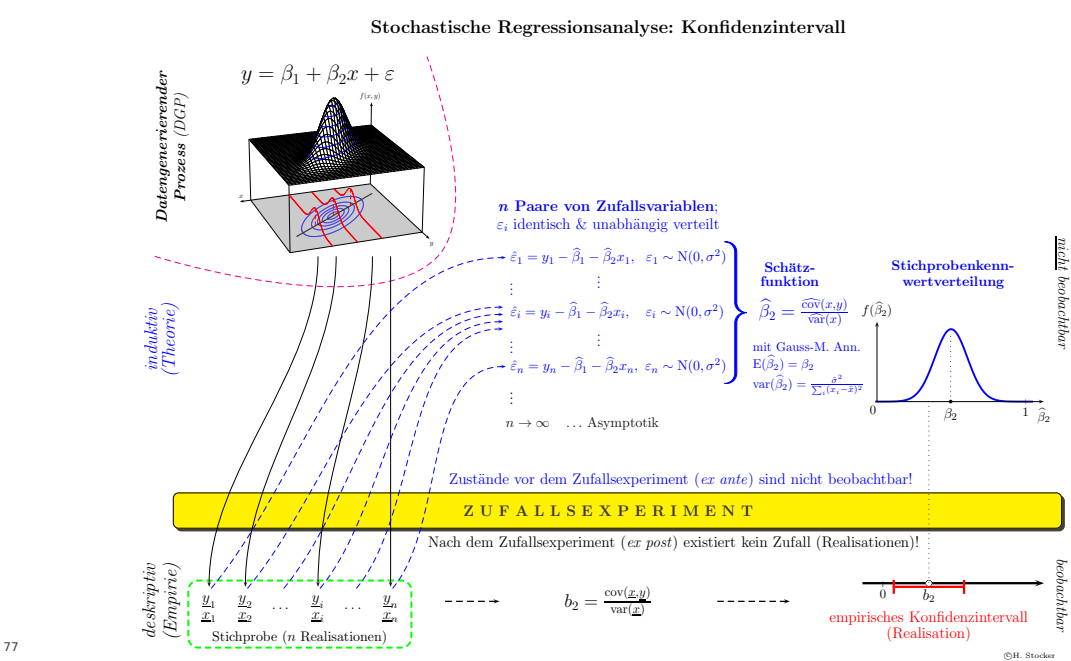
- Theoretisches Konfidenzintervall

$$\Pr\left[\widehat{\beta}_h - t^{\text{crit}}_{\alpha/2,df} \cdot \hat{\sigma}_{\widehat{\beta}_h} \;\leq\; \beta_h \;\leq\; \widehat{\beta}_h + t^{\text{crit}}_{\alpha/2,df} \cdot \hat{\sigma}_{\widehat{\beta}_h}\right] = 1 - \alpha$$

- Empirisches Konfidenzintervall (Realisation)

$$\left[b_h - t^{\text{crit}}_{\alpha/2,df}\, s_{\widehat{\beta}_h},\;\; b_h + t^{\text{crit}}_{\alpha/2,df}\, s_{\widehat{\beta}_h}\right] \quad \text{bzw.} \quad b_h \pm t^{\text{crit}}_{\alpha/2,df}\, s_{\widehat{\beta}_h}$$

Stochastische Regressionsanalyse: Konfidenzintervall



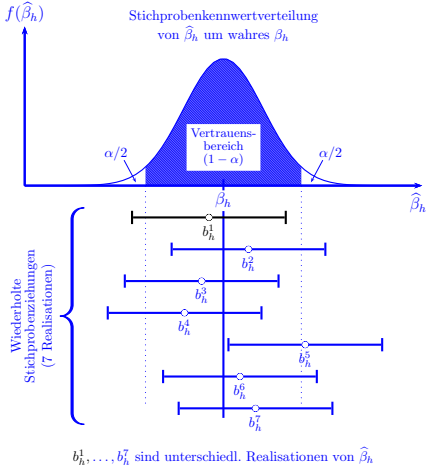
©H. Stocker

Konfidenzintervalle (CI)

Interpretation: *‘repeated sampling’*

Wenn wir – hypothetisch – unendlich viele unabhängige Stichproben ziehen würden, und für jede dieser Stichproben ein 95% Konfidenzintervall (CI) berechnen würden, so könnten wir darauf vertrauen, dass 95% dieser Konfidenzintervalle den wahren Parameter β_h überdecken werden.

Eine Schätzung ist umso genauer, je enger das CI ist!



Konfidenzintervalle (CI)

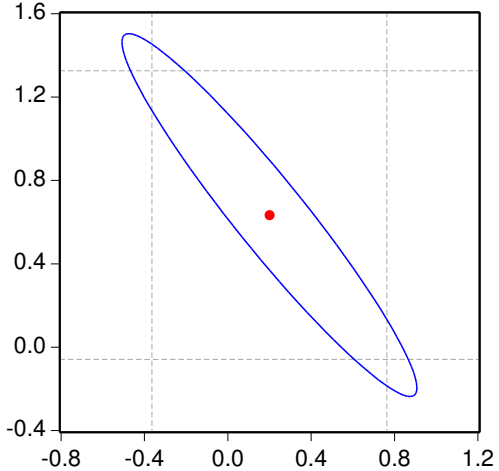
Ein Konfidenzintervall ist *ceteris paribus* umso *enger*, . . .

- je kleiner der Standardfehler ist, d.h.
  - umso kleiner die Varianz der Störterme σ^2 ist,
  - je größer die Stichprobe ist (auch über kritischen Wert $t^{\text{crit}}_{\alpha/2,df}$).
  - umso größer die Streuung der x (d.h. $\sum_i (x_i - \bar{x})^2$) ist,
  - umso kleiner die Korrelation zwischen den Regressoren ist,

- je kleiner das frei gewählte Konfidenzniveau $1 - \alpha$ ist (d.h. ein 90% Konfidenzintervall ist ceteris paribus enger als ein 99% Konfidenzintervall). Dieser Effekt wirkt sich über den kritischen Wert $t^{\text{crit}}_{\alpha/2,df}$ auf die Breite aus.

Konfidenzintervalle (CI)

Konfidenzellipsen für zwei Regressionskoeffizienten:



Konfidenzintervalle (CI)

Dualität zwischen Konfidenzintervallen und Hypothesentests:

- Konfidenzintervalle und Hypothesentests beruhen auf der gleichen Information und führen zu gleichen Schlussfolgerungen!
- Ein Konfidenzintervall ist die Menge aller Punkte, für die die entsprechenden Nullhypothese *nicht* abgelehnt werden kann; bzw. etwas genauer
- ein Konfidenzintervall ist ein Intervall auf $\mathbb{R}$, welches alle reellen Zahlen θ_0 enthält, für die die Nullhypothese $\theta = \theta_0$ durch einen entsprechenden Hypothesentest *nicht* abgelehnt werden kann.
- d.h., wenn das CI die H_0 überdeckt kann die H_0 *nicht* abgelehnt werden $\rightarrow$ wir lernen nichts Neues :-(

  wenn H_0 außerhalb CI $\rightarrow$:-)

  (das gilt analog auch für Konfidenzellipsen)

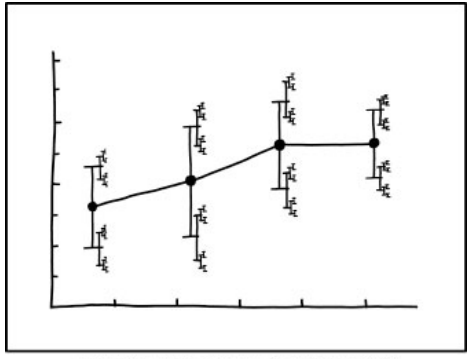
Hypothesentests

Hinweis: Für einen kritischen Überblick über die Verwendung von Indikatoren für die Unsicherheit von Schätzungen siehe

Imbens, Guido W. (2021) *“Statistical Significance, p-Values, and the Reporting of Uncertainty”* Journal of Economic Perspectives, 35 (3),
157-74 (Wirtschaftsnobelpreis 2021)

`https://www.aeaweb.org/articles?id=10.1257/jep.35.3.157`.

Konfidenzintervalle



Quelle: `https://xkcd.com/2110/`

Hypothesentests: F-Tests

Grundlagen der Ökonometrie

herbert.stocker@uibk.ac.at

www.hsto.info/econometrics

F-Tests

- Bisher wurde eine einzelne lineare Hypothese mittels t-Test getestet.
- Es können auch mehrere lineare Hypothesen *simultan* getestet werden $\rightarrow \mathrm{F}$-Test

F-Tests

- Bisher wurde eine einzelne lineare Hypothese mittels t-Test getestet.
- Es können auch mehrere lineare Hypothesen *simultan* getestet werden $\rightarrow$ F-Test
  - F-total-Statistik: *simultane* Signifikanz aller *Steigungs*koeffizienten
  - allgemeine F-Tests: beliebige lineare Tests
  - Chow-Tests auf Strukturbruch: stammen zwei Teilgruppen aus der gleichen Grundgesamtheit?

F-Tests

- Bisher wurde eine einzelne lineare Hypothese mittels t-Test getestet.
- Es können auch mehrere lineare Hypothesen *simultan* getestet werden
  $\rightarrow \mathrm{F}$-Test
  - F-total-Statistik: *simultane* Signifikanz aller *Steigungs*koeffizienten
  - allgemeine F-Tests: beliebige lineare Tests
  - Chow-Tests auf Strukturbruch: stammen zwei Teilgruppen aus der gleichen Grundgesamtheit?
- alles bisher Gesagte zu Tests bleibt gültig!
  (p-Werte, Neyman-Pearson mit Typ I & Typ II Fehler, Power, . . .)

F-total-Statistik

- ein Test auf *gemeinsame* Signifikanz *aller Steigungs*koeffizienten

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots \beta_k x_{ik} + \varepsilon_i$$

F-total-Statistik

- ein Test auf *gemeinsame* Signifikanz *aller Steigungs*koeffizienten

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots \beta_k x_{ik} + \varepsilon_i$$

- Startpunkt: Varianzzerlegung

$$\underbrace{\sum (y_i - \bar{y})^2}_{\text{TSS}} = \underbrace{\sum (\hat{y}_i - \bar{y})^2}_{\text{ESS}} + \underbrace{\sum \hat{\varepsilon}_i^2}_{\text{SSR}}$$

F-total-Statistik

- ein Test auf *gemeinsame* Signifikanz *aller Steigungs*koeffizienten

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots \beta_k x_{ik} + \varepsilon_i$$

- Startpunkt: Varianzzerlegung

$$\underbrace{\sum (y_i - \bar{y})^2}_{\text{TSS}} = \underbrace{\sum (\hat{y}_i - \bar{y})^2}_{\text{ESS}} + \underbrace{\sum \hat{\varepsilon}_i^2}_{\text{SSR}}$$

- Ergebnis der theoretischen Statistik:
  das Verhältnis zweier *unabhängig* χ^2-verteilten ZV (V_1, V_2), jeweils dividiert durch die Freiheitsgrade (ν_1, ν_2), ist F-verteilt:

$$F = \frac{V_1/\nu_1}{V_2/\nu_2} \sim \mathrm{F}_{\nu_1,\nu_2}$$

F-total-Statistik

- **ANOVA-Tafel** (*'ANalysis Of VAriance Table'*)

	Sum of Squares	df	Mean Square
Regression	ESS	$k-1$	ESS/$(k-1)$
Residuen	SSR	$n-k$	SSR/$(n-k)$
Total	TSS	$n-1$	

F-total-Statistik

- **ANOVA-Tafel** (*'ANalysis Of VAriance Table'*)

	Sum of Squares	df	Mean Square
Regression	ESS	$k-1$	ESS/$(k-1)$
Residuen	SSR	$n-k$	SSR/$(n-k)$
Total	TSS	$n-1$	

- H_0: $\beta_2 = \beta_3 = \cdots = \beta_k = 0$

  d.h. wenn H_0 wahr: $y = \beta_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \; = \; \beta_1 + \varepsilon$

F-total-Statistik

- **ANOVA-Tafel** (*'ANalysis Of VAriance Table'*)

	Sum of Squares	df	Mean Square
Regression	ESS	$k-1$	ESS/$(k-1)$
Residuen	SSR	$n-k$	SSR/$(n-k)$
Total	TSS	$n-1$	

- H_0: $\beta_2 = \beta_3 = \cdots = \beta_k = 0$

  d.h. wenn H_0 wahr: $y = \beta_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \; = \; \beta_1 + \varepsilon$

- Also

$$F = \frac{\frac{\mathrm{ESS}}{k-1}}{\frac{\mathrm{SSR}}{n-k}} \overset{\mathrm{H}_0}{\sim} \mathrm{F}(k-1, n-k)$$

F-total-Statistik

- Empirische Teststatistik:

$$F^{\text{emp}} = \frac{\frac{\text{ESS}}{k-1}}{\frac{\text{SSR}}{n-k}}$$

F-total-Statistik

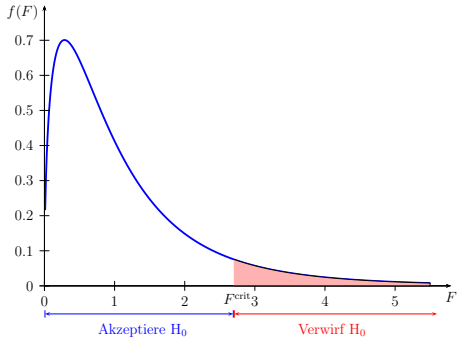
- Empirische Teststatistik:

$$F^{\text{emp}} = \frac{\frac{\text{ESS}}{k-1}}{\frac{\text{SSR}}{n-k}}$$

- Wenn gemeinsamer Erklärungsbeitrag der Regressoren gering (groß): $\text{ESS}/(k-1)$ klein (groß) im Verhältnis zu $\text{SSR}/(n-k) \Rightarrow F^{\text{emp}}$ klein (groß) !

Autobeispiel:

	Preis
Constant	$22\,692.730^{***}$
	(310.012)
Alter	$-2\,017.964^{***}$
	(175.375)
km	-0.023^{***}
	(0.006)
Observations	60
R^2	0.929
F Statistic	371.300^{***} (df = 2; 57)
p-value	$< 2.2e\text{-}16$
Note:	*p<0.1; **p<0.05; ***p<0.01



F – Tabelle

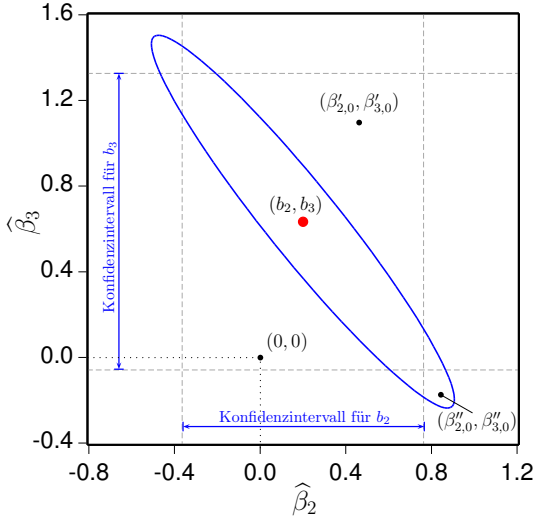
Tabelle: $F_{(q_1,q_2)}$-Verteilung für ein Signifikanzniveau von 5%

Nenner-	Zähler Freiheitsgrade						
Freiheitsgr.	1	2	3	5	10	15	30
1	161.4	199.5	215.7	230.2	241.9	245.9	250.1
2	18.51	19.00	19.16	19.30	19.40	19.43	19.46
3	10.13	9.552	9.277	9.013	8.786	8.703	8.617
4	7.709	6.944	6.591	6.256	5.964	5.858	5.746
5	6.608	5.786	5.409	5.050	4.735	4.619	4.496
6	5.987	5.143	4.757	4.387	4.060	3.938	3.808
7	5.591	4.737	4.347	3.972	3.637	3.511	3.376
8	5.318	4.459	4.066	3.687	3.347	3.218	3.079
9	5.117	4.256	3.863	3.482	3.137	3.006	2.864
10	4.965	4.103	3.708	3.326	2.978	2.845	2.700
15	4.543	3.682	3.287	2.901	2.544	2.403	2.247
20	4.351	3.493	3.098	2.711	2.348	2.203	2.039
30	4.171	3.316	2.922	2.534	2.165	2.015	1.841
50	4.034	3.183	2.790	2.400	2.026	1.871	1.687
55	4.016	3.165	2.773	2.383	2.008	1.852	1.666
57	4.010	*3.159*	2.766	2.377	2.001	1.846	1.659
60	4.001	3.150	2.758	2.368	1.993	1.836	1.649
80	3.960	3.111	2.719	2.329	1.951	1.793	1.602
100	3.936	3.087	2.696	2.305	1.927	1.768	1.573
1000	3.851	3.005	2.614	2.223	1.840	1.676	1.471

Für $k = 3$ und $n = 60$: $k - 1 = 2,\ n - k = 57$: $\mathrm{F}^{\mathrm{crit}}_{0.05,2,57} = 3.159$

R: `qf(p = 0.95, df1 = 2, df2 = 57) ## 3.158843`

Einzelhypothesen vs. gemeinsame Hypothese: (Konfidenzellipse)



$\mathbb{F}$ Statistiken: allgemeine lineare Restriktionen

- Geschachtelte Hypothesen (‘*nested tests*’): Vergleich zweier Modelle, wobei eines *durch lineare Parameterrestriktionen* in das andere übergeführt werden kann.

$\mathbb{F}$ Statistiken: allgemeine lineare Restriktionen

- Geschachtelte Hypothesen (‘*nested tests*’): Vergleich zweier Modelle, wobei eines *durch lineare Parameterrestriktionen* in das andere übergeführt werden kann.
- Nullhypothese: lineare Parameterrestriktion(en), z.B. für

  PRF: $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$

F Statistiken: allgemeine lineare Restriktionen

- Geschachtelte Hypothesen (‘*nested tests*’): Vergleich zweier Modelle, wobei eines *durch lineare Parameterrestriktionen* in das andere übergeführt werden kann.
- Nullhypothese: lineare Parameterrestriktion(en), z.B. für

  PRF: $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$

  - $\mathrm{H}_0\colon \quad \beta_2 = 0 \quad$ und $\quad \beta_3 = 0$
    restricted: $\Rightarrow \; y = \beta_1 + \varepsilon^*, \qquad (q = 2)$

$\mathbb{F}$ Statistiken: allgemeine lineare Restriktionen

- Geschachtelte Hypothesen (‘*nested tests*’): Vergleich zweier Modelle, wobei eines *durch lineare Parameterrestriktionen* in das andere übergeführt werden kann.
- Nullhypothese: lineare Parameterrestriktion(en), z.B. für

  PRF: $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$

  - $\mathrm{H}_0: \quad \beta_2 = 0 \quad$ und $\quad \beta_3 = 0$
    restricted: $\Rightarrow \; y = \beta_1 + \varepsilon^*, \qquad (q = 2)$
  - $\mathrm{H}_0: \; \beta_2 = \beta_3 \;$ $(q = 1)$
    restricted: $y = \beta_1 + \beta_2 (x_2 + x_3) + \varepsilon^*, \qquad (q = 1)$

F Statistiken: allgemeine lineare Restriktionen

- Geschachtelte Hypothesen (‘*nested tests*’): Vergleich zweier Modelle, wobei eines *durch lineare Parameterrestriktionen* in das andere übergeführt werden kann.
- Nullhypothese: lineare Parameterrestriktion(en), z.B. für

  PRF: $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$

  - $\mathrm{H}_0\colon \quad \beta_2 = 0 \quad$ und $\quad \beta_3 = 0$
    restricted: $\Rightarrow \; y = \beta_1 + \varepsilon^*, \qquad (q = 2)$
  - $\mathrm{H}_0: \; \beta_2 = \beta_3 \;\; (q = 1)$
    restricted: $y = \beta_1 + \beta_2 (x_2 + x_3) + \varepsilon^*, \qquad (q = 1)$
  - $\mathrm{H}_0: \; \beta_2 + \beta_3 = 1$
    restricted: $y - x_3 = \beta_1 + \beta_2 (x_2 - x_3) + \varepsilon^*, \qquad (q = 1)$

$\mathbb{F}$ Statistiken: allgemeine lineare Restriktionen

- Geschachtelte Hypothesen (‘*nested tests*’): Vergleich zweier Modelle, wobei eines *durch lineare Parameterrestriktionen* in das andere übergeführt werden kann.
- Nullhypothese: lineare Parameterrestriktion(en), z.B. für

  PRF: $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$

  - H_0: $\beta_2 = 0$ und $\beta_3 = 0$
    restricted: $\Rightarrow\ y = \beta_1 + \varepsilon^*, \qquad (q = 2)$
  - H_0 : $\beta_2 = \beta_3$ $(q = 1)$
    restricted: $y = \beta_1 + \beta_2 (x_2 + x_3) + \varepsilon^*, \qquad (q = 1)$
  - H_0 : $\beta_2 + \beta_3 = 1$
    restricted: $y - x_3 = \beta_1 + \beta_2 (x_2 - x_3) + \varepsilon^*, \qquad (q = 1)$

- Vergleich der SSR des ‘restringierten’ Modells (SST_r) mit der SSR des ‘*nicht* restringierten’ Modells (SST_u)

F Statistiken: allgemeine lineare Restriktionen

- Theoretische Teststatistik:

$$\hat{F} = \frac{(\mathsf{SSR}_r - \mathsf{SSR}_u)/q}{\mathsf{SSR}_u/(n-k)} \overset{\mathrm{H}_0}{\sim} \mathrm{F}_{q,n-k}$$

mit

F Statistiken: allgemeine lineare Restriktionen

- Theoretische Teststatistik:

$$\hat{F} = \frac{(\mathsf{SSR}_r - \mathsf{SSR}_u)/q}{\mathsf{SSR}_u/(n-k)} \overset{\mathrm{H}_0}{\sim} \mathrm{F}_{q,n-k}$$

mit

- SSR_r: Quadratsumme der Residuen des *restringierten* Modells,

F Statistiken: allgemeine lineare Restriktionen

- Theoretische Teststatistik:

$$\hat{F} = \frac{(\mathrm{SSR}_r - \mathrm{SSR}_u)/q}{\mathrm{SSR}_u/(n-k)} \overset{\mathrm{H}_0}{\sim} \mathrm{F}_{q,n-k}$$

mit

- SSR_r: Quadratsumme der Residuen des *restringierten* Modells,
- SSR_u: Quadratsumme der Residuen des *nicht restringierten* Modells (‘*unrestricted*’),

F Statistiken: allgemeine lineare Restriktionen

- Theoretische Teststatistik:

$$\hat{F} = \frac{(\mathsf{SSR}_r - \mathsf{SSR}_u)/q}{\mathsf{SSR}_u/(n-k)} \overset{\mathrm{H}_0}{\sim} \mathrm{F}_{q,n-k}$$

mit

- SSR_r: Quadratsumme der Residuen des *restringierten* Modells,
- SSR_u: Quadratsumme der Residuen des *nicht restringierten* Modells (‘*unrestricted*’),
- q: Zählerfreiheitsgrade, Anzahl der Restriktionen (Anzahl ‘=’ Zeichen)

F Statistiken: allgemeine lineare Restriktionen

- Theoretische Teststatistik:

$$\hat{F} = \frac{(\mathrm{SSR}_r - \mathrm{SSR}_u)/q}{\mathrm{SSR}_u/(n-k)} \overset{\mathrm{H}_0}{\sim} \mathrm{F}_{q,n-k}$$

mit

- SSR_r: Quadratsumme der Residuen des *restringierten* Modells,
- SSR_u: Quadratsumme der Residuen des *nicht restringierten* Modells (‘*unrestricted*’),
- q: Zählerfreiheitsgrade, Anzahl der Restriktionen (Anzahl ‘=’ Zeichen)
- $n-k$: Nennerfreiheitsgrade

F Statistiken: allgemeine lineare Restriktionen

- Theoretische Teststatistik:

$$\hat{F} = \frac{(\mathrm{SSR}_r - \mathrm{SSR}_u)/q}{\mathrm{SSR}_u/(n-k)} \overset{\mathrm{H}_0}{\sim} \mathrm{F}_{q,n-k}$$

  mit
  - SSR_r: Quadratsumme der Residuen des *restringierten* Modells,
  - SSR_u: Quadratsumme der Residuen des *nicht restringierten* Modells (‘*unrestricted*’),
  - q: Zählerfreiheitsgrade, Anzahl der Restriktionen (Anzahl ‘=’ Zeichen)
  - $n-k$: Nennerfreiheitsgrade
- Große F^{emp}-Statistik ist Evidenz gegen H_0

  H_0 wird verworfen, wenn F^{emp} in Verwerfungsbereich fällt!

F Statistiken: Cobb-Douglas Beispiel $H_0 : \alpha = \beta$

	log(Q)	
	unrestricted	restricted
Constant	2.481^{***}	2.390^{***}
	(0.129)	(0.307)
log(K)	0.640^{***}	
	(0.035)	
log(L)	0.257^{***}	
	(0.027)	
I(log(K) + log(L))		0.380^{***}
		(0.058)
Observations	25	25
R^2	0.942	0.650
Sum of Squared Resid.	0.187	1.122
F Statistic	177.220^{***} (df = 2; 22)	42.626^{***} (df = 1; 23)
Note:		*p<0.1; **p<0.05; ***p<0.01

F Statistiken: Cobb-Douglas Beispiel $H_0 : \alpha = \beta$

	log(Q)	
	unrestricted	restricted
Constant	2.481^{***}	2.390^{***}
	(0.129)	(0.307)
log(K)	0.640^{***}	
	(0.035)	
log(L)	0.257^{***}	
	(0.027)	
I(log(K) + log(L))		0.380^{***}
		(0.058)
Observations	25	25
R^2	0.942	0.650
Sum of Squared Resid.	0.187	1.122
F Statistic	177.220^{***} (df = 2; 22)	42.626^{***} (df = 1; 23)
Note:		*p<0.1; **p<0.05; ***p<0.01

$$F^{\text{emp}} = \frac{(\text{SSR}_r - \text{SSR}_u)/q}{\text{SSR}_u/(n-k)} = \frac{(1.122 - 0.187)/1}{0.187/(25-3)} = 109.93$$

Kritischer Wert: $F^{\text{crit}}_{0.05,1,22} = 4.301 \quad \Rightarrow$ Nullhypothese verwerfen!

$\mathbb{F}$ Statistiken: allgemeine lineare Restriktionen

Beispiel: Cobb-Douglas Produktionsfunktion $Q = AL^{\alpha}K^{\beta}$, $\mathrm{H}_0 \colon \alpha = \beta$

: z.B. `anova(equ_r, equ_u)` (← equation-Objekte!)

```
cd <- read.csv("https://www.uibk.ac.at/econometrics/data/prodfunkt.csv")

eq_u <- lm(log(Q) ~ log(K) + log(L), data = cd)    ## 2 Regressoren
eq_r <- lm(log(Q) ~ I(log(K) + log(L)), data = cd) ## 1 Regressor!
anova(eq_r, eq_u)

## Analysis of Variance Table
## Model 1: log(Q) ~ I(log(K) + log(L))
## Model 2: log(Q) ~ log(K) + log(L)
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     23 1.12208
## 2     22 0.18711  1   0.93497 109.93 5.067e-10 ***
```

F Statistiken: allgemeine lineare Restriktionen

Beispiel: Cobb-Douglas Produktionsfunktion $Q = AL^{\alpha}K^{\beta}$, $\mathrm{H}_0 \colon \alpha = \beta$

: Alternativ

```
library(car)
linearHypothesis(eq_u, c("log(K) = log(L)"))
##
## Linear hypothesis test
## Hypothesis:
## log(K) - log(L) = 0
##
## Model 1: restricted model
## Model 2: log(Q) ~ log(K) + log(L)
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     23 1.12208
## 2     22 0.18711  1   0.93497 109.93 5.067e-10 ***
```

$\mathbb{F}$ Statistiken: allgemeine lineare Restriktionen

Beispiel: Cobb-Douglas Produktionsfunktion $Q = AL^{\alpha}K^{\beta}$, $\mathbf{H}_0 \colon \alpha = \beta$

```
F = 109.93, p-value = 5.067e-10 ***
```

Interpretation: Wenn die H_0 wahr ist (d.h. sich die Koeffizienten in der Grundgesamtheit nicht unterscheiden) und wenn alle Annahmen erfüllt sind (u.a. das Modell richtig spezifiziert ist), dann ist es bei wiederholten Stichprobenziehungen extrem unwahrscheinlich neuerlich einen derart extremen Wert der Teststatistik zu erhalten!

$\Rightarrow$ wir verwerfen die Nullhypothese $\mathrm{H}_0 \colon \alpha = \beta$ auf dem 1% Niveau!

F Statistiken: Chow Strukturbruchtest

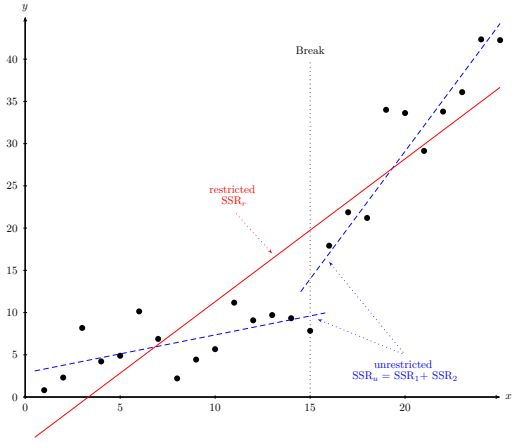
Chow Test: unterscheiden sich die Regressionskoeffizienten zwischen zwei Gruppen?

1. getrennte Regressionen für beide Gruppen $\rightarrow$ *unrestricted*

$$\text{SSR}_u = \text{SSR}_1 + \text{SSR}_2$$

2. eine Regression über alle Beobachtungen: $\text{SSR}_r \rightarrow$ *restricted*
3. Chow Teststatistik:

$$\hat{F} = \frac{(\text{SSR}_r - \text{SSR}_u)/k}{\text{SSR}_u/(n-2k)} \overset{\text{H}_0}{\sim} \text{F}_{k,n-2k}$$



F Statistiken: Chow Strukturbruchtest

Beispiel: Chow Test (s. Abbildung vorne)

	unrestr. 1	unrestr. 2	restricted
Constant	2.863^{*}	-22.398^{**}	-5.627^{**}
	(1.424)	(7.804)	(2.377)
x	0.449^{**}	2.616^{***}	1.692^{***}
	(0.157)	(0.377)	(0.160)
Observations	15	10	25
Sum of Squared Resid.	89.317	93.815	764.471
F Statistic	8.224^{**}	48.134^{***}	111.920^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01

F Statistiken: Chow Strukturbruchtest

Beispiel: Chow Test (s. Abbildung vorne)

	unrestr. 1	unrestr. 2	restricted
Constant	2.863*	−22.398**	−5.627**
	(1.424)	(7.804)	(2.377)
x	0.449**	2.616***	1.692***
	(0.157)	(0.377)	(0.160)
Observations	15	10	25
Sum of Squared Resid.	89.317	93.815	764.471
F Statistic	8.224**	48.134***	111.920***
Note:			*p<0.1; **p<0.05; ***p<0.01

$$F^{\text{emp}} = \frac{(\text{SSR}_r - \text{SSR}_u)/k}{\text{SSR}_u/(n-2k)} = \frac{(764.471 - (89.317 + 93.815))/2}{(89.317 + 93.815)/(25-4)} = 33.331$$

F Statistiken: Chow Strukturbruchtest

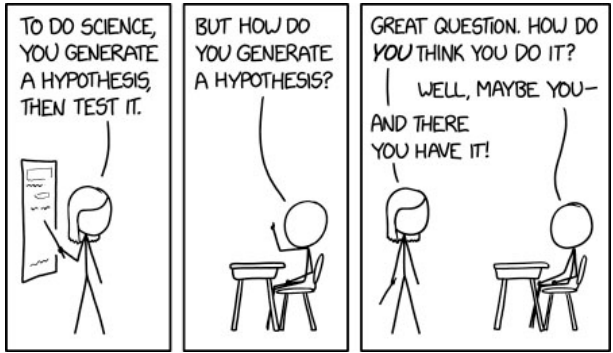
Beispiel: Chow Test (s. Abbildung vorne)

	unrestr. 1	unrestr. 2	restricted
Constant	2.863^{*}	-22.398^{**}	-5.627^{**}
	(1.424)	(7.804)	(2.377)
x	0.449^{**}	2.616^{***}	1.692^{***}
	(0.157)	(0.377)	(0.160)
Observations	15	10	25
Sum of Squared Resid.	89.317	93.815	764.471
F Statistic	8.224^{**}	48.134^{***}	111.920^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01

$$F^{\text{emp}} = \frac{(\text{SSR}_r - \text{SSR}_u)/k}{\text{SSR}_u/(n-2k)} = \frac{(764.471 - (89.317 + 93.815))/2}{(89.317 + 93.815)/(25-4)} = 33.331$$

Kritischer Wert: $F^{\text{crit}}_{0.05,2,21} = 3.4668 \quad \Rightarrow$ Nullhypothese verwerfen!

Hypothesen . . .



Quelle: `https://xkcd.com/2569/`

Thanks!