

# Wahrscheinlichkeitsrechnung und Statistik für Biologen **Wiederholung**

Noémie Becker & Dirk Metzler

[http://evol.bio.lmu.de/\\_statgen](http://evol.bio.lmu.de/_statgen)

Sommersemester 2013

# Übersicht

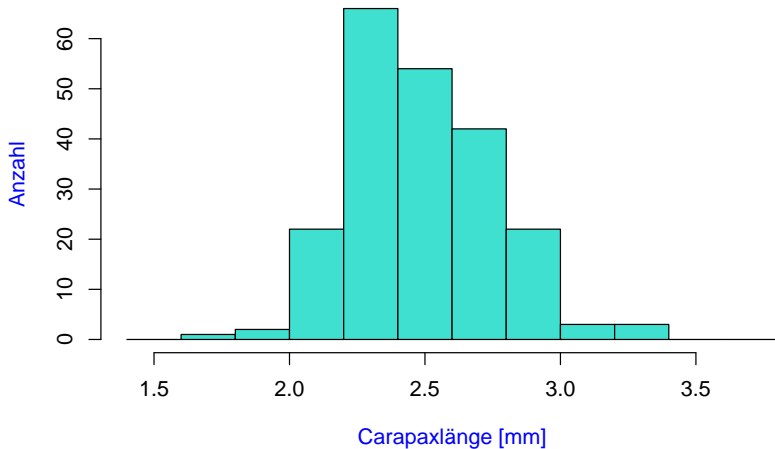
- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression
  - Lineare Zusammenhänge
  - $t$ -Test fuer lineare Zusammenhänge
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse
- 7 Versuchsplanung: Stichprobenumfang

# Übersicht

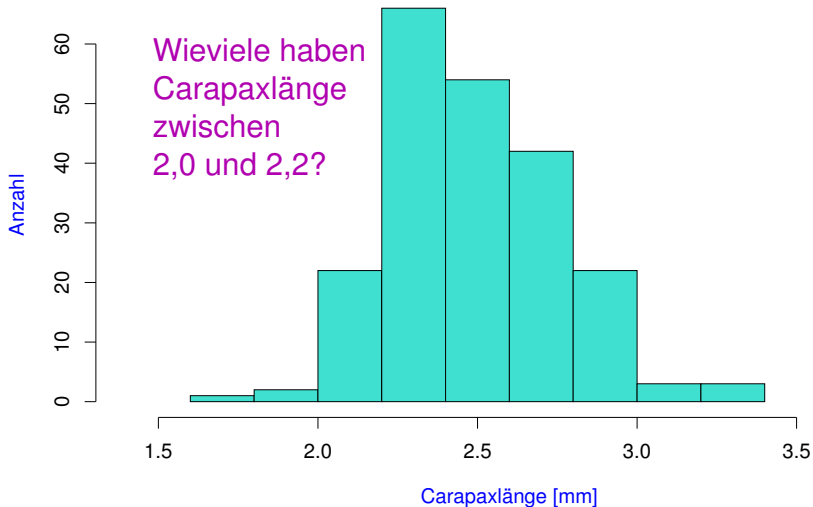
- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression
  - Lineare Zusammenhänge
  - $t$ -Test fuer lineare Zusammenhänge
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse
- 7 Versuchsplanung: Stichprobenumfang



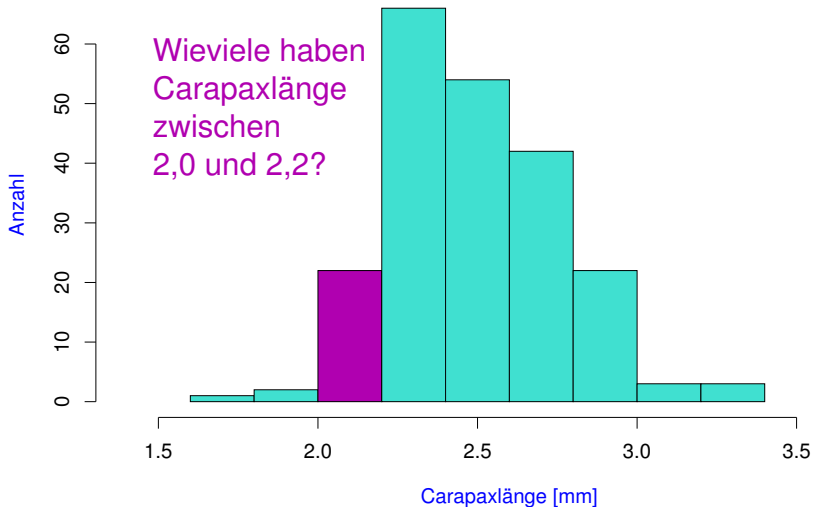
## Nichteiertragende Weibchen am 6. Sept. '88, n=215



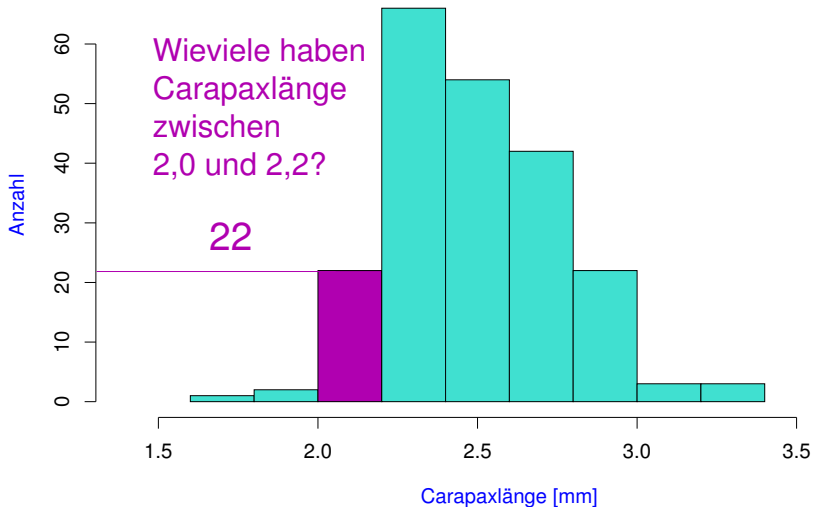
## Nichteiertragende Weibchen am 6. Sept. '88, n=215



## Nichteiertragende Weibchen am 6. Sept. '88, n=215

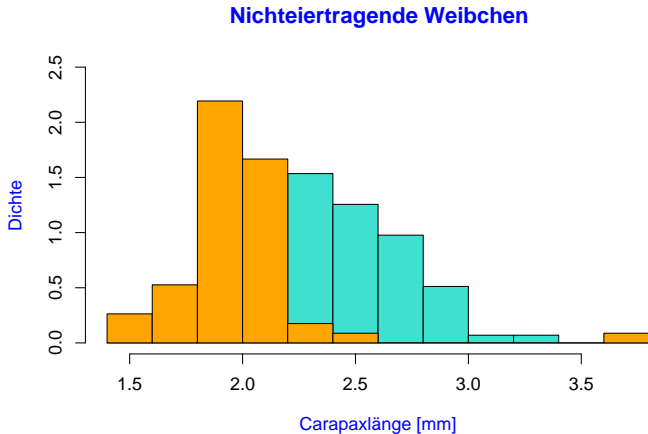


## Nichteiertragende Weibchen am 6. Sept. '88, n=215

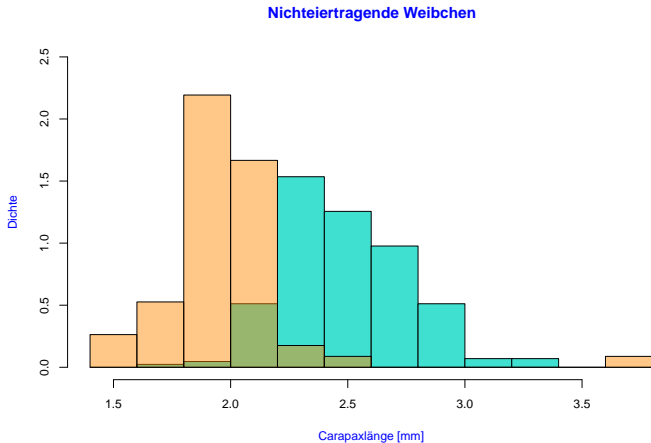




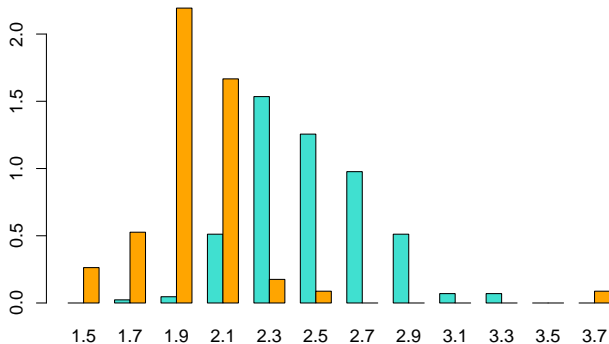
# Versuche, die Histogramme zusammen zu zeigen:



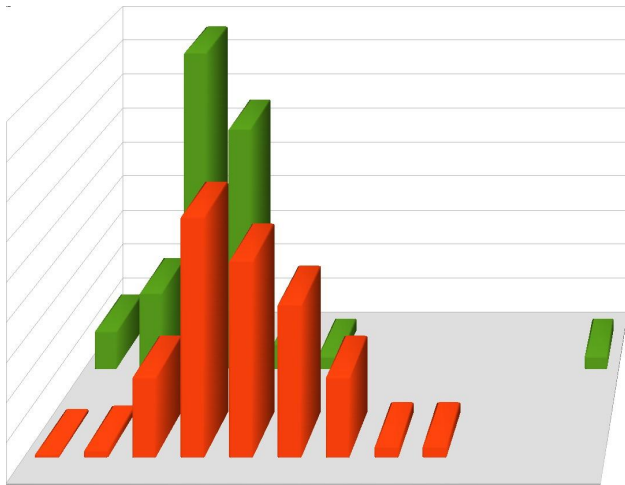
# Versuche, die Histogramme zusammen zu zeigen:



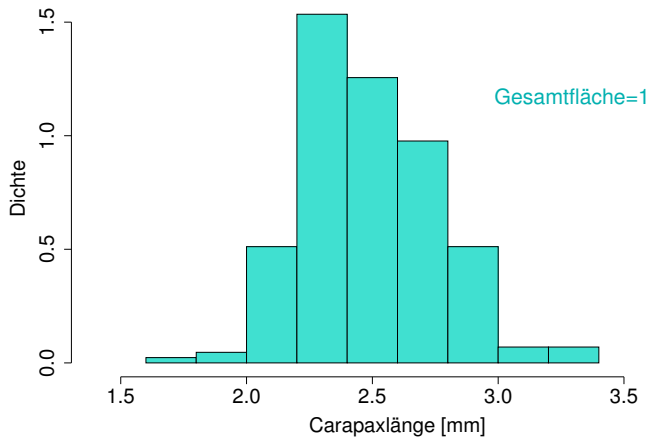
# Versuche, die Histogramme zusammen zu zeigen:



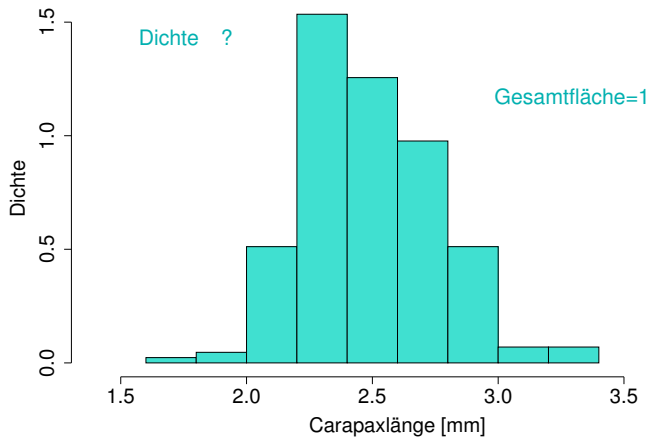
# Versuche, die Histogramme zusammen zu zeigen:



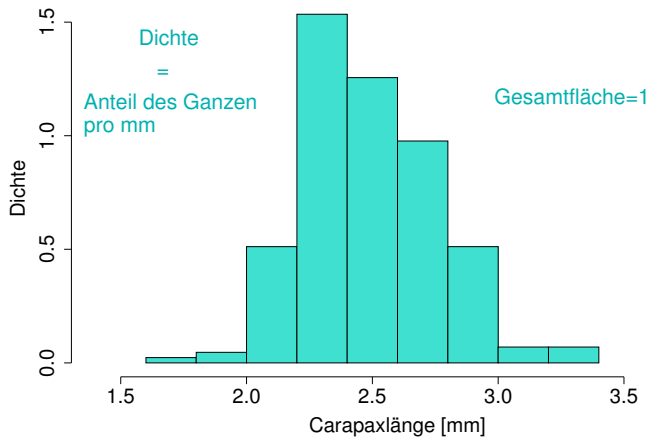
Nichteiertragende Weibchen am 6. Sept. '88, n=215



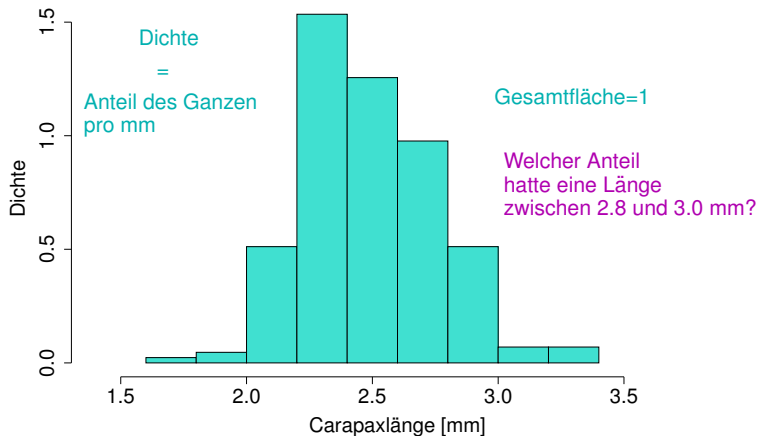
Nichteiertragende Weibchen am 6. Sept. '88, n=215



Nichteiertragende Weibchen am 6. Sept. '88, n=215

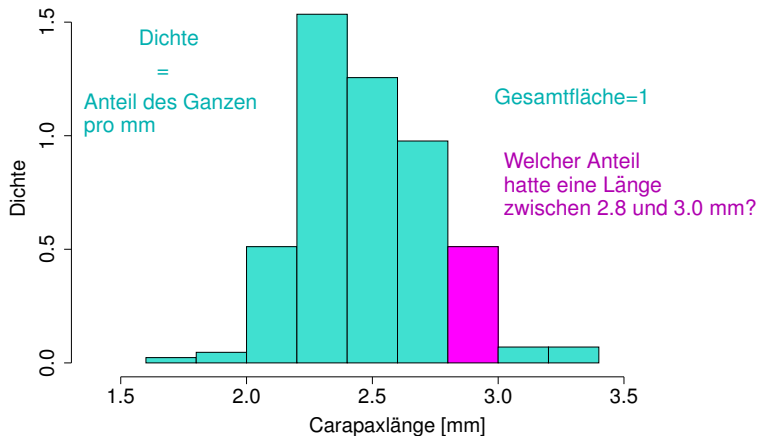


Nichteiertragende Weibchen am 6. Sept. '88, n=215

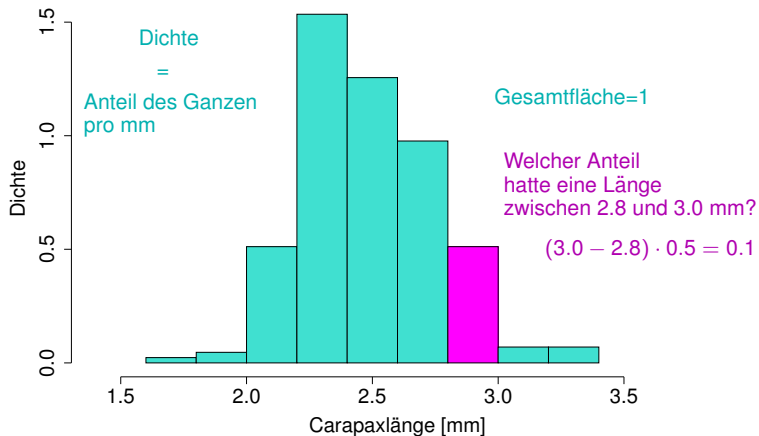




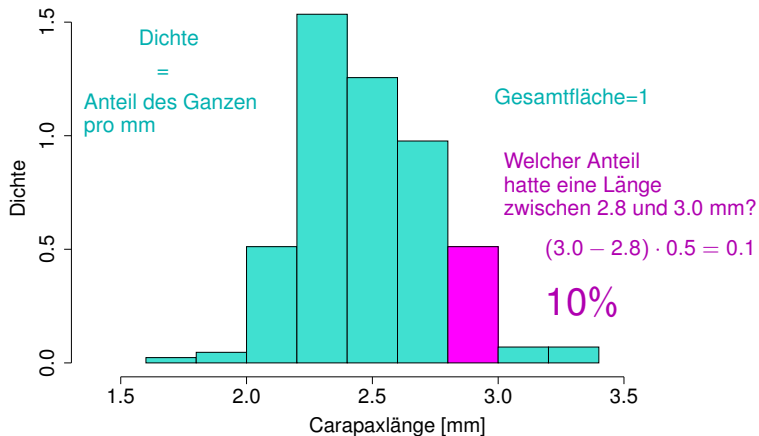
Nichteiertragende Weibchen am 6. Sept. '88, n=215



Nichteiertragende Weibchen am 6. Sept. '88, n=215

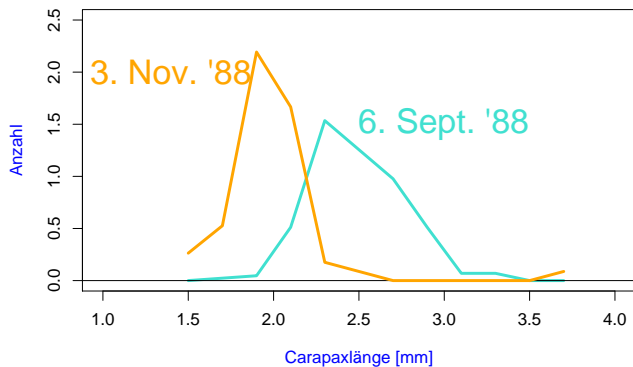


Nichteiertragende Weibchen am 6. Sept. '88, n=215



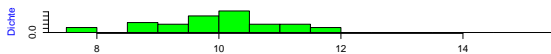
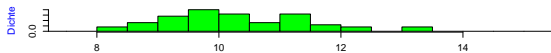
# Zwei und mehr Dichtepolygone in einem Plot

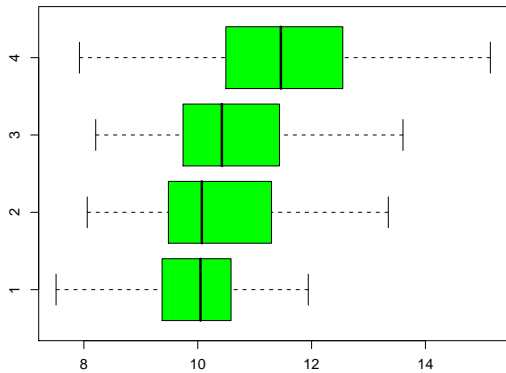
## Nichteiertragende Weibchen



# Beispiel:

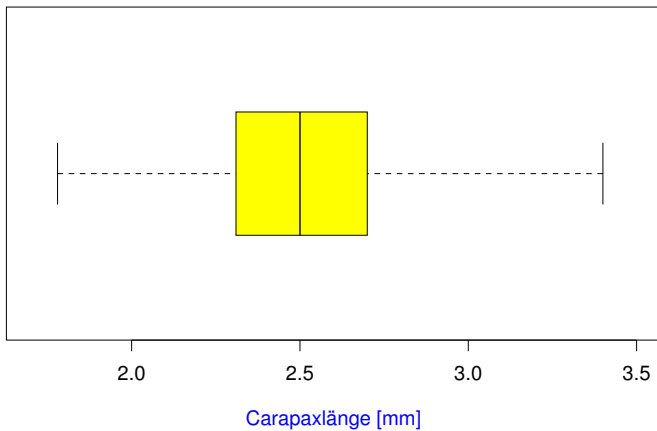
## Vergleich von mehreren Gruppen





# Der Boxplot

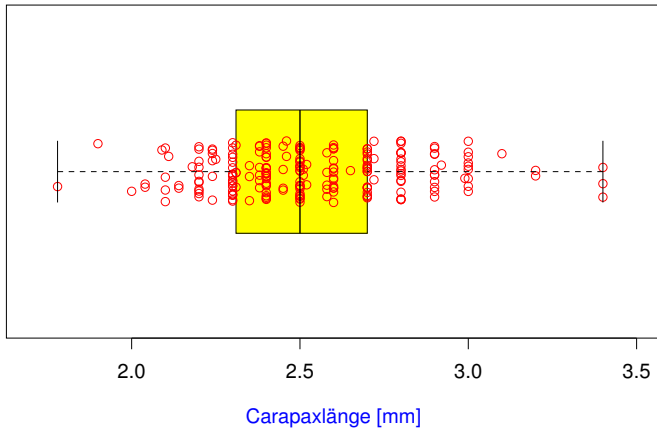
## Boxplot, einfache Ausführung





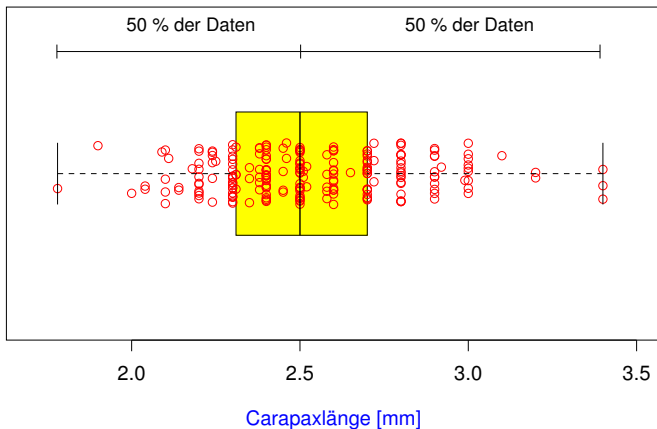
# Der Boxplot

## Boxplot, einfache Ausführung



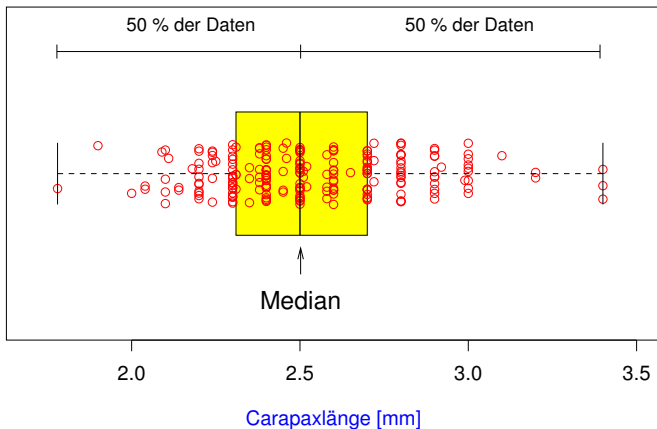
# Der Boxplot

## Boxplot, einfache Ausführung



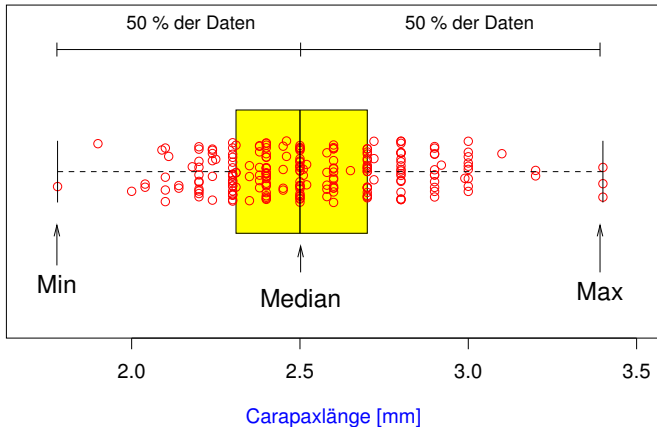
# Der Boxplot

## Boxplot, einfache Ausführung



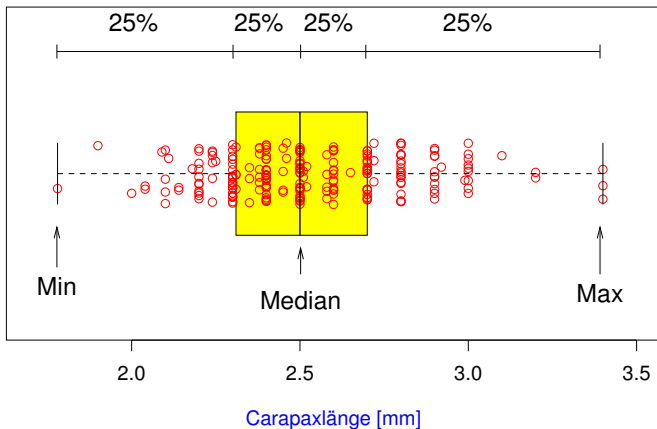
# Der Boxplot

## Boxplot, einfache Ausführung



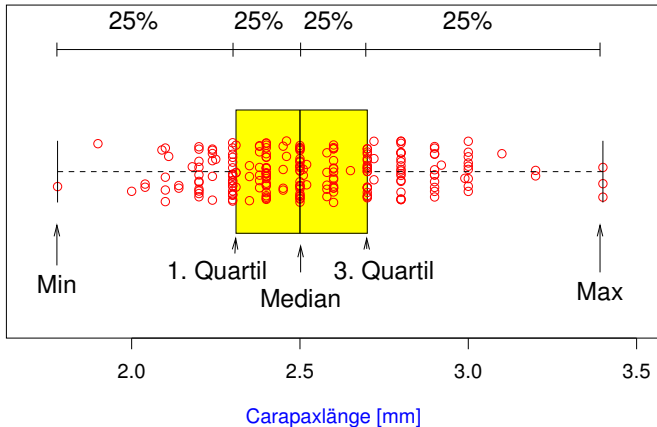
# Der Boxplot

## Boxplot, einfache Ausführung



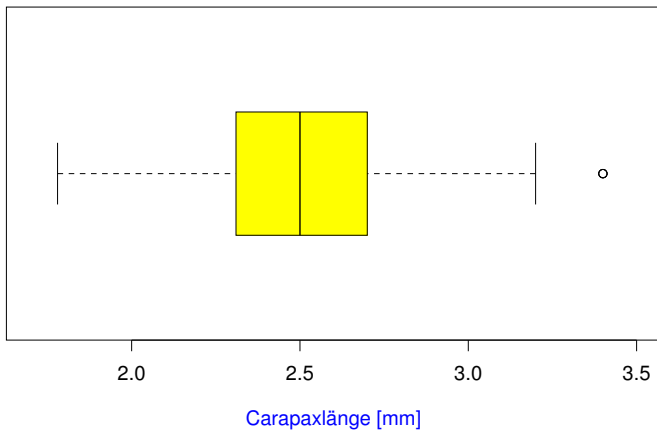
# Der Boxplot

## Boxplot, einfache Ausführung



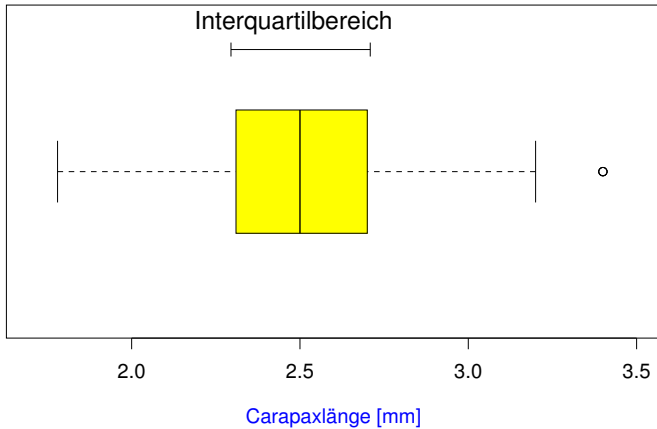
# Der Boxplot

## Boxplot, Standardausführung



# Der Boxplot

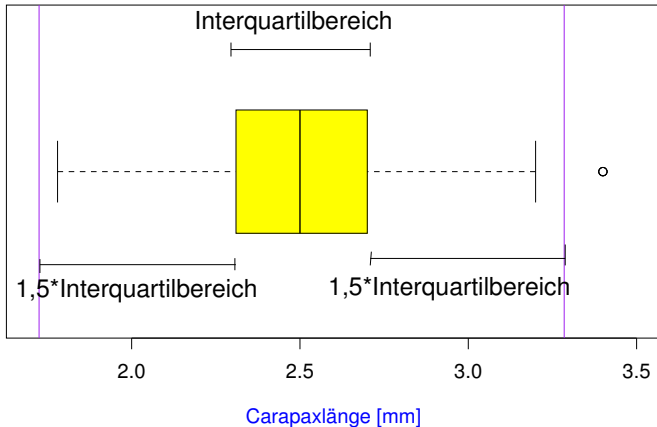
## Boxplot, Standardausführung





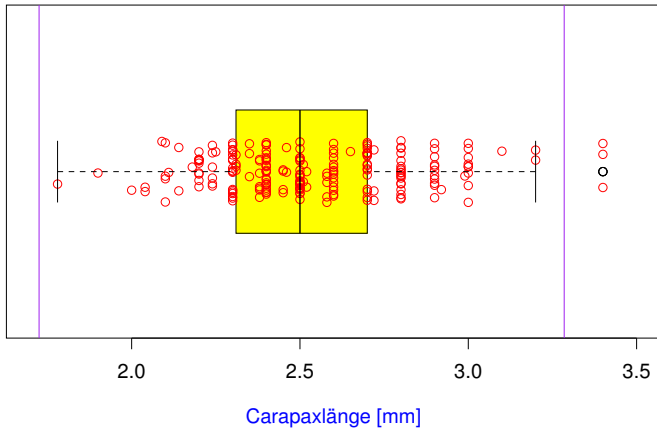
# Der Boxplot

## Boxplot, Standardausführung



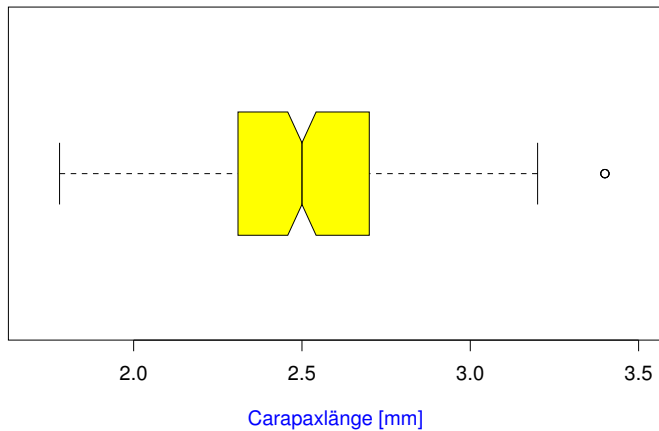
# Der Boxplot

## Boxplot, Standardausführung



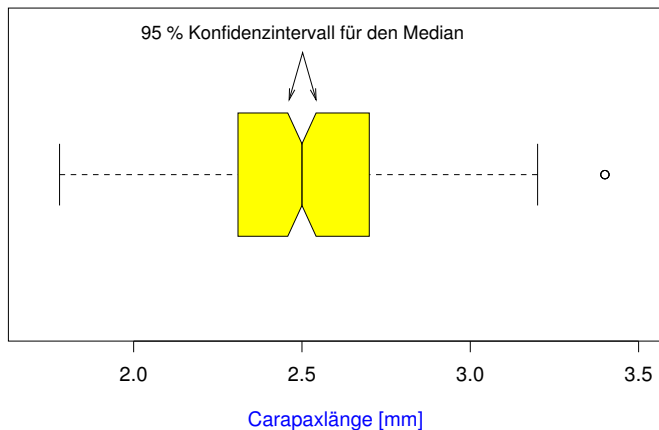
# Der Boxplot

## Boxplot, Profiausstattung



# Der Boxplot

## Boxplot, Profiausstattung



Es ist oft möglich,  
das Wesentliche  
an einer Stichprobe  
  
mit ein paar Zahlen  
zusammenzufassen.

Wesentlich:

1. Wie groß?

2. Wie variabel?

Wesentlich:

1. Wie groß?

Lageparameter

2. Wie variabel?

Wesentlich:

1. Wie groß?

Lageparameter

2. Wie variabel?

Streuungsparameter



Eine Möglichkeit  
kennen wir schon  
aus dem Boxplot:

# Lageparameter

## Der Median

Lageparameter

Der Median

Streuungsparameter

Lageparameter

Der Median

Streuungsparameter

Der Quartilabstand ( $Q_3 - Q_1$ )

## Der **Median**:

die Hälfte der Beobachtungen sind kleiner,  
die Hälfte sind größer.

Der **Median**:

die Hälfte der Beobachtungen sind kleiner,  
die Hälfte sind größer.

Der Median ist  
das **50%-Quantil**  
der Daten.

# Die Quartile

Das erste Quartil,  $Q_1$ :

# Die Quartile

**Das erste Quartil,  $Q_1$ :**  
ein Viertel der Beobachtungen  
sind kleiner,  
drei Viertel sind größer.



# Die Quartile

Das erste Quartil,  $Q_1$ :  
ein Viertel der Beobachtungen  
sind kleiner,  
drei Viertel sind größer.

$Q_1$  ist das  
**25%-Quantil**  
der Daten.

# Die Quartile

Das dritte Quartil,  $Q_3$ :

# Die Quartile

**Das dritte Quartil,  $Q_3$ :**  
drei Viertel der Beobachtungen  
sind kleiner,  
ein Viertel sind größer.

# Die Quartile

Das dritte Quartil,  $Q_3$ :  
drei Viertel der Beobachtungen  
sind kleiner,  
ein Viertel sind größer.

$Q_3$  ist das  
75%-Quantil  
der Daten.

Am häufigsten werden benutzt:

Lageparameter

Der Mittelwert  $\bar{x}$

Am häufigsten werden benutzt:

Lageparameter

Der Mittelwert  $\bar{x}$

Streuungsparameter

Die Standardabweichung  $s$

# Der Mittelwert

(engl. *mean*)

## NOTATION:

Wenn die Beobachtungen

$x_1, x_2, x_3, \dots, x_n$

heißen,

schreibt man oft

$\bar{x}$

für den Mittelwert.



## DEFINITION:

Mittelwert

=

$$\frac{\text{Summe der Messwerte}}{\text{Anzahl der Messwerte}}$$

## DEFINITION:

Mittelwert

=

$$\frac{\text{Summe}}{\text{Anzahl}}$$

# DEFINITION:

Der Mittelwert von  $x_1, x_2, \dots, x_n$  als Formel:

# DEFINITION:

Der Mittelwert von  $x_1, x_2, \dots, x_n$  als Formel:

$$\bar{x} = (x_1 + x_2 + \dots + x_n)/n$$

# DEFINITION:

Der Mittelwert von  $x_1, x_2, \dots, x_n$  als Formel:

$$\begin{aligned}\bar{x} &= (x_1 + x_2 + \dots + x_n)/n \\ &= \frac{1}{n} \sum_{i=1}^n x_i\end{aligned}$$

# Geometrische Bedeutung des Mittelwerts: Der Schwerpunkt

# Die Standardabweichung

# Die Standardabweichung

Wie weit weicht  
eine typische Beobachtung  
vom  
Mittelwert  
ab ?



Die **Standardabweichung**  $\sigma$  (“sigma”)  
ist ein  
etwas komisches  
gewichtetes Mittel  
der Abweichungsbeträge

Die **Standardabweichung**  $\sigma$  (“sigma”)  
ist ein  
etwas komisches  
gewichtetes Mittel  
der Abweichungsbeträge

und zwar

$$\sigma = \sqrt{\text{Summe}(\text{Abweichungen}^2)/n}$$

bzw. bei Stichproben

$$\sigma = \sqrt{\text{Summe}(\text{Abweichungen}^2)/(n - 1)}$$

als Schätzung der Standardabweichung in der  
Gesamtpopulation.

Die **Standardabweichung** von  $x_1, x_2, \dots, x_n$  als Formel:

Die **Standardabweichung** von  $x_1, x_2, \dots, x_n$  als Formel:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Die **Standardabweichung** von  $x_1, x_2, \dots, x_n$  als Formel:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  heißt **Varianz**.

Die **Standardabweichung** von  $x_1, x_2, \dots, x_n$  als Formel:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  heißt **Varianz**.

Bei Stichproben eher üblich:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Die **Standardabweichung** von  $x_1, x_2, \dots, x_n$  als Formel:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  heißt **Varianz**.

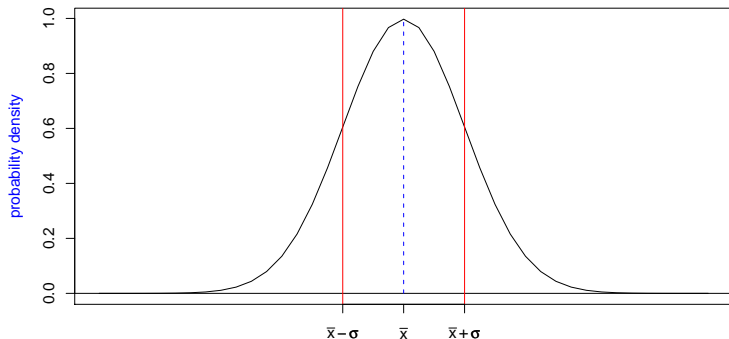
Bei Stichproben eher üblich:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

und  $\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ .

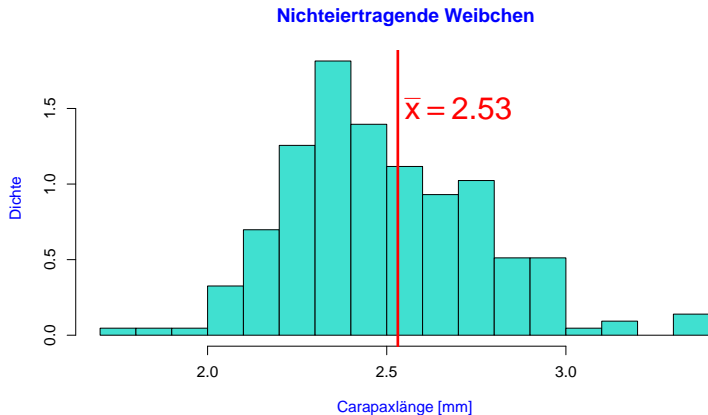
# Faustregel für die Standardabweichung

Bei ungefähr glockenförmigen (also eingipfligen und symmetrischen) Verteilungen liegen ca. 2/3 der Verteilung zwischen  $\bar{x} - \sigma$  und  $\bar{x} + \sigma$ .

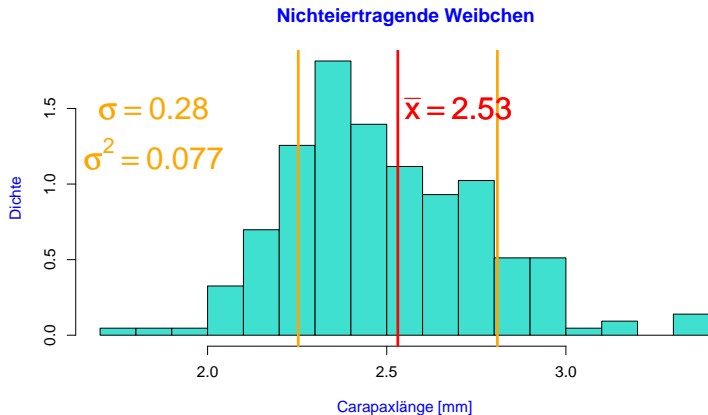




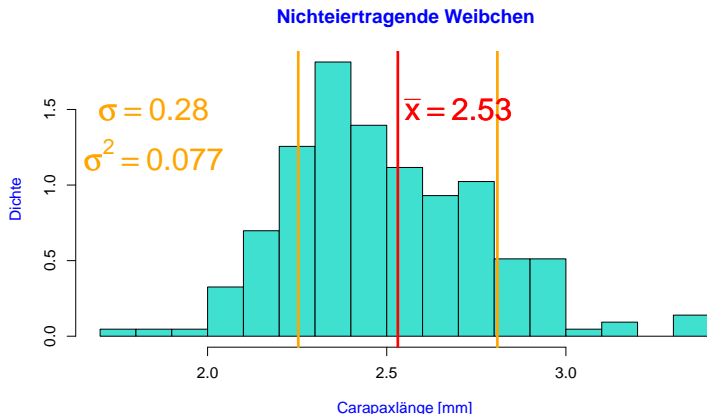
# Standardabweichung der Carapaxlängen nichteierttragender Weibchen vom 6.9.88



# Standardabweichung der Carapaxlängen nichteierttragender Weibchen vom 6.9.88



# Standardabweichung der Carapaxlängen nichteierttragender Weibchen vom 6.9.88



Hier liegt der Anteil zwischen  $\bar{x} - \sigma$  und  $\bar{x} + \sigma$  bei 72%.

# Varianz der Carapaxlängen nichteiertragender Weibchen vom 6.9.88

Alle Carapaxlängen im Meer:  $\mathcal{X} = (X_1, X_2, \dots, X_N)$ .

# Varianz der Carapaxlängen nichtteiertragender Weibchen vom 6.9.88

Alle Carapaxlängen im Meer:  $\mathcal{X} = (X_1, X_2, \dots, X_N)$ .

Carapaxlängen in unserer Stichprobe:  $\mathcal{S} = (S_1, S_2, \dots, S_{n=215})$

# Varianz der Carapaxlängen nichteiertragender Weibchen vom 6.9.88

Alle Carapaxlängen im Meer:  $\mathcal{X} = (X_1, X_2, \dots, X_N)$ .

Carapaxlängen in unserer Stichprobe:  $\mathcal{S} = (S_1, S_2, \dots, S_{n=215})$

Stichprobenvarianz:

$$\sigma_S^2 = \frac{1}{n} \sum_{i=1}^{215} (S_i - \bar{S})^2 \approx 0,0768$$

# Varianz der Carapaxlängen nichteiertragender Weibchen vom 6.9.88

Alle Carapaxlängen im Meer:  $\mathcal{X} = (X_1, X_2, \dots, X_N)$ .

Carapaxlängen in unserer Stichprobe:  $\mathcal{S} = (S_1, S_2, \dots, S_{n=215})$

Stichprobenvarianz:

$$\sigma_S^2 = \frac{1}{n} \sum_{i=1}^{215} (S_i - \bar{S})^2 \approx 0,0768$$

Können wir 0,0768 als Schätzwert für die Varianz  $\sigma_X^2$  in der ganzen Population verwenden?

# Varianz der Carapaxlängen nichtteiertragender Weibchen vom 6.9.88

Alle Carapaxlängen im Meer:  $\mathcal{X} = (X_1, X_2, \dots, X_N)$ .

Carapaxlängen in unserer Stichprobe:  $\mathcal{S} = (S_1, S_2, \dots, S_{n=215})$

Stichprobenvarianz:

$$\sigma_S^2 = \frac{1}{n} \sum_{i=1}^{215} (S_i - \bar{S})^2 \approx 0,0768$$

Können wir 0,0768 als Schätzwert für die Varianz  $\sigma_X^2$  in der ganzen Population verwenden?

Ja, können wir machen.



# Varianz der Carapaxlängen nichtteiertragender Weibchen vom 6.9.88

Alle Carapaxlängen im Meer:  $\mathcal{X} = (X_1, X_2, \dots, X_N)$ .

Carapaxlängen in unserer Stichprobe:  $\mathcal{S} = (S_1, S_2, \dots, S_{n=215})$

Stichprobenvarianz:

$$\sigma_S^2 = \frac{1}{n} \sum_{i=1}^{215} (S_i - \bar{S})^2 \approx 0,0768$$

Können wir 0,0768 als Schätzwert für die Varianz  $\sigma_{\mathcal{X}}^2$  in der ganzen Population verwenden?

Ja, können wir machen. Allerdings ist  $\sigma_S^2$  im Durchschnitt um den Faktor  $\frac{n-1}{n}$  ( $= 214/215 \approx 0,995$ ) kleiner als  $\sigma_{\mathcal{X}}^2$

# Varianzbegriffe

Varianz in der Population:  $\sigma_X^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2$

Stichprobenvarianz:  $\sigma_S^2 = \frac{1}{n} \sum_{i=1}^n (S_i - \bar{S})^2$

# Varianzbegriffe

Varianz in der Population:  $\sigma_X^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2$

Stichprobenvarianz:  $\sigma_S^2 = \frac{1}{n} \sum_{i=1}^n (S_i - \bar{S})^2$

korrigierte Stichprobenvarianz:

$$s^2 = \frac{n}{n-1} \sigma_S^2$$

# Varianzbegriffe

Varianz in der Population:  $\sigma_X^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2$

Stichprobenvarianz:  $\sigma_S^2 = \frac{1}{n} \sum_{i=1}^n (S_i - \bar{S})^2$

korrigierte Stichprobenvarianz:

$$\begin{aligned} s^2 &= \frac{n}{n-1} \sigma_S^2 \\ &= \frac{n}{n-1} \cdot \frac{1}{n} \cdot \sum_{i=1}^n (S_i - \bar{S})^2 \end{aligned}$$

# Varianzbegriffe

Varianz in der Population:  $\sigma_X^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2$

Stichprobenvarianz:  $\sigma_S^2 = \frac{1}{n} \sum_{i=1}^n (S_i - \bar{S})^2$

korrigierte Stichprobenvarianz:

$$\begin{aligned} s^2 &= \frac{n}{n-1} \sigma_S^2 \\ &= \frac{n}{n-1} \cdot \frac{1}{n} \cdot \sum_{i=1}^n (S_i - \bar{S})^2 \\ &= \frac{1}{n-1} \cdot \sum_{i=1}^n (S_i - \bar{S})^2 \end{aligned}$$

Mit “Standardabweichung von  $S$ ” ist meistens das korrigierte  $s$  gemeint.

## Definition (Varianz, Kovarianz und Korrelation)

Die *Varianz* einer  $\mathbb{R}$ -wertigen Zufallsgröße  $X$  ist

$$\text{Var}X = \sigma_X^2 = \mathbb{E} [(X - \mathbb{E}X)^2].$$

$\sigma_X = \sqrt{\text{Var} X}$  ist die *Standardabweichung*.

## Definition (Varianz, Kovarianz und Korrelation)

Die *Varianz* einer  $\mathbb{R}$ -wertigen Zufallsgröße  $X$  ist

$$\text{Var}X = \sigma_X^2 = \mathbb{E} [(X - \mathbb{E}X)^2].$$

$\sigma_X = \sqrt{\text{Var} X}$  ist die *Standardabweichung*.

Ist  $Y$  eine weitere reellwertige Zufallsvariable, so ist

$$\text{Cov}(X, Y) = \mathbb{E} [(X - \mathbb{E}X) \cdot (Y - \mathbb{E}Y)]$$

die *Kovarianz* von  $X$  und  $Y$ .

## Definition (Varianz, Kovarianz und Korrelation)

Die *Varianz* einer  $\mathbb{R}$ -wertigen Zufallsgröße  $X$  ist

$$\text{Var}X = \sigma_X^2 = \mathbb{E} [(X - \mathbb{E}X)^2].$$

$\sigma_X = \sqrt{\text{Var} X}$  ist die *Standardabweichung*.

Ist  $Y$  eine weitere reellwertige Zufallsvariable, so ist

$$\text{Cov}(X, Y) = \mathbb{E} [(X - \mathbb{E}X) \cdot (Y - \mathbb{E}Y)]$$

die *Kovarianz* von  $X$  und  $Y$ .

Die *Korrelation* von  $X$  und  $Y$  ist

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}.$$



## Definition (Varianz, Kovarianz und Korrelation)

Ist  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  eine gepaarte Stichprobe der Zufallsgrößen  $X$  und  $Y$ , so schätzen wir die Varianzen  $\sigma_X^2$  und  $\sigma_Y^2$  durch

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{und} \quad s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2,$$

die Kovarianz  $\text{Cov}(X, Y)$  durch

$$\text{cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

und die Korrelation  $\text{Cor}(X, Y)$  durch

$$\text{cor}(x, y) = \frac{\text{cov}(x, y)}{s_x \cdot s_y}.$$

$x$	$y$
1	2
2	1
3	6

$x$	$y$
1	2
2	1
3	6

$$\bar{x} = 2, \quad \bar{y} = 3$$

$x$	$y$
1	2
2	1
3	6

$\bar{x} = 2, \quad \bar{y} = 3$

$$s_x = \sqrt{((1 - 2)^2 + (2 - 2)^2 + (3 - 2)^2) / 2} = 1$$

$x$	$y$
1	2
2	1
3	6

$\bar{x} = 2, \quad \bar{y} = 3$

$$s_x = \sqrt{((1 - 2)^2 + (2 - 2)^2 + (3 - 2)^2) / 2} = 1$$

$$s_y = \sqrt{((2 - 3)^2 + (1 - 3)^2 + (6 - 3)^2) / 2} \approx 2.65.$$

$x$	$y$
1	2
2	1
3	6

$\bar{x} = 2, \quad \bar{y} = 3$

$$s_x = \sqrt{((1 - 2)^2 + (2 - 2)^2 + (3 - 2)^2) / 2} = 1$$

$$s_y = \sqrt{((2 - 3)^2 + (1 - 3)^2 + (6 - 3)^2) / 2} \approx 2.65.$$

$$\text{COV}(x,y) =$$

$$((1 - 2) \cdot (2 - 3) + (2 - 2) \cdot (1 - 3) + (3 - 2) \cdot (6 - 3)) / 2 = 2$$

$x$	$y$
1	2
2	1
3	6

$\bar{x} = 2, \quad \bar{y} = 3$

$$s_x = \sqrt{((1 - 2)^2 + (2 - 2)^2 + (3 - 2)^2) / 2} = 1$$

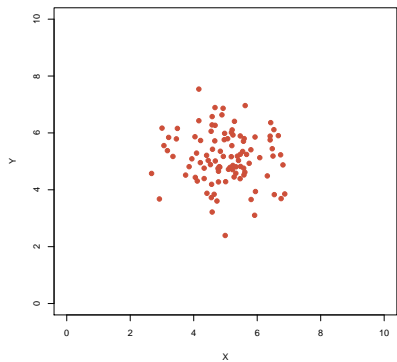
$$s_y = \sqrt{((2 - 3)^2 + (1 - 3)^2 + (6 - 3)^2) / 2} \approx 2.65.$$

$$\text{COV}(x,y) =$$

$$((1 - 2) \cdot (2 - 3) + (2 - 2) \cdot (1 - 3) + (3 - 2) \cdot (6 - 3)) / 2 = 2$$

$$\text{cor}(x,y) = \frac{\text{COV}(x,y)}{s_x \cdot s_y} = \frac{2}{1 \cdot 2.65} \approx 0.756$$

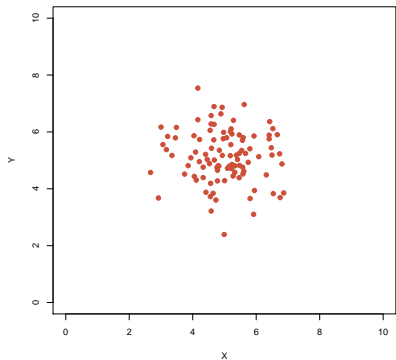
$$\sigma_X = 0.95, \sigma_Y = 0.92$$





$$\sigma_X = 0.95, \sigma_Y = 0.92$$

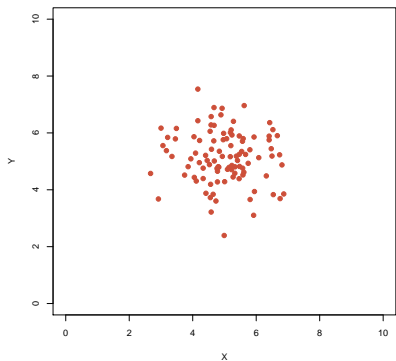
$$\text{Cov}(X, Y) = -0.06$$



$$\sigma_X = 0.95, \sigma_Y = 0.92$$

$$\text{Cov}(X, Y) = -0.06$$

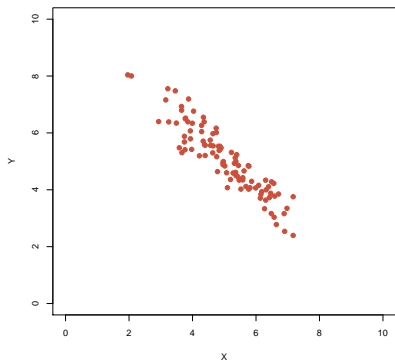
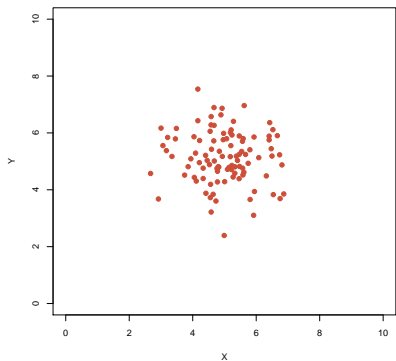
$$\text{Cor}(X, Y) = -0.069$$



$$\sigma_X = 0.95, \sigma_Y = 0.92$$

$$\text{Cov}(X, Y) = -0.06$$

$$\text{Cor}(X, Y) = -0.069$$

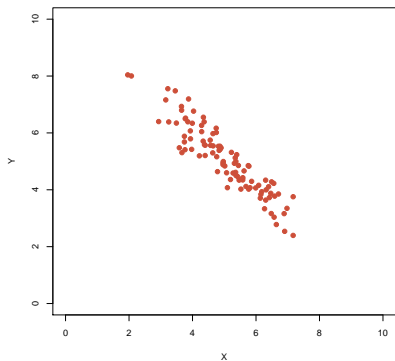
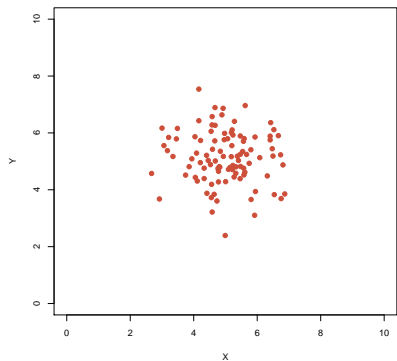


$$\sigma_X = 0.95, \sigma_Y = 0.92$$

$$\text{Cov}(X, Y) = -0.06$$

$$\text{Cor}(X, Y) = -0.069$$

$$\sigma_X = 1.13, \sigma_Y = 1.2$$



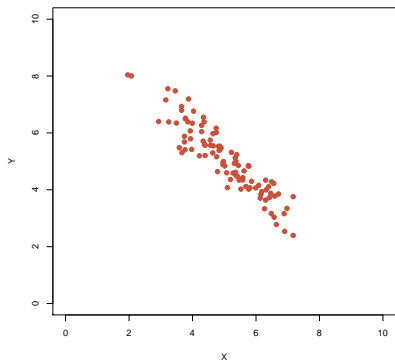
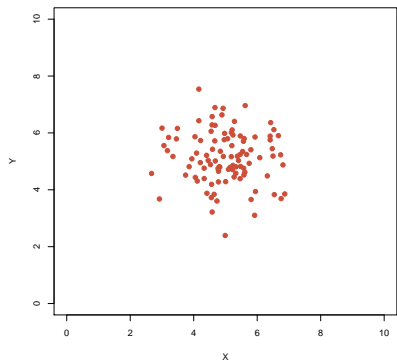
$$\sigma_X = 0.95, \sigma_Y = 0.92$$

$$\text{Cov}(X, Y) = -0.06$$

$$\text{Cor}(X, Y) = -0.069$$

$$\sigma_X = 1.13, \sigma_Y = 1.2$$

$$\text{Cov}(X, Y) = -1.26$$

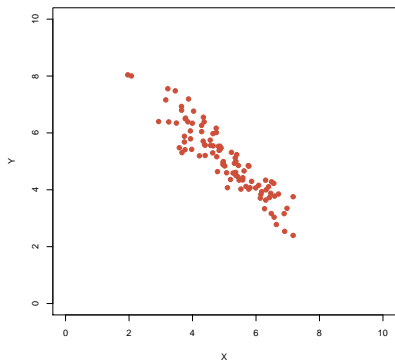
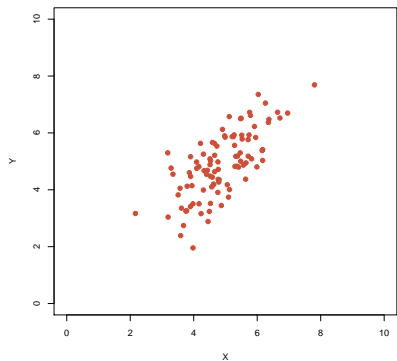


$$\sigma_X = 1.14, \sigma_Y = 0.78$$

$$\sigma_X = 1.13, \sigma_Y = 1.2$$

$$\text{Cov}(X, Y) = -1.26$$

$$\text{Cor}(X, Y) = -0.92$$



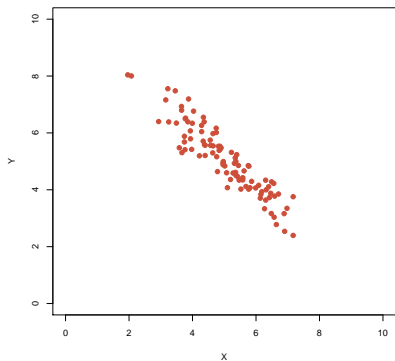
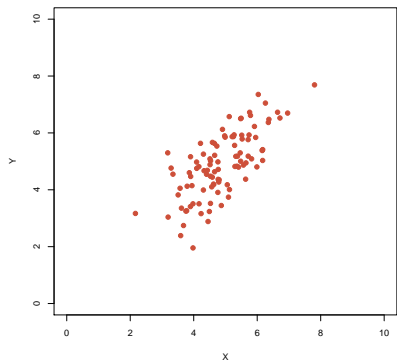
$$\sigma_X = 1.14, \sigma_Y = 0.78$$

$$\text{Cov}(X, Y) = 0.78$$

$$\sigma_X = 1.13, \sigma_Y = 1.2$$

$$\text{Cov}(X, Y) = -1.26$$

$$\text{Cor}(X, Y) = -0.92$$



$$\sigma_X = 1.14, \sigma_Y = 0.78$$

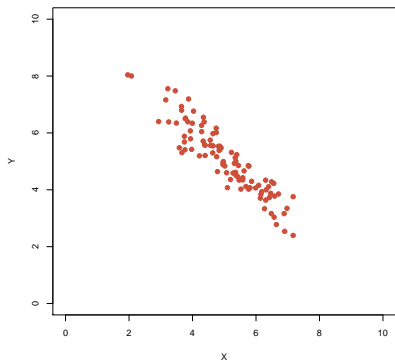
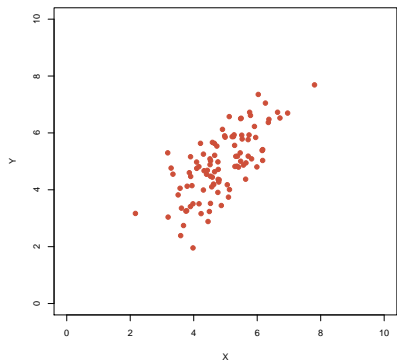
$$\text{Cov}(X, Y) = 0.78$$

$$\text{Cor}(X, Y) = 0.71$$

$$\sigma_X = 1.13, \sigma_Y = 1.2$$

$$\text{Cov}(X, Y) = -1.26$$

$$\text{Cor}(X, Y) = -0.92$$

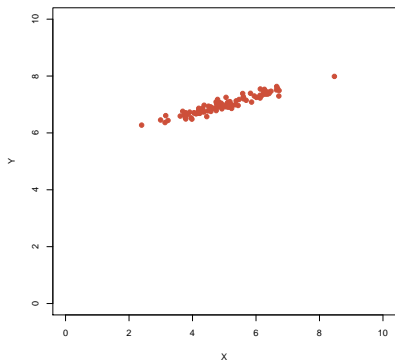
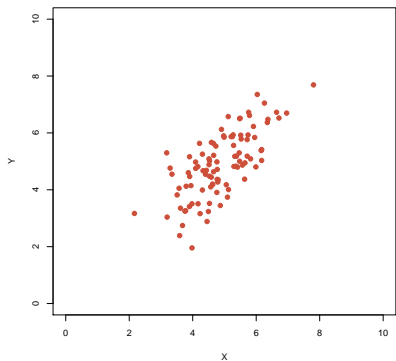




$$\sigma_X = 1.14, \sigma_Y = 0.78$$

$$\text{Cov}(X, Y) = 0.78$$

$$\text{Cor}(X, Y) = 0.71$$

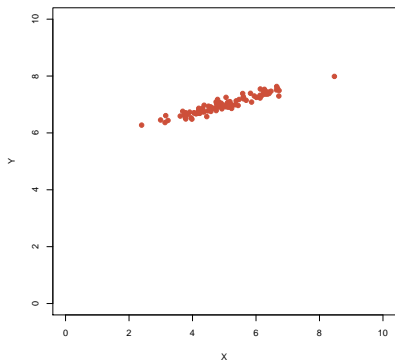
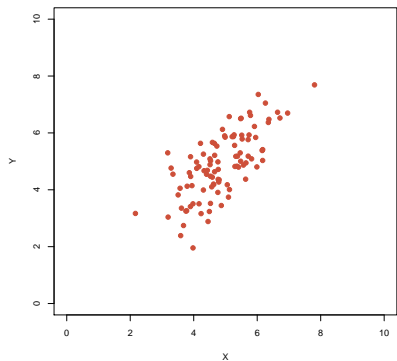


$$\sigma_X = 1.14, \sigma_Y = 0.78$$

$$\text{Cov}(X, Y) = 0.78$$

$$\text{Cor}(X, Y) = 0.71$$

$$\sigma_X = 1.03, \sigma_Y = 0.32$$



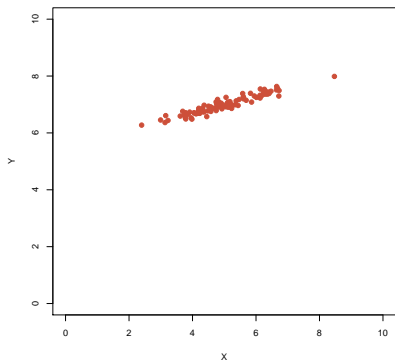
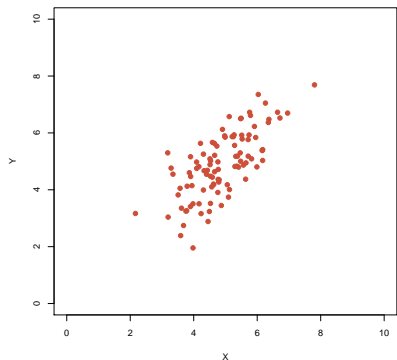
$$\sigma_X = 1.14, \sigma_Y = 0.78$$

$$\text{Cov}(X, Y) = 0.78$$

$$\text{Cor}(X, Y) = 0.71$$

$$\sigma_X = 1.03, \sigma_Y = 0.32$$

$$\text{Cov}(X, Y) = 0.32$$

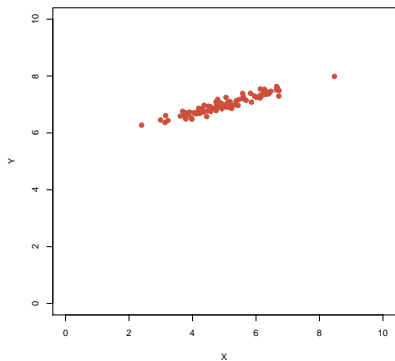
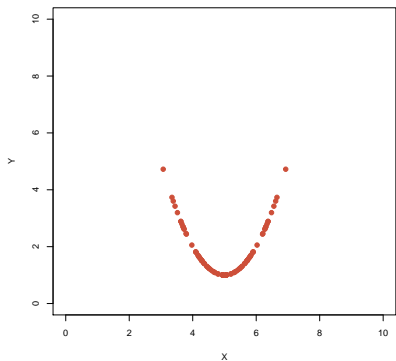


$$\sigma_X = 0.91, \sigma_Y = 0.88$$

$$\sigma_X = 1.03, \sigma_Y = 0.32$$

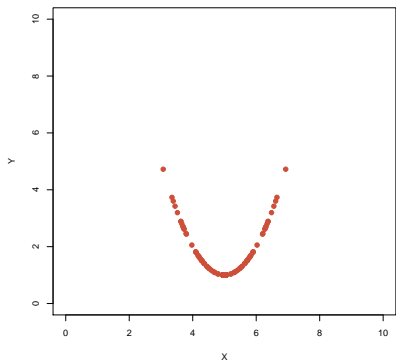
$$\text{Cov}(X, Y) = 0.32$$

$$\text{Cor}(X, Y) = 0.95$$



$$\sigma_X = 0.91, \sigma_Y = 0.88$$

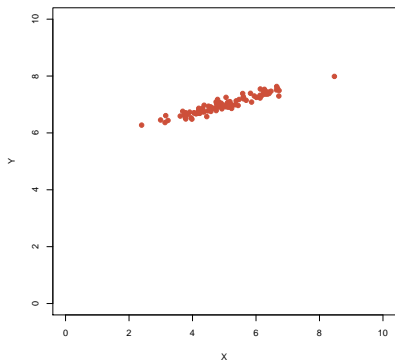
$$\text{Cov}(X, Y) = 0$$



$$\sigma_X = 1.03, \sigma_Y = 0.32$$

$$\text{Cov}(X, Y) = 0.32$$

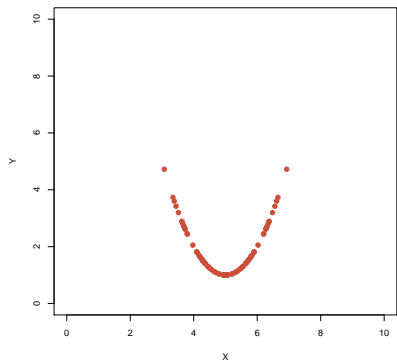
$$\text{Cor}(X, Y) = 0.95$$



$$\sigma_X = 0.91, \sigma_Y = 0.88$$

$$\text{Cov}(X, Y) = 0$$

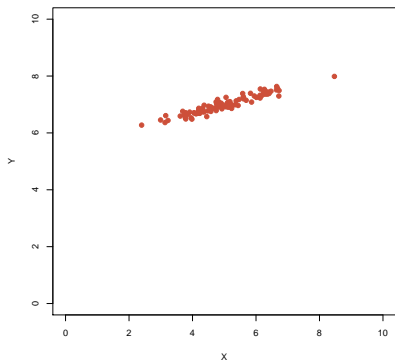
$$\text{Cor}(X, Y) = 0$$



$$\sigma_X = 1.03, \sigma_Y = 0.32$$

$$\text{Cov}(X, Y) = 0.32$$

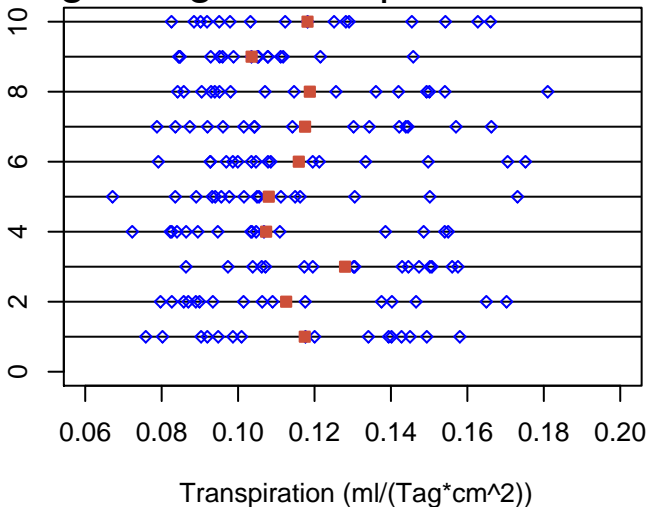
$$\text{Cor}(X, Y) = 0.95$$



# Übersicht

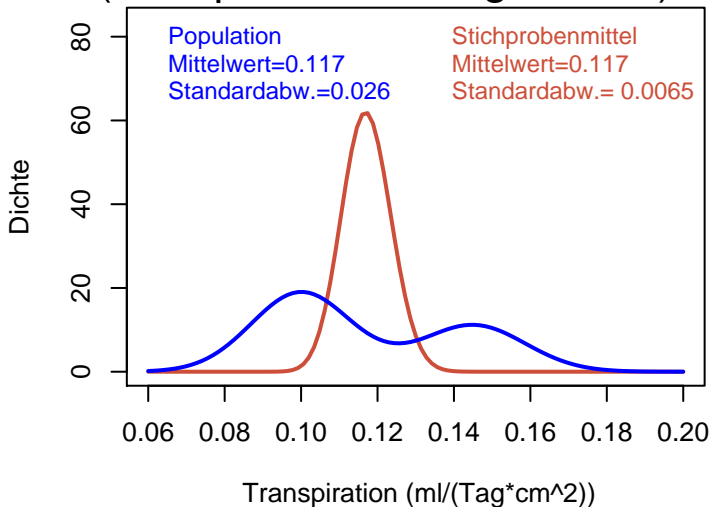
- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests**
- 3 Chi-Quadrat-Tests
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression
  - Lineare Zusammenhänge
  - $t$ -Test fuer lineare Zusammenhänge
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse
- 7 Versuchsplanung: Stichprobenumfang

# 10 Stichproben vom Umfang 16 und die zugehörigen Stichprobenmittel





# Verteilung der Stichprobenmittelwerte (Stichprobenumfang $n = 16$ )



# Die allgemeine Regel

Die Standardabweichung  
des Mittelwerts einer Stichprobe vom  
Umfang  $n$

# Die allgemeine Regel

Die Standardabweichung  
des Mittelwerts einer Stichprobe vom  
Umfang  $n$

ist

$$1/\sqrt{n}$$

mal

der Standardabweichung  
der Population.

Die Standardabweichung der Population  
bezeichnet man mit

$\sigma$   
(sigma).

Die Standardabweichung der Population  
bezeichnet man mit

$\sigma$   
(sigma).

Die Regel schreibt man häufig so:

$$\sigma(\bar{X}) = \frac{1}{\sqrt{n}}\sigma(X)$$

In der Praxis ist  
 $\sigma$   
unbekannt.

In der Praxis ist

$\sigma$

unbekannt.

Es wird durch

die Stichproben-Standardabweichung  $S$

geschätzt:

In der Praxis ist

$\sigma$

unbekannt.

Es wird durch

die Stichproben-Standardabweichung  $S$

geschätzt:

$$\sigma = ??$$



In der Praxis ist

$\sigma$

unbekannt.

Es wird durch

die Stichproben-Standardabweichung  $S$

geschätzt:

$$\sigma \approx S$$

$$s/\sqrt{n}$$

(die geschätzte  
Standardabweichung  
von  $\bar{x}$ )

nennt man den

*Standardfehler.*

$$s/\sqrt{n}$$

(die geschätzte  
Standardabweichung  
von  $\bar{x}$ )

nennt man den

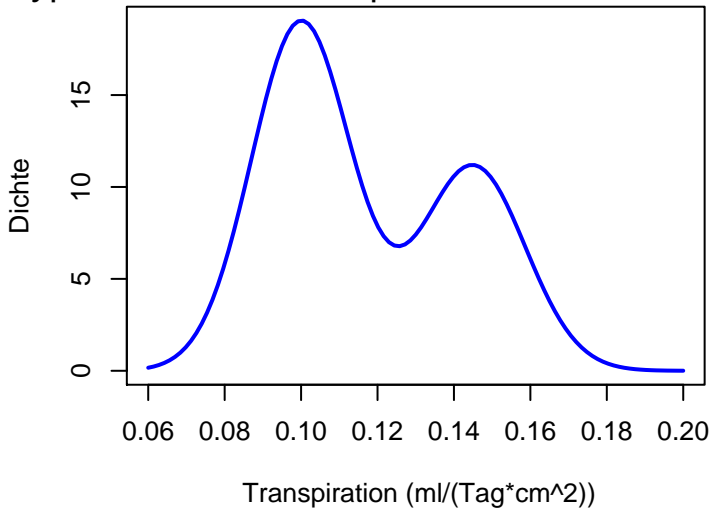
*Standardfehler.*

(Englisch: *standard error of the mean,*  
*standard error, SEM*)

Wir haben gesehen:

Auch wenn die Verteilung von  
 $x$  mehrgipfelig  
&  
asymmetrisch  
ist

# Hypothetische Transpirationsratenverteilung



ist die Verteilung von

$\bar{x}$

trotzdem

(annähernd)

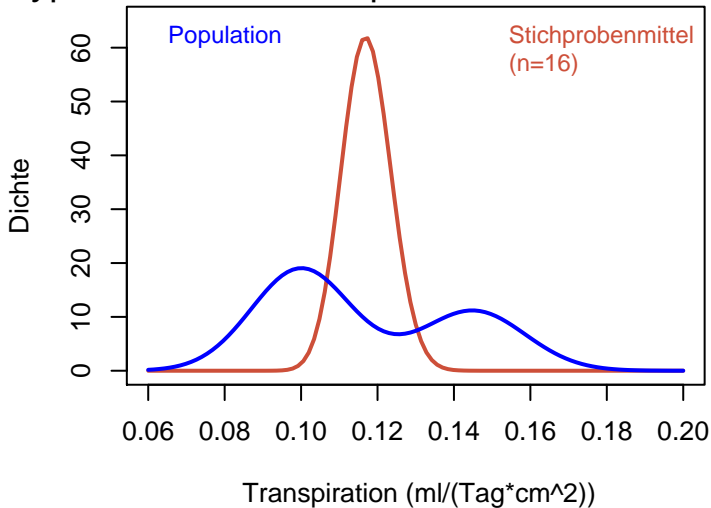
eingipfelig

&

symmetrisch

(wenn der Stichprobenumfang  $n$  nur groß genug ist)

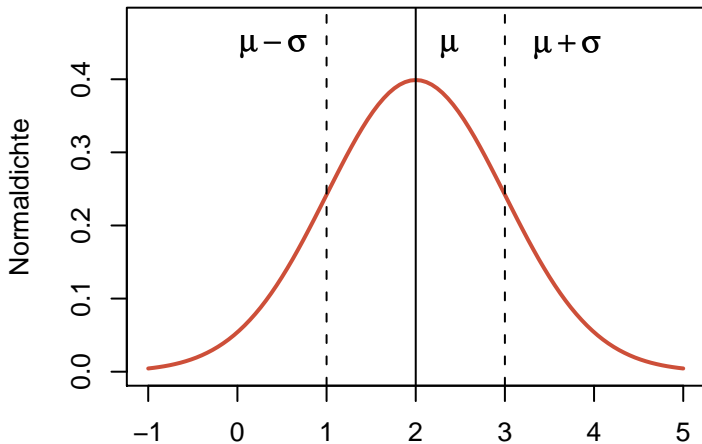
# Hypothetische Transpirationsratenverteilung



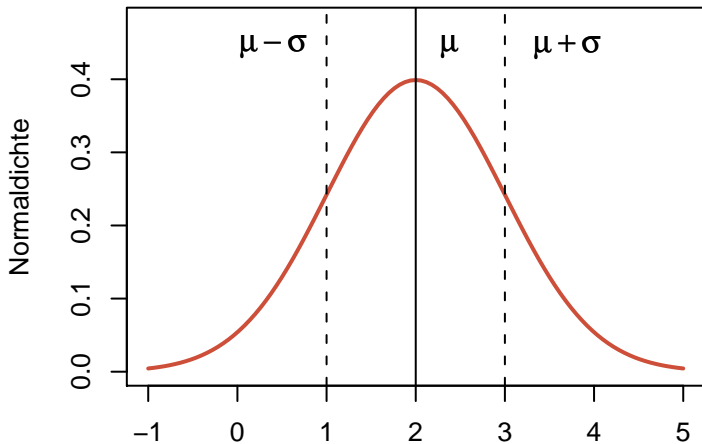
Die Verteilung von  $\bar{x}$   
hat annähernd  
eine ganz bestimmte Form:  
**die Normalverteilung.**



# Dichte der Normalverteilung



# Dichte der Normalverteilung



Die Normalverteilungsdichte heisst  
auch *Gauß'sche Glockenkurve*

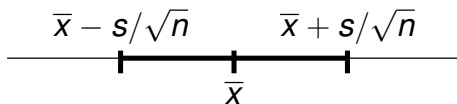
# Dichte der Normalverteilung



Die Normalverteilungsdichte heisst  
auch *Gauß'sche Glockenkurve*  
(nach Carl Friedrich Gauß, 1777-1855)

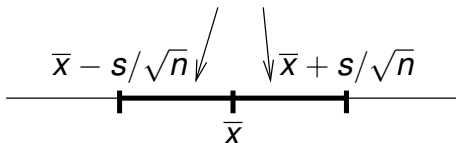
# Wichtige Folgerung

Wir betrachten das Intervall



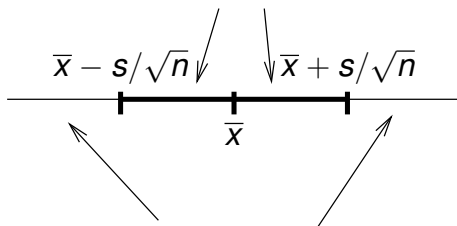
# Wichtige Folgerung

Mit Wahrscheinlichkeit ca.  $2/3$   
liegt  $\mu$  innerhalb dieses Intervalls



# Wichtige Folgerung

Mit Wahrscheinlichkeit ca.  $2/3$   
liegt  $\mu$  innerhalb dieses Intervalls



Mit Wahrscheinlichkeit ca.  $1/3$   
liegt  $\mu$  **ausserhalb** des Intervalls

Demnach:

Es kommt durchaus vor, dass  $\bar{x}$   
von  $\mu$   
um mehr als  
 $s/\sqrt{n}$  abweicht.

- Nehmen wir an, eine Population hat Mittelwert  $\mu$  und Standardabweichung  $\sigma$ .



- Nehmen wir an, eine Population hat Mittelwert  $\mu$  und Standardabweichung  $\sigma$ .
- Aus dieser Population ziehen wir eine Zufallsstichprobe vom Umfang  $n$ , mit Stichprobenmittelwert  $\bar{x}$ .

- Nehmen wir an, eine Population hat Mittelwert  $\mu$  und Standardabweichung  $\sigma$ .
- Aus dieser Population ziehen wir eine Zufallsstichprobe vom Umfang  $n$ , mit Stichprobenmittelwert  $\bar{x}$ .
- $\bar{x}$  ist eine Zufallsgröße

- Nehmen wir an, eine Population hat Mittelwert  $\mu$  und Standardabweichung  $\sigma$ .
- Aus dieser Population ziehen wir eine Zufallsstichprobe vom Umfang  $n$ , mit Stichprobenmittelwert  $\bar{x}$ .
- $\bar{x}$  ist eine Zufallsgröße mit Mittelwert  $\mu$  und Standardabweichung  $\sigma/\sqrt{n}$ .

- Nehmen wir an, eine Population hat Mittelwert  $\mu$  und Standardabweichung  $\sigma$ .
- Aus dieser Population ziehen wir eine Zufallsstichprobe vom Umfang  $n$ , mit Stichprobenmittelwert  $\bar{x}$ .
- $\bar{x}$  ist eine Zufallsgröße mit Mittelwert  $\mu$  und Standardabweichung  $\sigma/\sqrt{n}$ .
- Man schätzt die Standardabweichung von  $\bar{x}$  mit  $s/\sqrt{n}$ .

- Nehmen wir an, eine Population hat Mittelwert  $\mu$  und Standardabweichung  $\sigma$ .
- Aus dieser Population ziehen wir eine Zufallsstichprobe vom Umfang  $n$ , mit Stichprobenmittelwert  $\bar{x}$ .
- $\bar{x}$  ist eine Zufallsgröße mit Mittelwert  $\mu$  und Standardabweichung  $\sigma/\sqrt{n}$ .
- Man schätzt die Standardabweichung von  $\bar{x}$  mit  $s/\sqrt{n}$ .
- $s/\sqrt{n}$  nennt man den **Standardfehler**.

- Nehmen wir an, eine Population hat Mittelwert  $\mu$  und Standardabweichung  $\sigma$ .
- Aus dieser Population ziehen wir eine Zufallsstichprobe vom Umfang  $n$ , mit Stichprobenmittelwert  $\bar{x}$ .
- $\bar{x}$  ist eine Zufallsgröße mit Mittelwert  $\mu$  und Standardabweichung  $\sigma/\sqrt{n}$ .
- Man schätzt die Standardabweichung von  $\bar{x}$  mit  $s/\sqrt{n}$ .
- $s/\sqrt{n}$  nennt man den **Standardfehler**.
- Schwankungen in  $\bar{x}$  von der Größe  $s/\sqrt{n}$  kommen häufig vor.

- Nehmen wir an, eine Population hat Mittelwert  $\mu$  und Standardabweichung  $\sigma$ .
- Aus dieser Population ziehen wir eine Zufallsstichprobe vom Umfang  $n$ , mit Stichprobenmittelwert  $\bar{x}$ .
- $\bar{x}$  ist eine Zufallsgröße mit Mittelwert  $\mu$  und Standardabweichung  $\sigma/\sqrt{n}$ .
- Man schätzt die Standardabweichung von  $\bar{x}$  mit  $s/\sqrt{n}$ .
- $s/\sqrt{n}$  nennt man den **Standardfehler**.
- Schwankungen in  $\bar{x}$  von der Größe  $s/\sqrt{n}$  kommen häufig vor.  
Solche Schwankungen sind „**nicht signifikant**“: sie könnten Zufall sein.

## Allgemein gilt

Sind  $X_1, \dots, X_n$  unabhängig aus einer Normalverteilung mit Mittelwert  $\mu$  gezogen und ist

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2},$$

so ist

$$\frac{\bar{X} - \mu}{s/\sqrt{n}}$$

*t*-verteilt mit  $n - 1$  Freiheitsgraden (df=*degrees of freedom*).



## Allgemein gilt

Sind  $X_1, \dots, X_n$  unabhängig aus einer Normalverteilung mit Mittelwert  $\mu$  gezogen und ist

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2},$$

so ist

$$\frac{\bar{X} - \mu}{s/\sqrt{n}}$$

*t*-verteilt mit  $n - 1$  Freiheitsgraden (df=*degrees of freedom*).

Die t-Verteilung heißt auch *Student-Verteilung*, da Gosset sie unter diesem Pseudonym publiziert hat.

Wie (un)wahrscheinlich ist nun eine  
so große Abweichung wie 2.35 Standardfehler?

$$\Pr(T = 2.34) =$$

Wie (un)wahrscheinlich ist nun eine  
so große Abweichung wie 2.35 Standardfehler?

$$\Pr(T = 2.34) = 0$$

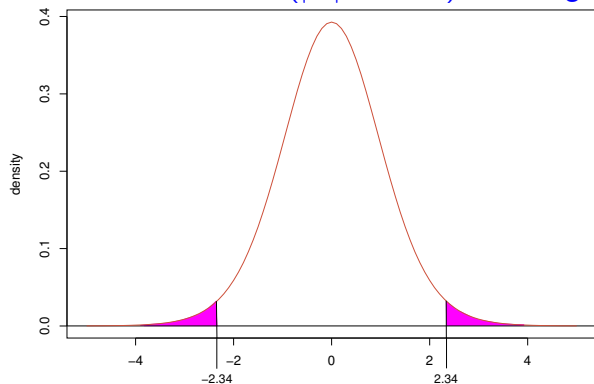
Wie (un)wahrscheinlich ist nun eine **mindestens**  
so große Abweichung wie 2.35 Standardfehler?

$\Pr(T = 2.34) = 0$       **Das bringt nichts!**

Wie (un)wahrscheinlich ist nun eine  
so große Abweichung wie 2.35 Standardfehler?

$$\Pr(T = 2.34) = 0 \quad \text{Das bringt nichts!}$$

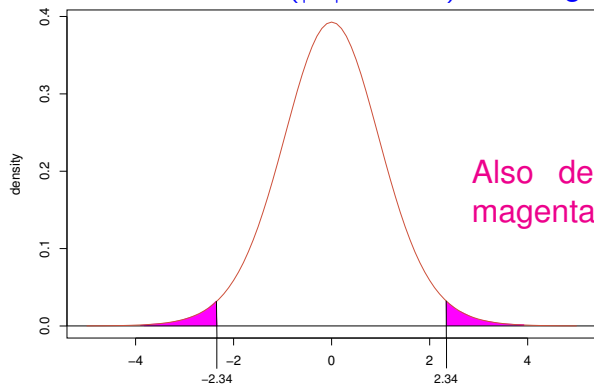
Zu berechnen ist  $\Pr(|T| \geq 2.34)$ , der sog.  $p$ -Wert.



Wie (un)wahrscheinlich ist nun eine  
so große Abweichung wie 2.35 Standardfehler?

$$\Pr(T = 2.34) = 0 \quad \text{Das bringt nichts!}$$

Zu berechnen ist  $\Pr(|T| \geq 2.34)$ , der sog.  $p$ -Wert.



Also der Gesamthalt der  
magentafarbenen Flächen.

Wir halten fest:

$$p\text{-Wert} = 0.03254$$

Wir halten fest:

$$p\text{-Wert} = 0.03254$$

d.h.: Wenn die **Nullhypothese** “alles nur Zufall”, also in diesem Fall die Hypothese  $\mu = 0$  gilt, dann ist eine mindestens so große Abweichung sehr unwahrscheinlich.



Wir halten fest:

$$p\text{-Wert} = 0.03254$$

d.h.: Wenn die **Nullhypothese** “alles nur Zufall”, also in diesem Fall die Hypothese  $\mu = 0$  gilt, dann ist eine mindestens so große Abweichung sehr unwahrscheinlich.

Wenn wir beschließen, dass wir die Nullhypothese immer verwerfen, wenn der  $p$ -Wert unterhalb einem **Signifikanzniveau** von 0.05 liegt, gilt:

Wir halten fest:

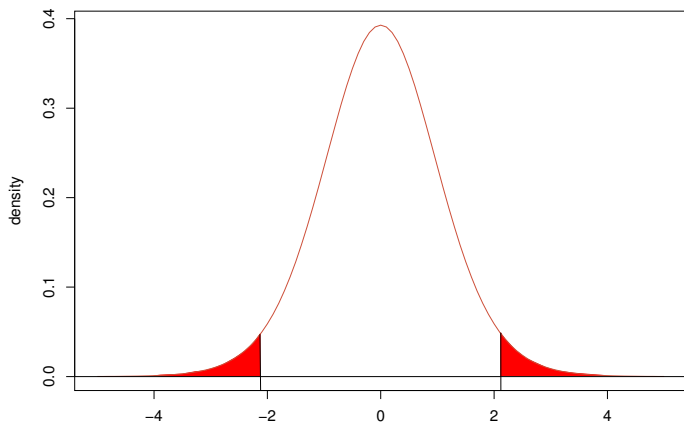
$$p\text{-Wert} = 0.03254$$

d.h.: Wenn die **Nullhypothese** “alles nur Zufall”, also in diesem Fall die Hypothese  $\mu = 0$  gilt, dann ist eine mindestens so große Abweichung sehr unwahrscheinlich.

Wenn wir beschließen, dass wir die Nullhypothese immer verwerfen, wenn der  $p$ -Wert unterhalb einem **Signifikanzniveau** von 0.05 liegt, gilt:

Falls die Nullhypothese zutrifft, ist die Wahrscheinlichkeit, dass wir sie zu Unrecht verwerfen, lediglich 0.05.

Wenn wir uns auf ein Signifikanzniveau von  $\alpha = 0.05$  festlegen, verwerfen wir die Nullhypothese also, wenn der  $t$ -Wert in den roten Bereich fällt:



(hier am Beispiel der  $t$ -Verteilung mit  $df = 16$  Freiheitsgraden)

# Statistische Tests: Die wichtigsten Begriffe

Nullhypothese  $H_0$  : wollen wir meistens verwerfen, denn sie besagt, dass die interessanten Auffälligkeiten in den Daten nur Zufallsschwankungen sind.

# Statistische Tests: Die wichtigsten Begriffe

Nullhypothese  $H_0$  : wollen wir meistens verwerfen, denn sie besagt, dass die interessanten Auffälligkeiten in den Daten nur Zufallsschwankungen sind.

Signifikanzniveau  $\alpha$  : Wahrscheinlichkeit, dass wir, falls die Nullhypothese gilt, diese zu Unrecht verwerfen.

# Statistische Tests: Die wichtigsten Begriffe

**Nullhypothese  $H_0$**  : wollen wir meistens verwerfen, denn sie besagt, dass die interessanten Auffälligkeiten in den Daten nur Zufallsschwankungen sind.

**Signifikanzniveau  $\alpha$**  : Wahrscheinlichkeit, dass wir, falls die Nullhypothese gilt, diese zu Unrecht verwerfen.

**Teststatistik** : Misst, auffällig unsere Daten von der Nullhypothese abweichen.

# Statistische Tests: Die wichtigsten Begriffe

**Nullhypothese  $H_0$**  : wollen wir meistens verwerfen, denn sie besagt, dass die interessanten Auffälligkeiten in den Daten nur Zufallsschwankungen sind.

**Signifikanzniveau  $\alpha$**  : Wahrscheinlichkeit, dass wir, falls die Nullhypothese gilt, diese zu Unrecht verwerfen.

**Teststatistik** : Misst, auffällig unsere Daten von der Nullhypothese abweichen.

**$p$ -Wert** : Für einen beobachteten Wert  $t$  der Teststatistik ist der  $p$ -Wert die Wahrscheinlichkeit, dass, falls die Nullhypothese gilt, die Teststatistik (etwa bei einer hypothetischen Wiederholung des Versuchs) einen mindestens so extremen Wert wie  $t$  annimmt. Dabei hängt es von der Art des Tests ab (z.B. einseitig/zweiseitig), was "extrem" heißt.

- Wir verwerfen  $H_0$ , falls der  $p$ -Wert kleiner als  $\alpha$  wird. (Üblich ist  $\alpha = 0.05$ ).



- Wir verwerfen  $H_0$ , falls der  $p$ -Wert kleiner als  $\alpha$  wird. (Üblich ist  $\alpha = 0.05$ ).
- Damit ergibt sich, dass wir nur in einem Anteil  $\alpha$  der Fälle, in denen  $H_0$  gilt, diese (fälschlicherweise) verwerfen.

- Wir verwerfen  $H_0$ , falls der  $p$ -Wert kleiner als  $\alpha$  wird. (Üblich ist  $\alpha = 0.05$ ).
- Damit ergibt sich, dass wir nur in einem Anteil  $\alpha$  der Fälle, in denen  $H_0$  gilt, diese (fälschlicherweise) verwerfen.
- Auch wer immer nur Daten analysiert, in denen außer Zufallsschwankungen nichts steckt, wird in einem Anteil  $\alpha$  der Tests die Nullhypothese verwerfen.

- Wir verwerfen  $H_0$ , falls der  $p$ -Wert kleiner als  $\alpha$  wird. (Üblich ist  $\alpha = 0.05$ ).
- Damit ergibt sich, dass wir nur in einem Anteil  $\alpha$  der Fälle, in denen  $H_0$  gilt, diese (fälschlicherweise) verwerfen.
- Auch wer immer nur Daten analysiert, in denen außer Zufallsschwankungen nichts steckt, wird in einem Anteil  $\alpha$  der Tests die Nullhypothese verwerfen.
- Ein schwerer Verstoß gegen die Wissenschaftlichkeit ist daher, so lange statistische Tests durchzuführen, bis mal  $H_0$  auf einem Signifikanzniveau von 5% verworfen werden kann, und dann nur letzteres zu veröffentlichen.

Welche  $t$ -Werte sind “auf dem 5%-Niveau” signifikant?

Anzahl Freiheitsgrade	$ t  \geq \dots$
5	2.57
10	2.23
20	2.09
30	2.04
100	1.98
$\infty$	1.96

Welche  $t$ -Werte sind “auf dem 5%-Niveau” signifikant?

Anzahl Freiheitsgrade	$ t  \geq \dots$
5	2.57
10	2.23
20	2.09
30	2.04
100	1.98
$\infty$	1.96

```
> qt(0.025,df=c(5,10,20,30,100,1e100))
```

Welche  $t$ -Werte sind “auf dem 5%-Niveau” signifikant?

Anzahl Freiheitsgrade	$ t  \geq \dots$
5	2.57
10	2.23
20	2.09
30	2.04
100	1.98
$\infty$	1.96

```
> qt(0.025,df=c(5,10,20,30,100,1e100))
[1] -2.570582 -2.228139 -2.085963 -2.042272 -1.983972
-1.959964
```

- Wir möchten belegen, dass eine Abweichung in den Daten vermutlich nicht allein auf Zufallsschwankung beruht.

- Wir möchten belegen, dass eine Abweichung in den Daten vermutlich nicht allein auf Zufallsschwankung beruht.
- Dazu spezifizieren wir zunächst eine **Nullhypothese  $H_0$** , d.h. wir konkretisieren, was “allein auf Zufall beruhen” bedeutet.



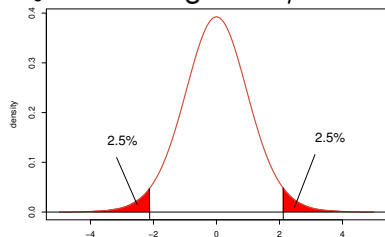
- Wir möchten belegen, dass eine Abweichung in den Daten vermutlich nicht allein auf Zufallsschwankung beruht.
- Dazu spezifizieren wir zunächst eine **Nullhypothese  $H_0$** , d.h. wir konkretisieren, was “allein auf Zufall beruhen” bedeutet.
- Dann versuchen wir zu zeigen: Wenn  **$H_0$**  gilt, dann ist eine **Abweichung** wie, die mindestens so groß sind wie die beobachtete, sehr unwahrscheinlich.

- Wir möchten belegen, dass eine Abweichung in den Daten vermutlich nicht allein auf Zufallsschwankung beruht.
- Dazu spezifizieren wir zunächst eine **Nullhypothese  $H_0$** , d.h. wir konkretisieren, was “allein auf Zufall beruhen” bedeutet.
- Dann versuchen wir zu zeigen: Wenn  **$H_0$**  gilt, dann ist eine **Abweichung** wie, die mindestens so groß sind wie die beobachtete, sehr unwahrscheinlich.
- Wenn uns das gelingt, verwerfen wir  **$H_0$** .

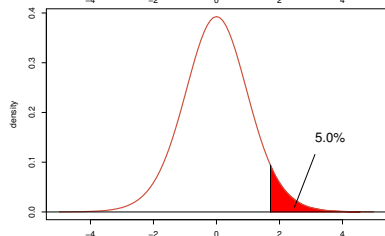
- Wir möchten belegen, dass eine Abweichung in den Daten vermutlich nicht allein auf Zufallsschwankung beruht.
- Dazu spezifizieren wir zunächst eine **Nullhypothese  $H_0$** , d.h. wir konkretisieren, was “allein auf Zufall beruhen” bedeutet.
- Dann versuchen wir zu zeigen: Wenn  $H_0$  gilt, dann ist eine **Abweichung** wie, die mindestens so groß sind wie die beobachtete, sehr unwahrscheinlich.
- Wenn uns das gelingt, verwerfen wir  $H_0$ .
- Was wir als **Abweichung** auffassen, sollten klar sein, bevor wir die Daten sehen.

# Zweiseitig oder einseitig testen?

Wir beobachten einen Wert  $x$ , der deutlich größer als der  $H_0$ -Erwartungswert  $\mu$  ist.



$$p\text{-Wert} = \Pr_{H_0}(|X - \mu| \geq |x - \mu|)$$



$$p\text{-Wert} = \Pr_{H_0}(X \geq x)$$

# Reine Lehre des statistischen Testens

- Formuliere eine **Nullhypothese  $H_0$** , z.B.  $\mu = 0$ .

# Reine Lehre des statistischen Testens

- Formuliere eine **Nullhypothese**  $H_0$ , z.B.  $\mu = 0$ .
- Lege ein **Signifikanzniveau**  $\alpha$  fest; üblich ist  $\alpha = 0.05$ .

# Reine Lehre des statistischen Testens

- Formuliere eine **Nullhypothese**  $H_0$ , z.B.  $\mu = 0$ .
- Lege ein **Signifikanzniveau**  $\alpha$  fest; üblich ist  $\alpha = 0.05$ .
- Lege ein Ereignis  $\mathcal{A}$  (“extreme Abweichung”) fest, so dass

$$\Pr_{H_0}(\mathcal{A}) = \alpha$$

(oder zumindest  $\Pr_{H_0}(\mathcal{A}) \leq \alpha$ ).

# Reine Lehre des statistischen Testens

- Formuliere eine **Nullhypothese**  $H_0$ , z.B.  $\mu = 0$ .
- Lege ein **Signifikanzniveau**  $\alpha$  fest; üblich ist  $\alpha = 0.05$ .
- Lege ein Ereignis  $\mathcal{A}$  ("extreme Abweichung") fest, so dass

$$\Pr_{H_0}(\mathcal{A}) = \alpha$$

(oder zumindest  $\Pr_{H_0}(\mathcal{A}) \leq \alpha$ ).

z.B.  $\mathcal{A} = \{\bar{X} > q\}$  oder  $\mathcal{A} = \{|\bar{X} - \mu| > r\}$



# Reine Lehre des statistischen Testens

- Formuliere eine **Nullhypothese**  $H_0$ , z.B.  $\mu = 0$ .
- Lege ein **Signifikanzniveau**  $\alpha$  fest; üblich ist  $\alpha = 0.05$ .
- Lege ein Ereignis  $\mathcal{A}$  ("extreme Abweichung") fest, so dass

$$\Pr_{H_0}(\mathcal{A}) = \alpha$$

(oder zumindest  $\Pr_{H_0}(\mathcal{A}) \leq \alpha$ ).

z.B.  $\mathcal{A} = \{\bar{X} > q\}$  oder  $\mathcal{A} = \{|\bar{X} - \mu| > r\}$

allgemein:  $\mathcal{A} = \{\mathbf{p}\text{-Wert} \leq \alpha\}$

# Reine Lehre des statistischen Testens

- Formuliere eine **Nullhypothese**  $H_0$ , z.B.  $\mu = 0$ .
- Lege ein **Signifikanzniveau**  $\alpha$  fest; üblich ist  $\alpha = 0.05$ .
- Lege ein Ereignis  $\mathcal{A}$  ("extreme Abweichung") fest, so dass

$$\Pr_{H_0}(\mathcal{A}) = \alpha$$

(oder zumindest  $\Pr_{H_0}(\mathcal{A}) \leq \alpha$ ).

z.B.  $\mathcal{A} = \{\bar{X} > q\}$  oder  $\mathcal{A} = \{|\bar{X} - \mu| > r\}$

allgemein:  $\mathcal{A} = \{\mathbf{p}\text{-Wert} \leq \alpha\}$

- **ERST DANN:** Betrachte die Daten und überprüfe, **ob**  $\mathcal{A}$  eintritt.

# Reine Lehre des statistischen Testens

- Formuliere eine **Nullhypothese**  $H_0$ , z.B.  $\mu = 0$ .
- Lege ein **Signifikanzniveau**  $\alpha$  fest; üblich ist  $\alpha = 0.05$ .
- Lege ein Ereignis  $\mathcal{A}$  (“extreme Abweichung”) fest, so dass

$$\Pr_{H_0}(\mathcal{A}) = \alpha$$

(oder zumindest  $\Pr_{H_0}(\mathcal{A}) \leq \alpha$ ).

z.B.  $\mathcal{A} = \{\bar{X} > q\}$  oder  $\mathcal{A} = \{|\bar{X} - \mu| > r\}$

allgemein:  $\mathcal{A} = \{\text{p-Wert} \leq \alpha\}$

- **ERST DANN:** Betrachte die Daten und überprüfe, **ob**  $\mathcal{A}$  eintritt.
- Dann ist die Wahrscheinlichkeit, dass  $H_0$  verworfen wird, wenn  $H_0$  eigentlich richtig ist (“**Fehler erster Art**”), lediglich  $\alpha$ .

# Verstöße gegen die reine Lehre

“Beim zweiseitigen Testen kam ein  $p$ -Wert von 0.06 raus. Also hab ich einseitig getestet, da hat's dann funktioniert.”

# Verstöße gegen die reine Lehre

“Beim zweiseitigen Testen kam ein  $p$ -Wert von 0.06 raus. Also hab ich einseitig getestet, da hat's dann funktioniert.”

genauso problematisch:

# Verstöße gegen die reine Lehre

“Beim zweiseitigen Testen kam ein  $p$ -Wert von 0.06 raus. Also hab ich einseitig getestet, da hat’s dann funktioniert.”

genauso problematisch:

“Beim ersten Blick auf die Daten habe ich sofort gesehen, dass  $\bar{x}$  größer ist als  $\mu_{H_0}$ . Also habe ich gleich einseitig getestet”

## Wichtig

Die Entscheidung ob einseitig oder zweiseitig getestet wird darf nicht von den konkreten Daten, die zum Test verwendet werden, abhängen.

## Wichtig

Die Entscheidung ob einseitig oder zweiseitig getestet wird darf nicht von den konkreten Daten, die zum Test verwendet werden, abhängen.

Allgemeiner: Ist  $\mathcal{A}$  das Ereignis, dass zum Verwerfen von  $H_0$  führt (falls es eintritt), so muss die Festlegung von  $\mathcal{A}$  stattfinden bevor man in den Daten herumgeschnüffelt hat.



Die **Wahl von  $\mathcal{A}$**  sollte von der **Alternative  $H_1$**  abhängen, also für das, was wir eigentlich zeigen wollen, indem wir  $H_0$  durch einen Test verwerfen. Es muss gelten:

$$\Pr_{H_0}(\mathcal{A}) = \alpha$$

und

$$\Pr_{H_1}(\mathcal{A}) = \text{möglichst groß,}$$

Die **Wahl von  $\mathcal{A}$**  sollte von der **Alternative  $H_1$**  abhängen, also für das, was wir eigentlich zeigen wollen, indem wir  $H_0$  durch einen Test verwerfen. Es muss gelten:

$$\Pr_{H_0}(\mathcal{A}) = \alpha$$

und

$$\Pr_{H_1}(\mathcal{A}) = \text{möglichst groß,}$$

damit die **W'keit eines Fehlers zweiter Art**, dass also  $H_0$  nicht verworfen wird, obwohl  $H_1$  zutrifft, möglichst klein.

# Beispiele

- Wenn wir von Anfang an unsere Vermutung belegen wollten, dass sich die Trauerschnäpper bei grünem Licht stärker auf eine Richtung konzentrieren als bei blauem, dürfen wir einseitig testen.

# Beispiele

- Wenn wir von Anfang an unsere Vermutung belegen wollten, dass sich die Trauerschnäpper bei grünem Licht stärker auf eine Richtung konzentrieren als bei blauem, dürfen wir einseitig testen.
- Wenn dann aber noch so deutlich herauskommt, dass die Richtungswahl bei blauem Licht deutlicher war, so ist das dann nicht als signifikant zu betrachten.

# Beispiele

- Wenn wir von Anfang an unsere Vermutung belegen wollten, dass sich die Trauerschnäpper bei grünem Licht stärker auf eine Richtung konzentrieren als bei blauem, dürfen wir einseitig testen.
- Wenn dann aber noch so deutlich herauskommt, dass die Richtungswahl bei blauem Licht deutlicher war, so ist das dann nicht als signifikant zu betrachten.
- Wenn wir von Anfang an die Vermutung belegen wollten, dass der Kork an der Nordseite des Baumes dicker war, dürfen wir einseitig testen.

# Beispiele

- Wenn wir von Anfang an unsere Vermutung belegen wollten, dass sich die Trauerschnäpper bei grünem Licht stärker auf eine Richtung konzentrieren als bei blauem, dürfen wir einseitig testen.
- Wenn dann aber noch so deutlich herauskommt, dass die Richtungswahl bei blauem Licht deutlicher war, so ist das dann nicht als signifikant zu betrachten.
- Wenn wir von Anfang an die Vermutung belegen wollten, dass der Kork an der Nordseite des Baumes dicker war, dürfen wir einseitig testen.
- Wenn dann aber noch so deutlich herauskommt, dass der Kork im Westen dicker ist, ist das nicht mehr signifikant.

Angenommen,  $H_0$  wird auf dem 5%-Niveau verworfen. Welche Aussage gilt dann?

- Die Nullhypothese ist falsch.
- 
-

Angenommen,  $H_0$  wird auf dem 5%-Niveau verworfen. Welche Aussage gilt dann?

- ~~Die Nullhypothese ist falsch.~~
- 
-



Angenommen,  $H_0$  wird auf dem 5%-Niveau verworfen. Welche Aussage gilt dann?

- ~~Die Nullhypothese ist falsch.~~
- $H_0$  ist mit 95%-iger Wahrscheinlichkeit falsch.
-

Angenommen,  $H_0$  wird auf dem 5%-Niveau verworfen. Welche Aussage gilt dann?

- ~~Die Nullhypothese ist falsch.~~
- ~~$H_0$  ist mit 95% iger Wahrscheinlichkeit falsch.~~
-

Angenommen,  $H_0$  wird auf dem 5%-Niveau verworfen. Welche Aussage gilt dann?

- ~~Die Nullhypothese ist falsch.~~
- ~~$H_0$  ist mit 95%-iger Wahrscheinlichkeit falsch.~~
- Falls die Nullhypothese wahr ist, beobachtet man ein so extremes Ergebnis nur in 5% der Fälle.

Angenommen,  $H_0$  wird auf dem 5%-Niveau verworfen. Welche Aussage gilt dann?

- ~~Die Nullhypothese ist falsch.~~
- ~~$H_0$  ist mit 95% iger Wahrscheinlichkeit falsch.~~
- Falls die Nullhypothese wahr ist, beobachtet man ein so extremes Ergebnis nur in 5% der Fälle. ✓

Angenommen,  $H_0$  konnte durch den Test nicht verworfen werden. Welche Aussagen sind dann richtig?

- Wir müssen die Alternative  $H_1$  verwerfen.
- 
- 
- 
- 
-

Angenommen,  $H_0$  konnte durch den Test nicht verworfen werden. Welche Aussagen sind dann richtig?

- ~~Wir müssen die Alternative  $H_1$  verwerfen.~~
- 
- 
- 
- 
-

Angenommen,  $H_0$  konnte durch den Test nicht verworfen werden. Welche Aussagen sind dann richtig?

- ~~Wir müssen die Alternative  $H_1$  verwerfen.~~
- $H_0$  ist wahr.
- 
- 
- 
-

Angenommen,  $H_0$  konnte durch den Test nicht verworfen werden. Welche Aussagen sind dann richtig?

- ~~Wir müssen die Alternative  $H_1$  verwerfen.~~
- ~~$H_0$  ist wahr.~~
- 
- 
- 
-



Angenommen,  $H_0$  konnte durch den Test nicht verworfen werden. Welche Aussagen sind dann richtig?

- ~~Wir müssen die Alternative  $H_1$  verwerfen.~~
- ~~$H_0$  ist wahr.~~
- $H_0$  ist wahrscheinlich wahr.
- 
- 
-

Angenommen,  $H_0$  konnte durch den Test nicht verworfen werden. Welche Aussagen sind dann richtig?

- ~~Wir müssen die Alternative  $H_1$  verwerfen.~~
- ~~$H_0$  ist wahr.~~
- ~~$H_0$  ist wahrscheinlich wahr.~~
- 
- 
-

Angenommen,  $H_0$  konnte durch den Test nicht verworfen werden. Welche Aussagen sind dann richtig?

- ~~Wir müssen die Alternative  $H_1$  verwerfen.~~
- ~~$H_0$  ist wahr.~~
- ~~$H_0$  ist wahrscheinlich wahr.~~
- Es ist ungefährlich, davon auszugehen, dass  $H_0$  zutrifft.
- 
-

Angenommen,  $H_0$  konnte durch den Test nicht verworfen werden. Welche Aussagen sind dann richtig?

- ~~Wir müssen die Alternative  $H_1$  verwerfen.~~
- ~~$H_0$  ist wahr.~~
- ~~$H_0$  ist wahrscheinlich wahr.~~
- ~~Es ist ungefährlich, davon auszugehen, dass  $H_0$  zutrifft.~~
- 
-

Angenommen,  $H_0$  konnte durch den Test nicht verworfen werden. Welche Aussagen sind dann richtig?

- ~~Wir müssen die Alternative  $H_1$  verwerfen.~~
- ~~$H_0$  ist wahr.~~
- ~~$H_0$  ist wahrscheinlich wahr.~~
- ~~Es ist ungefährlich, davon auszugehen, dass  $H_0$  zutrifft.~~
- Auch wenn  $H_0$  zutrifft, ist es nicht sehr unwahrscheinlich, dass unsere Teststatistik einen mindestens so extrem erscheinenden Wert annimmt.
-

Angenommen,  $H_0$  konnte durch den Test nicht verworfen werden. Welche Aussagen sind dann richtig?

- ~~Wir müssen die Alternative  $H_1$  verwerfen.~~
- ~~$H_0$  ist wahr.~~
- ~~$H_0$  ist wahrscheinlich wahr.~~
- ~~Es ist ungefährlich, davon auszugehen, dass  $H_0$  zutrifft.~~
- Auch wenn  $H_0$  zutrifft, ist es nicht sehr unwahrscheinlich, dass unsere Teststatistik einen mindestens so extrem erscheinenden Wert annimmt. ✓
-

Angenommen,  $H_0$  konnte durch den Test nicht verworfen werden. Welche Aussagen sind dann richtig?

- ~~Wir müssen die Alternative  $H_1$  verwerfen.~~
- ~~$H_0$  ist wahr.~~
- ~~$H_0$  ist wahrscheinlich wahr.~~
- ~~Es ist ungefährlich, davon auszugehen, dass  $H_0$  zutrifft.~~
- Auch wenn  $H_0$  zutrifft, ist es nicht sehr unwahrscheinlich, dass unsere Teststatistik einen mindestens so extrem erscheinenden Wert annimmt. ✓
- Die Nullhypothese ist in dieser Hinsicht mit den Daten verträglich.

Angenommen,  $H_0$  konnte durch den Test nicht verworfen werden. Welche Aussagen sind dann richtig?

- ~~Wir müssen die Alternative  $H_1$  verwerfen.~~
- ~~$H_0$  ist wahr.~~
- ~~$H_0$  ist wahrscheinlich wahr.~~
- ~~Es ist ungefährlich, davon auszugehen, dass  $H_0$  zutrifft.~~
- Auch wenn  $H_0$  zutrifft, ist es nicht sehr unwahrscheinlich, dass unsere Teststatistik einen mindestens so extrem erscheinenden Wert annimmt. ✓
- Die Nullhypothese ist in dieser Hinsicht mit den Daten verträglich. ✓



# Frage

Hipparion:  
Laubfresser  $\longrightarrow$  Grasfresser

# Frage

Hipparion:

Laubfresser  $\longrightarrow$  Grasfresser

andere Nahrung  $\longrightarrow$  andere Zähne?

# Frage

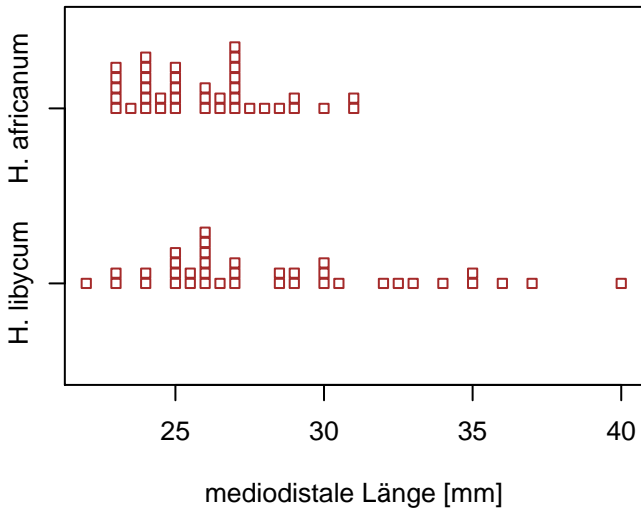
Hipparion:

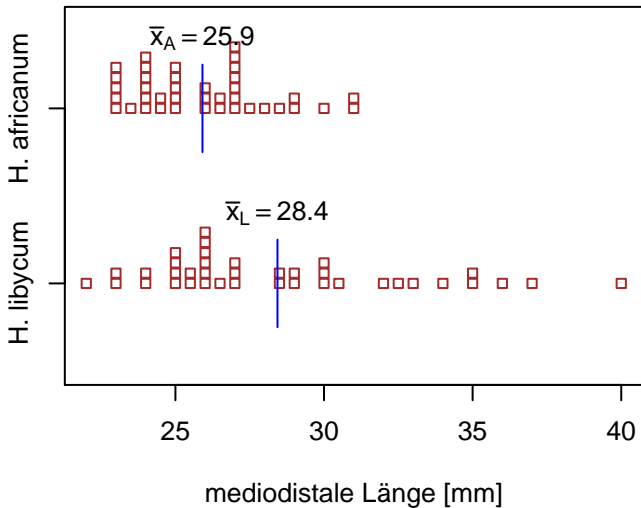
Laubfresser  $\longrightarrow$  Grasfresser

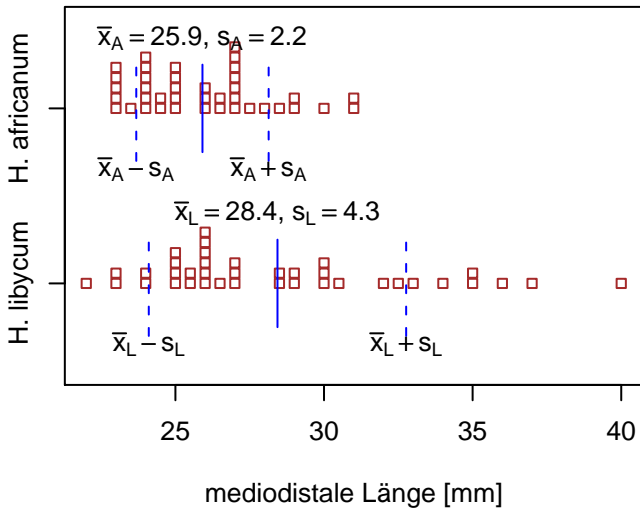
andere Nahrung  $\longrightarrow$  andere Zähne?

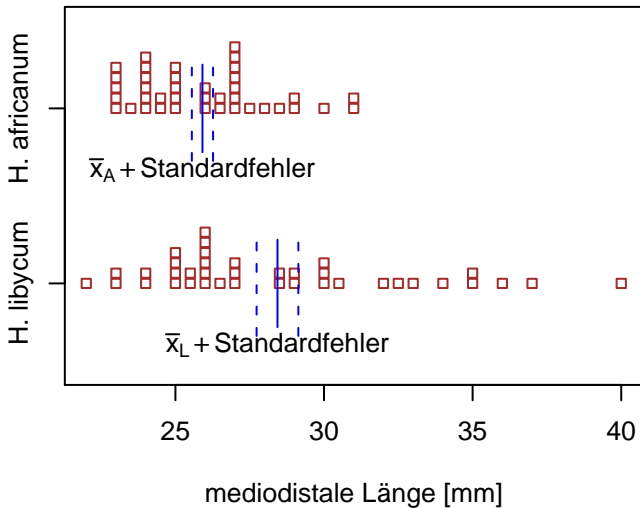
Messungen: mesiodistale Länge

distal = von der Mittellinie weg









Wir beobachten ( $n_A = 39$ ,  $n_L = 38$ ):

$$\bar{x}_A = 25,9, s_A = 2,2,$$



Wir beobachten ( $n_A = 39$ ,  $n_L = 38$ ):

$\bar{x}_A = 25,9$ ,  $s_A = 2,2$ , unser Schätzwert für die Streuung von  $\bar{x}_A$  ist also  $f_A = s_A / \sqrt{n_A} = 0,36$  (Standardfehler)

Wir beobachten ( $n_A = 39$ ,  $n_L = 38$ ):

$\bar{x}_A = 25,9$ ,  $s_A = 2,2$ , unser Schätzwert für die Streuung von  $\bar{x}_A$  ist also  $f_A = s_A = 2,2/\sqrt{n_A} = 0,36$  (Standardfehler),

$\bar{x}_L = 28,4$ ,  $s_L = 4,3$ ,

Wir beobachten ( $n_A = 39$ ,  $n_L = 38$ ):

$\bar{x}_A = 25,9$ ,  $s_A = 2,2$ , unser Schätzwert für die Streuung von  $\bar{x}_A$  ist also  $f_A = s_A / \sqrt{n_A} = 0,36$  (Standardfehler),

$\bar{x}_L = 28,4$ ,  $s_L = 4,3$ , unser Schätzwert für die Streuung von  $\bar{x}_L$  ist also  $f_L = s_L / \sqrt{n_L} = 0,70$ .

Wir beobachten ( $n_A = 39$ ,  $n_L = 38$ ):

$\bar{x}_A = 25,9$ ,  $s_A = 2,2$ , unser Schätzwert für die Streuung von  $\bar{x}_A$  ist also  $f_A = s_A = 2,2/\sqrt{n_A} = 0,36$  (Standardfehler),

$\bar{x}_L = 28,4$ ,  $s_L = 4,3$ , unser Schätzwert für die Streuung von  $\bar{x}_L$  ist also  $f_L = s_L = 2,2/\sqrt{n_L} = 0,70$ .

Ist die beobachtete Abweichung  $\bar{x}_L - \bar{x}_A = 2,5$  mit der **Nullhypothese** verträglich, dass  $\mu_L = \mu_A$ ?

# t-Statistik

Ist die beobachtete Abweichung  $\bar{x}_L - \bar{x}_A = 2,5$  mit der **Nullhypothese** verträglich, dass  $\mu_L = \mu_A$ ?

# t-Statistik

Ist die beobachtete Abweichung  $\bar{x}_L - \bar{x}_A = 2,5$  mit der **Nullhypothese** verträglich, dass  $\mu_L = \mu_A$ ?

Wir schätzen die Streuung von  $\bar{x}_L - \bar{x}_A$  durch  $f$ , wo

$$f^2 = f_L^2 + f_A^2$$

# t-Statistik

Ist die beobachtete Abweichung  $\bar{x}_L - \bar{x}_A = 2,5$  mit der **Nullhypothese** verträglich, dass  $\mu_L = \mu_A$ ?

Wir schätzen die Streuung von  $\bar{x}_L - \bar{x}_A$  durch  $f$ , wo

$$f^2 = f_L^2 + f_A^2$$

und bilden  $t = \frac{\bar{x}_L - \bar{x}_A}{f}$ .

# t-Statistik

Ist die beobachtete Abweichung  $\bar{x}_L - \bar{x}_A = 2,5$  mit der **Nullhypothese** verträglich, dass  $\mu_L = \mu_A$ ?

Wir schätzen die Streuung von  $\bar{x}_L - \bar{x}_A$  durch  $f$ , wo

$$f^2 = f_L^2 + f_A^2$$

und bilden  $t = \frac{\bar{x}_L - \bar{x}_A}{f}$ .

Wenn die Nullhypothese zutrifft, ist  $t$  (approximativ) Student-verteilt mit  $g$  Freiheitsgraden  
(wobei  $g$  aus den Daten geschätzt wird.)



## Theorem (Welch-t-Test, die Varianzen dürfen ungleich sein)

Seien  $X_1, \dots, X_n$  und  $Y_1, \dots, Y_m$  unabhängige normalverteilte Zufallsvariablen mit (möglicherweise verschiedenen) Varianzen  $\text{Var}X_i = \sigma_X^2$  und  $\text{Var}Y_j = \sigma_Y^2$ . Seien  $s_X$  und  $s_Y$  die aus den Stichproben berechneten Standardabweichungen. Unter der Nullhypothese gleicher Mittelwerten  $\mathbb{E}X_i = \mathbb{E}Y_j$  ist die Statistik

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$$

**ungefähr**  $t$ -verteilt mit  $\frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{s_X^4}{n^2 \cdot (n-1)} + \frac{s_Y^4}{m^2 \cdot (m-1)}}$  Freiheitsgraden.

## Theorem (Welch-t-Test, die Varianzen dürfen ungleich sein)

Seien  $X_1, \dots, X_n$  und  $Y_1, \dots, Y_m$  unabhängige normalverteilte Zufallsvariablen mit (möglicherweise verschiedenen) Varianzen  $\text{Var}X_i = \sigma_X^2$  und  $\text{Var}Y_i = \sigma_Y^2$ . Seien  $s_X$  und  $s_Y$  die aus den Stichproben berechneten Standardabweichungen. Unter der Nullhypothese gleicher Mittelwerte  $\mathbb{E}X_i = \mathbb{E}Y_j$  ist die Statistik

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$$

**ungefähr**  $t$ -verteilt mit  $\frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{s_X^4}{n^2 \cdot (n-1)} + \frac{s_Y^4}{m^2 \cdot (m-1)}}$  Freiheitsgraden.

(Diese Approximation für die Freiheitsgrade brauchen Sie sich nicht zu merken. R übernimmt das für Sie.)

## Theorem (zwei-Stichproben t-Test, ungepaart mit gleichen Varianzen)

Seien  $X_1, \dots, X_n$  und  $Y_1, \dots, Y_m$  unabhängige normalverteilte Zufallsvariablen mit der selben Varianz  $\sigma^2$ . Als **gepoolte Stichprobenvarianz** definieren wir

$$s_p^2 = \frac{(n-1) \cdot s_X^2 + (m-1) \cdot s_Y^2}{m+n-2}.$$

Unter der Nullhypothese gleicher Erwartungswerte  $\mu_X = \mu_Y$  folgt die Statistik

$$t = \frac{\bar{X} - \bar{Y}}{s_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

einer  $t$ -Verteilung mit  $n + m - 2$  mit Freiheitsgraden.

# Übersicht

- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests**
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression
  - Lineare Zusammenhänge
  - $t$ -Test fuer lineare Zusammenhänge
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse
- 7 Versuchsplanung: Stichprobenumfang

# Übersicht

- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests**
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression
  - Lineare Zusammenhänge
  - $t$ -Test fuer lineare Zusammenhänge
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse
- 7 Versuchsplanung: Stichprobenumfang

## Szenario:

- Ein Experiment habe  $r$  mögliche Ausgänge (z.B.  $r = 6$  beim Werfen eines Würfels).

## Szenario:

- Ein Experiment habe  $r$  mögliche Ausgänge (z.B.  $r = 6$  beim Werfen eines Würfels).
- Unter der Nullhypothese  $H_0$  habe Ausgang  $i$  Wahrscheinlichkeit  $p_i$ .

## Szenario:

- Ein Experiment habe  $r$  mögliche Ausgänge (z.B.  $r = 6$  beim Werfen eines Würfels).
- Unter der Nullhypothese  $H_0$  habe Ausgang  $i$  Wahrscheinlichkeit  $p_i$ .
- Unter  $n$  unabhängigen Wiederholungen des Experiments beobachten wir  $B_i$  mal Ausgang  $i$ .



## Szenario:

- Ein Experiment habe  $r$  mögliche Ausgänge (z.B.  $r = 6$  beim Werfen eines Würfels).
- Unter der Nullhypothese  $H_0$  habe Ausgang  $i$  Wahrscheinlichkeit  $p_i$ .
- Unter  $n$  unabhängigen Wiederholungen des Experiments beobachten wir  $B_i$  mal Ausgang  $i$ . Unter  $H_0$  erwarten wir  $E_i := \mathbb{E}[B_i] = np_i$  mal Ausgang  $i$  zu beobachten.
-

## Szenario:

- Ein Experiment habe  $r$  mögliche Ausgänge (z.B.  $r = 6$  beim Werfen eines Würfels).
- Unter der Nullhypothese  $H_0$  habe Ausgang  $i$  Wahrscheinlichkeit  $p_i$ .
- Unter  $n$  unabhängigen Wiederholungen des Experiments beobachten wir  $B_i$  mal Ausgang  $i$ . Unter  $H_0$  erwarten wir  $E_i := \mathbb{E}[B_i] = np_i$  mal Ausgang  $i$  zu beobachten.
- 

Frage: Geben die Beobachtungen Anlass, an der Nullhypothese zu zweifeln?

Erwarte  $E_i = np_i$  mal Ausgang  $i$ , beobachte  $B_i$  mal.

Geben diese Beobachtungen Anlass, an der Nullhypothese zu zweifeln?

Erwarte  $E_i = np_i$  mal Ausgang  $i$ , beobachte  $B_i$  mal.

Geben diese Beobachtungen Anlass, an der Nullhypothese zu zweifeln?

Vorgehen:

- Berechne  $X^2 = \sum_i \frac{(B_i - E_i)^2}{E_i}$

Erwarte  $E_i = np_i$  mal Ausgang  $i$ , beobachte  $B_i$  mal.

Geben diese Beobachtungen Anlass, an der Nullhypothese zu zweifeln?

Vorgehen:

- Berechne  $X^2 = \sum_i \frac{(B_i - E_i)^2}{E_i}$
- $X^2$  ist unter (approximativ, sofern  $n$  genügend groß)  $\chi^2_{r-1}$ -verteilt („Chi-Quadrat-verteilt mit  $r - 1$  Freiheitsgraden“)

Erwarte  $E_i = np_i$  mal Ausgang  $i$ , beobachte  $B_i$  mal.

Geben diese Beobachtungen Anlass, an der Nullhypothese zu zweifeln?

Vorgehen:

- Berechne  $X^2 = \sum_i \frac{(B_i - E_i)^2}{E_i}$
- $X^2$  ist unter (approximativ, sofern  $n$  genügend groß)  $\chi^2_{r-1}$ -verteilt („Chi-Quadrat-verteilt mit  $r - 1$  Freiheitsgraden“)
- Lehne  $H_0$  zum Signifikanzniveau  $\alpha$  ab, wenn  $X^2 \geq q_{1-\alpha}$ , wo  $q_{1-\alpha}$  das  $(1 - \alpha)$ -Quantil der  $\chi^2$ -Verteilung mit  $r - 1$  Freiheitsgraden ist.

**95%-Quantil der  $\chi^2$ -Verteilung in Abhängigkeit der Anzahl Freiheitsgrade**

F.g.	1	2	3	4	5	6	7	8	9
Quantil	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92

Beispiel: Unter 12.000 Würfeln eines Würfels beobachten wir folgende Häufigkeiten der Augenzahlen:

$i$	1	2	3	4	5	6
$B_i$	2014	2000	2017	1925	1998	2046

Beispiel: Unter 12.000 Würfeln eines Würfels beobachten wir folgende Häufigkeiten der Augenzahlen:

$i$	1	2	3	4	5	6
$B_i$	2014	2000	2017	1925	1998	2046

Ist der Würfel fair ( $H_0: p_1 = \dots = p_6 = 1/6$ )?



Beispiel: Unter 12.000 Würfeln eines Würfels beobachten wir folgende Häufigkeiten der Augenzahlen:

$i$	1	2	3	4	5	6
$B_i$	2014	2000	2017	1925	1998	2046

Ist der Würfel fair ( $H_0: p_1 = \dots = p_6 = 1/6$ )?

Es ist  $E_1 = \dots = E_6 = 12.000 \cdot 1/6 = 2000$ ,

$$\begin{aligned} \chi^2 &= \frac{(2014 - 2000)^2}{2000} + \frac{(2000 - 2000)^2}{2000} + \frac{(2017 - 2000)^2}{2000} \\ &+ \frac{(1925 - 2000)^2}{2000} + \frac{(1998 - 2000)^2}{2000} + \frac{(2046 - 2000)^2}{2000} = 4,115. \end{aligned}$$

Beispiel: Unter 12.000 Würfeln eines Würfels beobachten wir folgende Häufigkeiten der Augenzahlen:

$i$	1	2	3	4	5	6
$B_i$	2014	2000	2017	1925	1998	2046

Ist der Würfel fair ( $H_0: p_1 = \dots = p_6 = 1/6$ )?

Es ist  $E_1 = \dots = E_6 = 12.000 \cdot 1/6 = 2000$ ,

$$\begin{aligned} \chi^2 = & \frac{(2014 - 2000)^2}{2000} + \frac{(2000 - 2000)^2}{2000} + \frac{(2017 - 2000)^2}{2000} \\ & + \frac{(1925 - 2000)^2}{2000} + \frac{(1998 - 2000)^2}{2000} + \frac{(2046 - 2000)^2}{2000} = 4,115. \end{aligned}$$

Das 95%-Quantil der  $\chi^2$ -Verteilung mit 5 Freiheitsgraden ist  $9,49 > 4,115$ , wir lehnen  $H_0$  nicht ab (zum Signifikanzniveau 5%).

**95%-Quantil der  $\chi^2$ -Verteilung in Abh.keit d. Anz. Freiheitsgrade**

F.g.	1	2	3	4	5	6	7	8	9	10
Quantil	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31

Beispiel: Unter 12.000 Würfeln eines Würfels beobachten wir folgende Häufigkeiten der Augenzahlen:

$i$	1	2	3	4	5	6
$B_i$	2014	2000	2017	1925	1998	2046

Ist der Würfel fair ( $H_0: p_1 = \dots = p_6 = 1/6$ )?

Es ist  $E_1 = \dots = E_6 = 12.000 \cdot 1/6 = 2000$ ,

$$\begin{aligned} \chi^2 &= \frac{(2014 - 2000)^2}{2000} + \frac{(2000 - 2000)^2}{2000} + \frac{(2017 - 2000)^2}{2000} \\ &+ \frac{(1925 - 2000)^2}{2000} + \frac{(1998 - 2000)^2}{2000} + \frac{(2046 - 2000)^2}{2000} = 4,115. \end{aligned}$$

Das 95%-Quantil der  $\chi^2$ -Verteilung mit 5 Freiheitsgraden ist  $9,49 > 4,115$ , wir lehnen  $H_0$  nicht ab (zum Signifikanzniveau 5%).

### 95%-Quantil der $\chi^2$ -Verteilung in Abh.keit d. Anz. Freiheitsgrade

F.g.	1	2	3	4	5	6	7	8	9	10
Quantil	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31

Bemerkung:  $\chi^2_{5}([4,115, \infty)) = 0,533$ , d.h. wir finden einen

$p$ -Wert von 53%, der Test gibt keinen Anlass zu Zweifel an  $H_0$ .

# Übersicht

- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests**
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression
  - Lineare Zusammenhänge
  - $t$ -Test fuer lineare Zusammenhänge
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse
- 7 Versuchsplanung: Stichprobenumfang

Beispiel: 48 Teilnehmer eines Management-Kurses entscheiden über Beförderung:

	Weiblich	Männlich	Summe
Befördern	14	21	35
Ablegen	10	3	13
Summe	24	24	48

Beispiel: 48 Teilnehmer eines Management-Kurses entscheiden über Beförderung:

	Weiblich	Männlich	Summe
Befördern	14	21	35
Ablegen	10	3	13
Summe	24	24	48

Kann das Zufall sein? Testen wir  $H_0$  : „Geschlecht und Beförderungsentscheidung sind unabhängig“.

Beispiel: 48 Teilnehmer eines Management-Kurses entscheiden über Beförderung:

	Weiblich	Männlich	Summe
Befördern	14	21	35
Ablegen	10	3	13
Summe	24	24	48

Kann das Zufall sein? Testen wir  $H_0$  : „Geschlecht und Beförderungsentscheidung sind unabhängig“.

Anteil Weiblich= $24/48=0.5$ , Anteil befördert= $35/48=0.73$ , also erwartete Zahlen unter  $H_0$ :

Beispiel: 48 Teilnehmer eines Management-Kurses entscheiden über Beförderung:

	Weiblich	Männlich	Summe
Befördern	14	21	35
Ablegen	10	3	13
Summe	24	24	48

Kann das Zufall sein? Testen wir  $H_0$  : „Geschlecht und Beförderungsentscheidung sind unabhängig“.

Anteil Weiblich= $24/48=0.5$ , Anteil befördert= $35/48=0.73$ , also erwartete Zahlen unter  $H_0$ :

	Weiblich	Männlich	Summe
Befördern	17.5 ( $= 48 \cdot \frac{24}{48} \cdot \frac{35}{48}$ )	17.5 ( $= 48 \cdot \frac{24}{48} \cdot \frac{35}{48}$ )	35
Ablegen	6.5 ( $= 48 \cdot \frac{24}{48} \cdot \frac{13}{48}$ )	6.5 ( $= 48 \cdot \frac{24}{48} \cdot \frac{13}{48}$ )	13
Summe	24	24	48



$H_0$  : „Geschlecht und Beförderungsentscheidung sind unabhängig“

Beobachtete Anzahlen:

	Weiblich	Männlich	Summe
Befördern	14	21	35
Ablegen	10	3	13
Summe	24	24	48

Unter  $H_0$  erwartete Anzahlen:

	Weiblich	Männlich	Summe
Befördern	17.5	17.5	35
Ablegen	6.5	6.5	13
Summe	24	24	48

$H_0$  : „Geschlecht und Beförderungsentscheidung sind unabhängig“

Beobachtete Anzahlen:

	Weiblich	Männlich	Summe
Befördern	14	21	35
Ablegen	10	3	13
Summe	24	24	48

Unter  $H_0$  erwartete Anzahlen:

	Weiblich	Männlich	Summe
Befördern	17.5	17.5	35
Ablegen	6.5	6.5	13
Summe	24	24	48

Die  $X^2$ -Statistik ist

$$\chi^2 = \frac{(17.5 - 14)^2}{17.5} + \frac{(21 - 17.5)^2}{17.5} + \frac{(10 - 6.5)^2}{6.5} + \frac{(3 - 6.5)^2}{6.5} = 5.17.$$

$H_0$  : „Geschlecht und Beförderungsentscheidung sind unabhängig“

Beobachtete Anzahlen:

	Weiblich	Männlich	Summe
Befördern	14	21	35
Ablegen	10	3	13
Summe	24	24	48

Unter  $H_0$  erwartete Anzahlen:

	Weiblich	Männlich	Summe
Befördern	17.5	17.5	35
Ablegen	6.5	6.5	13
Summe	24	24	48

Die  $\chi^2$ -Statistik ist

$$\chi^2 = \frac{(17.5 - 14)^2}{17.5} + \frac{(21 - 17.5)^2}{17.5} + \frac{(10 - 6.5)^2}{6.5} + \frac{(3 - 6.5)^2}{6.5} = 5.17.$$

Unter  $H_0$  ist  $\chi^2$  (approximativ)  $\chi^2$ -verteilt mit einem Freiheitsgrad ( $1 = 4 - 1 - 1 - 1 = (2 - 1) \cdot (2 - 1)$ ): 4 Zellen, ein Freiheitsgrad geht für die feste Gesamtsumme, einer für das (prinzipiell) unbekannte Geschlechterverhältnis und einer für die (prinzipiell) unbekannte Beförderungswahrscheinlichkeit „verloren“.

**95%-Quantil der  $\chi^2$ -Verteilung in Abh.keit d. Anz. Freiheitsgrade**

F.g.	1	2	3	4	5	6	7	8	9	10
Quantil	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31

Wir können  $H_0$  zum Signifikanzniveau 5% ablehnen.

$H_0$  : „Geschlecht und Beförderungsentscheidung sind unabhängig“

Beobachtete Anzahlen:

	Weiblich	Männlich	Summe
Befördern	14	21	35
Ablegen	10	3	13
Summe	24	24	48

Unter  $H_0$  erwartete Anzahlen:

	Weiblich	Männlich	Summe
Befördern	17.5	17.5	35
Ablegen	6.5	6.5	13
Summe	24	24	48

Die  $\chi^2$ -Statistik ist

$$\chi^2 = \frac{(17.5 - 14)^2}{17.5} + \frac{(21 - 17.5)^2}{17.5} + \frac{(10 - 6.5)^2}{6.5} + \frac{(3 - 6.5)^2}{6.5} = 5.17.$$

Unter  $H_0$  ist  $\chi^2$  (approximativ)  $\chi^2$ -verteilt mit einem Freiheitsgrad ( $1 = 4 - 1 - 1 - 1 = (2 - 1) \cdot (2 - 1)$ ): 4 Zellen, ein Freiheitsgrad geht für die feste Gesamtsumme, einer für das (prinzipiell) unbekannte Geschlechterverhältnis und einer für die (prinzipiell) unbekannte Beförderungswahrscheinlichkeit „verloren“.

**95%-Quantil der  $\chi^2$ -Verteilung in Abh.keit d. Anz. Freiheitsgrade**

F.g.	1	2	3	4	5	6	7	8	9	10
Quantil	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31

Wir können  $H_0$  zum Signifikanzniveau 5% ablehnen.

(Es ist  $\chi^2_1([5.17, \infty)) = 0.023$ , d.h. wir finden einen  $p$ -Wert von ca. 2%.)

## Chi-Quadrat-Test auf Unabhängigkeit, allgemeine Situation:

- 2 Merkmale mit  $r$  bzw.  $s$  Ausprägungen ( $r \times s$ -Kontingenztafel),  $n$  Beobachtungen
- Bestimme erwartete Anzahlen unter  $H_0$  als Produkt der (normierten) Zeilen- und Spaltensummen
- $X^2$  ist unter  $H_0$  (approximativ)  $\chi^2$ -verteilt mit  $rs - 1 - (r - 1) - (s - 1) = (r - 1)(s - 1)$  Freiheitsgraden.

Chi-Quadrat-Test auf Unabhängigkeit, allgemeine Situation:

- 2 Merkmale mit  $r$  bzw.  $s$  Ausprägungen ( $r \times s$ -Kontingenztafel),  $n$  Beobachtungen
- Bestimme erwartete Anzahlen unter  $H_0$  als Produkt der (normierten) Zeilen- und Spaltensummen
- $X^2$  ist unter  $H_0$  (approximativ)  $\chi^2$ -verteilt mit  $rs - 1 - (r - 1) - (s - 1) = (r - 1)(s - 1)$  Freiheitsgraden.

Bemerkung: Im  $2 \times 2$ -Fall kann man auch Fishers exakten Test verwenden (zuma, wenn  $n$  recht klein).

# Übersicht

- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests**
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression
  - Lineare Zusammenhänge
  - $t$ -Test fuer lineare Zusammenhänge
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse
- 7 Versuchsplanung: Stichprobenumfang

$k$ : Anzahl Gruppen

$m$ : Anzahl Modellparameter, die geschätzt werden

Anzahl Freiheitsgrade:

$$df = k - m - 1$$



$k$ : Anzahl Gruppen

$m$ : Anzahl Modellparameter, die geschätzt werden

Anzahl Freiheitsgrade:

$$df = k - m - 1$$

Beispiel: Test auf Hardy-Weinberg-Gleichgewicht mit drei Genotypen AA, Aa, aa:

$k = 3$  Gruppen,

$k$ : Anzahl Gruppen

$m$ : Anzahl Modellparameter, die geschätzt werden

Anzahl Freiheitsgrade:

$$df = k - m - 1$$

Beispiel: Test auf Hardy-Weinberg-Gleichgewicht mit drei Genotypen AA, Aa, aa:

$k = 3$  Gruppen,  $m = 1$  Parameter wird geschätzt, nämlich der Anteil der Allele vom Typ A.

$k$ : Anzahl Gruppen

$m$ : Anzahl Modellparameter, die geschätzt werden

Anzahl Freiheitsgrade:

$$df = k - m - 1$$

Beispiel: Test auf Hardy-Weinberg-Gleichgewicht mit drei Genotypen AA, Aa, aa:

$k = 3$  Gruppen,  $m = 1$  Parameter wird geschätzt, nämlich der Anteil der Allele vom Typ A.

$$\Rightarrow sfd = 3 - 1 - 1 = 1$$

# Übersicht

- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle**
- 5 Lineare Regression
  - Lineare Zusammenhänge
  - $t$ -Test fuer lineare Zusammenhänge
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse
- 7 Versuchsplanung: Stichprobenumfang

Wir beobachten in einer Stichprobe der Größe  $n$  aus einer Population  $X$  Exemplare mit einer gewissen Eigenschaft (z.B. „ist männlich“) und möchten den (unbekannten) Anteil  $\theta$  dieser Eigenschaft in der Gesamtpopulation schätzen.

Wir beobachten in einer Stichprobe der Größe  $n$  aus einer Population  $X$  Exemplare mit einer gewissen Eigenschaft (z.B. „ist männlich“) und möchten den (unbekannten) Anteil  $\theta$  dieser Eigenschaft in der Gesamtpopulation schätzen.

Der offensichtliche Schätzer ist  $\hat{\theta} := \frac{X}{n}$   
(und dies ist auch der ML-Schätzer).

Wir beobachten in einer Stichprobe der Größe  $n$  aus einer Population  $X$  Exemplare mit einer gewissen Eigenschaft (z.B. „ist männlich“) und möchten den (unbekannten) Anteil  $\theta$  dieser Eigenschaft in der Gesamtpopulation schätzen.

Der offensichtliche Schätzer ist  $\hat{\theta} := \frac{X}{n}$   
(und dies ist auch der ML-Schätzer).

Frage: Wie verlässlich ist die Schätzung?

Wir beobachten in einer Stichprobe der Größe  $n$  aus einer Population  $X$  Exemplare mit einer gewissen Eigenschaft (z.B. „ist männlich“) und möchten den (unbekannten) Anteil  $\theta$  dieser Eigenschaft in der Gesamtpopulation schätzen.

Der offensichtliche Schätzer ist  $\hat{\theta} := \frac{X}{n}$   
(und dies ist auch der ML-Schätzer).

Frage: Wie verlässlich ist die Schätzung?

Gewünscht: Ein in Abhängigkeit von den Beobachtungen konstruiertes (und möglichst kurzes) Intervall  $[\hat{\theta}_u, \hat{\theta}_o]$  mit der Eigenschaft

$$\mathbb{P}_\theta \left( [\hat{\theta}_u, \hat{\theta}_o] \text{ überdeckt } \theta \right) \geq 1 - \alpha$$

für *jede Wahl* von  $\theta$ .



Wir beobachten in einer Stichprobe der Größe  $n$  aus einer Population  $X$  Exemplare mit einer gewissen Eigenschaft (z.B. „ist männlich“) und möchten den (unbekannten) Anteil  $\theta$  dieser Eigenschaft in der Gesamtpopulation schätzen.

Der offensichtliche Schätzer ist  $\hat{\theta} := \frac{X}{n}$   
(und dies ist auch der ML-Schätzer).

Frage: Wie verlässlich ist die Schätzung?

Gewünscht: Ein in Abhängigkeit von den Beobachtungen konstruiertes (und möglichst kurzes) Intervall  $[\hat{\theta}_u, \hat{\theta}_o]$  mit der Eigenschaft

$$\mathbb{P}_\theta \left( [\hat{\theta}_u, \hat{\theta}_o] \text{ überdeckt } \theta \right) \geq 1 - \alpha$$

für *jede Wahl* von  $\theta$ .

Ein solches Intervall heißt ein *Konfidenzintervall* (zum Irrtumsniveau  $\alpha$ ), engl. *confidence interval*.

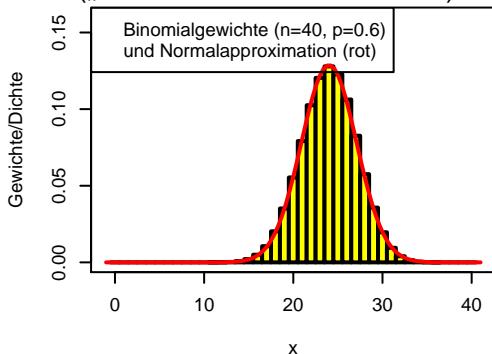
Für gegebenes  $\theta$  ist  $X$  Binomial( $n, \theta$ )-verteilt,  
 $\mathbb{E}[X] = n\theta$ ,  $\text{Var}[X] = n\theta(1 - \theta)$ .

Für gegebenes  $\theta$  ist  $X$  Binomial( $n, \theta$ )-verteilt,

$$\mathbb{E}[X] = n\theta, \text{Var}[X] = n\theta(1 - \theta).$$

Für (genügend) großes  $n$  ist  $X$  ungefähr normalverteilt  
mit Mittelwert  $n\theta$  und Varianz  $n\theta(1 - \theta)$

(„zentraler Grenzwertsatz“):

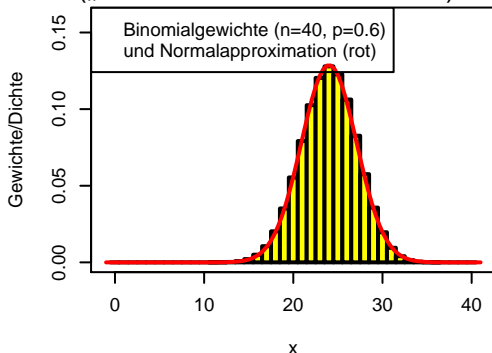


Für gegebenes  $\theta$  ist  $X$  Binomial( $n, \theta$ )-verteilt,

$$\mathbb{E}[X] = n\theta, \text{Var}[X] = n\theta(1 - \theta).$$

Für (genügend) großes  $n$  ist  $X$  ungefähr normalverteilt mit Mittelwert  $n\theta$  und Varianz  $n\theta(1 - \theta)$

(„zentraler Grenzwertsatz“):



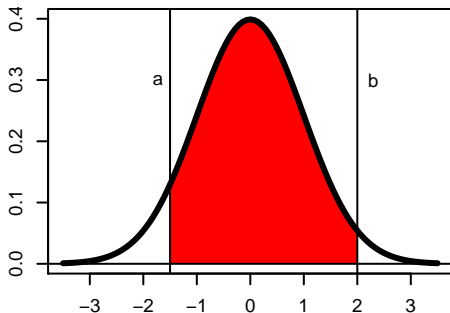
Also ist  $\hat{\theta} = \frac{X}{n}$  (ungefähr) normalverteilt mit Mittelwert  $\theta$  und Varianz  $\frac{1}{n}\theta(1 - \theta)$

$\hat{\theta} = \frac{X}{n}$  ist (ungefähr) normalverteilt  
mit Mittelwert  $\theta$  und Varianz  $\frac{1}{n}\theta(1 - \theta)$ :

$$\mathbb{P}_{\theta} \left( a \leq \frac{\hat{\theta} - \theta}{\sqrt{\frac{1}{n}\theta(1 - \theta)}} \leq b \right) \approx \mathbb{P}(a \leq Z \leq b)$$

(mit standard-normalverteiletem  $Z$ )

Dichte der Standard-Normalverteilung



Man schätzt die (unbekannte) Streuung von  $\hat{\theta}$   
durch  $\sqrt{\frac{1}{n}\hat{\theta}(1 - \hat{\theta})}$ :

Man schätzt die (unbekannte) Streuung von  $\hat{\theta}$   
 durch  $\sqrt{\frac{1}{n}\hat{\theta}(1-\hat{\theta})}$ :

Wähle  $z_\alpha$  so dass  $\mathbb{P}(-z_\alpha \leq Z \leq z_\alpha)$ , dann ist

$$\left[ \hat{\theta} - z_\alpha \frac{\sqrt{\hat{\theta}(1-\hat{\theta})}}{\sqrt{n}}, \hat{\theta} + z_\alpha \frac{\sqrt{\hat{\theta}(1-\hat{\theta})}}{\sqrt{n}} \right]$$

ein (approximatives) Konfidenzintervall für  $\theta$  zum  
 Irrtumsniveau  $\alpha$ .

( $-z_\alpha$  is das  $\alpha/2$ -Quantil der Normalverteilung, mit R:  
 $-qnorm(\alpha/2)$ ,  $z_{0,05} \doteq 1,96$ ,  $z_{0,01} \doteq 2,58$ .)

In einem Fang von 53 Porzellankrabben waren 23 Weibchen und 30 Männchen, d.h. der Männchenanteil in der Stichprobe war  $30/53 = 0,57$ .



In einem Fang von 53 Porzellankrabben waren 23 Weibchen und 30 Männchen, d.h. der Männchenanteil in der Stichprobe war  $30/53 = 0,57$ .

(Approximatives) 95%-Konfidenzintervall für  $\theta$ , den Männchenanteil in der Gesamtpopulation:

In einem Fang von 53 Porzellankrabben waren 23 Weibchen und 30 Männchen, d.h. der Männchenanteil in der Stichprobe war  $30/53 = 0,57$ .

(Approximatives) 95%-Konfidenzintervall für  $\theta$ , den Männchenanteil in der Gesamtpopulation:

$$\begin{aligned} I &= \left[ \frac{30}{53} - 1.96 \sqrt{\frac{(30/53)(23/53)}{53}}, \frac{30}{53} + 1.96 \sqrt{\frac{(30/53)(23/53)}{53}} \right] \\ &= [0.43, 0.70] \end{aligned}$$

## Anmerkungen

$\left[ \hat{\theta} \pm z_{\alpha} \frac{\sqrt{\hat{\theta}(1-\hat{\theta})}}{\sqrt{n}} \right]$  ist ein (approximatives) Konfidenzintervall für  $\theta$   
zum Irrtumsniveau  $\alpha$ .

## Anmerkungen

$\left[ \hat{\theta} \pm z_{\alpha} \frac{\sqrt{\hat{\theta}(1-\hat{\theta})}}{\sqrt{n}} \right]$  ist ein (approximatives) Konfidenzintervall für  $\theta$   
zum Irrtumsniveau  $\alpha$ .

- Für die Gültigkeit der Approximation muss  $n$  genügend groß und  $\theta$  nicht zu nahe an 0 oder 1 sein. (Eine häufig zitierte „Faustregel“ ist “ $n\theta(1 - \theta) \geq 9$ ”.)

## Anmerkungen

$\left[ \hat{\theta} \pm z_{\alpha} \frac{\sqrt{\hat{\theta}(1-\hat{\theta})}}{\sqrt{n}} \right]$  ist ein (approximatives) Konfidenzintervall für  $\theta$  zum Irrtumsniveau  $\alpha$ .

- Für die Gültigkeit der Approximation muss  $n$  genügend groß und  $\theta$  nicht zu nahe an 0 oder 1 sein. (Eine häufig zitierte „Faustregel“ ist “ $n\theta(1 - \theta) \geq 9$ ”.)
- Die Philosophie der Konfidenzintervalle entstammt der *frequentistischen* Interpretation der Statistik: Für jede Wahl des Parameters  $\theta$  würden wir bei häufiger Wiederholung des Experiments finden, dass in (ca.)  $(1 - \alpha) \cdot 100\%$  der Fälle das (zufällige) Konfidenzintervall den „wahren“ (festen) Parameter  $\theta$  überdeckt.

## Anmerkungen

$\left[ \hat{\theta} \pm z_{\alpha} \frac{\sqrt{\hat{\theta}(1-\hat{\theta})}}{\sqrt{n}} \right]$  ist ein (approximatives) Konfidenzintervall für  $\theta$  zum Irrtumsniveau  $\alpha$ .

- Für die Gültigkeit der Approximation muss  $n$  genügend groß und  $\theta$  nicht zu nahe an 0 oder 1 sein. (Eine häufig zitierte „Faustregel“ ist „ $n\theta(1 - \theta) \geq 9$ “.)
- Die Philosophie der Konfidenzintervalle entstammt der *frequentistischen* Interpretation der Statistik: Für jede Wahl des Parameters  $\theta$  würden wir bei häufiger Wiederholung des Experiments finden, dass in (ca.)  $(1 - \alpha) \cdot 100\%$  der Fälle das (zufällige) Konfidenzintervall den „wahren“ (festen) Parameter  $\theta$  überdeckt.
- Formulierungen, die sich auf eine Wahrscheinlichkeitsverteilung des Parameters  $\theta$  beziehen (beispielsweise: „Wie ~~wahrscheinlich~~ ist es, dass  $\theta \leq 0,3$ ?“), sind in der frequentistischen Interpretation sinnlos. (Dies ist anders in der *Bayesschen Interpretation*.)

Wir haben schon früher gesehen, dass das Konfidenzintervall für  $\theta$  eine zu geringe Überdeckungswahrscheinlichkeit hat, falls  $\theta$  nahe bei 0 oder nahe bei 1 ist. Ein sehr einfacher und in der Praxis gut funktionierender Trick ist, bei der Schätzung der Wahrscheinlichkeit  $\theta$  die Zahl der “Erfolge”  $x$  um 1 und die Gesamtzahl  $n$  um 2 zu erhöhen. Mit

$$\tilde{\theta} := (x + 1)/(n + 2)$$

erhält man dann ein relativ zuverlässiges 95%-Konfidenzintervall

$$\left[ \tilde{\theta} - 1.96 \cdot \sqrt{\tilde{\theta} \cdot (1 - \tilde{\theta})/n}, \tilde{\theta} + 1.96 \cdot \sqrt{\tilde{\theta} \cdot (1 + \tilde{\theta})/n} \right].$$

## Konfidenzintervalle für den Mittelwert im normalen Modell

$x_1, \dots, x_n$  u.a. Beobachtungen aus  $\mathcal{N}(\mu, \sigma^2)$ ,  $\mu, \sigma$  unbekannt.



## Konfidenzintervalle für den Mittelwert im normalen Modell

$x_1, \dots, x_n$  u.a. Beobachtungen aus  $\mathcal{N}(\mu, \sigma^2)$ ,  $\mu, \sigma$  unbekannt.

$$\bar{x} = \frac{1}{n}(x_1 + \dots + x_n), \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

## Konfidenzintervalle für den Mittelwert im normalen Modell

$x_1, \dots, x_n$  u.a. Beobachtungen aus  $\mathcal{N}(\mu, \sigma^2)$ ,  $\mu, \sigma$  unbekannt.

$$\bar{x} = \frac{1}{n}(x_1 + \dots + x_n), \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Wir wissen:

$\frac{\bar{x} - \mu}{s/\sqrt{n}}$  ist Student- $t$ -verteilt mit  $n - 1$  Freiheitsgraden, d.h.

$$\mathbb{P} \left( \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right| \geq q_{1-\alpha/2} \right) = \alpha$$

( $q_{1-\alpha/2}$  das  $1 - \alpha/2$ -Quantil der Student-Verteilung mit  $n - 1$  Freiheitsgraden).

## Konfidenzintervalle für den Mittelwert im normalen Modell

$x_1, \dots, x_n$  u.a. Beobachtungen aus  $\mathcal{N}(\mu, \sigma^2)$ ,  $\mu, \sigma$  unbekannt.

$$\bar{x} = \frac{1}{n}(x_1 + \dots + x_n), \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Wir wissen:

$\frac{\bar{x} - \mu}{s/\sqrt{n}}$  ist Student- $t$ -verteilt mit  $n - 1$  Freiheitsgraden, d.h.

$$\mathbb{P} \left( \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right| \geq q_{1-\alpha/2} \right) = \alpha$$

( $q_{1-\alpha/2}$  das  $1 - \alpha/2$ -Quantil der Student-Verteilung mit  $n - 1$  Freiheitsgraden).

$$\left[ \bar{x} - q_{1-\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + q_{1-\alpha/2} \frac{s}{\sqrt{n}} \right]$$

ist ein Konfidenzintervall für  $\mu$  zum Sicherheitsniveau  $1 - \alpha$ .

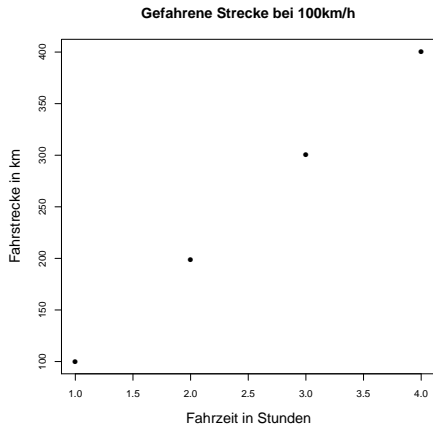
# Übersicht

- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression**
  - Lineare Zusammenhänge
  - $t$ -Test fuer lineare Zusammenhänge
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse
- 7 Versuchsplanung: Stichprobenumfang

# Übersicht

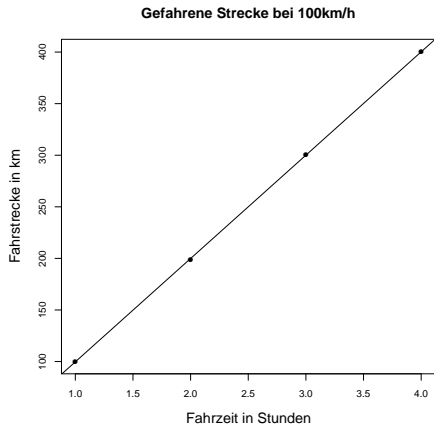
- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression**
  - **Lineare Zusammenhänge**
  - $t$ -Test fuer lineare Zusammenhänge
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse
- 7 Versuchsplanung: Stichprobenumfang

# Erst mal ohne Zufallsschwankungen



Gemessene Fahrstrecke bei exakt 100 km/h

# Erst mal ohne Zufallsschwankungen



Gemessene Fahrstrecke bei exakt 100 km/h

Zusammenhang:

Strecke  $s$  in km, Zeit  $t$  in Stunden

$$s = 100 \frac{\text{km}}{\text{h}} \cdot t$$

Problem und Lösung:

- Problem: Strecke ist schwer zu messen.



## Problem und Lösung:

- Problem: Strecke ist schwer zu messen.
- Beobachtung: Zeit ist leicht zu messen (Blick auf Uhr)

## Problem und Lösung:

- Problem: Strecke ist schwer zu messen.
- Beobachtung: Zeit ist leicht zu messen (Blick auf Uhr)
- Lösung: Linearer Zusammenhang zwischen Strecke und Zeit ermöglicht leichte Berechnung der Strecke ( $\rightsquigarrow$  Problem gelöst)



photo (c) by Jörg Hempel

Englisch: Grif-  
fon Vulture  
*Gyps fulvus*  
Gänsegeier

## Beispiel: Herzfrequenz und Stoffwechselrate beim Gänsegeier

- Frage: Was ist die Stoffwechselrate bei Gänsegeiern im Alltag (zB im Flug)?

## Beispiel: Herzfrequenz und Stoffwechselrate beim Gänsegeier

- Frage: Was ist die Stoffwechselrate bei Gänsegeiern im Alltag (zB im Flug)?
- Problem: Stoffwechselrate ist aufwändig zu messen (eigentlich nur im Labor)

## Beispiel: Herzfrequenz und Stoffwechselrate beim Gänsegeier

- Frage: Was ist die Stoffwechselrate bei Gänsegeiern im Alltag (zB im Flug)?
- Problem: Stoffwechselrate ist aufwändig zu messen (eigentlich nur im Labor)
- Beobachtung: Herzfrequenz ist leicht zu messen.

## Beispiel: Herzfrequenz und Stoffwechselrate beim Gänsegeier

- Frage: Was ist die Stoffwechselrate bei Gänsegeiern im Alltag (zB im Flug)?
- Problem: Stoffwechselrate ist aufwändig zu messen (eigentlich nur im Labor)
- Beobachtung: Herzfrequenz ist leicht zu messen.
- Lösung: Nutze linearen Zusammenhang zwischen Stoffwechselrate und Herzfrequenz.

## Beispiel: Herzfrequenz und Stoffwechselrate beim Gänsegeier

- Frage: Was ist die Stoffwechselrate bei Gänsegeiern im Alltag (zB im Flug)?
- Problem: Stoffwechselrate ist aufwändig zu messen (eigentlich nur im Labor)
- Beobachtung: Herzfrequenz ist leicht zu messen.
- Lösung: Nutze linearen Zusammenhang zwischen Stoffwechselrate und Herzfrequenz.
- Komplikation: Der lineare Zusammenhang ist nicht deterministisch, sondern zufallsbehaftet auf Grund von Messfehlern und da Stoffwechselrate von der „Tagesform“ abhängt.

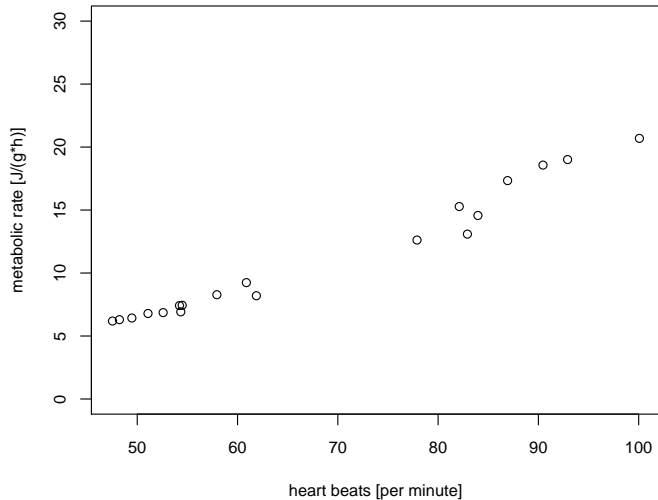


## Beispiel: Herzfrequenz und Stoffwechselrate beim Gänsegeier

- Frage: Was ist die Stoffwechselrate bei Gänsegeiern im Alltag (zB im Flug)?
- Problem: Stoffwechselrate ist aufwändig zu messen (eigentlich nur im Labor)
- Beobachtung: Herzfrequenz ist leicht zu messen.
- Lösung: Nutze linearen Zusammenhang zwischen Stoffwechselrate und Herzfrequenz.
- Komplikation: Der lineare Zusammenhang ist nicht deterministisch, sondern zufallsbehaftet auf Grund von Messfehlern und da Stoffwechselrate von der „Tagesform“ abhängt.
- Lineare Regression löst diese Komplikation in Wohlgefallen auf.

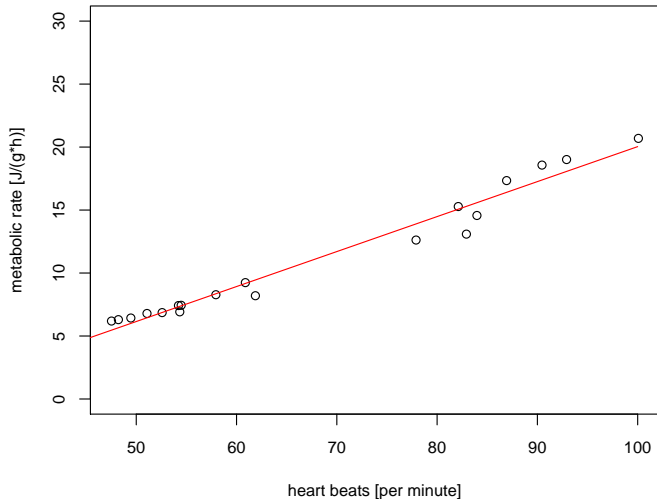
# Beispiel: Herzfrequenz und Stoffwechselrate beim Gänsegeier

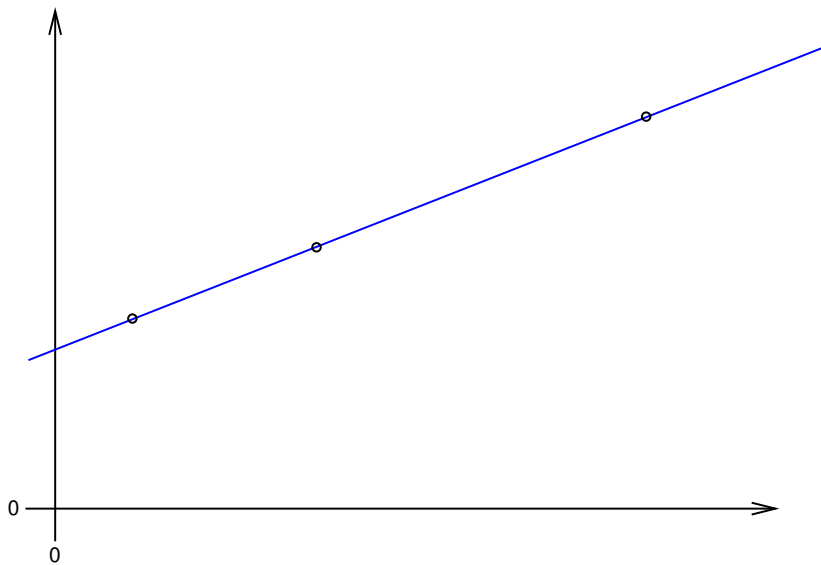
griffon vulture, 17.05.99, 16 degrees C

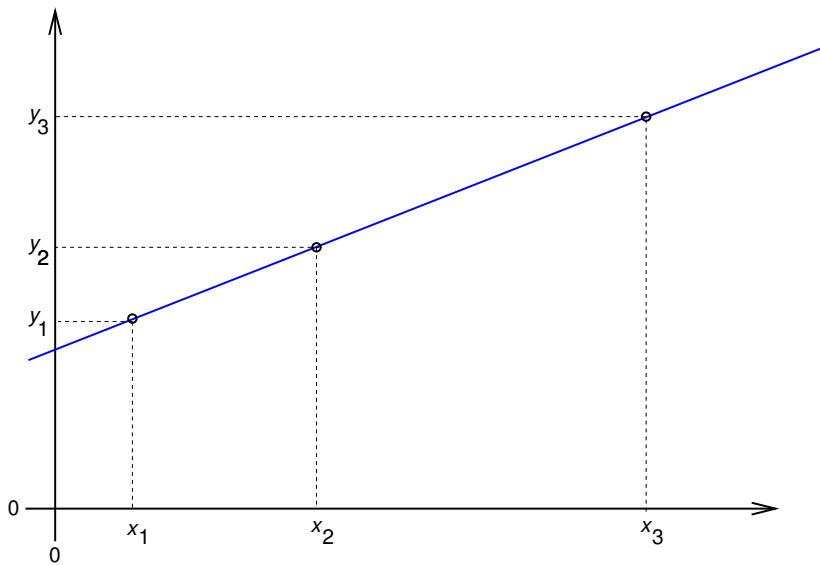


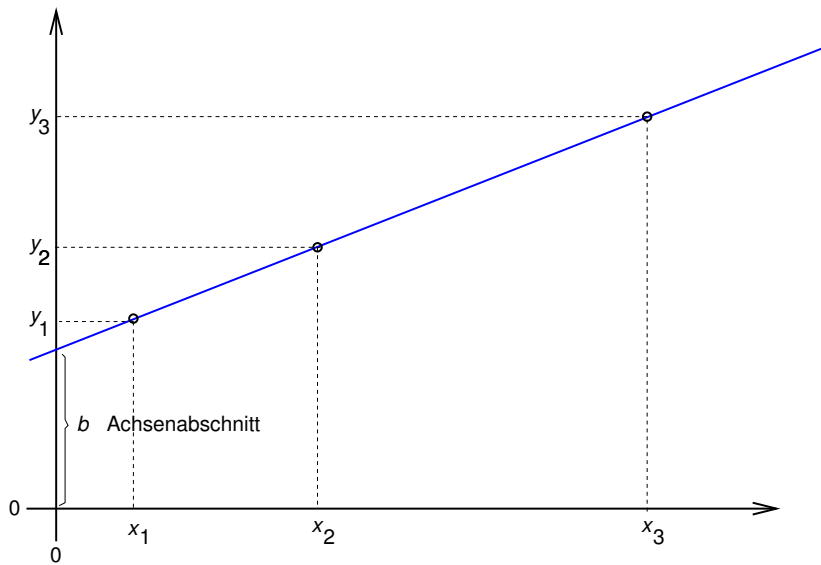
# Beispiel: Herzfrequenz und Stoffwechselrate beim Gänsegeier

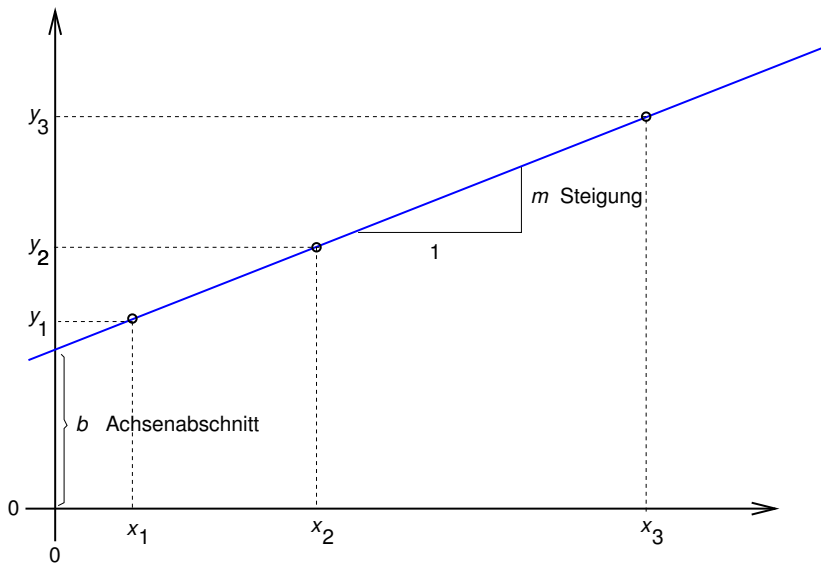
griffon vulture, 17.05.99, 16 degrees C

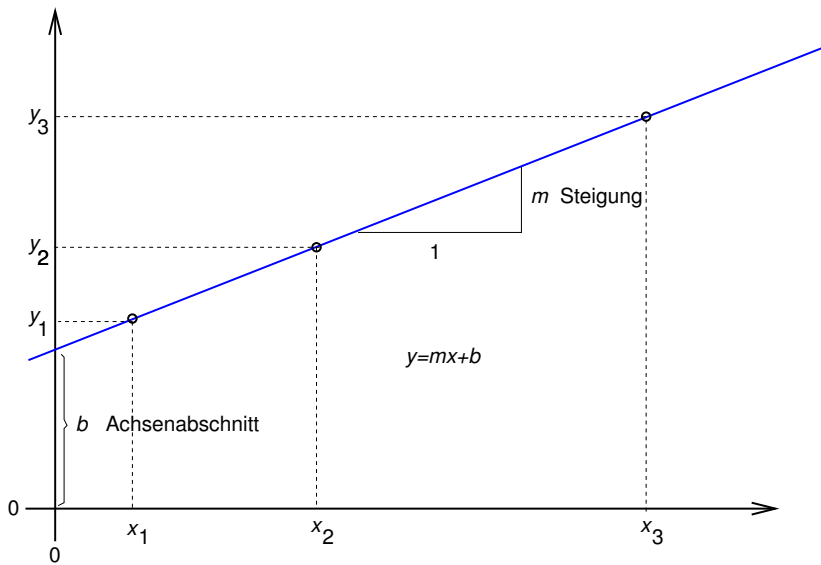




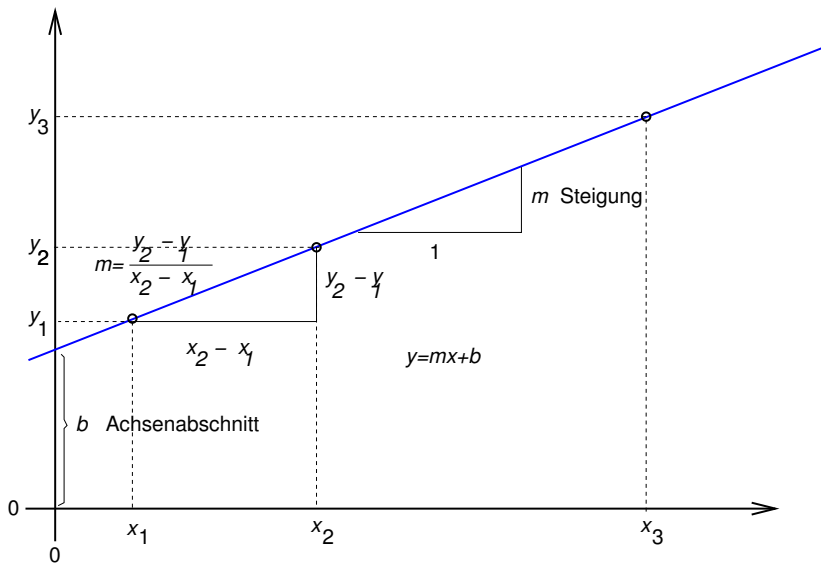


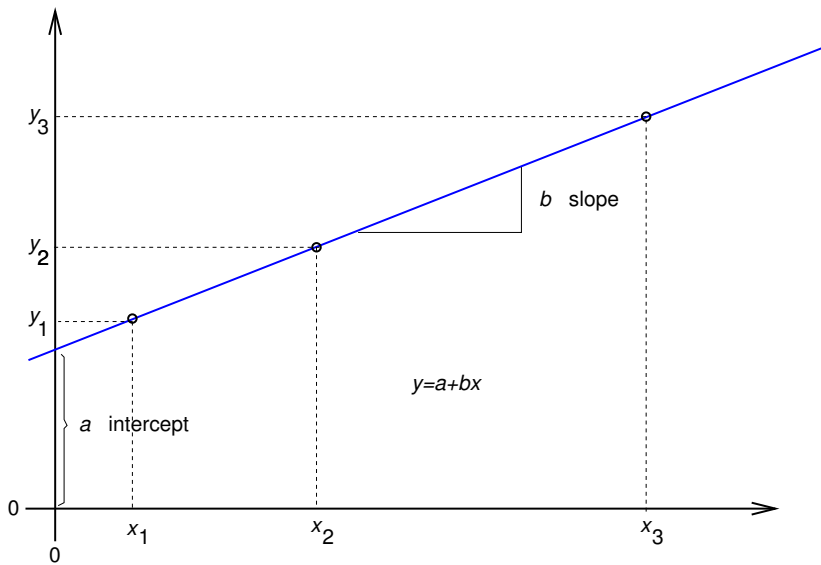


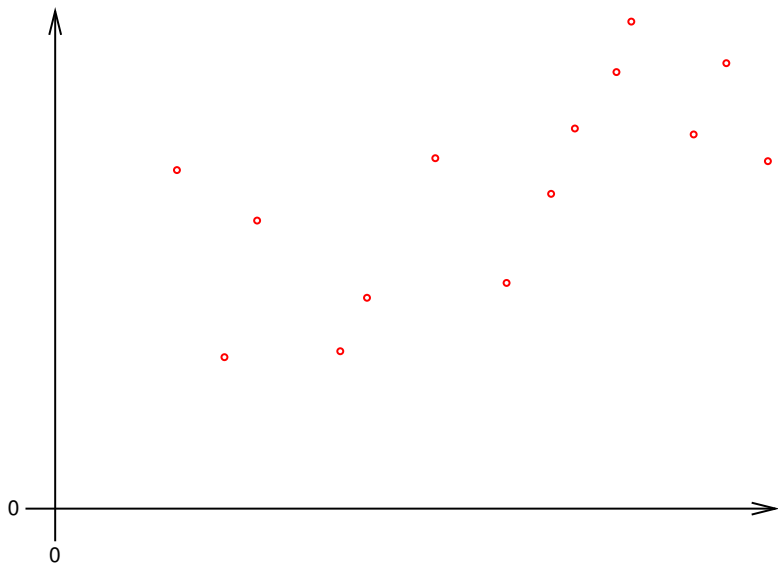


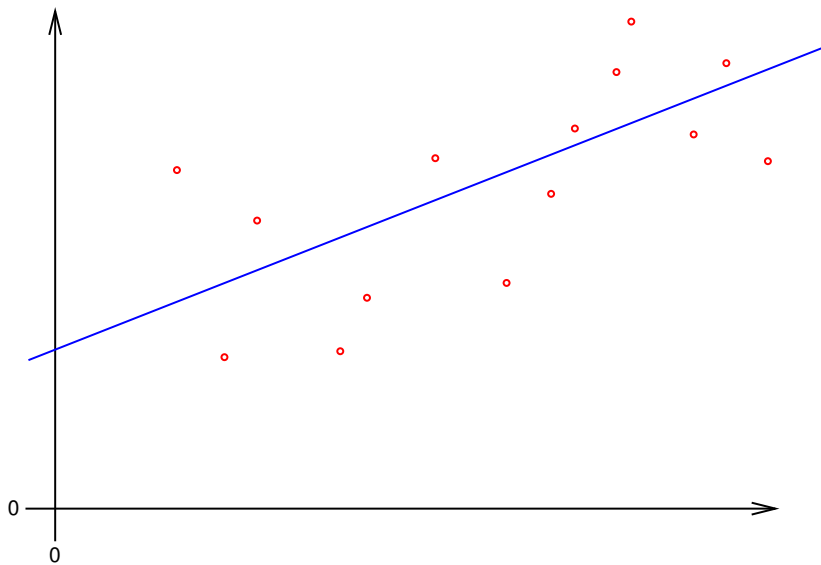


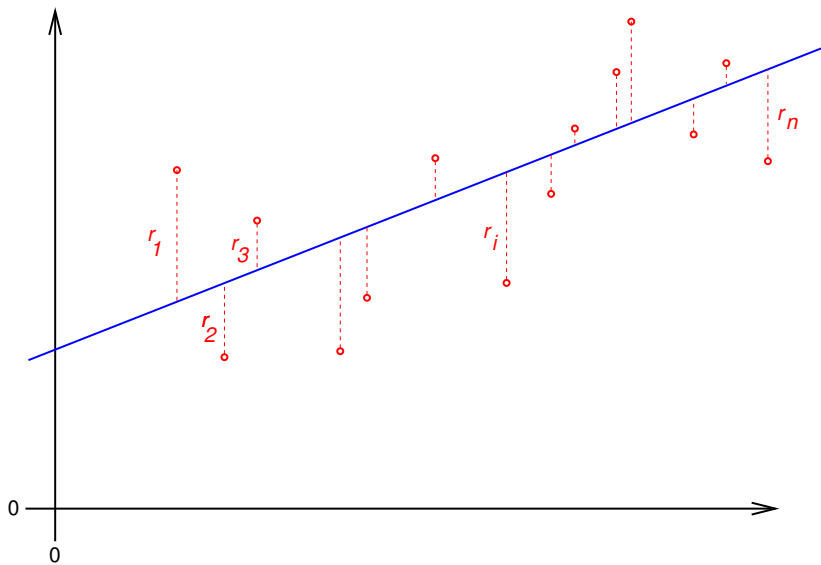


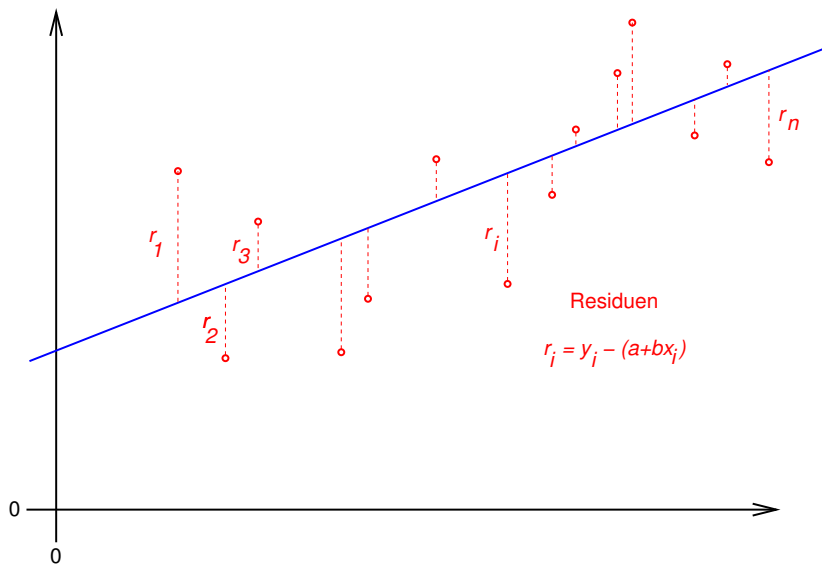


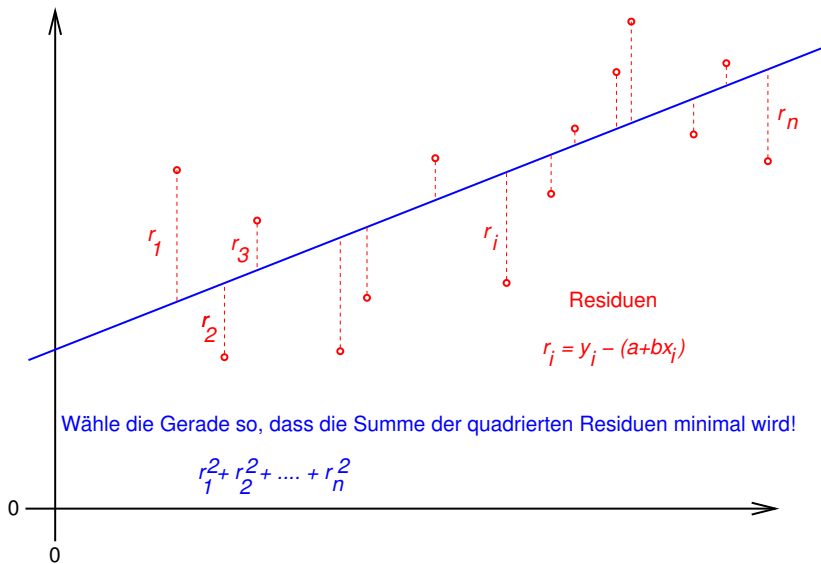












Definiere die Regressionsgerade

$$y = \hat{a} + \hat{b} \cdot x$$

durch die Minimierung der Summe der quadrierten Residuen:

$$(\hat{a}, \hat{b}) = \arg \min_{(a,b)} \sum_i (y_i - (a + b \cdot x_i))^2$$

Dahinter steckt die Modellvorstellung, dass Werte  $a, b$  existieren, so dass für alle Datenpaare  $(x_i, y_i)$  gilt

$$y_i = a + b \cdot x_i + \varepsilon_i,$$

wobei alle  $\varepsilon_i$  unabhängig und normalverteilt sind und alle dieselbe Varianz  $\sigma^2$  haben.



gegebene Daten:

**Y**

**X**

$y_1$

$x_1$

$y_2$

$x_2$

$y_3$

$x_3$

$\vdots$

$\vdots$

$y_n$

$x_n$

gegebene Daten:

<b>Y</b>	<b>X</b>
$y_1$	$x_1$
$y_2$	$x_2$
$y_3$	$x_3$
$\vdots$	$\vdots$
$y_n$	$x_n$

Modell: es gibt Zahlen  
 $a$ ,  $b$ ,  $\sigma^2$ , so dass

$$y_1 = a + b \cdot x_1 + \varepsilon_1$$

$$y_2 = a + b \cdot x_2 + \varepsilon_2$$

$$y_3 = a + b \cdot x_3 + \varepsilon_3$$

$$\vdots \quad \vdots$$

$$y_n = a + b \cdot x_n + \varepsilon_n$$

gegebene Daten:

<b>Y</b>	<b>X</b>
$y_1$	$x_1$
$y_2$	$x_2$
$y_3$	$x_3$
$\vdots$	$\vdots$
$y_n$	$x_n$

Modell: es gibt Zahlen  
 $a, b, \sigma^2$ , so dass

$$\begin{aligned}y_1 &= a + b \cdot x_1 + \varepsilon_1 \\y_2 &= a + b \cdot x_2 + \varepsilon_2 \\y_3 &= a + b \cdot x_3 + \varepsilon_3 \\&\vdots \\&\vdots \\y_n &= a + b \cdot x_n + \varepsilon_n\end{aligned}$$

Dabei sind  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  unabhängig  $\sim \mathcal{N}(0, \sigma^2)$ .

gegebene Daten:

<b>Y</b>	<b>X</b>
$y_1$	$x_1$
$y_2$	$x_2$
$y_3$	$x_3$
$\vdots$	$\vdots$
$y_n$	$x_n$

Modell: es gibt Zahlen  
 $a, b, \sigma^2$ , so dass

$$\begin{aligned}y_1 &= a + b \cdot x_1 + \varepsilon_1 \\y_2 &= a + b \cdot x_2 + \varepsilon_2 \\y_3 &= a + b \cdot x_3 + \varepsilon_3 \\&\vdots \\y_n &= a + b \cdot x_n + \varepsilon_n\end{aligned}$$

Dabei sind  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  unabhängig  $\sim \mathcal{N}(0, \sigma^2)$ .

$\Rightarrow y_1, y_2, \dots, y_n$  sind unabhängig  $y_i \sim \mathcal{N}(a + b \cdot x_i, \sigma^2)$ .

gegebene Daten:

<b>Y</b>	<b>X</b>
$y_1$	$x_1$
$y_2$	$x_2$
$y_3$	$x_3$
$\vdots$	$\vdots$
$y_n$	$x_n$

Modell: es gibt Zahlen  
 $a, b, \sigma^2$ , so dass

$$y_1 = a + b \cdot x_1 + \varepsilon_1$$

$$y_2 = a + b \cdot x_2 + \varepsilon_2$$

$$y_3 = a + b \cdot x_3 + \varepsilon_3$$

$$\vdots \quad \vdots$$

$$y_n = a + b \cdot x_n + \varepsilon_n$$

Dabei sind  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  unabhängig  $\sim \mathcal{N}(0, \sigma^2)$ .

$\Rightarrow y_1, y_2, \dots, y_n$  sind unabhängig  $y_i \sim \mathcal{N}(a + b \cdot x_i, \sigma^2)$ .

$a, b, \sigma^2$  sind unbekannt, aber **nicht zufällig**.

Wir schätzen  $a$  und  $b$ , indem wir

$$(\hat{a}, \hat{b}) := \arg \min_{(a,b)} \sum_i (y_i - (a + b \cdot x_i))^2 \quad \text{berechnen.}$$

Wir schätzen  $a$  und  $b$ , indem wir

$$(\hat{a}, \hat{b}) := \arg \min_{(a,b)} \sum_i (y_i - (a + b \cdot x_i))^2 \quad \text{berechnen.}$$

## Theorem

$\hat{a}$  und  $\hat{b}$  sind gegeben durch

$$\hat{b} = \frac{\sum_i (y_i - \bar{y}) \cdot (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i y_i \cdot (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

und

$$\hat{a} = \bar{y} - \hat{b} \cdot \bar{x}.$$

Wir schätzen  $a$  und  $b$ , indem wir

$$(\hat{a}, \hat{b}) := \arg \min_{(a,b)} \sum_i (y_i - (a + b \cdot x_i))^2 \quad \text{berechnen.}$$

### Theorem

$\hat{a}$  und  $\hat{b}$  sind gegeben durch

$$\hat{b} = \frac{\sum_i (y_i - \bar{y}) \cdot (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i y_i \cdot (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

und

$$\hat{a} = \bar{y} - \hat{b} \cdot \bar{x}.$$

### Bitte merken:

Die Gerade  $y = \hat{a} + \hat{b} \cdot x$  geht genau durch den Schwerpunkt der Punktwolke  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ .



# Lineare Regression in R

- In R wird die Gerade mit `lm()` (linear model) gefunden.

# Lineare Regression in R

- In R wird die Gerade mit  $\text{lm}()$  (linear model) gefunden.
- Die Abhängigkeit der Größe  $Y$  von  $X$  drückt man mit  $Y \sim X$  aus.

# Lineare Regression in R

- In R wird die Gerade mit `lm()` (linear model) gefunden.
- Die Abhängigkeit der Größe  $Y$  von  $X$  drückt man mit  $Y \sim X$  aus.
- `lm()` bewertet ausserdem die Genauigkeit für die Schätzung der Steigung und des  $y$ -Achsenabschnittes der Geraden.

```
> model <- lm(metabol~heartbpm,data=vulture,  
              subset=day=="17.05.")
```

```
> summary(model)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.2026	-0.2555	0.1005	0.6393	1.1834

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-7.73522	0.84543	-9.149	5.60e-08 ***
heartbpm	0.27771	0.01207	23.016	2.98e-14 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.912 on 17 degrees of freedom

Multiple R-squared: 0.9689, Adjusted R-squared: 0.9671

F-statistic: 529.7 on 1 and 17 DF, p-value: 2.979e-14

Folgendes muss man aus der Ausgabe herauslesen:

- Die Geradengleichung:

$$\text{metabol} = -7.73522 + 0.27771 * \text{heartbpm}$$

Folgendes muss man aus der Ausgabe herauslesen:

- Die Geradengleichung:

$$\text{metabol} = -7.73522 + 0.27771 * \text{heartbpm}$$

- Auch für nicht-gemessene Herzfrequenzen lässt sich hieraus die Stoffwechselrate schätzen. Beispiel: Bei Herzfrequenz 70 ist die Stoffwechselrate in etwa gleich

$$-7.73522 + 0.27771 * 70 = 11.70448$$

Die Streuung dieses Schätzers ist die geschätzte Standardabweichung der Residuen, also **0.912**.

Folgendes muss man aus der Ausgabe herauslesen:

- Die Geradengleichung:

$$\text{metabol} = -7.73522 + 0.27771 * \text{heartbpm}$$

- Auch für nicht-gemessene Herzfrequenzen lässt sich hieraus die Stoffwechselrate schätzen. Beispiel: Bei Herzfrequenz 70 ist die Stoffwechselrate in etwa gleich

$$-7.73522 + 0.27771 * 70 = 11.70448$$

Die Streuung dieses Schätzers ist die geschätzte Standardabweichung der Residuen, also **0.912**.

- Somit ist das Problem gelöst: Aus Messungen der Herzfrequenz können wir nun die Stoffwechselrate schätzen.

Es stellen sich noch ein paar Fragen:

- Ist das Modell unnötig kompliziert? (D.h. lässt sich die Parameterzahl reduzieren?)
- Passt das Modell auf die Daten? (Oder gibt es deutliche Abweichungen vom linearer Zusammenhang? Müssen weitere Parameter berücksichtigt werden?)



Kompliziertheit des Modells: Die Frage ist, kann auf den Steigungsparameter oder auf den y-Achsenabschnitt verzichten? Vielleicht ist einer dieser Werte so nahe bei 0, dass er gar nicht signifikant davon verschieden ist.

Kompliziertheit des Modells: Die Frage ist, kann auf den Steigungsparameter oder auf den y-Achsenabschnitt verzichten? Vielleicht ist einer dieser Werte so nahe bei 0, dass er gar nicht signifikant davon verschieden ist.

R liefert folgende Antworten:

- $H_0$ : Steigung gleich 0.  
Diese Nullhypothese hat den p-Wert  $5.60e - 08$ .

Kompliziertheit des Modells: Die Frage ist, kann auf den Steigungsparameter oder auf den y-Achsenabschnitt verzichten? Vielleicht ist einer dieser Werte so nahe bei 0, dass er gar nicht signifikant davon verschieden ist.

R liefert folgende Antworten:

- $H_0$ : Steigung gleich 0.  
Diese Nullhypothese hat den p-Wert  $5.60e - 08$ .
- $H_0$ : y-Achsenabschnitt gleich 0.  
Diese Nullhypothese hat den p-Wert  $2.98e - 14$ .

Kompliziertheit des Modells: Die Frage ist, kann auf den Steigungsparameter oder auf den y-Achsenabschnitt verzichten? Vielleicht ist einer dieser Werte so nahe bei 0, dass er gar nicht signifikant davon verschieden ist.

R liefert folgende Antworten:

- $H_0$ : Steigung gleich 0.  
Diese Nullhypothese hat den p-Wert  $5.60e - 08$ .
- $H_0$ : y-Achsenabschnitt gleich 0.  
Diese Nullhypothese hat den p-Wert  $2.98e - 14$ .

Beide Parameter sind signifikant von 0 verschieden. Würde man einen auf 0 setzen, wäre das Modell signifikant schlechter.

**Aber Vorsicht:** Ist die zu einer Variablen gehörende Steigung nicht signifikant ( $p\text{-Wert} > 0.05$ ), heißt das nicht unbedingt, dass das Modell besser wird, wenn man die Variable entfernt.

**Aber Vorsicht:** Ist die zu einer Variablen gehörende Steigung nicht signifikant ( $p\text{-Wert} > 0.05$ ), heißt das nicht unbedingt, dass das Modell besser wird, wenn man die Variable entfernt.

Auch hier gilt: **Aus nicht-Signifikanz kann man so gut wie gar nichts folgern.**

**Aber Vorsicht:** Ist die zu einer Variablen gehörende Steigung nicht signifikant ( $p\text{-Wert} > 0.05$ ), heißt das nicht unbedingt, dass das Modell besser wird, wenn man die Variable entfernt.

Auch hier gilt: **Aus nicht-Signifikanz kann man so gut wie gar nichts folgern.**

Um zu entscheiden, welche nicht-signifikanten Variablen man im Modell lassen sollte, kann man neben biologischem Sachverstand z.B. AIC-Werte berücksichtigen oder eine Kreuzvalidierung durchführen.

# Übersicht

- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression**
  - Lineare Zusammenhänge
  - **t-Test fuer lineare Zusammenhänge**
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse
- 7 Versuchsplanung: Stichprobenumfang



Modell:

$$Y = a + b \cdot X + \varepsilon \quad \text{mit } \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Modell:

$$Y = a + b \cdot X + \varepsilon \quad \text{mit } \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Wie berechnet man die Signifikanz eines Zusammenhangs zwischen dem *erklärenden Merkmal*  $X$  und der *Zielgröße*  $Y$ ?

Modell:

$$Y = a + b \cdot X + \varepsilon \quad \text{mit } \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Wie berechnet man die Signifikanz eines Zusammenhangs zwischen dem *erklärenden Merkmal*  $X$  und der *Zielgröße*  $Y$ ?

Anders formuliert: Mit welchem Test können wir der Nullhypothese  $b = 0$  zu Leibe rücken?

Modell:

$$Y = a + b \cdot X + \varepsilon \quad \text{mit } \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Wie berechnet man die Signifikanz eines Zusammenhangs zwischen dem *erklärenden Merkmal*  $X$  und der *Zielgröße*  $Y$ ?

Anders formuliert: Mit welchem Test können wir der Nullhypothese  $b = 0$  zu Leibe rücken?

Wir haben  $b$  durch  $\hat{b}$  geschätzt (und gehen jetzt mal von  $\hat{b} \neq 0$  aus). Könnte das wahre  $b$  auch 0 sein?

Modell:

$$Y = a + b \cdot X + \varepsilon \quad \text{mit } \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Wie berechnet man die Signifikanz eines Zusammenhangs zwischen dem *erklärenden Merkmal*  $X$  und der *Zielgröße*  $Y$ ?

Anders formuliert: Mit welchem Test können wir der Nullhypothese  $b = 0$  zu Leibe rücken?

Wir haben  $b$  durch  $\hat{b}$  geschätzt (und gehen jetzt mal von  $\hat{b} \neq 0$  aus). Könnte das wahre  $b$  auch 0 sein?

Wie groß ist der Standardfehler unserer Schätzung  $\hat{b}$ ?

$$y_i = a + b \cdot x_i + \varepsilon \quad \text{mit } \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

nicht zufällig:  $a, b, x_i, \sigma^2$       zufällig:  $\varepsilon, y_i$

$$\text{var}(y_i) = \text{var}(a + b \cdot x_i + \varepsilon) = \text{var}(\varepsilon) = \sigma^2$$

und  $y_1, y_2, \dots, y_n$  sind stochastisch unabhängig.

$$y_i = a + b \cdot x_i + \varepsilon \quad \text{mit } \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

nicht zufällig:  $a, b, x_i, \sigma^2$       zufällig:  $\varepsilon, y_i$

$$\text{var}(y_i) = \text{var}(a + b \cdot x_i + \varepsilon) = \text{var}(\varepsilon) = \sigma^2$$

und  $y_1, y_2, \dots, y_n$  sind stochastisch unabhängig.

$$\hat{b} = \frac{\sum_i y_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

$$y_i = a + b \cdot x_i + \varepsilon \quad \text{mit } \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

nicht zufällig:  $a, b, x_i, \sigma^2$       zufällig:  $\varepsilon, y_i$

$$\text{var}(y_i) = \text{var}(a + b \cdot x_i + \varepsilon) = \text{var}(\varepsilon) = \sigma^2$$

und  $y_1, y_2, \dots, y_n$  sind stochastisch unabhängig.

$$\hat{b} = \frac{\sum_i y_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

$$\text{var}(\hat{b}) = \text{var}\left(\frac{\sum_i y_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}\right) = \frac{\text{var}(\sum_i y_i (x_i - \bar{x}))}{(\sum_i (x_i - \bar{x})^2)^2}$$



$$y_i = a + b \cdot x_i + \varepsilon \quad \text{mit } \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

nicht zufällig:  $a, b, x_i, \sigma^2$       zufällig:  $\varepsilon, y_i$

$$\text{var}(y_i) = \text{var}(a + b \cdot x_i + \varepsilon) = \text{var}(\varepsilon) = \sigma^2$$

und  $y_1, y_2, \dots, y_n$  sind stochastisch unabhängig.

$$\hat{b} = \frac{\sum_i y_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

$$\begin{aligned} \text{var}(\hat{b}) &= \text{var}\left(\frac{\sum_i y_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}\right) = \frac{\text{var}(\sum_i y_i (x_i - \bar{x}))}{(\sum_i (x_i - \bar{x})^2)^2} \\ &= \frac{\sum_i \text{var}(y_i) (x_i - \bar{x})^2}{(\sum_i (x_i - \bar{x})^2)^2} = \sigma^2 \cdot \frac{\sum_i (x_i - \bar{x})^2}{(\sum_i (x_i - \bar{x})^2)^2} \end{aligned}$$

$$y_i = a + b \cdot x_i + \varepsilon \quad \text{mit } \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

nicht zufällig:  $a, b, x_i, \sigma^2$       zufällig:  $\varepsilon, y_i$

$$\text{var}(y_i) = \text{var}(a + b \cdot x_i + \varepsilon) = \text{var}(\varepsilon) = \sigma^2$$

und  $y_1, y_2, \dots, y_n$  sind stochastisch unabhängig.

$$\hat{b} = \frac{\sum_i y_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

$$\begin{aligned} \text{var}(\hat{b}) &= \text{var}\left(\frac{\sum_i y_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}\right) = \frac{\text{var}(\sum_i y_i (x_i - \bar{x}))}{(\sum_i (x_i - \bar{x})^2)^2} \\ &= \frac{\sum_i \text{var}(y_i) (x_i - \bar{x})^2}{(\sum_i (x_i - \bar{x})^2)^2} = \sigma^2 \cdot \frac{\sum_i (x_i - \bar{x})^2}{(\sum_i (x_i - \bar{x})^2)^2} \\ &= \sigma^2 / \sum_i (x_i - \bar{x})^2 \end{aligned}$$

Tatsächlich ist  $\hat{b}$  Normalverteilt mit Mittelwert  $b$  und

$$\text{var}(\hat{b}) = \sigma^2 / \sum_i (x_i - \bar{x})^2$$

Tatsächlich ist  $\hat{b}$  Normalverteilt mit Mittelwert  $b$  und

$$\text{var}(\hat{b}) = \sigma^2 / \sum_i (x_i - \bar{x})^2$$

**Problem:** Wir kennen  $\sigma^2$  nicht.

Tatsächlich ist  $\hat{b}$  Normalverteilt mit Mittelwert  $b$  und

$$\text{var}(\hat{b}) = \sigma^2 / \sum_i (x_i - \bar{x})^2$$

**Problem:** Wir kennen  $\sigma^2$  nicht.

Wir schätzen  $\sigma^2$  mit Hilfe der beobachteten Residuenvarianz durch

$$s^2 := \frac{\sum_i (y_i - \hat{a} - \hat{b} \cdot x_i)^2}{n - 2}$$

Tatsächlich ist  $\hat{b}$  Normalverteilt mit Mittelwert  $b$  und

$$\text{var}(\hat{b}) = \sigma^2 / \sum_i (x_i - \bar{x})^2$$

**Problem:** Wir kennen  $\sigma^2$  nicht.

Wir schätzen  $\sigma^2$  mit Hilfe der beobachteten Residuenvarianz durch

$$s^2 := \frac{\sum_i (y_i - \hat{a} - \hat{b} \cdot x_i)^2}{n - 2}$$

Zu beachten ist, dass durch  $n - 2$  geteilt wird. Das hat damit zu tun, dass zwei Modellparameter  $a$  und  $b$  bereits geschätzt wurden, und somit 2 Freiheitsgrade verloren gegangen sind.

$$\text{var}(\hat{b}) = \sigma^2 / \sum_i (x_i - \bar{x})^2$$

Schätze  $\sigma^2$  durch

$$s^2 = \frac{\sum_i (y_i - \hat{a} - \hat{b} \cdot x_i)^2}{n - 2}.$$

$$\text{var}(\hat{b}) = \sigma^2 / \sum_i (x_i - \bar{x})^2$$

Schätze  $\sigma^2$  durch

$$s^2 = \frac{\sum_i (y_i - \hat{a} - \hat{b} \cdot x_i)^2}{n - 2}.$$

Dann ist

$$\frac{\hat{b} - b}{s / \sqrt{\sum_i (x_i - \bar{x})^2}}$$

Student- $t$ -verteilt mit  $n - 2$  Freiheitsgraden und wir können den  $t$ -Test anwenden, um die Nullhypothese  $b = 0$  zu testen.



$$\text{var}(\hat{b}) = \sigma^2 / \sum_i (x_i - \bar{x})^2$$

Schätze  $\sigma^2$  durch

$$s^2 = \frac{\sum_i (y_i - \hat{a} - \hat{b} \cdot x_i)^2}{n - 2}$$

Dann ist

$$\frac{\hat{b} - b}{s / \sqrt{\sum_i (x_i - \bar{x})^2}}$$

Student- $t$ -verteilt mit  $n - 2$  Freiheitsgraden und wir können den  $t$ -Test anwenden, um die Nullhypothese  $b = 0$  zu testen.

Verwerfe  $H_0$ : „ $b = 0$ “ zum Signifikanzniveau  $\alpha$ , wenn

$\left| \frac{\hat{b}}{s / \sqrt{\sum_i (x_i - \bar{x})^2}} \right| \geq q_{1-\alpha/2}$ , wo  $q_{1-\alpha/2}$  das  $(1 - \alpha/2)$ -Quantil der Student-Verteilung mit  $n - 2$  Freiheitsgraden ist.

# Beispiel: Rothirsch (*Cervus elaphus*)

Theorie: Hirschkühe können das Geschlecht ihrer Nachkommen beeinflussen.

# Beispiel: Rothirsch (*Cervus elaphus*)

Theorie: Hirschkühe können das Geschlecht ihrer Nachkommen beeinflussen.

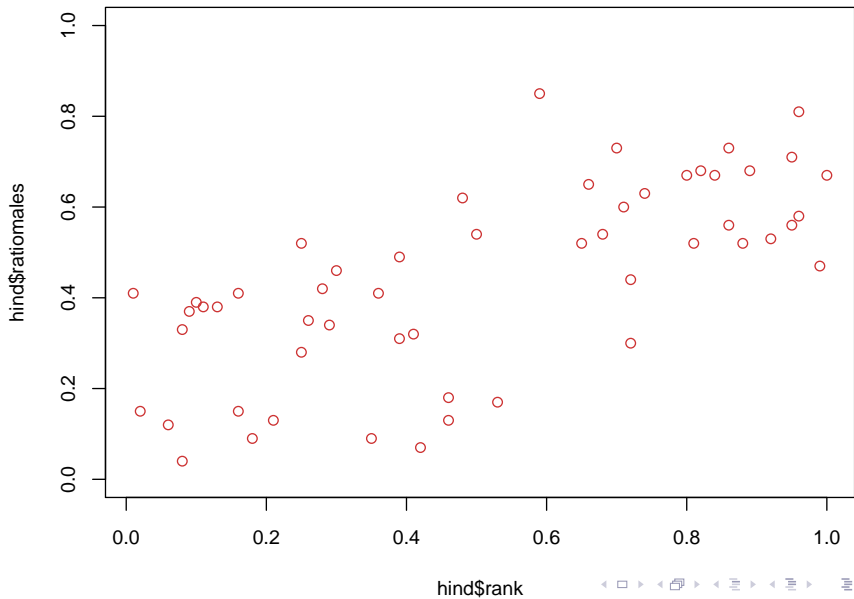
Unter dem Gesichtspunkt evolutionär stabiler Strategien ist zu erwarten, dass schwache Tiere eher zu weiblichem und starke Tiere eher zu männlichem Nachwuchs tendieren.

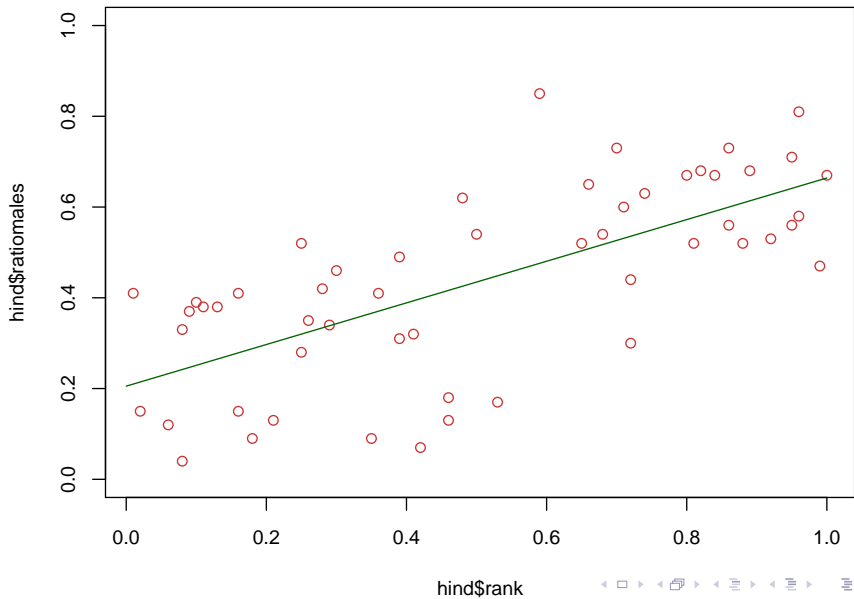


Clutton-Brock, T. H. , Albon, S. D., Guinness, F. E. (1986)  
Great expectations: dominance, breeding success and  
offspring sex ratios in red deer.  
*Anim. Behav.* **34**, 460—471.

```
> hind
  rank rationales
1 0.01      0.41
2 0.02      0.15
3 0.06      0.12
4 0.08      0.04
5 0.08      0.33
6 0.09      0.37
.  .        .
.  .        .
.  .        .
52 0.96     0.81
53 0.99     0.47
54 1.00     0.67
```

**ACHTUNG: Simulierte Daten, die sich an den Daten aus der Originalpublikation lediglich orientieren.**





```
> mod <- lm(ratiomales~rank,data=hind)
```

```
> summary(mod)
```

```
Call:
```

```
lm(formula = ratiomales ~ rank, data = hind)
```

```
Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.32798	-0.09396	0.02408	0.11275	0.37403

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	0.20529	0.04011	5.119	4.54e-06	***
rank	0.45877	0.06732	6.814	9.78e-09	***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.154 on 52 degrees of freedom
```

```
Multiple R-squared: 0.4717, Adjusted R-squared: 0.4616
```

```
F-statistic: 46.44 on 1 and 52 DF,  p-value: 9.78e-09
```

# Übersicht

- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression**
  - Lineare Zusammenhänge
  - $t$ -Test fuer lineare Zusammenhänge
  - **Überprüfen der Annahmen und Modellwahl**
- 6 Varianzanalyse
- 7 Versuchsplanung: Stichprobenumfang



Passt das Modell auf die Daten?, z.B.

$$Y = a + b \cdot X + \varepsilon \quad \text{mit } \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Passt das Modell auf die Daten?, z.B.

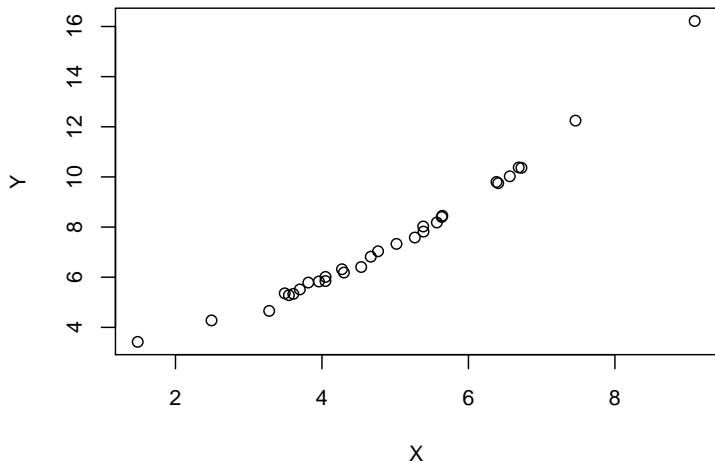
$$Y = a + b \cdot X + \varepsilon \quad \text{mit } \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Wenn dieses Modell zutrifft, dann müssen die Residuen

$$y_i - (\hat{a} + \hat{b} \cdot x_i)$$

normalverteilt aussehen und dürfen keinen offensichtlichen Abhängigkeiten mit  $X$  oder  $\hat{a} + \hat{b} \cdot X$  aufweisen.

Beispiel: Wird der Zusammenhang zwischen  $X$  und  $Y$  hinreichend gut durch eine lineare Gleichung  $Y_i = a + b \cdot X_i + \varepsilon_i$  beschrieben?



```
> mod <- lm(Y ~ X)
> summary(mod)
```

Call:

```
lm(formula = Y ~ X)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.49984	-0.26727	-0.13472	0.01344	1.82718

Coefficients:

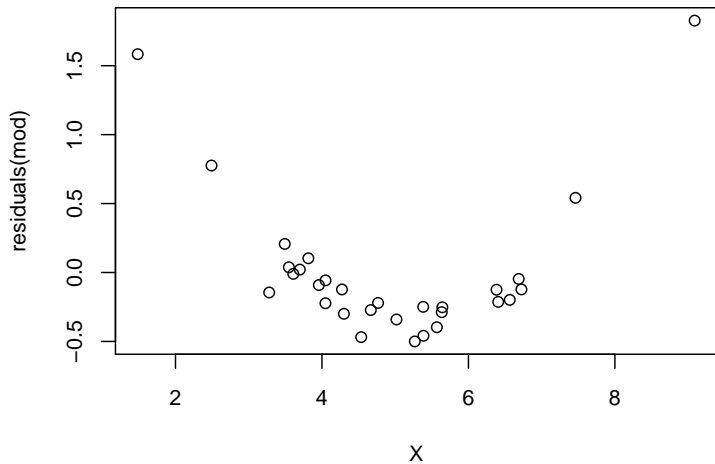
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.61118	0.33295	-1.836	0.077 .
X	1.65055	0.06472	25.505	<2e-16 ***

---

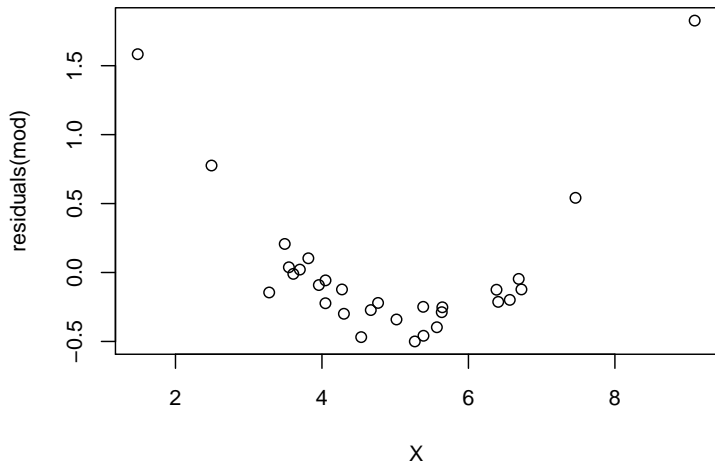
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1

Residual standard error: 0.5473 on 28 degrees of freedom

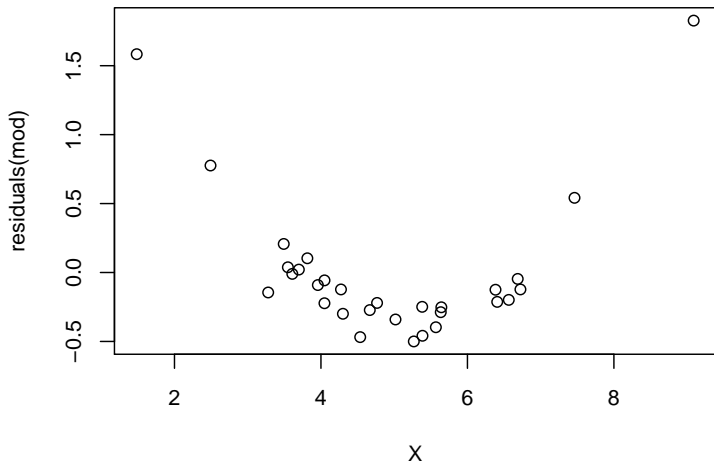
```
> plot(X,residuals(mod))
```



```
> plot(X,residuals(mod))
```



```
> plot(X,residuals(mod))
```



Die Resiuden sind offensichtlich tendenziell größer für kleine und für große  $X$ -Werte als für mittlere  $X$ -Werte. Das darf nicht sein!

Idee: Vielleicht kann man an die Punkte  $(x_i, y_i)$  statt einer Geraden besser einen Ausschnitt aus einer Parabel anpassen.



Idee: Vielleicht kann man an die Punkte  $(x_i, y_i)$  statt einer Geraden besser einen Ausschnitt aus einer Parabel anpassen.  
Also: Passe besser ein Modell der folgenden Form an:

$$Y_i = a + b \cdot X_i + c \cdot X_i^2 + \varepsilon_i$$

Idee: Vielleicht kann man an die Punkte  $(x_i, y_i)$  statt einer Geraden besser einen Ausschnitt aus einer Parabel anpassen.  
Also: Passe besser ein Modell der folgenden Form an:

$$Y_i = a + b \cdot X_i + c \cdot X_i^2 + \varepsilon_i$$

Ist das überhaupt noch ein lineares Modell?

Idee: Vielleicht kann man an die Punkte  $(x_i, y_i)$  statt einer Geraden besser einen Ausschnitt aus einer Parabel anpassen.  
Also: Passe besser ein Modell der folgenden Form an:

$$Y_i = a + b \cdot X_i + c \cdot X_i^2 + \varepsilon_i$$

Ist das überhaupt noch ein lineares Modell? Ja: Sei  $Z = X^2$ , dann ist  $Y$  linear in  $X$  und  $Z$ .

Idee: Vielleicht kann man an die Punkte  $(x_i, y_i)$  statt einer Geraden besser einen Ausschnitt aus einer Parabel anpassen.  
Also: Passe besser ein Modell der folgenden Form an:

$$Y_i = a + b \cdot X_i + c \cdot X_i^2 + \varepsilon_i$$

Ist das überhaupt noch ein lineares Modell? Ja: Sei  $Z = X^2$ , dann ist  $Y$  linear in  $X$  und  $Z$ .

In R:

```
> Z <- X^2  
> mod2 <- lm(Y ~ X+Z)
```

```
> summary(mod2)
```

```
Call:
```

```
lm(formula = Y ~ X + Z)
```

```
Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.321122	-0.060329	0.007706	0.075337	0.181965

```
Coefficients:
```

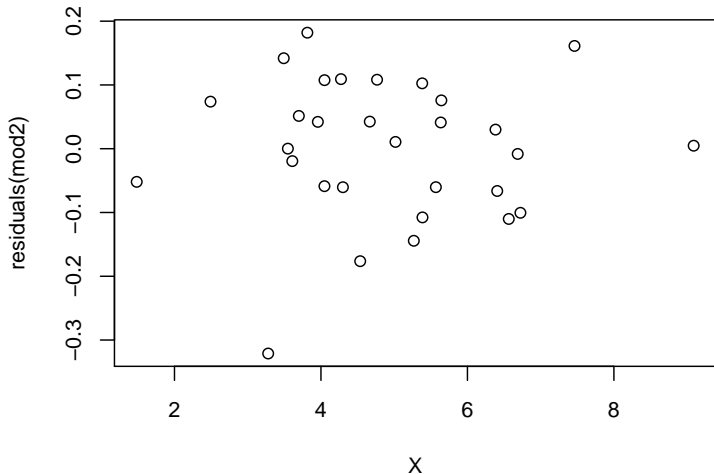
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	2.933154	0.158825	18.468	<2e-16	***
X	0.150857	0.061921	2.436	0.0217	*
Z	0.144156	0.005809	24.817	<2e-16	***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Für dieses Modell gibt es keinen erkennbaren Zusammenhang mehr zwischen  $X$  und den Residuen:

```
plot(X, residuals(mod2))
```



Wie gut passt die Normalverteilungsannahme auf das Modell  $Y_i = a + b \cdot X_i + \varepsilon_i$ ?

Wie gut passt die Normalverteilungsannahme auf das Modell  $Y_i = a + b \cdot X_i + \varepsilon_i$ ?

Sind die Residuen  $r_i = Y_i - (\hat{a} + \hat{b} \cdot X_i)$  einigermaßen normalverteilt?



Wie gut passt die Normalverteilungsannahme auf das Modell  $Y_i = a + b \cdot X_i + \varepsilon_i$ ?

Sind die Residuen  $r_i = Y_i - (\hat{a} + \hat{b} \cdot X_i)$  einigermaßen normalverteilt?

Graphische Methode: vergleiche die theoretischen Quantile der Standardnormalverteilung  $\mathcal{N}(0, 1)$  mit denen der Residuen.

Wie gut passt die Normalverteilungsannahme auf das Modell  $Y_i = a + b \cdot X_i + \varepsilon_i$ ?

Sind die Residuen  $r_i = Y_i - (\hat{a} + \hat{b} \cdot X_i)$  einigermaßen normalverteilt?

Graphische Methode: vergleiche die theoretischen Quantile der Standardnormalverteilung  $\mathcal{N}(0, 1)$  mit denen der Residuen.

Hintergrund: trägt man die Quantile von  $\mathcal{N}(\mu, \sigma^2)$  gegen die von  $\mathcal{N}(0, 1)$  auf, so erhält man die Gerade  $y(x) = \mu + \sigma \cdot x$ .

Wie gut passt die Normalverteilungsannahme auf das Modell  $Y_i = a + b \cdot X_i + \varepsilon_i$ ?

Sind die Residuen  $r_i = Y_i - (\hat{a} + \hat{b} \cdot X_i)$  einigermaßen normalverteilt?

Graphische Methode: vergleiche die theoretischen Quantile der Standardnormalverteilung  $\mathcal{N}(0, 1)$  mit denen der Residuen.

Hintergrund: trägt man die Quantile von  $\mathcal{N}(\mu, \sigma^2)$  gegen die von  $\mathcal{N}(0, 1)$  auf, so erhält man die Gerade  $y(x) = \mu + \sigma \cdot x$ . (Grund: Ist  $X$  standardnormalverteilt und  $Y = a + b \cdot X$ , so ist  $Y$  normalverteilt mit Mittelwert  $a$  and Varianz  $b^2$ .)

Bevor man die Modellanpassung mit `lm()` vornimmt, muss man erst mal überprüfen, ob die Normalverteilungsannahme erfüllt ist.

~~Bevor man die Modellanpassung mit  $\text{lm}()$  vornimmt, muss man erst mal überprüfen, ob die Normalverteilungsannahme erfüllt ist.~~

~~Bevor man die Modellanpassung mit  $\text{lm}()$  vornimmt, muss man erst mal überprüfen, ob die Normalverteilungsannahme erfüllt ist.~~

Um zu überprüfen ob die Normalverteilungsannahme eines linearen Modells gilt, benötigt man die Residuen.

~~Bevor man die Modellanpassung mit  $\text{lm}()$  vornimmt, muss man erst mal überprüfen, ob die Normalverteilungsannahme erfüllt ist.~~

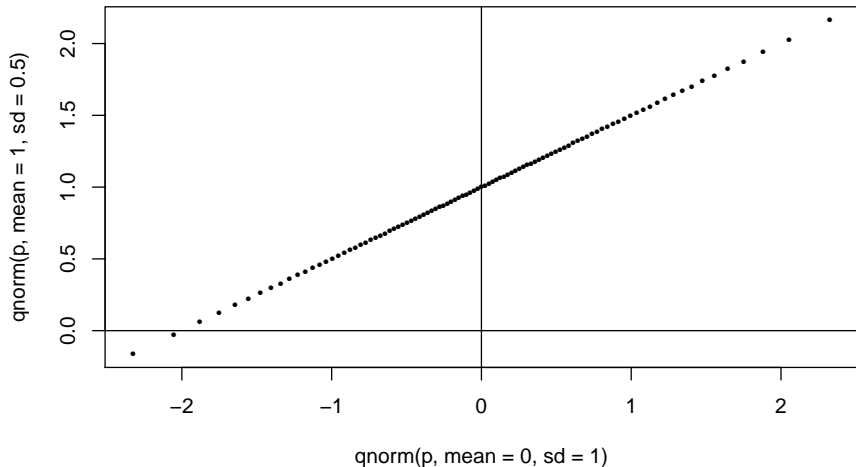
Um zu überprüfen ob die Normalverteilungsannahme eines linearen Modells gilt, benötigt man die Residuen. Um die Residuen zu berechnen, muss man zuerst die Modellanpassung berechnen (in R mit  $\text{lm}()$ ).

~~Bevor man die Modellanpassung mit  $\text{lm}()$  vornimmt, muss man erst mal überprüfen, ob die Normalverteilungsannahme erfüllt ist.~~

Um zu überprüfen ob die Normalverteilungsannahme eines linearen Modells gilt, benötigt man die Residuen. Um die Residuen zu berechnen, muss man zuerst die Modellanpassung berechnen (in R mit  $\text{lm}()$ ). Erst dann kann man die Normalverteilungsannahmen überprüfen und entscheiden, ob man bei dem Modell bleibt oder das Modell noch verändern muss.



```
p <- seq(from=0,to=1,by=0.01)
plot(qnorm(p,mean=0,sd=1),qnorm(p,mean=1,sd=0.5),
     pch=16,cex=0.5)
abline(v=0,h=0)
```



Trägt man statt der theoretischen Quantile die empirischen Quantile der Residuen gegen die Quantile einer Normalverteilung auf, so liegen die Werte nicht genau auf einer Geraden, aber in der Nähe einer Geraden.

Trägt man statt der theoretischen Quantile die empirischen Quantile der Residuen gegen die Quantile einer Normalverteilung auf, so liegen die Werte nicht genau auf einer Geraden, aber in der Nähe einer Geraden.

Falls keine *systematischen* Abweichungen von einer imaginären Geraden erkennbar sind: **Normalverteilungsannahme ist akzeptabel**

Trägt man statt der theoretischen Quantile die empirischen Quantile der Residuen gegen die Quantile einer Normalverteilung auf, so liegen die Werte nicht genau auf einer Geraden, aber in der Nähe einer Geraden.

Falls keine *systematischen* Abweichungen von einer imaginären Geraden erkennbar sind: **Normalverteilungsannahme ist akzeptabel**

Falls *systematische* Abweichungen von einer imaginären Geraden deutlich erkennbar sind: **Normalverteilungsannahme ist heikel, möglicherweise sollten Variablen umskaliert andere Variablen hinzugezogen werden.**

Beispiel:

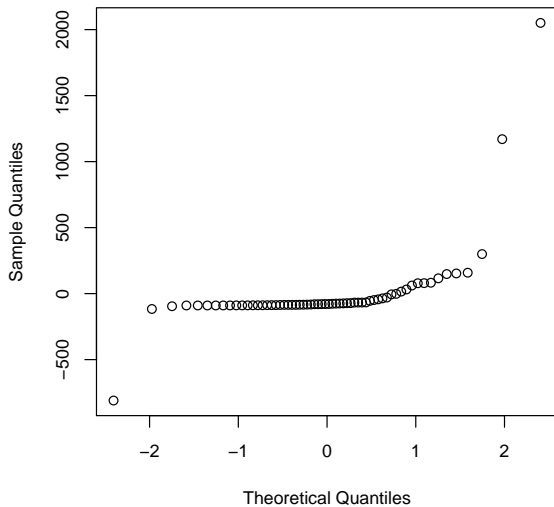
Daten: Typisches Körpergewicht [kg] und Gehirngewicht [g] von 62 Säugetierarten (und 3 Dinosaurierarten)

Beispiel:

Daten: Typisches Körpergewicht [kg] und Gehirngewicht [g] von 62 Säugetierarten (und 3 Dinosaurierarten)

Sind die Residuen einigermaßen normalverteilt?

```
qqnorm(modell$residuals)
```

**Normal Q-Q Plot**

Die Residuen sind offensichtlich nicht normalverteilt. Zumindest gibt es extreme Ausreißer.



Die Residuen sind offensichtlich nicht normalverteilt. Zumindest gibt es extreme Ausreißer.

Wenn man sich aber die logarithmierten Größen ansieht, so erhält man normalverteilte Residuen:

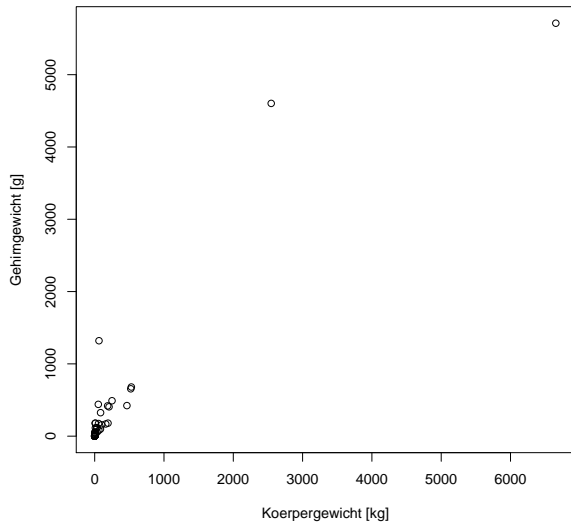
```
> logmodell <- lm(log(brain.weight.g)~log(weight.kg.),  
                  subset=extinct=="no")
```

# Daten: Typisches Körpergewicht [kg] und Gehirngewicht [g] von 62 Säugetierarten (und 3 Dinosaurierarten)

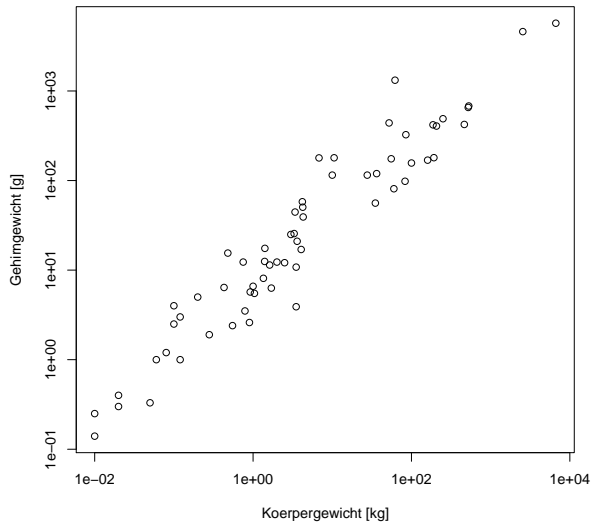
```
> data
```

```
  weight.kg. brain.weight.g      species extinct
1    6654.00    5712.00 african elephant   no
2         1.00         6.60                no
3         3.39        44.50                no
4         0.92         5.70                no
5   2547.00    4603.00  asian elephant   no
6    10.55    179.50                no
7         0.02         0.30                no
8   160.00    169.00                no
9         3.30        25.60                cat   no
10    52.16    440.00 chimpanzee   no
11     0.43         6.40                no
.         .         .                .
.         .         .                .
.         .         .                .
```

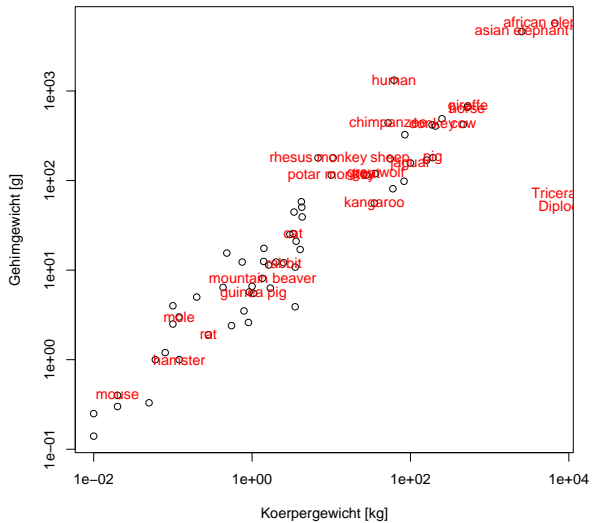
## typische Werte bei 62 Säugetierarten



## typische Werte bei 62 Säugetierarten



## typische Werte bei 62 Säugetierarten



```
> modell <- lm(brain.weight.g~weight.kg.,subset=extinct=="no")
> summary(modell)
```

Call:

```
lm(formula = brain.weight.g ~ weight.kg., subset = extinct ==
    "no")
```

Residuals:

Min	1Q	Median	3Q	Max
-809.95	-87.43	-78.55	-31.17	2051.05

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	89.91213	43.58134	2.063	0.0434 *
weight.kg.	0.96664	0.04769	20.269	<2e-16 ***

---

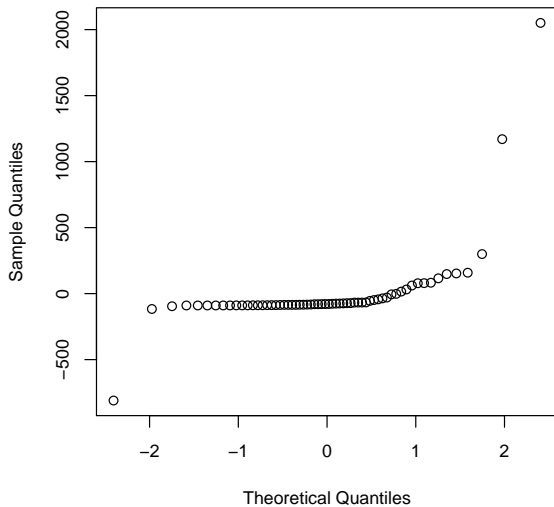
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 334.8 on 60 degrees of freedom

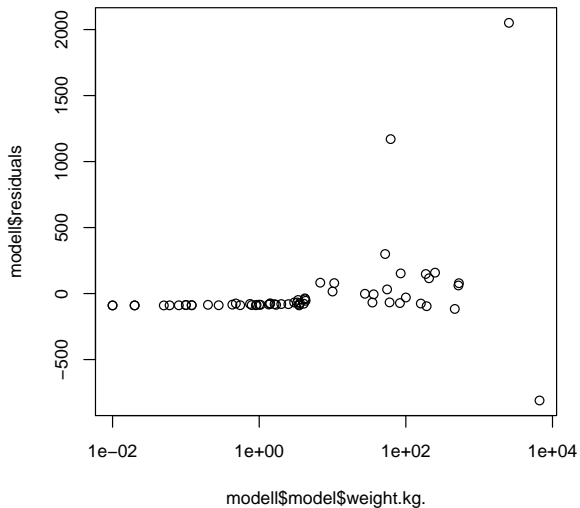
Multiple R-squared: 0.8726, Adjusted R-squared: 0.8704

F-statistic: 410.8 on 1 and 60 DF, p-value: < 2.2e-16

```
qqnorm(modell$residuals)
```

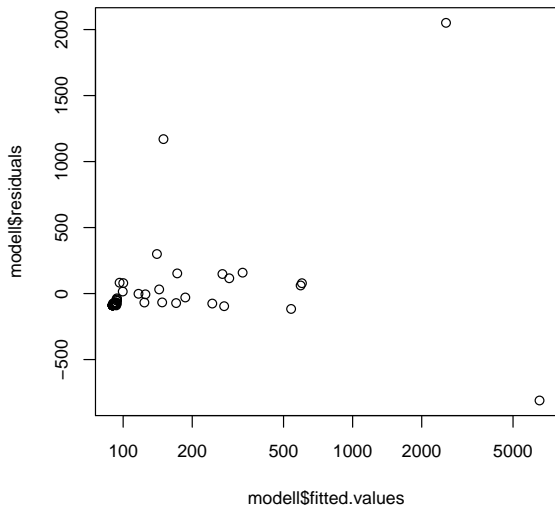
**Normal Q-Q Plot**

```
plot(modell$fitted.values,modell$residuals)
```

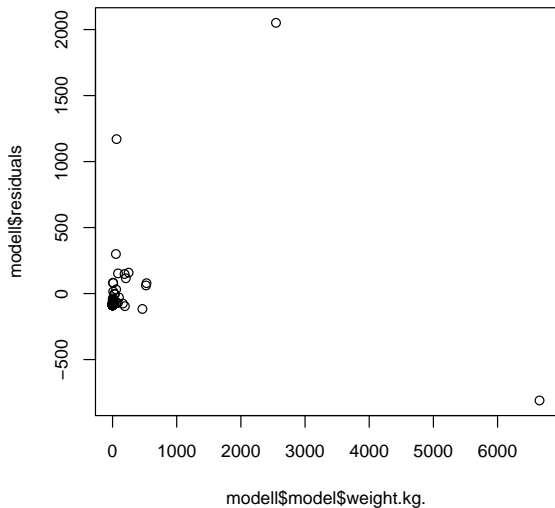




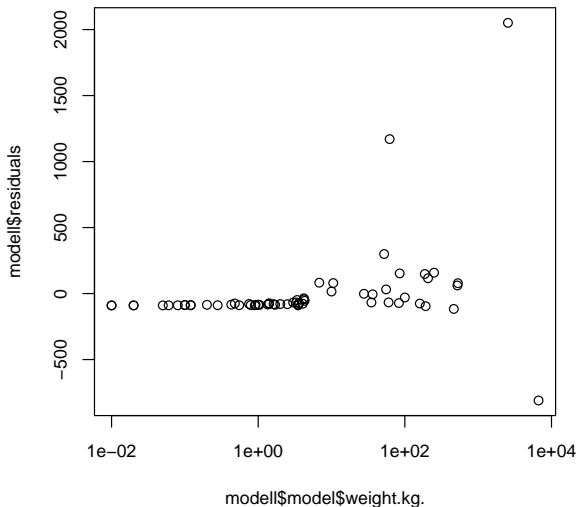
```
plot(modell$fitted.values,modell$residuals,log='x')
```



```
plot(modell$model$weight.kg.,modell$residuals)
```



```
plot(modell$model$weight.kg.,modell$residuals,log='x' )
```



Wir sehen, dass die Varianz der Residuen von den angepassten Werten bzw. dem Körpergewicht abhängt. Man sagt, es liegt *Heteroskedastizität* vor.

Wir sehen, dass die Varianz der Residuen von den angepassten Werten bzw. dem Körpergewicht abhängt. Man sagt, es liegt *Heteroskedastizität* vor.

Das Modell geht aber von *Homoskedastizität* aus, d.h. die Residuenvarianz soll von den erklärenden Merkmalen (dem Körpergewicht) und den angepassten Werten (annähernd) unabhängig sein.

Wir sehen, dass die Varianz der Residuen von den angepassten Werten bzw. dem Körpergewicht abhängt. Man sagt, es liegt *Heteroskedastizität* vor.

Das Modell geht aber von *Homoskedastizität* aus, d.h. die Residuenvarianz soll von den erklärenden Merkmalen (dem Körpergewicht) und den angepassten Werten (annähernd) unabhängig sein.

### **Varianzstabilisierende Transformation:**

Wie können wir die Körper- und Hirnmasse umskalieren, um Homoskedastizität zu erreichen?

Eigentlich ist es ja offensichtlich: Bei Elefanten kann das typischerweise 5 kg schwere Hirn je nach Individuum auch mal 500 g schwerer oder leichter sein. Wenn bei einer Tierart das Hirn typischerweise 5 g schwer ist, wird es nicht um 500 g variieren können, sondern vielleicht ebenfalls um 10%, also  $\pm 0.5$  g. Die Varianz ist hier also nicht additiv, sondern multiplikativ:

$$\text{Hirnmasse} = (\text{erwartete Hirnmasse}) \cdot \text{Zufall}$$

Das können wir aber in etwas mit additivem Zufallsterm umwandeln, indem wir auf beiden Seiten den (natürlichen) Logarithmus ziehen:

$$\log(\text{Hirnmasse}) = \log(\text{erwartete Hirnmasse}) + \log(\text{Zufall})$$

```
> logmodell <- lm(log(brain.weight.g)~log(weight.kg.),subset=
> summary(logmodell)
```

Call:

```
lm(formula = log(brain.weight.g) ~ log(weight.kg.), subset = c
    "no")
```

Residuals:

Min	1Q	Median	3Q	Max
-1.68908	-0.51262	-0.05016	0.46023	1.97997

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.11067	0.09794	21.55	<2e-16 ***
log(weight.kg.)	0.74985	0.02888	25.97	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

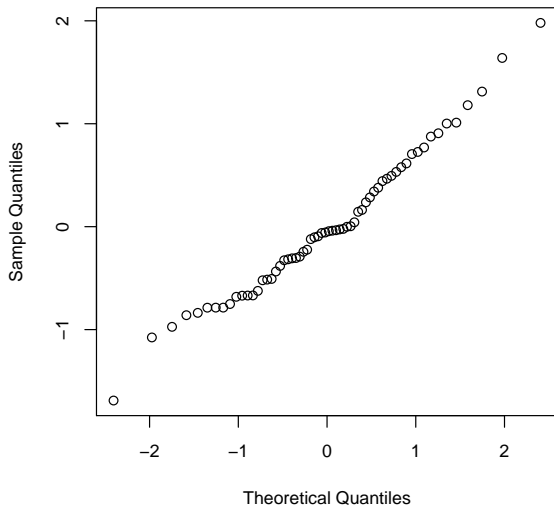
Residual standard error: 0.7052 on 60 degrees of freedom

Multiple R-squared: 0.9183, Adjusted R-squared: 0.9169

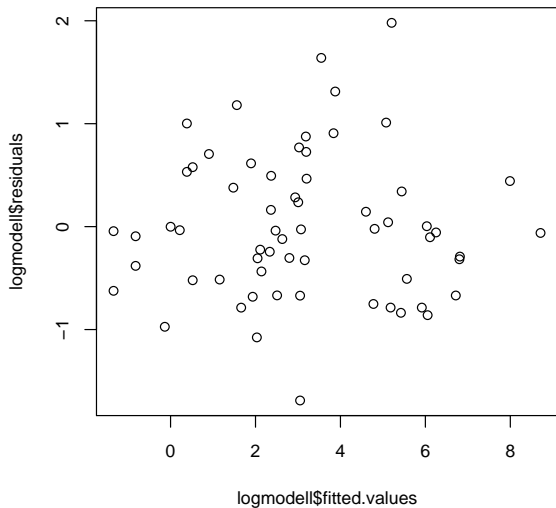
F-statistic: 674.3 on 1 and 60 DF. p-value: < 2.2e-16



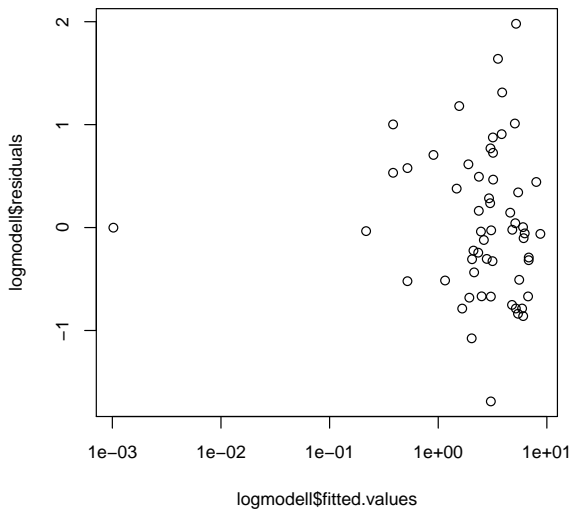
```
qqnorm(modell$residuals)
```

**Normal Q-Q Plot**

```
plot(logmodell$fitted.values, logmodell$residuals)
```

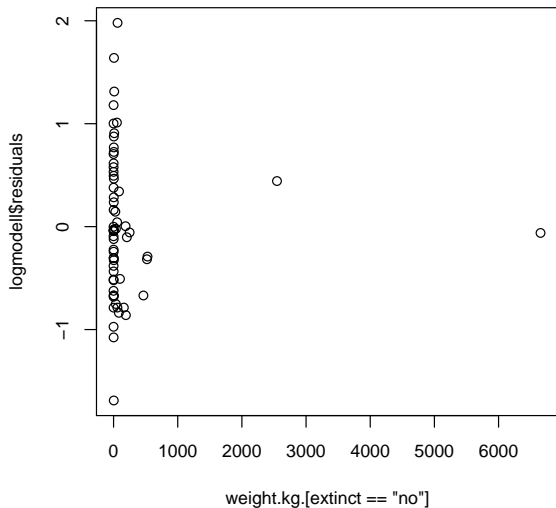


```
plot(logmodell$fitted.values, logmodell$residuals, log='x')
```

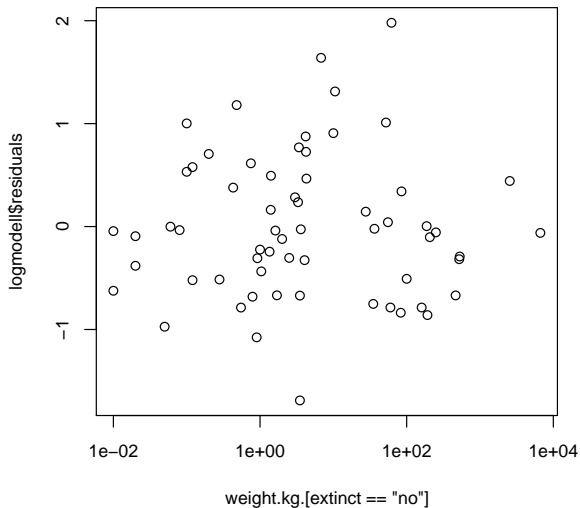


)

```
plot(weight.kg.[extinct=='no'], logmodell$residuals)
```

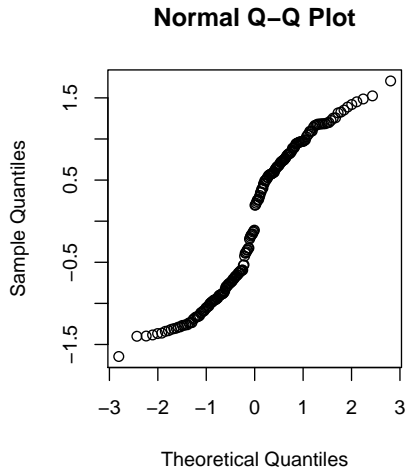
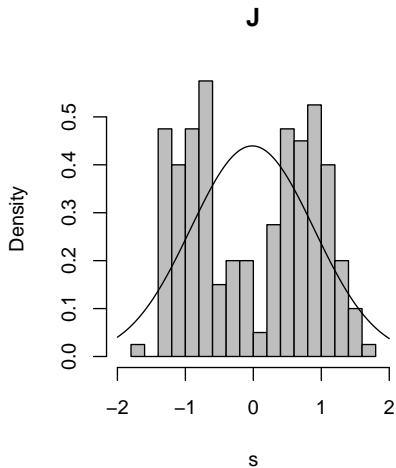


```
plot(weight.kg.[extinct='no'], logmodell$residuals, log='x')
```

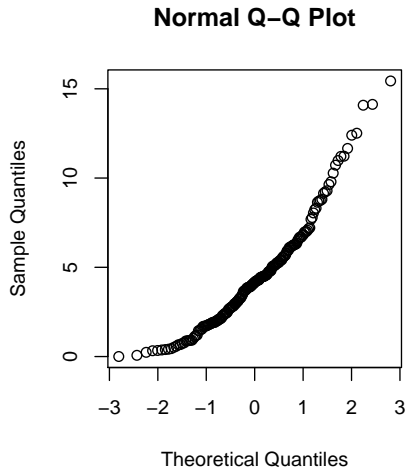
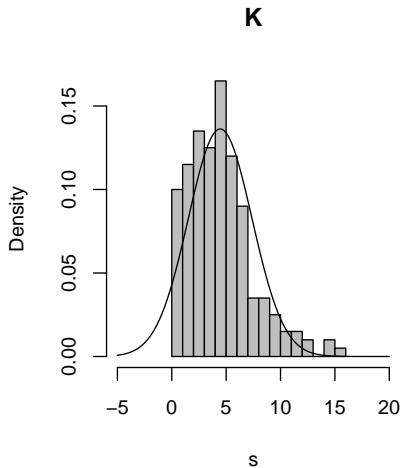


)

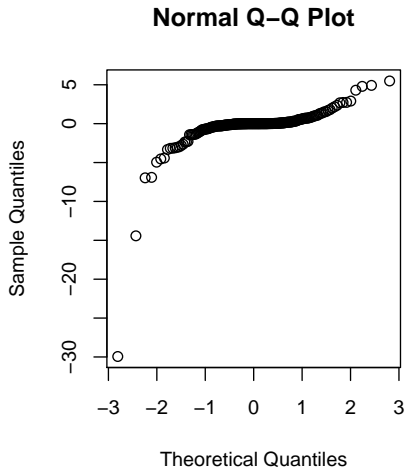
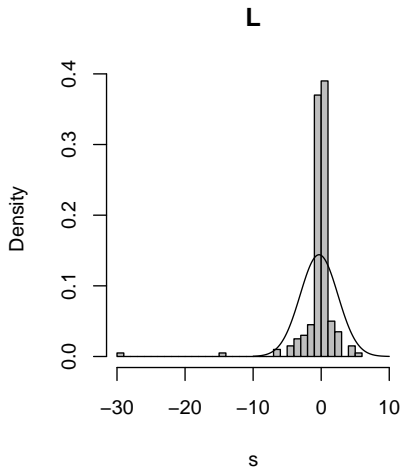
# Beispiele von Histogrammen und qqnorm-plots einiger Verteilungen



# Beispiele von Histogrammen und qqnorm-plots einiger Verteilungen

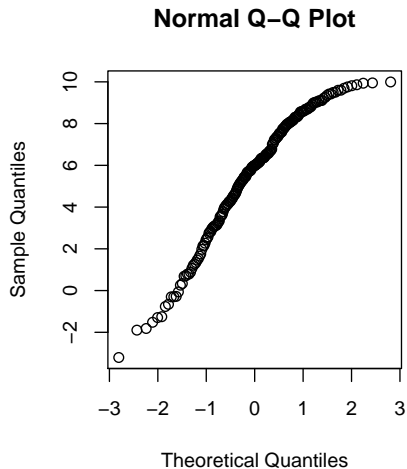
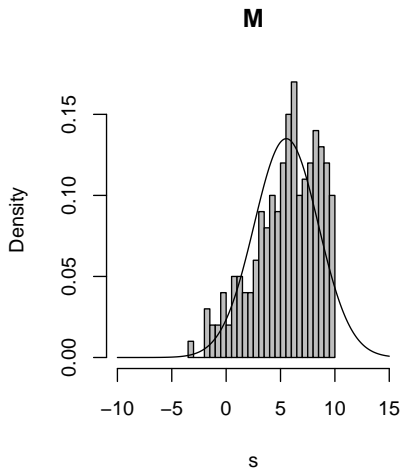


# Beispiele von Histogrammen und qqnorm-plots einiger Verteilungen

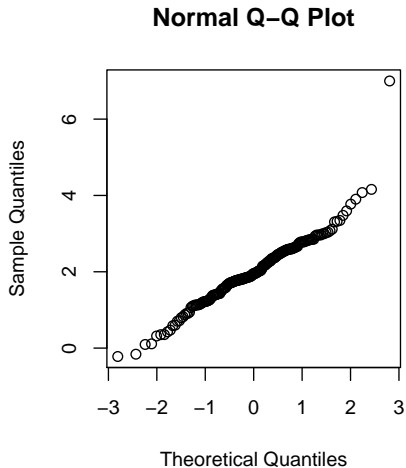
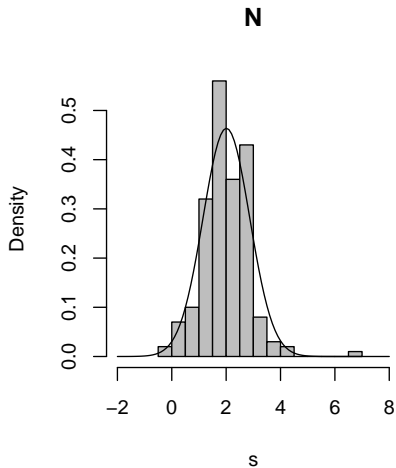




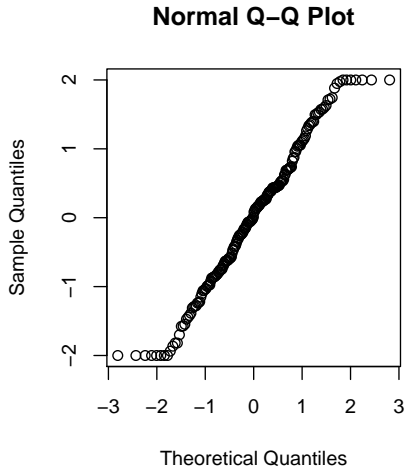
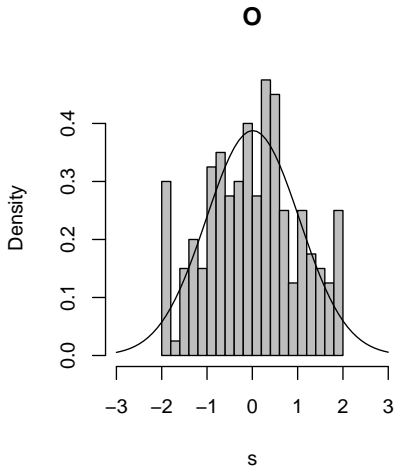
# Beispiele von Histogrammen und qqnorm-plots einiger Verteilungen



# Beispiele von Histogrammen und qqnorm-plots einiger Verteilungen



# Beispiele von Histogrammen und qqnorm-plots einiger Verteilungen

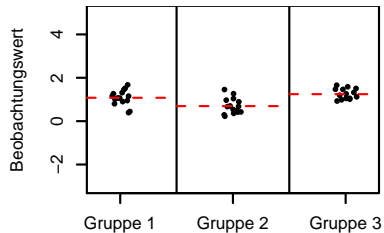
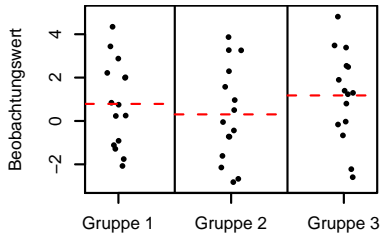


# Übersicht

- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression
  - Lineare Zusammenhänge
  - $t$ -Test fuer lineare Zusammenhänge
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse**
- 7 Versuchsplanung: Stichprobenumfang

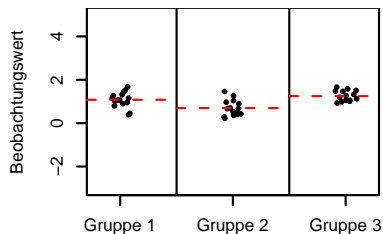
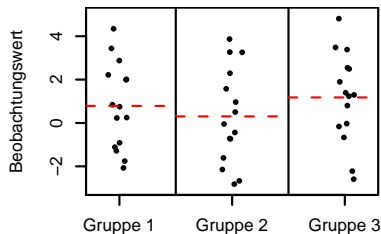
# Varianzanalyse: „Logo“

Wir beobachten unterschiedliche Gruppenmittelwerte:



# Varianzanalyse: „Logo“

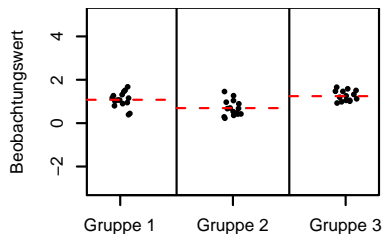
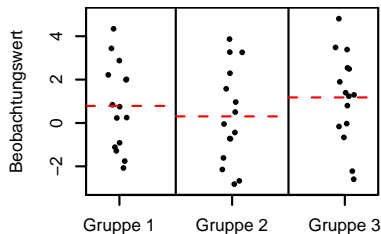
Wir beobachten unterschiedliche Gruppenmittelwerte:



Sind die beobachteten Unterschiede der Gruppenmittelwerte ernst zu nehmen — oder könnte das alles Zufall sein?

# Varianzanalyse: „Logo“

Wir beobachten unterschiedliche Gruppenmittelwerte:

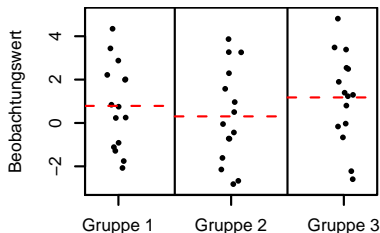


Sind die beobachteten Unterschiede der Gruppenmittelwerte ernst zu nehmen — oder könnte das alles Zufall sein?

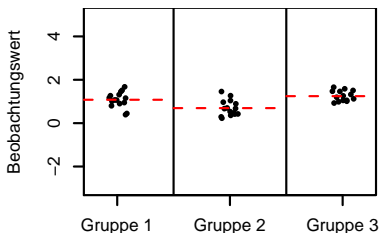
Das hängt vom Verhältnis der Variabilität der Gruppenmittelwerte und der Variabilität der Beobachtungen innerhalb der Gruppen ab:

# Varianzanalyse: „Logo“

Wir beobachten unterschiedliche Gruppenmittelwerte:



Variabilität innerhalb  
der Gruppen groß



Variabilität innerhalb  
der Gruppen klein

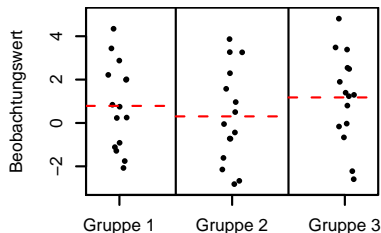
Sind die beobachteten Unterschiede der Gruppenmittelwerte ernst zu nehmen — oder könnte das alles Zufall sein?

Das hängt vom Verhältnis der Variabilität der Gruppenmittelwerte und der Variabilität der Beobachtungen innerhalb der Gruppen ab:

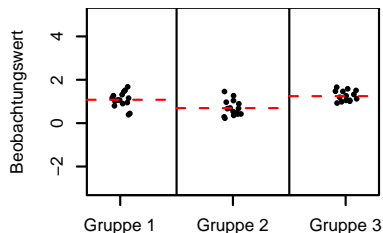


# Varianzanalyse: „Logo“

Wir beobachten unterschiedliche Gruppenmittelwerte:



Variabilität innerhalb  
der Gruppen groß



Variabilität innerhalb  
der Gruppen klein

Sind die beobachteten Unterschiede der Gruppenmittelwerte ernst zu nehmen — oder könnte das alles Zufall sein?

Das hängt vom Verhältnis der Variabilität der Gruppenmittelwerte und der Variabilität der Beobachtungen innerhalb der Gruppen ab: die Varianzanalyse gibt eine (quantitative) Antwort.

## Beispiel

Blutgerinnungszeit bei Ratten unter 4 versch. Behandlungen

Gruppe	Beobachtung							
1	62	60	63	59				
2	63	67	71	64	65	66		
3	68	66	71	67	68	68		
4	56	62	60	61	63	64	63	59

## Beispiel

Blutgerinnungszeit bei Ratten unter 4 versch. Behandlungen

Gruppe	Beobachtung
1	62 60 63 59
2	63 67 71 64 65 66
3	68 66 71 67 68 68
4	56 62 60 61 63 64 63 59

Globalmittelwert  $\bar{x}_{..} = 64$ ,

Gruppenmittelwerte  $\bar{x}_{1.} = 61$ ,  $\bar{x}_{2.} = 66$ ,  $\bar{x}_{3.} = 68$ ,  $\bar{x}_{4.} = 61$ .

# Beispiel

## Blutgerinnungszeit bei Ratten unter 4 versch. Behandlungen

Gr.	$\bar{x}_j$	Beobachtung							
1	61	62	60	63	59				
		$(62 - 61)^2$	$(60 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$				
2	66	63	67	71	64	65	66		
		$(63 - 66)^2$	$(67 - 66)^2$	$(71 - 66)^2$	$(64 - 66)^2$	$(65 - 66)^2$	$(66 - 66)^2$		
3	68	68	66	71	67	68	68		
		$(68 - 68)^2$	$(66 - 68)^2$	$(71 - 68)^2$	$(67 - 68)^2$	$(68 - 68)^2$	$(68 - 68)^2$		
4	61	56	62	60	61	63	64	63	59
		$(56 - 61)^2$	$(62 - 61)^2$	$(60 - 61)^2$	$(61 - 61)^2$	$(63 - 61)^2$	$(64 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$

Globalmittelwert  $\bar{x}_{..} = 64$ ,

Gruppenmittelwerte  $\bar{x}_1 = 61$ ,  $\bar{x}_2 = 66$ ,  $\bar{x}_3 = 68$ ,  $\bar{x}_4 = 61$ .

# Beispiel

## Blutgerinnungszeit bei Ratten unter 4 versch. Behandlungen

Gr.	$\bar{x}_j$	Beobachtung							
1	61	62	60	63	59				
		$(62 - 61)^2$	$(60 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$				
2	66	63	67	71	64	65	66		
		$(63 - 66)^2$	$(67 - 66)^2$	$(71 - 66)^2$	$(64 - 66)^2$	$(65 - 66)^2$	$(66 - 66)^2$		
3	68	68	66	71	67	68	68		
		$(68 - 68)^2$	$(66 - 68)^2$	$(71 - 68)^2$	$(67 - 68)^2$	$(68 - 68)^2$	$(68 - 68)^2$		
4	61	56	62	60	61	63	64	63	59
		$(56 - 61)^2$	$(62 - 61)^2$	$(60 - 61)^2$	$(61 - 61)^2$	$(63 - 61)^2$	$(64 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$

Globalmittelwert  $\bar{x}_{..} = 64$ ,

Gruppenmittelwerte  $\bar{x}_1 = 61$ ,  $\bar{x}_2 = 66$ ,  $\bar{x}_3 = 68$ ,  $\bar{x}_4 = 61$ .

Die roten Werte (ohne die Quadrate) heißen **Residuen**: die „Restvariabilität“ der Beobachtungen, die das Modell nicht erklärt.

# Beispiel

## Blutgerinnungszeit bei Ratten unter 4 versch. Behandlungen

Gr.	$\bar{x}_j$	Beobachtung							
1	61	62	60	63	59				
		$(62 - 61)^2$	$(60 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$				
2	66	63	67	71	64	65	66		
		$(63 - 66)^2$	$(67 - 66)^2$	$(71 - 66)^2$	$(64 - 66)^2$	$(65 - 66)^2$	$(66 - 66)^2$		
3	68	68	66	71	67	68	68		
		$(68 - 68)^2$	$(66 - 68)^2$	$(71 - 68)^2$	$(67 - 68)^2$	$(68 - 68)^2$	$(68 - 68)^2$		
4	61	56	62	60	61	63	64	63	59
		$(56 - 61)^2$	$(62 - 61)^2$	$(60 - 61)^2$	$(61 - 61)^2$	$(63 - 61)^2$	$(64 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$

Globalmittelwert  $\bar{x}_{..} = 64$ ,

Gruppenmittelwerte  $\bar{x}_1 = 61$ ,  $\bar{x}_2 = 66$ ,  $\bar{x}_3 = 68$ ,  $\bar{x}_4 = 61$ .

Die roten Werte (ohne die Quadrate) heißen Residuen: die „Restvariabilität“ der Beobachtungen, die das Modell nicht erklärt.

Quadratsumme innerhalb der Gruppen:

$$SS_{\text{innerh}} = 112,$$

# Beispiel

## Blutgerinnungszeit bei Ratten unter 4 versch. Behandlungen

Gr.	$\bar{x}_j$	Beobachtung							
1	61	62	60	63	59				
		$(62 - 61)^2$	$(60 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$				
2	66	63	67	71	64	65	66		
		$(63 - 66)^2$	$(67 - 66)^2$	$(71 - 66)^2$	$(64 - 66)^2$	$(65 - 66)^2$	$(66 - 66)^2$		
3	68	68	66	71	67	68	68		
		$(68 - 68)^2$	$(66 - 68)^2$	$(71 - 68)^2$	$(67 - 68)^2$	$(68 - 68)^2$	$(68 - 68)^2$		
4	61	56	62	60	61	63	64	63	59
		$(56 - 61)^2$	$(62 - 61)^2$	$(60 - 61)^2$	$(61 - 61)^2$	$(63 - 61)^2$	$(64 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$

Globalmittelwert  $\bar{x}_{..} = 64$ ,

Gruppenmittelwerte  $\bar{x}_1 = 61$ ,  $\bar{x}_2 = 66$ ,  $\bar{x}_3 = 68$ ,  $\bar{x}_4 = 61$ .

Die roten Werte (ohne die Quadrate) heißen Residuen: die „Restvariabilität“ der Beobachtungen, die das Modell nicht erklärt.

Quadratsumme innerhalb der Gruppen:

$ss_{\text{innerh}} = 112$ , 20 Freiheitsgrade

# Beispiel

## Blutgerinnungszeit bei Ratten unter 4 versch. Behandlungen

Gr.	$\bar{x}_j$	Beobachtung							
1	61	62	60	63	59				
		$(62 - 61)^2$	$(60 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$				
2	66	63	67	71	64	65	66		
		$(63 - 66)^2$	$(67 - 66)^2$	$(71 - 66)^2$	$(64 - 66)^2$	$(65 - 66)^2$	$(66 - 66)^2$		
3	68	68	66	71	67	68	68		
		$(68 - 68)^2$	$(66 - 68)^2$	$(71 - 68)^2$	$(67 - 68)^2$	$(68 - 68)^2$	$(68 - 68)^2$		
4	61	56	62	60	61	63	64	63	59
		$(56 - 61)^2$	$(62 - 61)^2$	$(60 - 61)^2$	$(61 - 61)^2$	$(63 - 61)^2$	$(64 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$

Globalmittelwert  $\bar{x}_{..} = 64$ ,

Gruppenmittelwerte  $\bar{x}_1 = 61$ ,  $\bar{x}_2 = 66$ ,  $\bar{x}_3 = 68$ ,  $\bar{x}_4 = 61$ .

Die roten Werte (ohne die Quadrate) heißen Residuen: die „Restvariabilität“ der Beobachtungen, die das Modell nicht erklärt.

Quadratsumme innerhalb der Gruppen:

$$ss_{\text{innerh}} = 112, 20 \text{ Freiheitsgrade}$$

Quadratsumme zwischen den Gruppen:

$$ss_{\text{zw}} = 4 \cdot (61 - 64)^2 + 6 \cdot (66 - 64)^2 + 6 \cdot (68 - 64)^2 + 8 \cdot (61 - 64)^2 = 228,$$



# Beispiel

## Blutgerinnungszeit bei Ratten unter 4 versch. Behandlungen

Gr.	$\bar{x}_j$	Beobachtung							
1	61	62	60	63	59				
		$(62 - 61)^2$	$(60 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$				
2	66	63	67	71	64	65	66		
		$(63 - 66)^2$	$(67 - 66)^2$	$(71 - 66)^2$	$(64 - 66)^2$	$(65 - 66)^2$	$(66 - 66)^2$		
3	68	68	66	71	67	68	68		
		$(68 - 68)^2$	$(66 - 68)^2$	$(71 - 68)^2$	$(67 - 68)^2$	$(68 - 68)^2$	$(68 - 68)^2$		
4	61	56	62	60	61	63	64	63	59
		$(56 - 61)^2$	$(62 - 61)^2$	$(60 - 61)^2$	$(61 - 61)^2$	$(63 - 61)^2$	$(64 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$

Globalmittelwert  $\bar{x}_{..} = 64$ ,

Gruppenmittelwerte  $\bar{x}_1 = 61$ ,  $\bar{x}_2 = 66$ ,  $\bar{x}_3 = 68$ ,  $\bar{x}_4 = 61$ .

Die roten Werte (ohne die Quadrate) heißen Residuen: die „Restvariabilität“ der Beobachtungen, die das Modell nicht erklärt.

Quadratsumme innerhalb der Gruppen:

$$ss_{\text{innerh}} = 112, 20 \text{ Freiheitsgrade}$$

Quadratsumme zwischen den Gruppen:

$$ss_{\text{zw}} = 4 \cdot (61 - 64)^2 + 6 \cdot (66 - 64)^2 + 6 \cdot (68 - 64)^2 + 8 \cdot (61 - 64)^2 = 228,$$

3 Freiheitsgrade

# Beispiel

## Blutgerinnungszeit bei Ratten unter 4 versch. Behandlungen

Gr.	$\bar{x}_j$	Beobachtung							
1	61	62	60	63	59				
		$(62 - 61)^2$	$(60 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$				
2	66	63	67	71	64	65	66		
		$(63 - 66)^2$	$(67 - 66)^2$	$(71 - 66)^2$	$(64 - 66)^2$	$(65 - 66)^2$	$(66 - 66)^2$		
3	68	68	66	71	67	68	68		
		$(68 - 68)^2$	$(66 - 68)^2$	$(71 - 68)^2$	$(67 - 68)^2$	$(68 - 68)^2$	$(68 - 68)^2$		
4	61	56	62	60	61	63	64	63	59
		$(56 - 61)^2$	$(62 - 61)^2$	$(60 - 61)^2$	$(61 - 61)^2$	$(63 - 61)^2$	$(64 - 61)^2$	$(63 - 61)^2$	$(59 - 61)^2$

Globalmittelwert  $\bar{x}_{..} = 64$ ,

Gruppenmittelwerte  $\bar{x}_1 = 61$ ,  $\bar{x}_2 = 66$ ,  $\bar{x}_3 = 68$ ,  $\bar{x}_4 = 61$ .

Die roten Werte (ohne die Quadrate) heißen Residuen: die „Restvariabilität“ der Beobachtungen, die das Modell nicht erklärt.

Quadratsumme innerhalb der Gruppen:

$$ss_{\text{innerh}} = 112, 20 \text{ Freiheitsgrade}$$

Quadratsumme zwischen den Gruppen:

$$ss_{\text{zw}} = 4 \cdot (61 - 64)^2 + 6 \cdot (66 - 64)^2 + 6 \cdot (68 - 64)^2 + 8 \cdot (61 - 64)^2 = 228,$$

3 Freiheitsgrade

$$F = \frac{ss_{\text{zw}}/3}{ss_{\text{innerh}}/20} = \frac{76}{5,6} = 13,57$$

## Beispiel: Blutgerinnungszeit bei Ratten unter 4 versch. Behandlungen

## ANOVA-Tafel („ANalysis Of VAriance“)

	Freiheits- grade (DF)	Quadrat- summe (SS)	mittlere summe (SS/DF)	Quadrat- summe (SS/DF)	<i>F</i> -Wert
Gruppe	3	228		76	13,57
Residuen	20	112		5,6	

## Beispiel: Blutgerinnungszeit bei Ratten unter 4 versch. Behandlungen

## ANOVA-Tafel („ANalysis Of VAriance“)

	Freiheits- grade (DF)	Quadrat- summe (SS)	mittlere Quadrat- summe (SS/DF)	<i>F</i> -Wert
Gruppe	3	228	76	13,57
Residuen	20	112	5,6	

Unter der Hypothese  $H_0$  „die Gruppenmittelwerte sind gleich“  
(und einer Normalverteilungsannahme an die Beobachtungen)

ist  $F$  Fisher-verteilt mit 3 und 20 Freiheitsgraden,

$$p = \text{Fisher}_{3,20}([13,57, \infty)) \leq 5 \cdot 10^{-5}.$$

## Beispiel: Blutgerinnungszeit bei Ratten unter 4 versch. Behandlungen

## ANOVA-Tafel („ANalysis Of VAriance“)

	Freiheits- grade (DF)	Quadrat- summe (SS)	mittlere Quadrat- summe (SS/DF)	<i>F</i> -Wert
Gruppe	3	228	76	13,57
Residuen	20	112	5,6	

Unter der Hypothese  $H_0$  „die Gruppenmittelwerte sind gleich“  
(und einer Normalverteilungsannahme an die Beobachtungen)

ist  $F$  Fisher-verteilt mit 3 und 20 Freiheitsgraden,

$$p = \text{Fisher}_{3,20}([13,57, \infty)) \leq 5 \cdot 10^{-5}.$$

Wir lehnen demnach  $H_0$  ab.

95%-Quantil der Fisher-Verteilung in Abhängigkeit der Anzahl Freiheitsgrade  
 ( $k_1$  Zähler-,  $k_2$  Nennerfreiheitsgrade)

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	24
1	161.45	199.5	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.8
2	18.51	19	19.16	19.25	19.3	19.33	19.35	19.37	19.38	19.4	19.4
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6	5.96	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.1	4.06	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.5	3.44	3.39	3.35	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.14
10	4.96	4.1	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.98
11	4.84	3.98	3.59	3.36	3.2	3.09	3.01	2.95	2.9	2.85	2.85
12	4.75	3.89	3.49	3.26	3.11	3	2.91	2.85	2.8	2.75	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.67
14	4.6	3.74	3.34	3.11	2.96	2.85	2.76	2.7	2.65	2.6	2.6
15	4.54	3.68	3.29	3.06	2.9	2.79	2.71	2.64	2.59	2.54	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.49
17	4.45	3.59	3.2	2.96	2.81	2.7	2.61	2.55	2.49	2.45	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.41
19	4.38	3.52	3.13	2.9	2.74	2.63	2.54	2.48	2.42	2.38	2.38
20	4.35	3.49	3.1	2.87	2.71	2.6	2.51	2.45	2.39	2.35	2.35

# F-Test

$n = n_1 + n_2 + \dots + n_l$  Beobachtungen in  $l$  Gruppen,  
 $X_{ij} = j$ -te Beobachtung in der  $i$ -ten Gruppe,  $j = 1, \dots, n_i$ .

# F-Test

$n = n_1 + n_2 + \dots + n_l$  Beobachtungen in  $l$  Gruppen,  
 $X_{ij} = j$ -te Beobachtung in der  $i$ -ten Gruppe,  $j = 1, \dots, n_i$ .

Modellannahme:  $X_{ij} = \mu_i + \varepsilon_{ij}$ ,

mit unabhängigen, normalverteilten  $\varepsilon_{ij}$ ,  $\mathbb{E}[\varepsilon_{ij}] = 0$ ,  $\text{Var}[\varepsilon_{ij}] = \sigma^2$



# F-Test

$n = n_1 + n_2 + \dots + n_l$  Beobachtungen in  $l$  Gruppen,  
 $X_{ij} = j$ -te Beobachtung in der  $i$ -ten Gruppe,  $j = 1, \dots, n_i$ .

Modellannahme:  $X_{ij} = \mu_i + \varepsilon_{ij}$ ,

mit unabhängigen, normalverteilten  $\varepsilon_{ij}$ ,  $\mathbb{E}[\varepsilon_{ij}] = 0$ ,  $\text{Var}[\varepsilon_{ij}] = \sigma^2$

( $\mu_i$  ist der „wahre“ Mittelwert innerhalb der  $i$ -ten Gruppe.)

# F-Test

$n = n_1 + n_2 + \dots + n_l$  Beobachtungen in  $l$  Gruppen,

$X_{ij}$  =  $j$ -te Beobachtung in der  $i$ -ten Gruppe,  $j = 1, \dots, n_i$ .

Modellannahme:  $X_{ij} = \mu_i + \varepsilon_{ij}$ ,

mit unabhängigen, normalverteilten  $\varepsilon_{ij}$ ,  $\mathbb{E}[\varepsilon_{ij}] = 0$ ,  $\text{Var}[\varepsilon_{ij}] = \sigma^2$

( $\mu_i$  ist der „wahre“ Mittelwert innerhalb der  $i$ -ten Gruppe.)

$\bar{X}_{..} = \frac{1}{n} \sum_{i=1}^l \sum_{j=1}^{n_i} X_{ij}$  (empirisches) „Globalmittel“

$\bar{X}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$  (empirischer) Mittelwert der  $i$ -ten Gruppe

## F-Test

$n = n_1 + n_2 + \dots + n_l$  Beobachtungen in  $l$  Gruppen,  
 $X_{ij}$  =  $j$ -te Beobachtung in der  $i$ -ten Gruppe,  $j = 1, \dots, n_i$ .

Modellannahme:  $X_{ij} = \mu_i + \varepsilon_{ij}$ ,

mit unabhängigen, normalverteilten  $\varepsilon_{ij}$ ,  $\mathbb{E}[\varepsilon_{ij}] = 0$ ,  $\text{Var}[\varepsilon_{ij}] = \sigma^2$   
 ( $\mu_i$  ist der „wahre“ Mittelwert innerhalb der  $i$ -ten Gruppe.)

$\bar{X}_{..} = \frac{1}{n} \sum_{i=1}^l \sum_{j=1}^{n_i} X_{ij}$  (empirisches) „Globalmittel“

$\bar{X}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$  (empirischer) Mittelwert der  $i$ -ten Gruppe

$SS_{\text{innerh}} = \sum_{i=1}^l \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2$  Quadratsumme innerhalb d. Gruppen,  
 $n - l$  Freiheitsgrade

$SS_{\text{zw}} = \sum_{i=1}^l n_i (\bar{X}_{i.} - \bar{X}_{..})^2$  Quadratsumme zwischen d. Gruppen,  
 $l - 1$  Freiheitsgrade

## F-Test

$n = n_1 + n_2 + \dots + n_l$  Beobachtungen in  $l$  Gruppen,  
 $X_{ij}$  =  $j$ -te Beobachtung in der  $i$ -ten Gruppe,  $j = 1, \dots, n_i$ .

Modellannahme:  $X_{ij} = \mu_i + \varepsilon_{ij}$ ,

mit unabhängigen, normalverteilten  $\varepsilon_{ij}$ ,  $\mathbb{E}[\varepsilon_{ij}] = 0$ ,  $\text{Var}[\varepsilon_{ij}] = \sigma^2$   
 ( $\mu_i$  ist der „wahre“ Mittelwert innerhalb der  $i$ -ten Gruppe.)

$\bar{X}_{..} = \frac{1}{n} \sum_{i=1}^l \sum_{j=1}^{n_i} X_{ij}$  (empirisches) „Globalmittel“

$\bar{X}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$  (empirischer) Mittelwert der  $i$ -ten Gruppe

$SS_{\text{innerh}} = \sum_{i=1}^l \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2$  Quadratsumme innerhalb d. Gruppen,  
 $n - l$  Freiheitsgrade

$SS_{\text{zw}} = \sum_{i=1}^l n_i (\bar{X}_{i.} - \bar{X}_{..})^2$  Quadratsumme zwischen d. Gruppen,  
 $l - 1$  Freiheitsgrade

$$F = \frac{SS_{\text{zw}} / (l - 1)}{SS_{\text{innerh}} / (n - l)}$$

## F-Test

$X_{ij}$  =  $j$ -te Beobachtung in der  $i$ -ten Gruppe,  $j = 1, \dots, n_i$ ,

Modellannahme:  $X_{ij} = \mu_i + \varepsilon_{ij}$ .  $\mathbb{E}[\varepsilon_{ij}] = 0$ ,  $\text{Var}[\varepsilon_{ij}] = \sigma^2$

$SS_{\text{innerh}} = \sum_{i=1}^I \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i\cdot})^2$  Quadratsumme innerhalb d. Gruppen,  
 $n - I$  Freiheitsgrade

$SS_{\text{zw}} = \sum_{i=1}^I n_i (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$  Quadratsumme zwischen d. Gruppen,  
 $I - 1$  Freiheitsgrade

$$F = \frac{SS_{\text{zw}} / (I - 1)}{SS_{\text{innerh}} / (n - I)}$$

Unter der Hypothese  $H_0 : \mu_1 = \dots = \mu_I$  („alle  $\mu_i$  sind gleich“) ist  $F$  Fisher-verteilt mit  $I - 1$  und  $n - I$  Freiheitsgraden (unabhängig vom tatsächlichen gemeinsamen Wert der  $\mu_i$ ).

$F$ -Test: Wir lehnen  $H_0$  zum Signifikanzniveau  $\alpha$  ab, wenn  $F \geq q_{1-\alpha}$ , wobei  $q_{1-\alpha}$  das  $(1 - \alpha)$ -Quantil der Fisher-Verteilung mit  $I - 1$  und  $n - I$  Freiheitsgraden ist.

# Übersicht

- 1 Deskriptive Statistik
- 2 Standardfehler und t-Tests
- 3 Chi-Quadrat-Tests
  - $\chi^2$ -Test für eine feste Verteilung
  - $\chi^2$ -Test auf Unabhängigkeit (oder Homogenität)
  - $\chi^2$ -Test für Modelle mit Parametern
- 4 Konfidenzintervalle
- 5 Lineare Regression
  - Lineare Zusammenhänge
  - $t$ -Test fuer lineare Zusammenhänge
  - Überprüfen der Annahmen und Modellwahl
- 6 Varianzanalyse
- 7 **Versuchsplanung: Stichprobenumfang**

Je größer die Stichprobenlänge ist,

- desto wahrscheinlicher wird ein vorhandener Unterschied durch einen statistischen Test angezeigt

Je größer die Stichprobenlänge ist,

- desto wahrscheinlicher wird ein vorhandener Unterschied durch einen statistischen Test angezeigt
- desto kleinere Unterschiede können durch statistische Tests erkannt werden



Je größer die Stichprobenlänge ist,

- desto wahrscheinlicher wird ein vorhandener Unterschied durch einen statistischen Test angezeigt
- desto kleinere Unterschiede können durch statistische Tests erkannt werden
- desto teurer wird der Versuch.

Je größer die Stichprobenlänge ist,

- desto wahrscheinlicher wird ein vorhandener Unterschied durch einen statistischen Test angezeigt
- desto kleinere Unterschiede können durch statistische Tests erkannt werden
- desto teurer wird der Versuch.

Je größer die Stichprobenlänge ist,

- desto wahrscheinlicher wird ein vorhandener Unterschied durch einen statistischen Test angezeigt
- desto kleinere Unterschiede können durch statistische Tests erkannt werden
- desto teurer wird der Versuch.

Es ist also wichtig, eine geeignete Stichprobenlänge zu wählen. Dazu muss man sich überlegen,

- welcher Unterschied durch die anzuwendenden Tests erkannt werden soll,

Je größer die Stichprobenlänge ist,

- desto wahrscheinlicher wird ein vorhandener Unterschied durch einen statistischen Test angezeigt
- desto kleinere Unterschiede können durch statistische Tests erkannt werden
- desto teurer wird der Versuch.

Es ist also wichtig, eine geeignete Stichprobenlänge zu wählen. Dazu muss man sich überlegen,

- welcher Unterschied durch die anzuwendenden Tests erkannt werden soll,
- wie groß die Variabilität in den Daten in etwa sein wird.

# Einstichproben-Tests

**Frage:** Ist der wahre Mittelwert gleich  $\mu_0$ ?

# Einstichproben-Tests

**Frage:** Ist der wahre Mittelwert gleich  $\mu_0$ ?

**Beispiel:** Kältestress-Toleranz bei Fruchtfliegen.



photo (c) André Karwath (Bild zeigt eine *Drosophila melanogaster*)

# Einstichproben-Tests

Die Chill-Coma Recovery Time (CCRT) ist die Zeit in Sekunden, nach der die Fliege nach einem Kältekoma wieder aufwacht. In früheren Versuchen wurde bei *Drosophila ananassae* aus Bangkok eine mittlere CCRT von 46 gemessen.

# Einstichproben-Tests

Die Chill-Coma Recovery Time (CCRT) ist die Zeit in Sekunden, nach der die Fliege nach einem Kältekoma wieder aufwacht. In früheren Versuchen wurde bei *Drosophila ananassae* aus Bangkok eine mittlere CCRT von 46 gemessen.

**Frage:** Ist die CCRT bei *Drosophila ananassae* aus Kathmandu (Nepal) verschieden von 46?

**Geplanter Test:**



# Einstichproben-Tests

Die Chill-Coma Recovery Time (CCRT) ist die Zeit in Sekunden, nach der die Fliege nach einem Kältekoma wieder aufwacht. In früheren Versuchen wurde bei *Drosophila ananassae* aus Bangkok eine mittlere CCRT von 46 gemessen.

**Frage:** Ist die CCRT bei *Drosophila ananassae* aus Kathmandu (Nepal) verschieden von 46?

**Geplanter Test:** (zweiseitiger) Einstichproben t-Test.

**Ziel:** Finde Unterschiede, die größer als  $d = 4$  sind. Signifikanzniveau  $\alpha = 5\%$ . Testmacht  $1 - \beta = 80\%$ .

# Einstichproben-Tests

Die Chill-Coma Recovery Time (CCRT) ist die Zeit in Sekunden, nach der die Fliege nach einem Kältekoma wieder aufwacht. In früheren Versuchen wurde bei *Drosophila ananassae* aus Bangkok eine mittlere CCRT von 46 gemessen.

**Frage:** Ist die CCRT bei *Drosophila ananassae* aus Kathmandu (Nepal) verschieden von 46?

**Geplanter Test:** (zweiseitiger) Einstichproben t-Test.

**Ziel:** Finde Unterschiede, die größer als  $d = 4$  sind.  
Signifikanzniveau  $\alpha = 5\%$ . Testmacht  $1 - \beta = 80\%$ .

**Vorwissen:** Standardabweichung bei Vortest war  $s = 11.9$

**Frage:** Bei wie vielen Fliegen muss ich die CCRT messen, um das Ziel zu erreichen?

Man benötigt:

- $d$  = Unterschied, den man mindestens erkennen können möchte. (engl: detection level)

Man benötigt:

- $d$  = Unterschied, den man mindestens erkennen können möchte. (engl: detection level)
- einen ungefähren Wert  $s$  für die Standardabweichung, die man in den Daten erwartet (oft ein Wert aus Vorversuchen).

Man benötigt:

- $d$  = Unterschied, den man mindestens erkennen können möchte. (engl: detection level)
- einen ungefähren Wert  $s$  für die Standardabweichung, die man in den Daten erwartet (oft ein Wert aus Vorversuchen).
- $\alpha = \Pr_{H_0}(H_0 \text{ wird (fälschlicherweise) verworfen})$ . Meist 5%.  
 $\alpha$  ist das Signifikanzniveau. Die Ws  $\alpha$  heißt auch Fehler 1.Art.

Man benötigt:

- $d$  = Unterschied, den man mindestens erkennen können möchte. (engl: detection level)
- einen ungefähren Wert  $s$  für die Standardabweichung, die man in den Daten erwartet (oft ein Wert aus Vorversuchen).
- $\alpha = \Pr_{H_0}(H_0 \text{ wird (fälschlicherweise) verworfen})$ . Meist 5%.  $\alpha$  ist das Signifikanzniveau. Die Ws  $\alpha$  heißt auch Fehler 1.Art.
- $\beta = \Pr_{\text{Alternative}}(H_0 \text{ wird (fälschlicherweise) nicht verworfen})$ . Der Index **Alternative** bedeutet hier, dass wir von einem tatsächlichen Unterschied  $\geq d$  ausgehen. Die Wahl von  $\beta$  hängt stark vom Problem ab.  $1 - \beta$  ist die Testmacht (engl. *power*). Die Ws  $\beta$  heißt auch Fehler 2.Art.

# Einstichproben-Tests

**Frage:** Stichprobenlänge für CCRT-Versuch?

**Lösung:** Es soll gelten:

$$\Pr_{\mu_0+d} \left( \frac{|X - \mu_0|}{s/\sqrt{n}} \leq t_{1-\alpha/2, n-1} \right) \leq \beta$$

wobei  $t_{1-\frac{\alpha}{2}, n-1} \leftarrow \text{qt}(1-\alpha/2, n-1)$  das  $(1 - \alpha/2)$ -Quantil ist.

# Einstichproben-Tests

**Frage:** Stichprobenlänge für CCRT-Versuch?

**Lösung:** Es soll gelten:

$$\Pr_{\mu_0+d} \left( \frac{|X - \mu_0|}{s/\sqrt{n}} \leq t_{1-\alpha/2, n-1} \right) \leq \beta$$

wobei  $t_{1-\frac{\alpha}{2}, n-1} \leftarrow \text{qt}(1-\alpha/2, n-1)$  das  $(1 - \alpha/2)$ -Quantil ist.

Unter den Verteilungsannahmen des  $t$ -Tests ist die Ungleichung erfüllt, falls

$$n \geq \frac{s^2 \cdot (t_{1-\frac{\alpha}{2}, n-1} + t_{1-\beta, n-1})^2}{d^2}.$$



# Einstichproben-Tests

**Frage:** Stichprobenlänge für CCRT-Versuch?

**Lösung:** Es soll gelten:

$$\Pr_{\mu_0+d} \left( \frac{|X - \mu_0|}{s/\sqrt{n}} \leq t_{1-\alpha/2, n-1} \right) \leq \beta$$

wobei  $t_{1-\frac{\alpha}{2}, n-1} \leftarrow \text{qt}(1-\alpha/2, n-1)$  das  $(1 - \alpha/2)$ -Quantil ist.

Unter den Verteilungsannahmen des  $t$ -Tests ist die Ungleichung erfüllt, falls

$$n \geq \frac{s^2 \cdot (t_{1-\frac{\alpha}{2}, n-1} + t_{1-\beta, n-1})^2}{d^2}.$$

Leider kann man nicht einfach einsetzen, da die rechte Seite von  $n$  abhängt.

# Einstichproben-Tests

**Frage:** Stichprobenlänge für CCRT-Versuch?

**Lösung:** Es soll gelten:

$$\Pr_{\mu_0+d} \left( \frac{|X - \mu_0|}{s/\sqrt{n}} \leq t_{1-\alpha/2, n-1} \right) \leq \beta$$

wobei  $t_{1-\frac{\alpha}{2}, n-1} \leftarrow \text{qt}(1-\alpha/2, n-1)$  das  $(1 - \alpha/2)$ -Quantil ist.

Unter den Verteilungsannahmen des  $t$ -Tests ist die Ungleichung erfüllt, falls

$$n \geq \frac{s^2 \cdot (t_{1-\frac{\alpha}{2}, n-1} + t_{1-\beta, n-1})^2}{d^2}.$$

Leider kann man nicht einfach einsetzen, da die rechte Seite von  $n$  abhängt.

Entweder probiert man herum und sucht das kleinste  $n$  wofür die Ungleichung gilt.

# Einstichproben-Tests

Oder man beginnt mit

$$n_0 = \frac{s^2 \cdot (z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2}{d^2}$$

wobei  $z_{1-\frac{\alpha}{2}} \leftarrow \text{qnorm}(1-\alpha/2)$  das  $(1 - \alpha/2)$ -Quantil  
und  $z_{1-\beta} \leftarrow \text{qnorm}(1-\beta)$  das  $(1 - \beta)$ -Quantil der  
Normalverteilung ist.

# Einstichproben-Tests

Oder man beginnt mit

$$n_0 = \frac{s^2 \cdot (z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2}{d^2}$$

wobei  $z_{1-\frac{\alpha}{2}} \leftarrow \text{qnorm}(1-\alpha/2)$  das  $(1 - \alpha/2)$ -Quantil und  $z_{1-\beta} \leftarrow \text{qnorm}(1-\beta)$  das  $(1 - \beta)$ -Quantil der Normalverteilung ist.

Die benötigte Stichprobenlänge findet man dann durch Iteration:

$$n_1 = \frac{s^2 \cdot (t_{1-\frac{\alpha}{2}, n_0-1} + t_{1-\beta, n_0-1})^2}{d^2}$$

$$n_2 = \frac{s^2 \cdot (t_{1-\frac{\alpha}{2}, n_1-1} + t_{1-\beta, n_1-1})^2}{d^2}$$

usw bis sich nichts mehr ändert.

# Einstichproben-Tests

Zurück zum Beispiel:

$$n_0 = \frac{s^2 \cdot (z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2}{d^2} = \frac{11.9^2 (z_{0.975} + z_{0.8})^2}{4^2} = 69.48 \sim 70$$

# Einstichproben-Tests

Zurück zum Beispiel:

$$n_0 = \frac{s^2 \cdot (z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2}{d^2} = \frac{11.9^2 (z_{0.975} + z_{0.8})^2}{4^2} = 69.48 \sim 70$$

$$n_1 = \frac{s^2 \cdot (t_{1-\frac{\alpha}{2}, n_0-1} + t_{1-\beta, n_0-1})^2}{d^2} = \frac{11.9^2 (t_{0.975, 69} + t_{0.8, 69})^2}{4^2} \\ = 71.47 \sim 72$$

# Einstichproben-Tests

Zurück zum Beispiel:

$$n_0 = \frac{s^2 \cdot (z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2}{d^2} = \frac{11.9^2 (z_{0.975} + z_{0.8})^2}{4^2} = 69.48 \sim 70$$

$$n_1 = \frac{s^2 \cdot (t_{1-\frac{\alpha}{2}, n_0-1} + t_{1-\beta, n_0-1})^2}{d^2} = \frac{11.9^2 (t_{0.975, 69} + t_{0.8, 69})^2}{4^2}$$

$$= 71.47 \sim 72$$

$$n_2 = \frac{s^2 \cdot (t_{1-\frac{\alpha}{2}, n_1-1} + t_{1-\beta, n_1-1})^2}{d^2} = \frac{11.9^2 (t_{0.975, 71} + t_{0.8, 71})^2}{4^2}$$

$$= 71.41 \sim 72$$

# Einstichproben-Tests

Zurück zum Beispiel:

$$n_0 = \frac{s^2 \cdot (z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2}{d^2} = \frac{11.9^2 (z_{0.975} + z_{0.8})^2}{4^2} = 69.48 \sim 70$$

$$n_1 = \frac{s^2 \cdot (t_{1-\frac{\alpha}{2}, n_0-1} + t_{1-\beta, n_0-1})^2}{d^2} = \frac{11.9^2 (t_{0.975, 69} + t_{0.8, 69})^2}{4^2}$$

$$= 71.47 \sim 72$$

$$n_2 = \frac{s^2 \cdot (t_{1-\frac{\alpha}{2}, n_1-1} + t_{1-\beta, n_1-1})^2}{d^2} = \frac{11.9^2 (t_{0.975, 71} + t_{0.8, 71})^2}{4^2}$$

$$= 71.41 \sim 72$$

**Antwort:** Die Stichprobenlänge für den CCRT-Versuch sollte mindestens  $n \geq 72$  sein.



# Einstichproben-Tests

**Bemerkung:** Bei einer Testmacht von 80% erhält man in ca. 80% der Fälle (also in 4 von 5 Fällen) keine Signifikanz, obwohl sich die wahren Mittelwerte um  $d$  unterscheiden.

# Einstichproben-Tests

**Bemerkung:** Bei einer Testmacht von 80% erhält man in ca. 80% der Fälle (also in 4 von 5 Fällen) keine Signifikanz, obwohl sich die wahren Mittelwerte um  $d$  unterscheiden. Wenn man den Versuch 5 mal durchführt, so erhält man im Schnitt nur 4 mal Signifikanz selbst wenn der wahre Unterschied in etwa  $d$  ist.

# Zweiseitiger Einstichproben t-Test

**Geplanter Test:** Zweiseitiger Einstichproben t-Test.

**Ziel:** Finde Unterschiede, die größer als  $d$  sind.  
Signifikanzniveau  $\alpha$ . Testmacht  $1 - \beta$ .

**Vorwissen:** Standardabweichung bei Vortest war  $s$

**Lösung:** Es soll gelten:

$$n \geq \frac{s^2 \cdot (t_{1-\frac{\alpha}{2}, n-1} + t_{1-\beta, n-1})^2}{d^2}$$

# Zweiseitiger ungepaarter Zweistichproben t-Test

**Geplanter Test:** Zweiseitiger ungepaarter Zweistichproben t-Test.

**Ziel:** Finde Unterschiede, die größer als  $d$  sind.  
Signifikanzniveau  $\alpha$ . Testmacht  $1 - \beta$ .

**Vorwissen:** Die Standardabweichungen in den beiden Stichproben sind in etwa  $s_1$  beziehungsweise  $s_2$ .

**Lösung:** In jeder Gruppe muss die Stichprobenlänge mindestens

$$n \geq \frac{(s_1^2 + s_2^2) \cdot (t_{1-\frac{\alpha}{2}, 2*n-2} + t_{1-\beta, 2*n-2})^2}{d^2}$$

sein.

# Zweiseitiger gepaarter Zweistichproben t-Test

**Geplanter Test:** Zweiseitiger gepaarter Zweistichproben t-Test.

**Ziel:** Finde Unterschiede, die größer als  $d$  sind.  
Signifikanzniveau  $\alpha$ . Testmacht  $1 - \beta$ .

**Vorwissen:** Standardabweichung der Differenz der beiden Stichproben ist in etwa  $s_d$ .

**Lösung:** In jeder Gruppe muss die Stichprobenlänge mindestens

$$n \geq \frac{s_d^2 \cdot (t_{1-\frac{\alpha}{2}, n-1} + t_{1-\beta, n-1})^2}{d^2}$$

sein.

# Einseitiger Einstichproben t-Test

**Geplanter Test:** Einseitiger Einstichproben t-Test.

**Ziel:** Finde Unterschiede, die größer als  $d$  sind.  
Signifikanzniveau  $\alpha$ . Testmacht  $1 - \beta$ .

**Vorwissen:** Standardabweichung bei Vortest war  $s$

**Lösung:** Es soll gelten:

$$n \geq \frac{s^2 \cdot (t_{1-\alpha, n-1} + t_{1-\beta, n-1})^2}{d^2}$$

# Einseitiger ungepaarter Zweistichproben t-Test

**Geplanter Test:** Einseitiger ungepaarter Zweistichproben t-Test.

**Ziel:** Finde Unterschiede, die größer als  $d$  sind.

Signifikanzniveau  $\alpha$ . Testmacht  $1 - \beta$ .

**Vorwissen:** Die Standardabweichungen in den beiden Stichproben sind in etwa  $s_1$  und  $s_2$ .

**Lösung:** In jeder Gruppe muss die Stichprobenlänge mindestens

$$n \geq \frac{(s_1^2 + s_2^2) \cdot (t_{1-\alpha, 2*n-2} + t_{1-\beta, 2*n-2})^2}{d^2}$$

sein.

# Einseitiger gepaarter Zweistichproben t-Test

**Geplanter Test:** Einseitiger gepaarter Zweistichproben t-Test.

**Ziel:** Finde Unterschiede, die größer als  $d$  sind.

Signifikanzniveau  $\alpha$ . Testmacht  $1 - \beta$ .

**Vorwissen:** Standardabweichung der Differenz der beiden Stichproben ist in etwa  $s_d$ .

**Lösung:** In jeder Gruppe muss die Stichprobenlänge mindestens

$$n \geq \frac{s_d^2 \cdot (t_{1-\alpha, n-1} + t_{1-\beta, n-1})^2}{d^2}$$

sein.



In R ermittelt man die benötigte Stichprobenlänge mit

```
power.t.test(n = , delta = , sd = , sig.level = ,  
             power = ,  
             type = c("two.sample", "one.sample", "paired"),  
             alternative = c("two.sided", "one.sided") )
```

In R ermittelt man die benötigte Stichprobenlänge mit

```
power.t.test(n = , delta = , sd = , sig.level = ,  
             power = ,  
             type = c("two.sample", "one.sample", "paired"),  
             alternative = c("two.sided", "one.sided") )
```

Die Argumente sind:

- $n$  = Stichprobenlänge (pro Gruppe bzw pro Stichprobe)

In R ermittelt man die benötigte Stichprobenlänge mit

```
power.t.test(n = , delta = , sd = , sig.level = ,  
             power = ,  
             type = c("two.sample", "one.sample", "paired"),  
             alternative = c("two.sided", "one.sided") )
```

Die Argumente sind:

- $n$  = Stichprobenlänge (pro Gruppe bzw pro Stichprobe)
- $\delta$  =  $d$  (minimale Differenz, detection level)

In R ermittelt man die benötigte Stichprobenlänge mit

```
power.t.test(n = , delta = , sd = , sig.level = ,  
             power = ,  
             type = c("two.sample", "one.sample", "paired"),  
             alternative = c("two.sided", "one.sided") )
```

Die Argumente sind:

- $n$  = Stichprobenlänge (pro Gruppe bzw pro Stichprobe)
- $\text{delta} = d$  (minimale Differenz, detection level)
- $\text{sd} = s$  (vermutete Standardabweichung pro Gruppe)

In R ermittelt man die benötigte Stichprobenlänge mit

```
power.t.test(n = , delta = , sd = , sig.level = ,  
             power = ,  
             type = c("two.sample", "one.sample", "paired"),  
             alternative = c("two.sided", "one.sided") )
```

Die Argumente sind:

- `n` = Stichprobenlänge (pro Gruppe bzw pro Stichprobe)
- `delta` =  $d$  (minimale Differenz, detection level)
- `sd` =  $s$  (vermutete Standardabweichung pro Gruppe)
- `sig.level` =  $\alpha$  (Signifikanzniveau)

In R ermittelt man die benötigte Stichprobenlänge mit

```
power.t.test(n = , delta = , sd = , sig.level = ,  
             power = ,  
             type = c("two.sample", "one.sample", "paired"),  
             alternative = c("two.sided", "one.sided") )
```

Die Argumente sind:

- $n$  = Stichprobenlänge (pro Gruppe bzw pro Stichprobe)
- $\text{delta} = d$  (minimale Differenz, detection level)
- $\text{sd} = s$  (vermutete Standardabweichung pro Gruppe)
- $\text{sig.level} = \alpha$  (Signifikanzniveau)
- $\text{power} = 1 - \beta$  (Testmacht)

In R ermittelt man die benötigte Stichprobenlänge mit

```
power.t.test(n = , delta = , sd = , sig.level = ,
             power = ,
             type = c("two.sample", "one.sample", "paired"),
             alternative = c("two.sided", "one.sided") )
```

Die Argumente sind:

- $n$  = Stichprobenlänge (pro Gruppe bzw pro Stichprobe)
- $\text{delta} = d$  (minimale Differenz, detection level)
- $\text{sd} = s$  (vermutete Standardabweichung pro Gruppe)
- $\text{sig.level} = \alpha$  (Signifikanzniveau)
- $\text{power} = 1 - \beta$  (Testmacht)

Genau eines der Argumente  $n, \text{delta}, \text{sd}, \text{sig.level}, \text{power}$  muss als NULL übergeben werden. Dieses wird dann berechnet.

Der Befehl `power.t.test()` kann auch dazu benutzt werden, die Testmacht zu berechnen, wenn man sich auf die Stichprobenlänge bereits festgelegt hat.



Der Befehl `power.t.test()` kann auch dazu benutzt werden, die Testmacht zu berechnen, wenn man sich auf die Stichprobenlänge bereits festgelegt hat.

Beispiel:

CCRT bei *D. ananassae*:  $n = 100$ ,  $d = 4$ ,  $s = 11.9$ ,  $\alpha = 5\%$

```
> power.t.test(n=100, delta=4, sd=11.9,  
+ sig.level=0.05, power=NULL,  
+ type="one.sample", alternative="two.sided")
```

One-sample t test power calculation

```
      n = 100  
delta = 4  
      sd = 11.9  
sig.level = 0.05  
      power = 0.9144375  
alternative = two.sided
```

# F-Test

Will man testen, ob die Mittelwerte bei 3 oder mehr Gruppen gleich sind, so verwendet man den F-Test. Um eine Aussage über die Stichprobenlänge treffen zu können, benötigt man die Variabilität innerhalb der Gruppen und die Variabilität zwischen den Gruppen (z.B. aus Vorversuchen).

# F-Test

Will man testen, ob die Mittelwerte bei 3 oder mehr Gruppen gleich sind, so verwendet man den F-Test. Um eine Aussage über die Stichprobenlänge treffen zu können, benötigt man die Variabilität innerhalb der Gruppen und die Variabilität zwischen den Gruppen (z.B. aus Vorversuchen).

Die Formel für die benötigte Stichprobe ist hier weniger übersichtlich. Deshalb konzentrieren wir uns auf die Berechnung mit R.

# F-Test

Will man testen, ob die Mittelwerte bei 3 oder mehr Gruppen gleich sind, so verwendet man den F-Test. Um eine Aussage über die Stichprobenlänge treffen zu können, benötigt man die Variabilität innerhalb der Gruppen und die Variabilität zwischen den Gruppen (z.B. aus Vorversuchen).

Die Formel für die benötigte Stichprobe ist hier weniger übersichtlich. Deshalb konzentrieren wir uns auf die Berechnung mit R.

Wir zeigen an folgendem Beispiel, wie man den R-Befehl `power.anova.test()` einsetzt, um die benötigte Stichprobenlänge zu ermitteln.

# Beispiel: Blutgerinnungszeit bei Ratten

**Frage:** Unterscheidet sich die Blutgerinnungszeit bei Ratten unter 4 verschiedenen Behandlungen?

# Beispiel: Blutgerinnungszeit bei Ratten

**Frage:** Unterscheidet sich die Blutgerinnungszeit bei Ratten unter 4 verschiedenen Behandlungen?

**Geplanter Test:** F-Test.

**Signifikanzniveau:**  $\alpha = 5\%$

**Testmacht:**  $1 - \beta = 90\%$ .

# Beispiel: Blutgerinnungszeit bei Ratten

**Frage:** Unterscheidet sich die Blutgerinnungszeit bei Ratten unter 4 verschiedenen Behandlungen?

**Geplanter Test:** F-Test.

**Signifikanzniveau:**  $\alpha = 5\%$

**Testmacht:**  $1 - \beta = 90\%$ .

**Vorwissen:** Standardabweichung innerhalb jeder Gruppe ist in etwa  $s_{\text{innerh}} = 2.4$ . Beachte:  $s_{\text{innerh}}^2 = ss_{\text{innerh}} / df_{\text{innerh}}$ .  
Standardabweichung zwischen den Gruppen ist in etwa  $s_{\text{zw}} = 1.2$ . Beachte:  $s_{\text{zw}}^2 = ss_{\text{zw}} / df_{\text{zw}}$ .

**Frage:** Bei wie vielen Ratten muss die Blutgerinnungszeit gemessen werden?

# Beispiel: Blutgerinnungszeit bei Ratten

```
> power.anova.test(groups=4, n=NULL, between.var=1.2^2,  
  within.var=2.4^2, sig.level=0.05, power=0.9)
```

Balanced one-way analysis of variance power calculation

```
groups = 4  
n = 19.90248  
between.var = 1.44  
within.var = 5.76  
sig.level = 0.05  
power = 0.9
```

NOTE: n is number in each group



# Beispiel: Blutgerinnungszeit bei Ratten

```
> power.anova.test(groups=4, n=NULL, between.var=1.2^2,  
  within.var=2.4^2, sig.level=0.05, power=0.9)
```

Balanced one-way analysis of variance power calculation

```
groups = 4  
n = 19.90248  
between.var = 1.44  
within.var = 5.76  
sig.level = 0.05  
power = 0.9
```

NOTE: n is number in each group

**Antwort:** Für jede der 4 Behandlungen braucht man mindestens 20 Ratten.

Viel Erfolg beim Lernen!