# Statistik zur Datenanalyse

Dr. Meike Wocken

**Vorlesung 3**

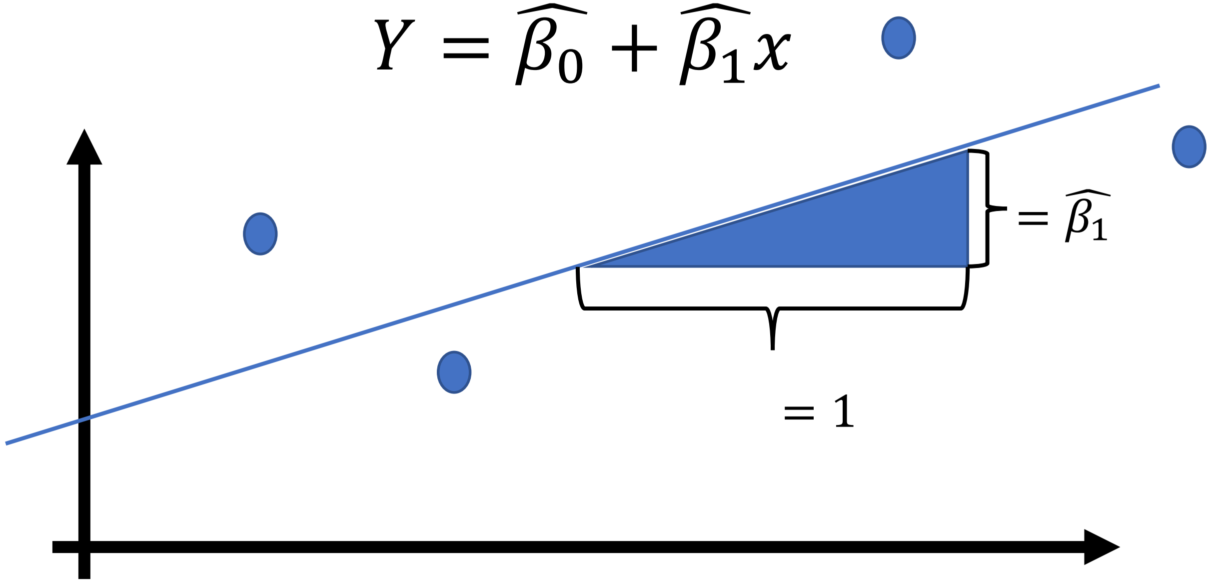
HS Bielefeld
Digitale Technologien (M.Sc.)
WiSe 2023/24

Meike.Wocken@codecentric.de

# Bis jetzt: Regressionsschätzer

Änderung von $x$ um eine Einheit bewirkt ceteris paribus (c.p.) eine durchschnittliche Änderung von $Y$ um $\hat{\beta}_1$ Einheiten

$$Y = \widehat{\beta_0} + \widehat{\beta_1}x$$



# Rückschluss von Stichprobe auf Verhalten in einer Grundgesamtheit

z.B. Untersuchung zum Anteil der Radfahrer unter den Studierenden.

*“Wer fährt mit dem Rad ins Büro?“*

2 Stichproben, n = 3

Stichprobe 1: Anteil Radfahrer = 1/3

Stichprobe 2: Anteil Radfahrer = 2/3

Der beobachtete Anteilswert hängt von der Stichprobe ab, der Schätzer ist somit auch eine Zufallsgröße.

# Punktschätzer

# Vorgehen Punktschätzer

*Schätzer für Erwartungswert, Varianz, Median, Quantile, Korrelationen, …*

- $\theta$ ist unbekannter Parameter der Grundgesamtheit
- Schätzer für $\theta$: $t = g(x_1, \dots, x_n)$
- $t$ ist eine Funktion, die abhängig ist von den independent, identical distributed (i.i.d.) Beobachtungen $x_1, \dots, x_n$ der Variablen $X$
- Beispiel: $\theta = E(X)$ ist der unbekannte Parameter,

$$t = g(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$E\left(\frac{1}{n}\sum x_i\right) = \frac{1}{n}E\left(\sum x_i\right) = \frac{1}{n} * n * E(X)$$

liefert dann den Schätzwert.

- Neue Stichprobe liefert veränderten Schätzwert! $T = g(X_1, \dots, X_n)$ ist selber eine Zufallsvariable als Funktion von Zufallsvariablen.

# Wie erhalten wir Schätzfunktion g(), um damit Schätzwerte zu ermitteln?

- Wir haben i.i.d. Beobachtungen $x_1, \ldots, x_n$
- Annahme: Beobachtungen stammen aus Normalverteilung

  $\mu, \sigma^2$ sind unbekannt
- Die Wahrscheinlichkeit, $x_1, \ldots, x_n$ zu beobachten ist

$$L(\mu, \sigma^2) = f(x_1, \ldots, x_n | \mu, \sigma^2) = f(x_1 | \mu, \sigma^2) \cdot \ldots \cdot f(x_n | \mu, \sigma^2)$$

$f$ *ist Dichtefunktion der Normalverteilung, es wird die i.i.d. Annahme ausgenutzt*

$L()$ ist **Likelihood** Funktion, die $\mu$ und $\sigma^2$ als Argumente besitzt.

# Maximum-Likelihood Prinzip

- Wähle $\mu$ und $\sigma^2$ so, dass die Wahrscheinlichkeit, genau $x_1, \ldots, x_n$ zu beobachten maximal wird

$$L(\widehat{\theta}) = \max_{\theta} L(\theta)$$

Dann ist der Schätzwert $\widehat{\theta}$ der Wert, der die plausibelste Erklärung für das Zustandekommen der Beobachtungen $x_1, \ldots, x_n$ liefert.

# Lokales Optimum

- Ableiten und Null setzen
- Produkt führt zu „unschönen“ Ausdrücken

daher: **maximiere die Log-Likelihood**

log() ist eine streng monoton wachsende Funktion, daher ist Maximum an gleicher Stelle wie bei der Likelihood.

$$L(\mu, \sigma^2) = f(x_1, \dots, x_n | \mu, \sigma^2) = f(x_1 | \mu, \sigma^2) \cdot \dots \cdot f(x_n | \mu, \sigma^2)$$

$$\ln L(\mu, \sigma^2) = \sum_{i=1}^{n} \ln f(x_i \mid \mu, \sigma^2)$$

*Siehe Skript für Ableitung und Null setzen.*

# Kleinste Quadrate Schätzer

- siehe letzte Vorlesung bei der Schätzung der Regressionsparameter
- Die Schätzer sind wieder Zufallsvariablen, allerdings (aufgrund der gemachten Annahmen) mit bekannter Verteilung!!

Bisher: Schätzung von Verteilungsparametern

Ab jetzt: Überprüfen von Annahmen über das Verhalten eines interessierenden Merkmals in der Grundgesamtheit einer Stichprobe.

# Einfaches Beispiel - Produktion

- Produktion Werkstück **<u>mit vorgegebener Länge $\boldsymbol{\mu_0}$</u>**
- Um die Einhaltung der Länge $X$ zu prüfen, werden regelmäßig Stichproben gezogen und arithmetisches Mittel berechnet.
- Annahme: $X$ ist Länge Werkstück
  $X \sim N(\mu_0, \sigma^2)$
  $\mu_0$ ist vorgegebene Sollgröße
  $\sigma$ ist Standardabweichung, wurde durch Vielzahl an Vorläufen ermittelt

Wie können wir das berechnete $\bar{x}$ aus Stichprobe nutzen, um zu prüfen, ob Länge des Werkstücks noch mit Sollwert übereinstimmt?

# Einfaches Beispiel - Produktion

- Hypothesentest!!

Hypothesen aufstellen: $H_0: \mu = \mu_0 \;\; H_1: \mu \neq \mu_0$

Ist $H_0$ wahr, so gilt $E(\bar{X}) = \mu_0$
Ist $H_0$ unwahr, so gilt $E(\bar{X}) \neq \mu_0$, es liegt also eine Abweichung zwischen Mittelwert und Sollwert vor!!

**Wie stark muss eine Abweichung sein, damit $H_0$ abgelehnt werden lann?**

# Entwicklung einer Prüfgröße

- Wir schauen uns die Differenz an:

$$\bar{X} - \mu_0$$

# Entwicklung einer Prüfgröße

- Wir schauen uns die Differenz an:

$$\bar{X} - \mu_0$$

Annahme: $X \sim N(\mu_0, \sigma^2)$

Daher gilt $(\bar{X} - \mu_0) \sim N(0, \sigma^2)$

# Entwicklung einer Prüfgröße

- Wir schauen uns die Differenz an:

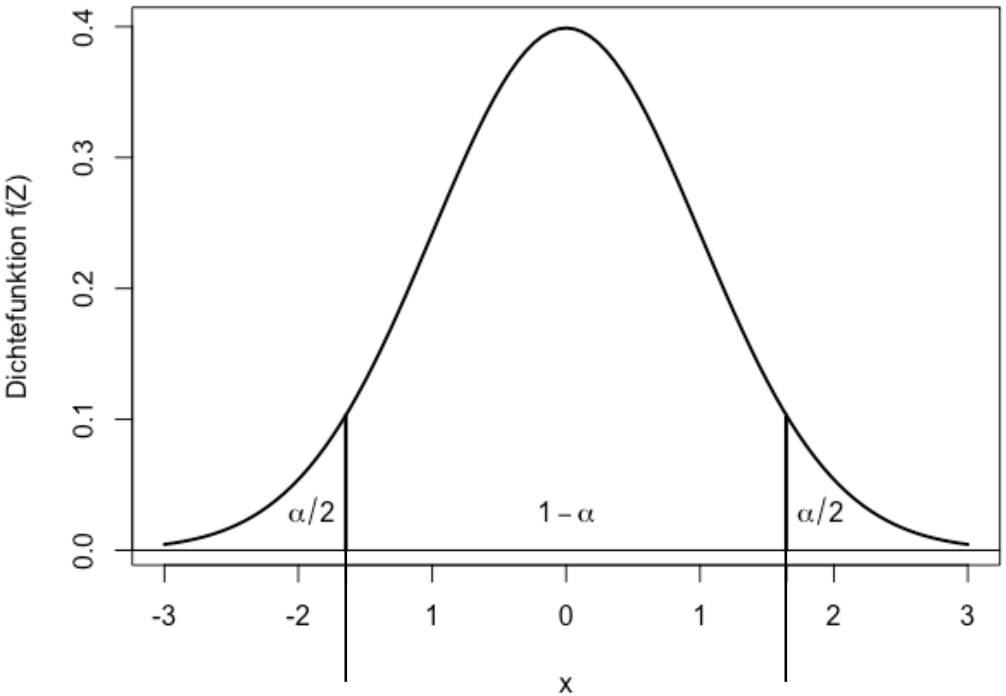
$$Z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

Annahme: $X \sim N(\mu_0, \sigma^2)$

# Beurteilung der Abweichung $\bar{X} - \mu_0$

- Es wird immer eine Abweichung zwischen $\bar{X}$ und $\mu_0$ geben, auch wenn $H_0 : \mu = \mu_0$ wahr ist (Produktionsstück hat im Durchschnitt Sollwert-Länge).

Die Größe $Z$ folgt der Standard-Normalverteilung. Unter der Annahme von $H_0$ beträgt die Wahrscheinlichkeit $(1 - \alpha) \cdot 100\%$, das $Z$ im Bereich zwischen $\left[-Z_{1-\frac{\alpha}{2}}, +Z_{1-\frac{\alpha}{2}}\right]$ liegt. D.h., solange $Z$ im mittleren Bereich liegt, spricht alles für die Gültigkeit von $H_0$.

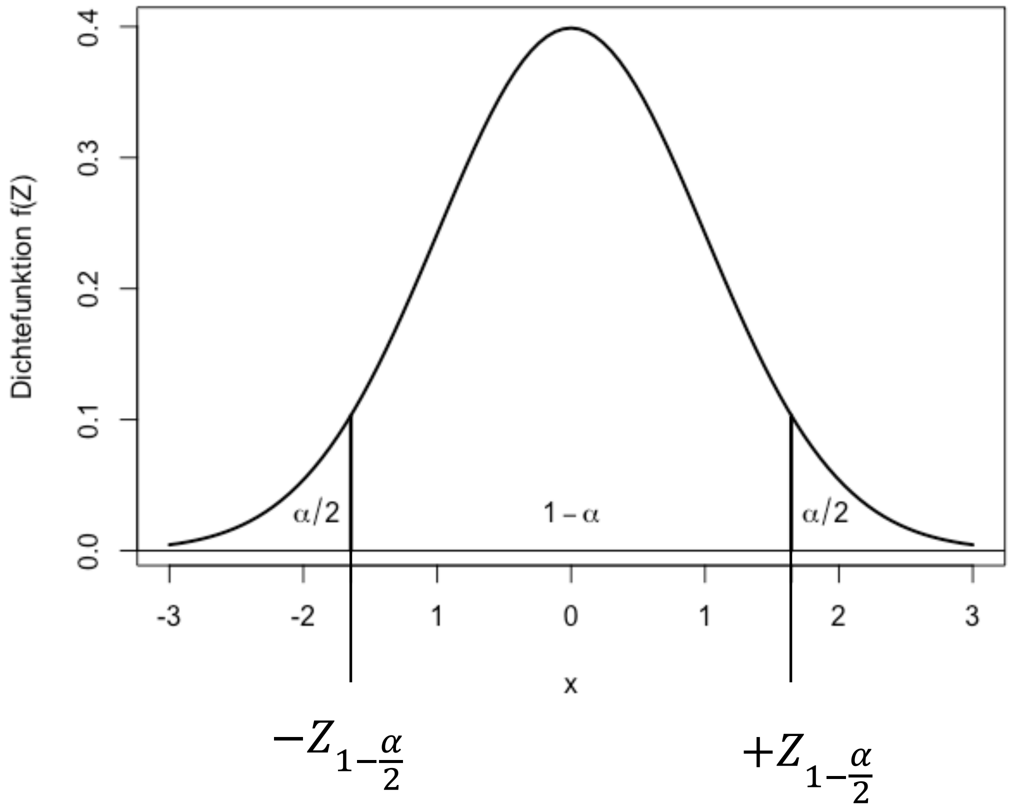



$-Z_{1-\frac{\alpha}{2}}$ $+Z_{1-\frac{\alpha}{2}}$

# Wie ist es zu beweten, wenn $Z$ Werte außerhalb des mittleren Bereiches liegen?

Die Wahrscheinlichkeit ist gering ($\alpha \cdot 100\%$) , dass ich $Z$ unter der Annahmen, dass $H_0$gültig ist, außen zu beobachten.

D.h. entweder beobachten wir damit ein sehr unwahrscheinliches Ereignis oder es ist ein Hinweis, dass die Annahme $H_0$ ungültig ist.



# Fehler 1. und 2. Art

- Fehler 1. Art:

$H_0$ wird verworfen, obwohl $H_0$ wahr ist

- Fehler 2. Art:

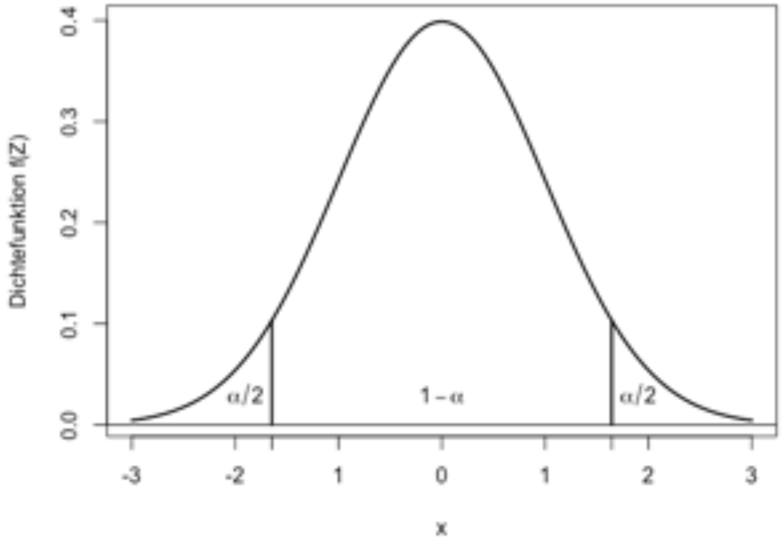
$H_0$ wird beibehalten, obwohl $H_1$ wahr ist.

Statistische Tests können nur den Fehler 1. Art kontrollieren:

Test zum Signifikanzniveau $\alpha, 0 < \alpha < 1$ mit

$$P(H_1 annehmen \,|H_0 wahr) \leq \alpha$$

Über das Signifikanzniveau $\alpha$ wird gesteuert, wie viel Wahrscheinlichkeit „rechts“ und „links“ bleibt.



# Zusammenfassung

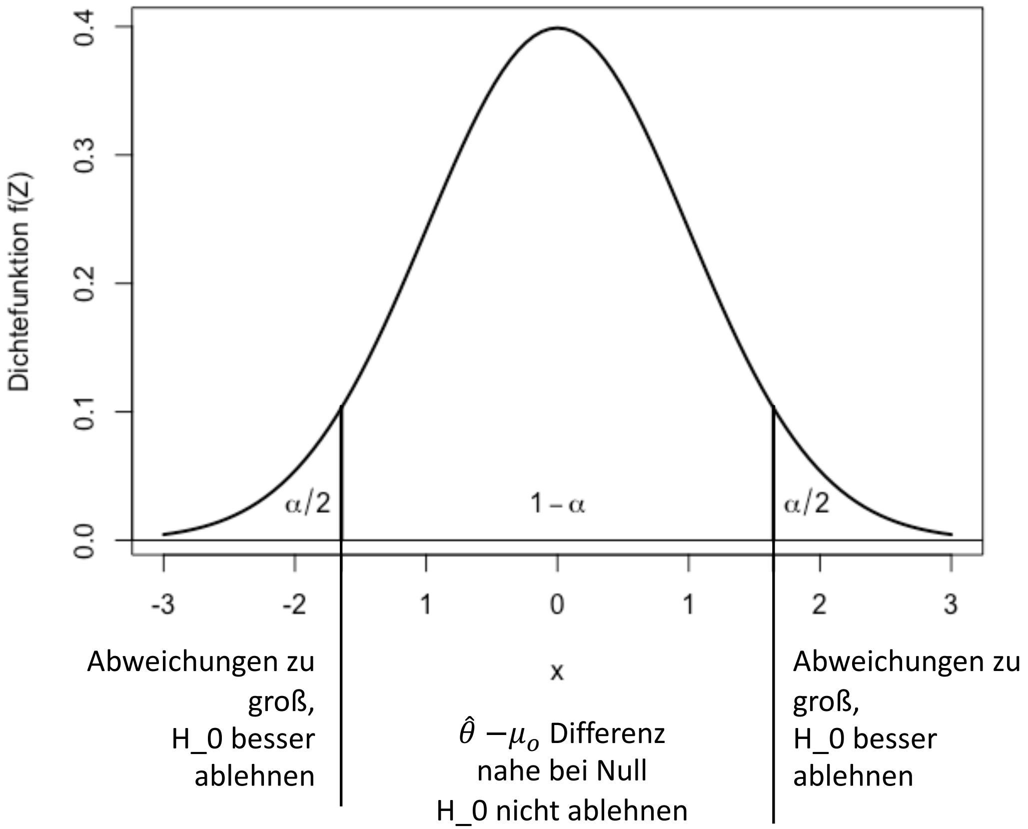
- Aussage zu einem Parameter $\theta = \mu$ gewünscht
- Schätzfkt. $\hat{\theta} = t(x_1, \dots, x_n) = \frac{\sum x_i}{n}$
- Zu prüfende Hypothesen aufstellen

$$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0$$

- Wenn $H_0$ wahr ist, dann gilt

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

Wir berechnen die Prüfgröße $Z$



# Aufgabe 1

1. Lebenshaltungskosten:

Ein Marktforschungsinstitut führt jährliche Untersuchungen zu den Lebenshaltungskosten durch. Die Kosten für einen bestimmten Warenkorb beliefen sich in den letzten Jahren auf durchschnittlich 600 EUR. Im Beispieljahr wurde in einer Stichprobe von

10 Testen von Hypothesen - 117 -

40 zufällig ausgewählten Kaufhäusern jeweils der aktuelle Preis des Warenkorbs bestimmt. Als Schätzer für den aktuellen Preis des Warenkorbs ergab sich ein mittlerer Preis von 605 EUR. Die Varianz $\sigma^2 = 225$ sei aufgrund langjähriger Erfahrung bekannt. Gehen Sie von einer Normalverteilung des Preises für den Warenkorb aus.

- Hat sich der Preis des Warenkorbs im Vergleich zu den Vorjahren signifikant zum Niveau $\alpha = 0,01$ erhöht? Wie lautet das zugehörige statistische Testproblem?

1) Zu prüfende Hypothesen aufstellen

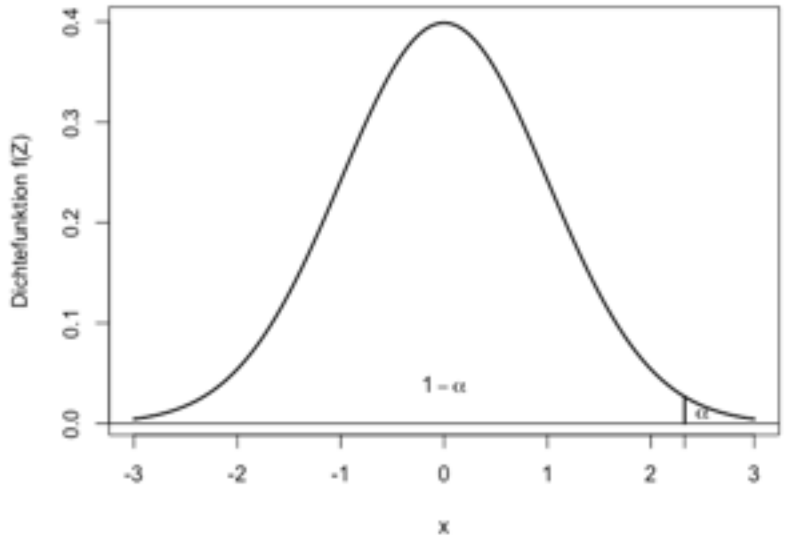
$$H_0: \mu = 600 \quad H_1: \mu > 600$$

**Einseitiger Test!!**

2) Prüfgröße berechnen:

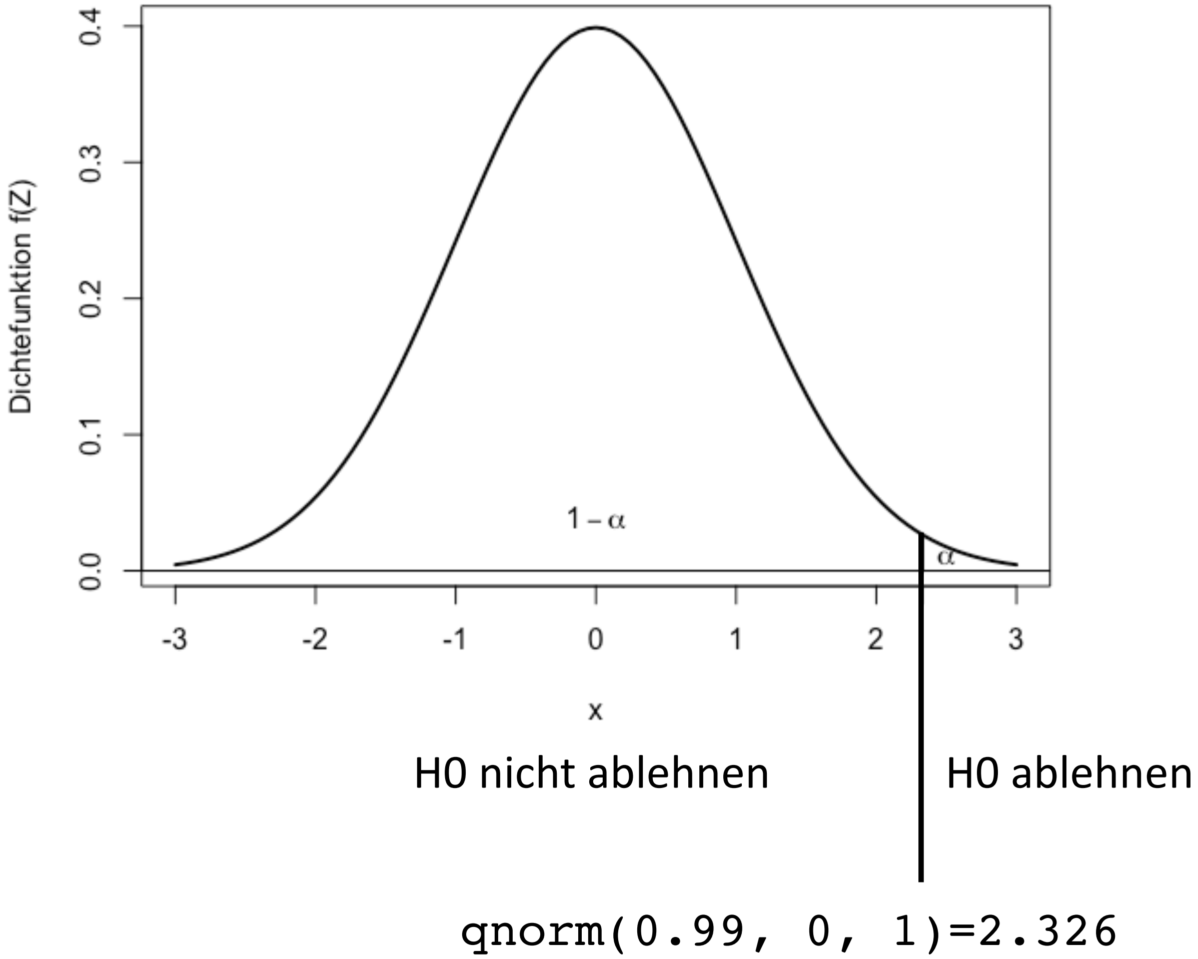
$$Z = \frac{605 - 600}{\sqrt{225/40}} = \frac{5}{15/\sqrt{40}} = 2,108$$

Unter $H_0$ gilt $Z \sim N(0,1)$



# Aufgabe 1

Da gilt $Z = 2{,}108 < 2{,}326$ kann Ho nicht abgelehnt werden.



# Aufgabe 2 – bitte selber rechnen

2. Abfüllanlage:

Der Output einer Abfüllanlage lag in allen Test seit der Inbetriebnahme im Mittel bei $26.60ml$ je abgefüllter Einheit mit einer Varianz von $0.025ml^2$. Der letzte Test lag einige Monate zurück. Nun wird eine Zufallststichprobe von 40 Einheiten entnommen und die Füllmenge überprüft. Der Mittelwert der Stichprobe liegt bei $26.65ml$ je abgefüllter Einheit. Die bisherigen Testergebnisse werden als Parameter der Grundgesamtheit angenommen, d.h. der Mittelwert der Grundgesamtheit beträgt $\mu_0 = 26.60ml$, die Varianz beträgt $\sigma_0^2 = 0.025ml^2$. Es kann eine Normalverteilung der Abfüllmenge unterstellt werden.

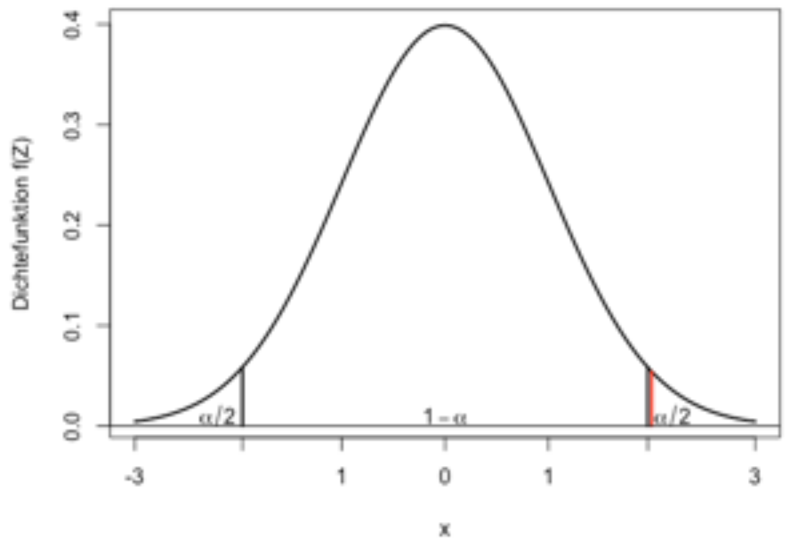
- Weicht die Abfüllmenge statistisch signifikant zum Niveau $\alpha = 0,05$ von den vorherigen Werten ab? Wie lautet das zugehörige statistische Testproblem?

```
qnorm(0.025,0,1) = -1.9599

qnorm(0.975,0,1) = 1.9599
```

# Aufgabe 2 - Lösung

- 1) Zu prüfende Hypothesen aufstellen

$$H_0\colon \mu = 26{,}60 \quad H_1\colon \mu \neq 26{,}60$$




**Zweiseitiger Test!!**

2) Prüfgröße berechnen:

$$Z = \frac{26{,}65 - 26{,}60}{\sqrt{0{,}025/40}} = \frac{26{,}65 - 26{,}60}{\sqrt{0{,}025}}\sqrt{40} = 2$$

Unter $H_0$ gilt $Z \sim N(0{,}1)$

Schwarze Linie:

`qnorm(0.025, 0, 1) = -1.9599 qnorm(0.975, 0, 1) = 1.9599`

**Rote Linie: Z=2**

H0 ist abzulehnen!

# Beispiel Aufgabe 8 (Vorlesung 2)

**Prognose für Gehalt**

Mittels Forward Algorithmus hatten wir bestes Modell ermittelt. Nimmt Anzahl von Hits um eine Einheit zu, so steigt c.p. das durchschnittliche Gehalt um 6,942T US-$.

**Ist der Effekt statistisch signifikant?**

$$H_0: \beta_{Hits} = 0 \quad H_1: \beta_{Hits} \neq 0$$

# R Output

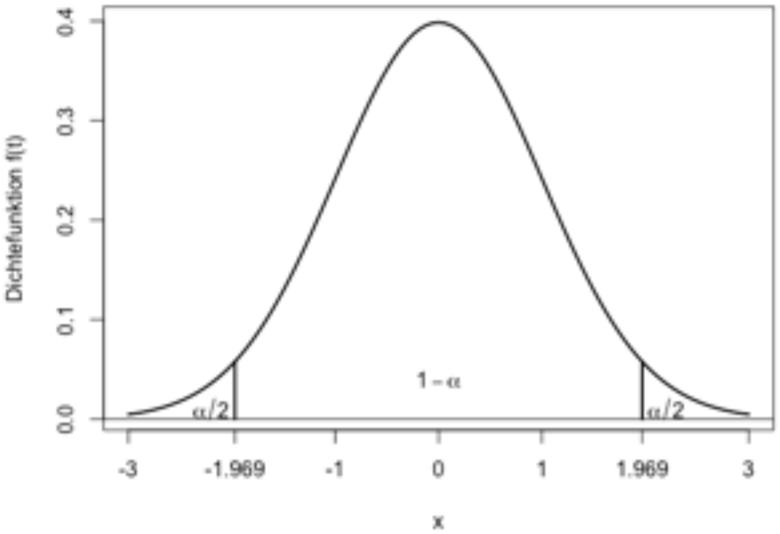
```
Call:
lm(formula = Salary ~ AtBat + Hits + Walks + CAtBat + CRuns +
    CRBI + CWalks + League + Division + PutOuts + Assists, data = Hitters)

Residuals:
   Min     1Q Median     3Q    Max
-932.2 -175.4  -29.2  130.4 1897.2

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  135.75122   71.34623   1.903 0.058223 .
AtBat         -2.12775    0.53746  -3.959 9.81e-05 ***
Hits           6.92370    1.64612   4.206 3.62e-05 ***
Walks          5.62028    1.59064   3.533 0.000488 ***
CAtBat        -0.13899    0.05609  -2.478 0.013870 *
CRuns          1.45533    0.39270   3.706 0.000259 ***
CRBI           0.78525    0.20978   3.743 0.000225 ***
CWalks        -0.82286    0.26361  -3.121 0.002010 **
LeagueN       43.11162   39.96612   1.079 0.281755
DivisionW   -111.14603   39.21835  -2.834 0.004970 **
PutOuts        0.28941    0.07478   3.870 0.000139 ***
Assists        0.26883    0.15816   1.700 0.090430 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 311.7 on 251 degrees of freedom
Multiple R-squared:  0.5426,    Adjusted R-squared:  0.5226
F-statistic: 27.07 on 11 and 251 DF,  p-value: < 2.2e-16
```

Parameterschätzer $\widehat{\beta}$ ist selber Zufallsvariable, abhängig von der Stichprobe.
Unter derAnnahme, das Homoskedastizität, Erwartungstreue $E(\widehat{\theta}) = \theta$ und Konsistenz ($n \uparrow \Rightarrow \widehat{\theta} \rightarrow \theta$, wenn die Anzahl an Beobachtungen groß wird, konvergiert der Schätzer gegen den wahren Wert) vorliegen (siehe Annahmen für lineare Modelle), können wir Aussagen zur Verteilung machen.

Prüfgröße:

$$t = \frac{\widehat{\beta_{hits}} - 0}{se(\widehat{\beta_{hits}})} = \frac{6{,}924 - 0}{1{,}646} = 4{,}206 \text{ (R-Output!)}$$

Die Prüfgröße folgt einer t-Verteilung mit n-2 Freiheitsgraden.



# R Output

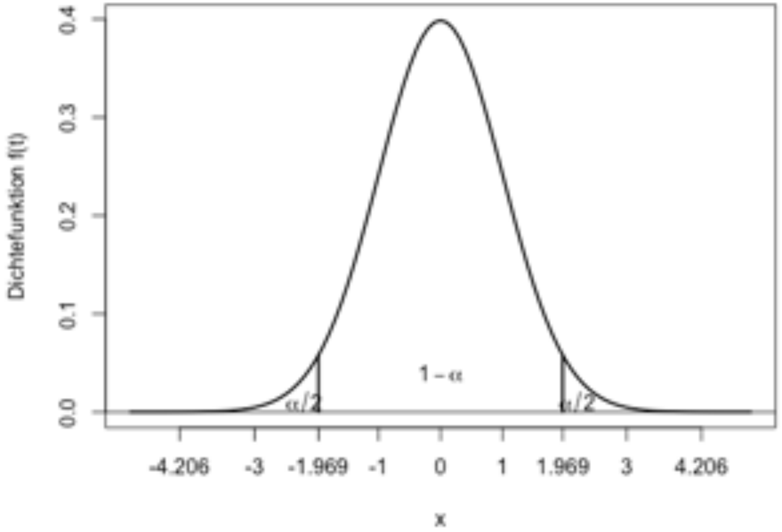
```
Call:
lm(formula = Salary ~ AtBat + Hits + Walks + CAtBat + CRuns +
    CRBI + CWalks + League + Division + PutOuts + Assists, data = Hitters)

Residuals:
   Min     1Q Median     3Q    Max
-932.2 -175.4  -29.2  130.4 1897.2

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  135.75122   71.34623   1.903 0.058223 .
AtBat         -2.12775    0.53746  -3.959 9.81e-05 ***
Hits           6.92370    1.64612   4.206 3.62e-05 ***
Walks          5.62028    1.59064   3.533 0.000488 ***
CAtBat        -0.13899    0.05609  -2.478 0.013870 *
CRuns          1.45533    0.39270   3.706 0.000259 ***
CRBI           0.78525    0.20978   3.743 0.000225 ***
CWalks        -0.82286    0.26361  -3.121 0.002010 **
LeagueN       43.11162   39.96612   1.079 0.281755
DivisionW   -111.14603   39.21835  -2.834 0.004970 **
PutOuts        0.28941    0.07478   3.870 0.000139 ***
Assists        0.26883    0.15816   1.700 0.090430 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 311.7 on 251 degrees of freedom
Multiple R-squared:  0.5426,    Adjusted R-squared:  0.5226
F-statistic: 27.07 on 11 and 251 DF,  p-value: < 2.2e-16
```

Prüfgröße:

$$t = \frac{\widehat{\beta_{hits}} - 0}{se(\widehat{\beta_{hits}})} = \frac{6{,}924 - 0}{1{,}646} = 4{,}206 \text{ (R-Output!)}$$

Für eine Entscheidung müssen wir entweder die Prüfgröße 4,206 mit dem $\left(1 - \frac{\alpha}{2}\right)$-Quantil der t-Verteilung vergleichen oder die Fläche, die rechts und links von der Prüfgröße liegt auswerten (p-Werte)

Die Prüfgröße folgt einer t-Verteilung mit n-2 Freiheitsgraden.



# p-Wert

Der p-Wert ist die Wahrscheinlichkeot, einen noch extremeren Wert als die berechnete Prüfgröße unter H0 zu erhalten. Ist p-Wert$< \alpha$, dann kann H0 zum Signifikanzniveau $\alpha$ abgelehnt werden.

In der Aufgabe:

Hits hat p-Wert = 3,62 10^-5 (siehe auch die *** im R Output), daher ist H0 abzulehnen, Hits hat einen statistisch signifikanten Einfluß.

Bei Heteroskedastizität und/oder Kollinearität im Modell sind Hypothesentests nicht verlässlich, aufgrund einer vergrößerten Standardabweichung der Schätzer.

# Kategoriale Variablen

```
Call:
lm(formula = Salary ~ AtBat + Hits + Walks + CAtBat + CRuns +
    CRBI + CWalks + League + Division + PutOuts + Assists, data = Hitters)

Residuals:
   Min     1Q Median     3Q    Max
-932.2 -175.4  -29.2  130.4 1897.2

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  135.75122   71.34623   1.903 0.058223 .
AtBat         -2.12775    0.53746  -3.959 9.81e-05 ***
Hits           6.92370    1.64612   4.206 3.62e-05 ***
Walks          5.62028    1.59064   3.533 0.000488 ***
CAtBat        -0.13899    0.05609  -2.478 0.013870 *
CRuns          1.45533    0.39270   3.706 0.000259 ***
CRBI           0.78525    0.20978   3.743 0.000225 ***
CWalks        -0.82286    0.26361  -3.121 0.002010 **
LeagueN       43.11162   39.96612   1.079 0.281755
DivisionW   -111.14603   39.21835  -2.834 0.004970 **
PutOuts        0.28941    0.07478   3.870 0.000139 ***
Assists        0.26883    0.15816   1.700 0.090430 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 311.7 on 251 degrees of freedom
Multiple R-squared:  0.5426,	Adjusted R-squared:  0.5226
F-statistic: 27.07 on 11 and 251 DF,  p-value: < 2.2e-16
```

- LeagueN ist 1, wenn League =N ist, sonst 0 (League = A).
- $H_0: \beta_{League} = 0$ kann nicht abgelehnt werden. Was bedeutet das?

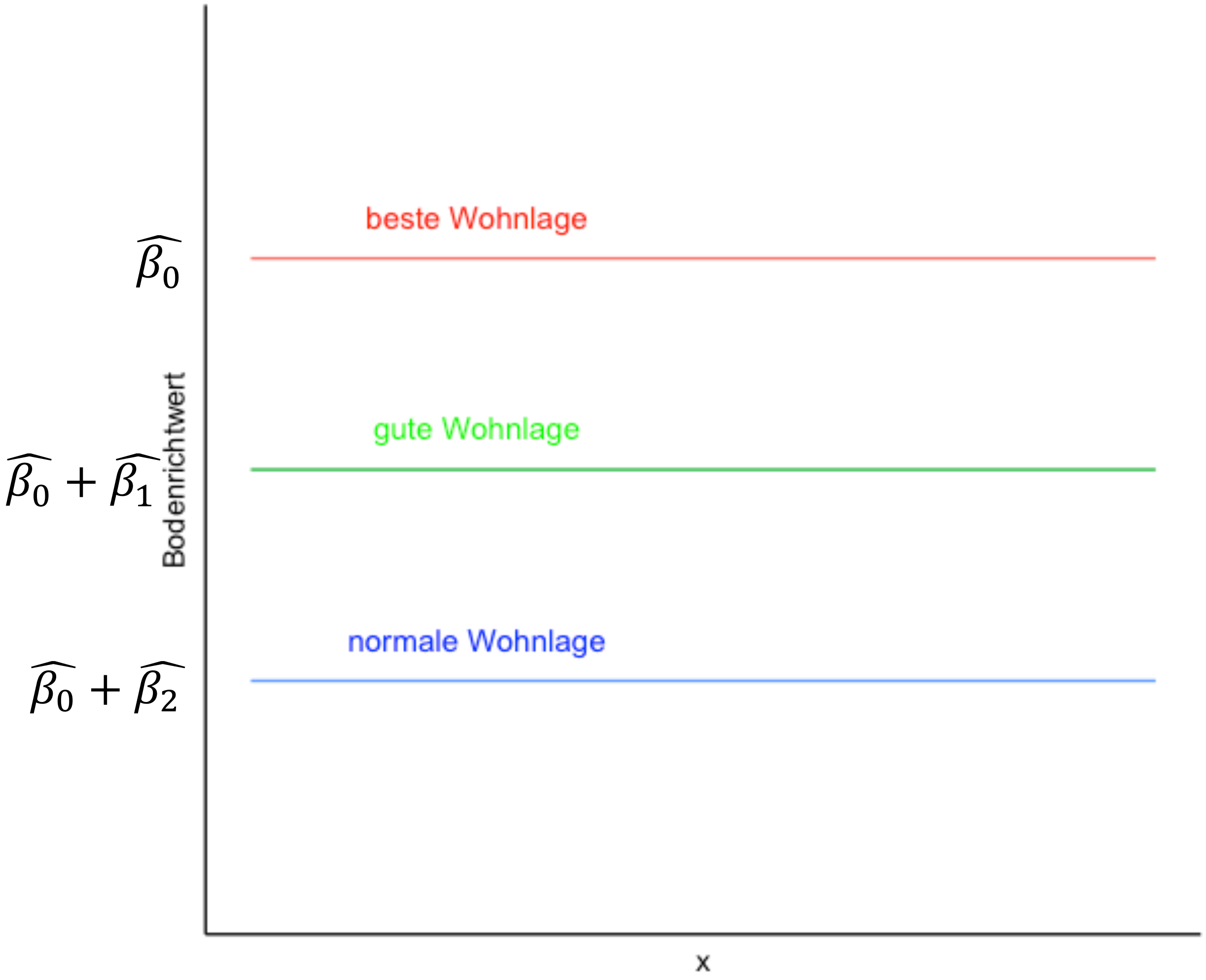
Der **Unterschied** zwischen den Leagues ist statistisch nicht signifikant!!

# Beispiel Bodenrichtwerte

- 3 Kategorien: 1 = beste Lage, 2 = gute Lage, 3 =normale Lage
- 2 Dummy Variablen (nicht 3!! Wir brauchen eine **Referenzkategorie**, gegen die Verglichen wird)

$$\widehat{Bodenrichtwert} = \widehat{\beta_0} + \widehat{\beta_1}\, guteWohnlage + \widehat{\beta_2}\, normaleWohnlage$$

$\widehat{\beta_1}$ und $\widehat{\beta_2}$ geben die Differenz an zur Referenzkategorie und ob der Unterschied zur Referenzkategorie statistisch signifikant ist.

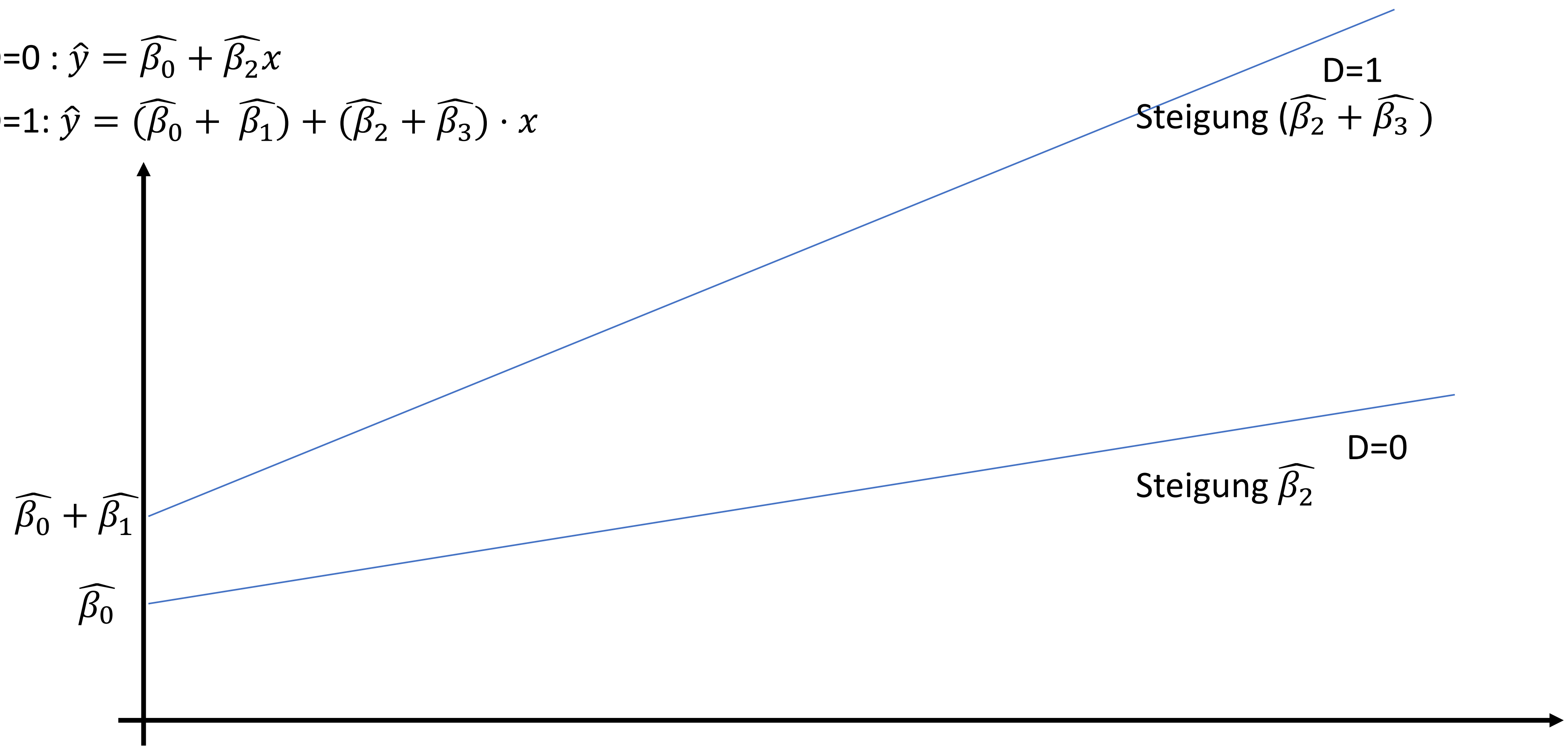


# Exkurs: Modellierung unterschiedlicher Steigungen mit Dummy Variablen

$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} \cdot D + \widehat{\beta_2} x + \widehat{\beta_3} \cdot D \cdot x$

D=0 : $\hat{y} = \widehat{\beta_0} + \widehat{\beta_2} x$

D=1: $\hat{y} = (\widehat{\beta_0} + \widehat{\beta_1}) + (\widehat{\beta_2} + \widehat{\beta_3}) \cdot x$



# F-Test auf Gesamtsignifikanz (Overall F test)

```
Call:
lm(formula = Salary ~ AtBat + Hits + Walks + CAtBat + CRuns +
    CRBI + CWalks + League + Division + PutOuts + Assists, data = Hitters)

Residuals:
   Min     1Q Median     3Q    Max
-932.2 -175.4  -29.2  130.4 1897.2

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  135.75122   71.34623   1.903 0.058223 .
AtBat         -2.12775    0.53746  -3.959 9.81e-05 ***
Hits           6.92370    1.64612   4.206 3.62e-05 ***
Walks          5.62028    1.59064   3.533 0.000488 ***
CAtBat        -0.13899    0.05609  -2.478 0.013870 *
CRuns          1.45533    0.39270   3.706 0.000259 ***
CRBI           0.78525    0.20978   3.743 0.000225 ***
CWalks        -0.82286    0.26361  -3.121 0.002010 **
LeagueN       43.11162   39.96612   1.079 0.281755
DivisionW   -111.14603   39.21835  -2.834 0.004970 **
PutOuts        0.28941    0.07478   3.870 0.000139 ***
Assists        0.26883    0.15816   1.700 0.090430 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 311.7 on 251 degrees of freedom
Multiple R-squared:  0.5426,	Adjusted R-squared:  0.5226
F-statistic: 27.07 on 11 and 251 DF,  p-value: < 2.2e-16
```

- Im R Output gibt es F-Test auf Gesamtsignifikanz für lineare Regressionen

- $H_0: \beta_1 = \cdots = \beta_p = 0$

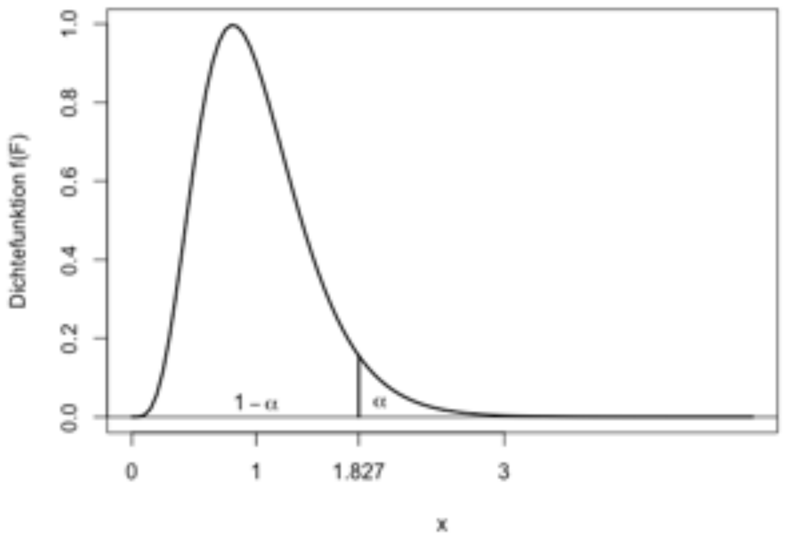
(die **erklärenden** Variablen können nichts erklären)

- $H_1$: mind. ein $\beta_i \neq 0, i \in 1, \ldots, p$
- Einseitiger Test

- Prüfgröße ist F-verteilt mit p und n-p-1 Freiheitsgraden

# F-Test auf Gesamtsignifikanz (Overall F test)

```
Call:
lm(formula = Salary ~ AtBat + Hits + Walks + CAtBat + CRuns +
    CRBI + CWalks + League + Division + PutOuts + Assists, data = Hitters)

Residuals:
   Min     1Q Median     3Q    Max
-932.2 -175.4  -29.2  130.4 1897.2

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  135.75122   71.34623   1.903 0.058223 .
AtBat         -2.12775    0.53746  -3.959 9.81e-05 ***
Hits           6.92370    1.64612   4.206 3.62e-05 ***
Walks          5.62028    1.59064   3.533 0.000488 ***
CAtBat        -0.13899    0.05609  -2.478 0.013870 *
CRuns          1.45533    0.39270   3.706 0.000259 ***
CRBI           0.78525    0.20978   3.743 0.000225 ***
CWalks        -0.82286    0.26361  -3.121 0.002010 **
LeagueN       43.11162   39.96612   1.079 0.281755
DivisionW   -111.14603   39.21835  -2.834 0.004970 **
PutOuts        0.28941    0.07478   3.870 0.000139 ***
Assists        0.26883    0.15816   1.700 0.090430 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 311.7 on 251 degrees of freedom
Multiple R-squared:  0.5426,    Adjusted R-squared:  0.5226
F-statistic: 27.07 on 11 and 251 DF,  p-value: < 2.2e-16
```




Prüfgröße 27,07 ist größer als das .95-Quantil der F-Verteilung (=1,827).
P-Wert ist sehr klein. H0 ist zum Signifikanzniveau $\alpha = 0.05$ abzulehnen!