

Biostatistik, WS 2017/18

Wiederholung

Matthias Birkner

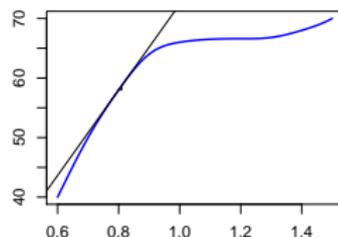
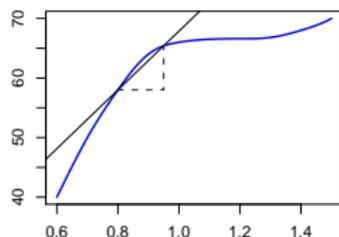
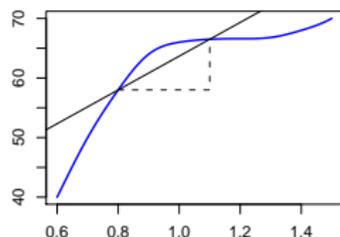
<http://www.staff.uni-mainz.de/birkner/Biostatistik1718/>

9.2.2018



JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

Differenzierbarkeit: Definition



Eine Funktion f sei auf einem Intervall I definiert, das x enthält.
 f heißt *differenzierbar* an der Stelle x , wenn der Grenzwert

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} =: f'(x)$$

existiert.

Anschauung: Existenz einer Tangente im Punkt x .

Erinnerung Das heißt: Für jedes $\varepsilon > 0$ gibt es ein $0 < h_0 (= h_0(\varepsilon, x, f))$, so dass für alle h mit $|h| \leq h_0$ gilt $\left| \frac{f(x+h) - f(x)}{h} - f'(x) \right| \leq \varepsilon$.

Beispiel: Exponentielles Wachstum

$N(t)$ = Populationsgröße zur Zeit t

Wenn sich alle Individuen einer Population (annähernd) unabhängig vermehren, sollte die Zunahme *proportional* zur aktuellen Größe sein, d.h. mit einem $\alpha \in \mathbb{R}$ gilt

$$N'(t) = \alpha N(t).$$

Dies ist eine (sehr einfache) *Differentialgleichung*.

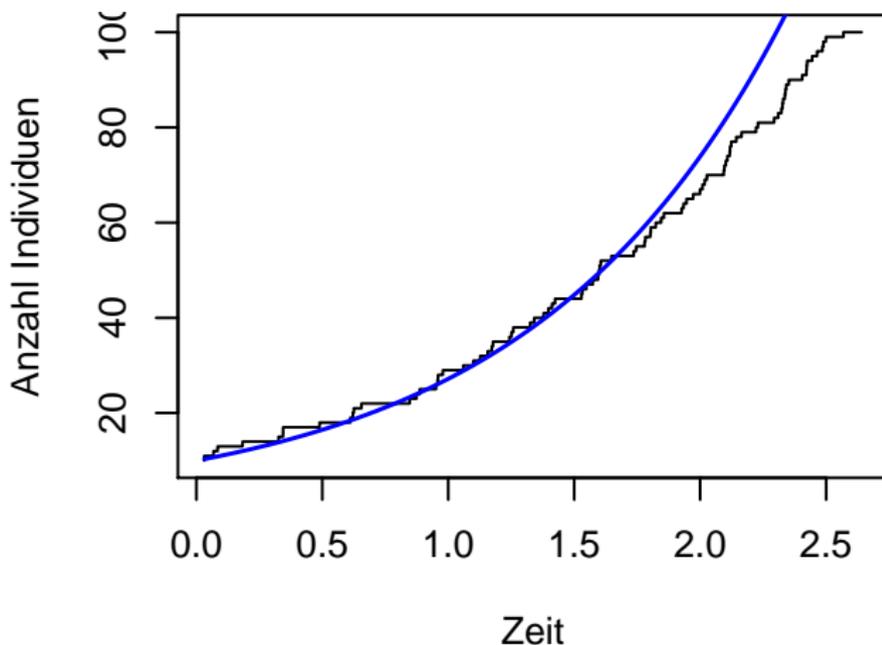
Wir wissen: Die Lösung ist $N(t) = N_0 e^{\alpha t}$

(denn $\frac{d}{dt} N_0 e^{\alpha t} = N_0 \frac{d}{dt} e^{\alpha t} = N_0 e^{\alpha t} \frac{d}{dt} (\alpha t) = \alpha N_0 e^{\alpha t}$)

Bericht: Eindeutigkeit der Lösung

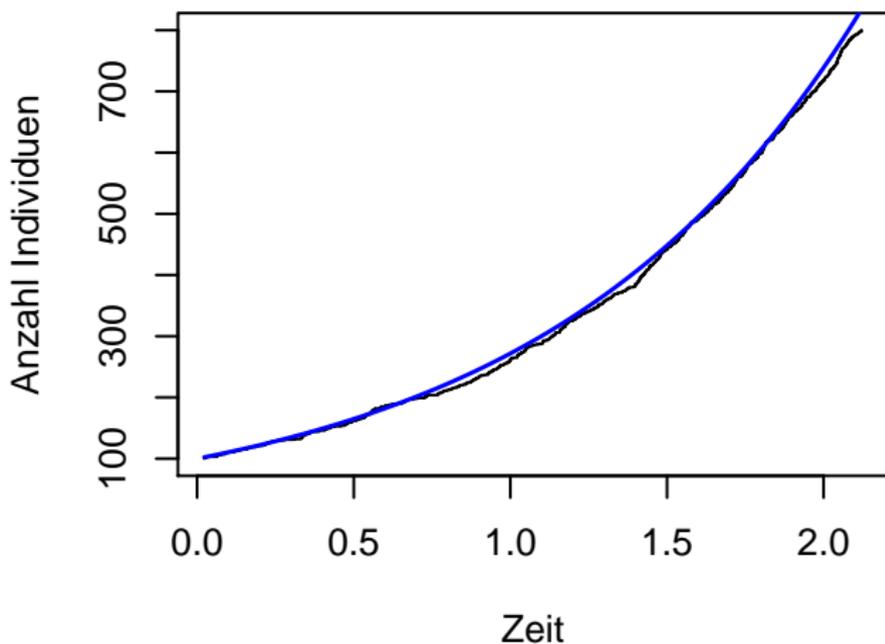
Exponentielles Wachstum, α heißt *Malthusischer Parameter* (nach Thomas Malthus, 1766-1834) oder *Wachstumsparameter* ($\alpha < 0$ entspricht einer schrumpfenden Population)

Exponentielles Wachstum als Approximation zufälliger Populationsdynamik



Starte mit 10 Individuen, jedes Ind. verdoppelt sich nach einer *zufälligen* Zeit (die Mittelwert 1 hat). Blaue Kurve: $10e^t$

Exponentielles Wachstum als Approximation zufälliger Populationsdynamik, II



Starte mit 100 Individuen, jedes Ind. verdoppelt sich nach einer *zufälligen* Zeit (die Mittelwert 1 hat). Blaue Kurve: $100e^t$

Stammfunktion, unbestimmtes Integral

f auf einem Intervall $I = [a, b]$ definiert.

F heißt *Stammfunktion* von f , wenn F auf I differenzierbar ist mit $F' = f$

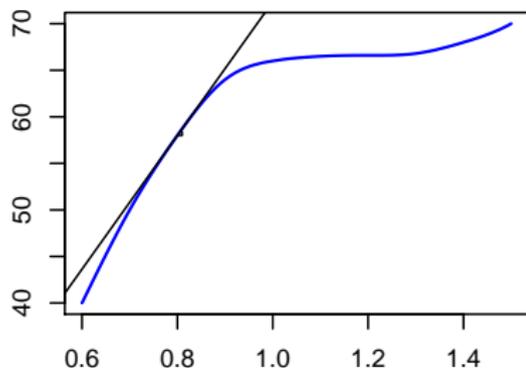
Hauptsatz der Integral- und Differentialrechnung

f stetig auf I , so ist $F(x) := \int_a^x f(x) dx$ eine Stammfunktion (und jede andere Stammfunktion \tilde{F} unterscheidet sich von dieser nur durch eine additive Konstante: $\tilde{F}(x) = F(x) + c$ für ein $c \in \mathbb{R}$).
Insbesondere: $\int_a^b f(x) dx = \tilde{F}(b) - \tilde{F}(a)$ für jede Stammfunktion \tilde{F} von f .

Man schreibt $\int f(x) dx$ für die (Menge aller) Stammfunktion(en) von f („unbestimmtes Integral von f “).

Beispiel:

$f(x)$ = Wegstrecke, die ein Zug nach Zeit x zurückgelegt hat.



$f'(x)$ = Geschwindigkeit zur Zeit x

Wir erhalten durch Integrieren der Geschwindigkeit die zurückgelegte Wegstrecke zurück (wenn wir die Zusatzinformation benutzen, dass die Wegstrecke zur Zeit $0 = 0$ war):

$$f(x) - \underbrace{f(0)}_{=0} = \int_0^x f'(y) dy$$

Als **Zufallsgröße oder Zufallsvariable** bezeichnet man das (Mess-)Ergebnis eines zufälligen Vorgangs.

Der **Wertebereich \mathcal{S}** (engl. state space) einer Zufallsgröße ist die Menge aller möglichen Werte.

Die **Verteilung einer Zufallsgröße X** weist jeder Menge $A \subseteq \mathcal{S}$ die **Wahrscheinlichkeit $\mathbb{P}(X \in A)$** zu, dass X einen Wert in A annimmt

Für Zufallsgrößen werden üblicherweise Großbuchstaben verwendet (z.B. X, Y, Z), für konkrete Werte Kleinbuchstaben.

Beispiel: Würfelwurf $W =$ Augenzahl des nächsten Würfelwurfs.

$$\mathcal{S} = \{1, 2, \dots, 6\}$$

$$\mathbb{P}(W = 1) = \dots = \mathbb{P}(W = 6) = \frac{1}{6}$$

$$(\mathbb{P}(W = x) = \frac{1}{6} \text{ für alle } x \in \{1, \dots, 6\})$$

Die Verteilung erhält man aus einer Symmetrieüberlegung
oder aus einer langen Würfelreihe.

Beispiel: Geschlecht X bei Neugeborenen.

$$\mathcal{S} = \{\text{„männlich“}, \text{„weiblich“}\}$$

Die Verteilung erhält man aus einer langen Beobachtungsreihe.

Beispiel: Körpergrößenverteilung in Deutschland.

Die Verteilung erhält man aus einer langen Messreihe.

Rechenregeln:

Sei X Zufallsgröße mit Wertebereich \mathcal{S} .

- $0 \leq \mathbb{P}(X \in A) \leq 1$ für jede Teilmenge $A \subseteq \mathcal{S}$
- $\mathbb{P}(X \in \mathcal{S}) = 1$
- Sind $A, B \subseteq \mathcal{S}$ disjunkt, d.h. $A \cap B = \emptyset$,

$$\mathbb{P}(X \in A \cup B) = \mathbb{P}(X \in A) + \mathbb{P}(X \in B),$$

insbesondere $\mathbb{P}(X \in A^c) = 1 - \mathbb{P}(X \in A)$ mit $A^c = \mathcal{S} \setminus A$

- Allgemein gilt

$$\mathbb{P}(X \in A \cup B) = \mathbb{P}(X \in A) + \mathbb{P}(X \in B) - \mathbb{P}(X \in A \cap B)$$

(„Einschluss-Ausschluss-Formel“)

Bedingte Wahrscheinlichkeit

Ws des Ereignisses $\{Y \in B\}$ unter der Bedingung $\{X \in A\}$

$$\mathbb{P}(Y \in B | X \in A) := \frac{\mathbb{P}(Y \in B, X \in A)}{\mathbb{P}(X \in A)} \quad (*)$$

„bedingte Ws von $\{Y \in B\}$ gegeben $\{X \in A\}$ “

Beachte:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B | X \in A)$$

(*) in Worten ausgedrückt:

Die Ws des Ereignisses $\{X \in A, Y \in B\}$ läßt sich in zwei Schritten berechnen:

- Zunächst muss das Ereignis $\{X \in A\}$ eintreten.
- Die Ws hiervon wird multipliziert mit der Ws von $\{Y \in B\}$, wenn man schon weiß, daß $\{X \in A\}$ eintritt.

Die Formel von Bayes

Seien X, Y Zufallsgrößen mit Wertebereichen \mathcal{S}_X bzw. \mathcal{S}_Y ,
 $A \subset \mathcal{S}_X$, $B \subset \mathcal{S}_Y$, dann gilt

$$\begin{aligned} & \mathbb{P}(Y \in B \mid X \in A) \\ &= \frac{\mathbb{P}(X \in A \mid Y \in B) \cdot \mathbb{P}(Y \in B)}{\mathbb{P}(X \in A \mid Y \in B) \cdot \mathbb{P}(Y \in B) + \mathbb{P}(X \in A \mid Y \in B^c) \cdot \mathbb{P}(Y \in B^c)} \end{aligned}$$

Denn

$$\text{Zähler} = \mathbb{P}(X \in A, Y \in B)$$

$$\begin{aligned} \text{Nenner} &= \mathbb{P}(X \in A, Y \in B) + \mathbb{P}(X \in A, Y \in B^c) \\ &= \mathbb{P}(X \in A, Y \in B \cup B^c) = \mathbb{P}(X \in A) \end{aligned}$$

Beispiel: Medizinische Reihenuntersuchung

Eine Krankheit komme bei 2% der Bevölkerung vor („Prävalenz 2%“),

ein Test schlage bei 95% der Kranken an („Sensitivität 95%“),
aber auch bei 10% der Gesunden („Spezifität 90%“)

Eine zufällig gewählte Person werde mit positivem Resultat getestet.

Wie wahrscheinlich ist es, dass sie tatsächlich krank ist?

Modell: X = Testergebnis ($S_X = \{\text{positiv, negativ}\}$),

Y = Gesundheitszustand ($S_Y = \{\text{gesund, krank}\}$) der Person

Gesucht $\mathbb{P}(Y = \text{krank} \mid X = \text{positiv}) = ?$

Wir wissen: $\mathbb{P}(Y = \text{krank}) = 0.02$, $\mathbb{P}(Y = \text{gesund}) = 0.98$,

$\mathbb{P}(X = \text{positiv} \mid Y = \text{krank}) = 0.95$,

$\mathbb{P}(X = \text{positiv} \mid Y = \text{gesund}) = 0.1$,

also $\mathbb{P}(Y = \text{krank} \mid X = \text{positiv}) = \frac{0.02 \cdot 0.95}{0.02 \cdot 0.95 + 0.98 \cdot 0.1} \doteq 0.162$

Stochastische Unabhängigkeit

Definition

Zwei Zufallsgrößen X und Y heißen (stochastisch) unabhängig, wenn für alle Ereignisse $\{X \in A\}$, $\{Y \in B\}$ gilt

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B)$$

Beispiel:

- Werfen zweier Würfel:
 $X =$ Augenzahl Würfel 1, $Y =$ Augenzahl Würfel 2.

$$\mathbb{P}(X = 2, Y = 5) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(X = 2) \cdot \mathbb{P}(Y = 5)$$

Stochastische Unabhängigkeit

In der Praxis wendet man häufig Resultate an, die Unabhängigkeit einer Stichprobe voraussetzen.

Beispiele:

- Für eine Studie wird eine zufällige Person in München und eine zufällige Person in Hamburg befragt. Die Antworten dürfen als unabhängig voneinander angenommen werden.
- Befragt man zwei Schwestern oder nahe verwandte (getrennt voneinander), so werden die Antworten nicht unabhängig voneinander sein.

Bernoulli-Verteilung

Als **Bernoulli-Experiment** bezeichnet man jeden zufälligen Vorgang mit exakt zwei möglichen Werten.

Diese werden üblicherweise mit 1 und 0 bezeichnet, beziehungsweise als „Erfolg“ und „Misserfolg“.

Bernoulli-Zufallsgröße X :

Zustandsraum $\mathcal{S} = \{0, 1\}$.

Verteilung:

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p$$

Der Parameter $p \in [0, 1]$ heißt **Erfolgswahrscheinlichkeit**.

Beispiele:

- Münzwurf: mögliche Werte sind „Kopf“ und „Zahl“.
- Hat die gesampelte Drosophila eine Mutation, die weiße Augen verursacht? Mögliche Antworten sind „Ja“ und „Nein“.
- Das Geschlecht einer Person: „männlich“ oder „weiblich“

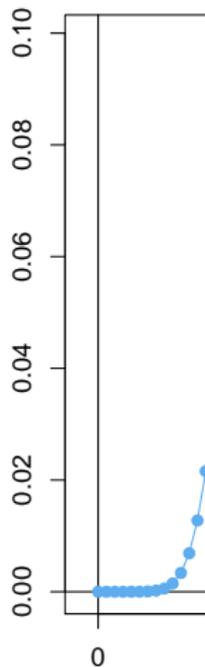
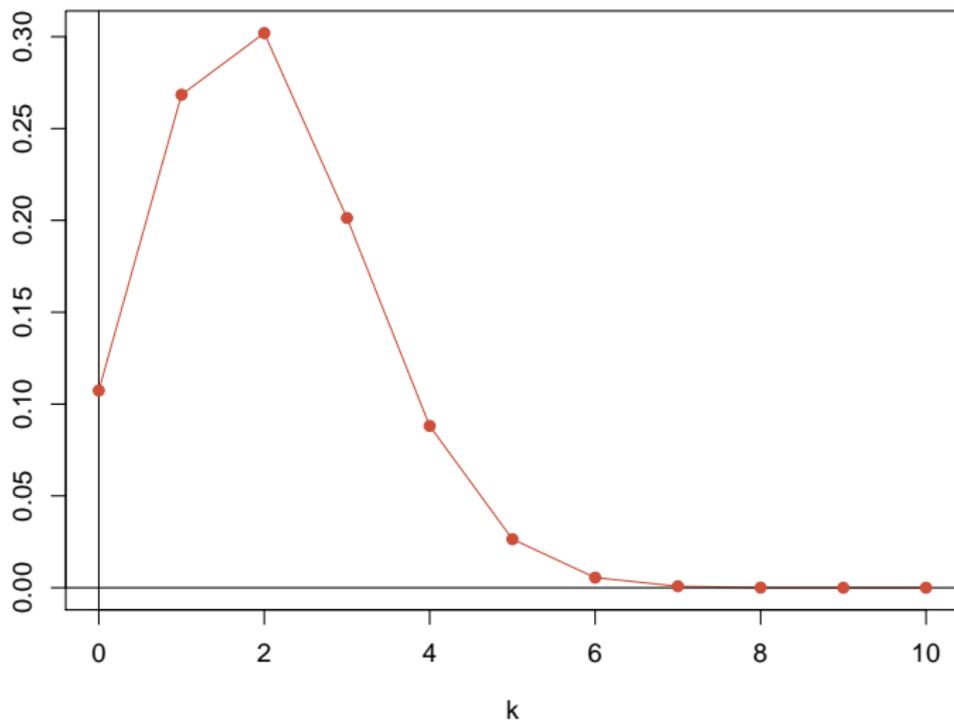
Binomialverteilung

Sei X die Anzahl der Erfolge bei n unabhängigen Versuchen mit Erfolgswahrscheinlichkeit von jeweils p . Dann gilt für $k \in \{0, 1, \dots, n\}$

$$\mathbb{P}(X = k) = \binom{n}{k} p^k \cdot (1 - p)^{n-k}$$

und X heißt *binomialverteilt*, kurz:

$$X \sim \text{bin}(n, p).$$

probabilities of $\text{bin}(n=10, p=0.2)$ 

Definition (Erwartungswert)

Sei X eine Zufallsvariable mit endlichem oder abzählbarem Wertebereich $S = \{a_1, a_2, a_3 \dots\} \subseteq \mathbb{R}$. Dann ist der *Erwartungswert* von X definiert durch

$$\mathbb{E}X = \sum_{a \in S} a \cdot \mathbb{P}(X = a)$$

Beispiele:

- Sei X Bernoulli-verteilt mit Erfolgswahrscheinlichkeit $p \in [0, 1]$. Dann gilt

$$\mathbb{E}X = 1 \cdot \mathbb{P}(X = 1) + 0 \cdot \mathbb{P}(X = 0) = \mathbb{P}(X = 1) = p$$

- Sei W die Augenzahl bei einem Würfelwurf. Dann gilt

$$\begin{aligned} \mathbb{E}W &= 1 \cdot \mathbb{P}(W = 1) + 2 \cdot \mathbb{P}(W = 2) + \dots + 6 \cdot \mathbb{P}(W = 6) \\ &= \frac{1 + \dots + 6}{6} = \frac{21}{6} = 3.5 \end{aligned}$$

Rechnen mit Erwartungswerten

Satz (Linearität der Erwartung)

Sind X und Y Zufallsvariablen mit Werten in \mathbb{R} und ist $a \in \mathbb{R}$, so gilt:

- $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}X$
- $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$

Folgerung: Für $X \sim \text{bin}(n, p)$ gilt $\mathbb{E}(X) = n \cdot p$

Satz (Nur für Unabhängige!)

Sind X und Y **stochastisch unabhängige** Zufallsvariablen mit Werten in \mathbb{R} , so gilt

- $\mathbb{E}(X \cdot Y) = \mathbb{E}X \cdot \mathbb{E}Y$.

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 .

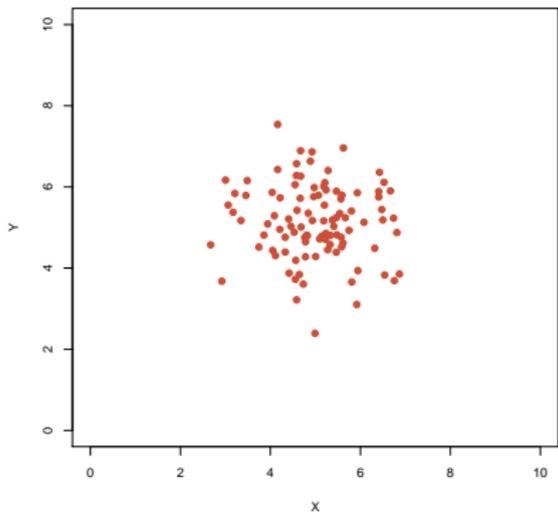
Die *Korrelation* von X und Y ist

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

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

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

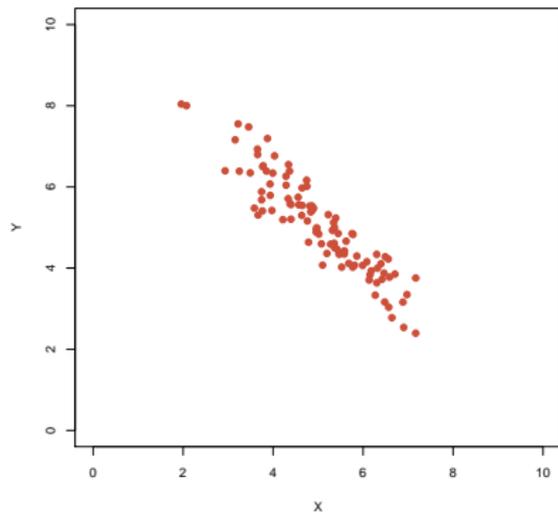
$$\text{Cor}(X, Y) = -0.069$$



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

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

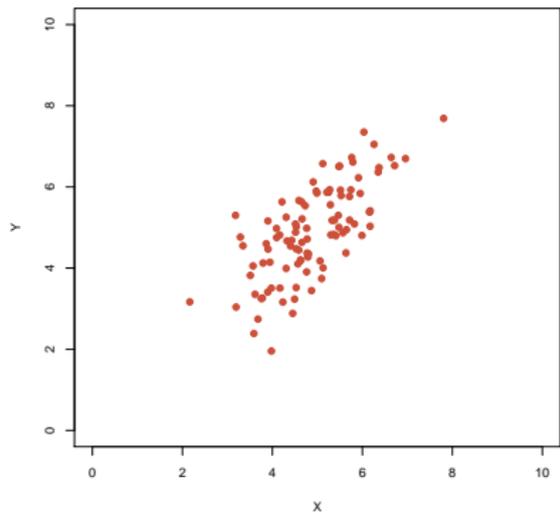
$$\text{Cor}(X, Y) = -0.92$$



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

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

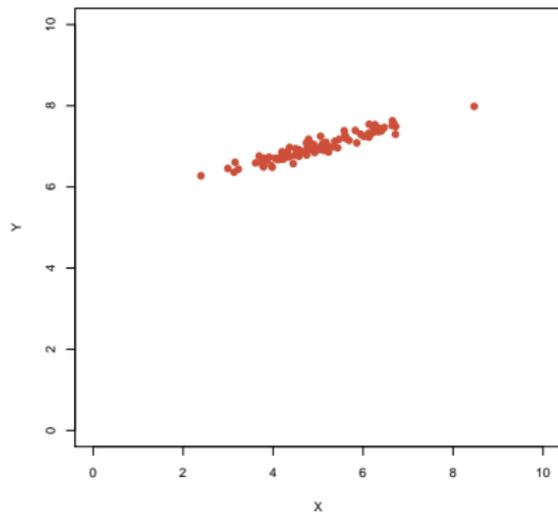
$$\text{Cor}(X, Y) = 0.71$$



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

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

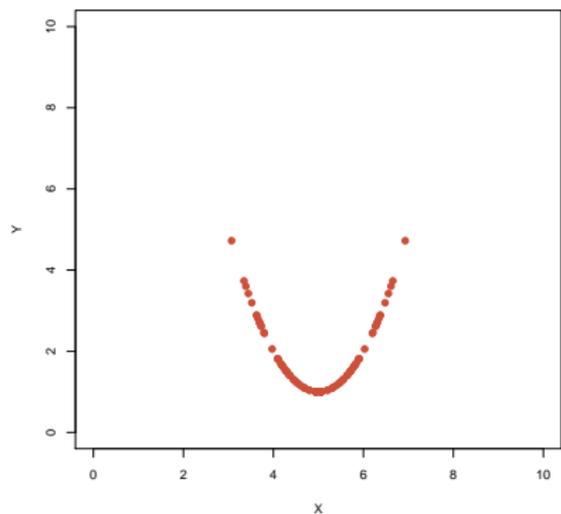
$$\text{Cor}(X, Y) = 0.95$$



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

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

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



Rechenregeln für Varianzen

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

- $\text{Var}X = \text{Cov}(X, X)$
- $\text{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$
- $\text{Var}(a \cdot X) = a^2 \cdot \text{Var}X$
- $\text{Var}(X + Y) = \text{Var}X + \text{Var}Y + 2 \cdot \text{Cov}(X, Y)$
- Sind (X, Y) stochastisch unabhängig, so folgt:

$$\text{Var}(X + Y) = \text{Var}X + \text{Var}Y$$

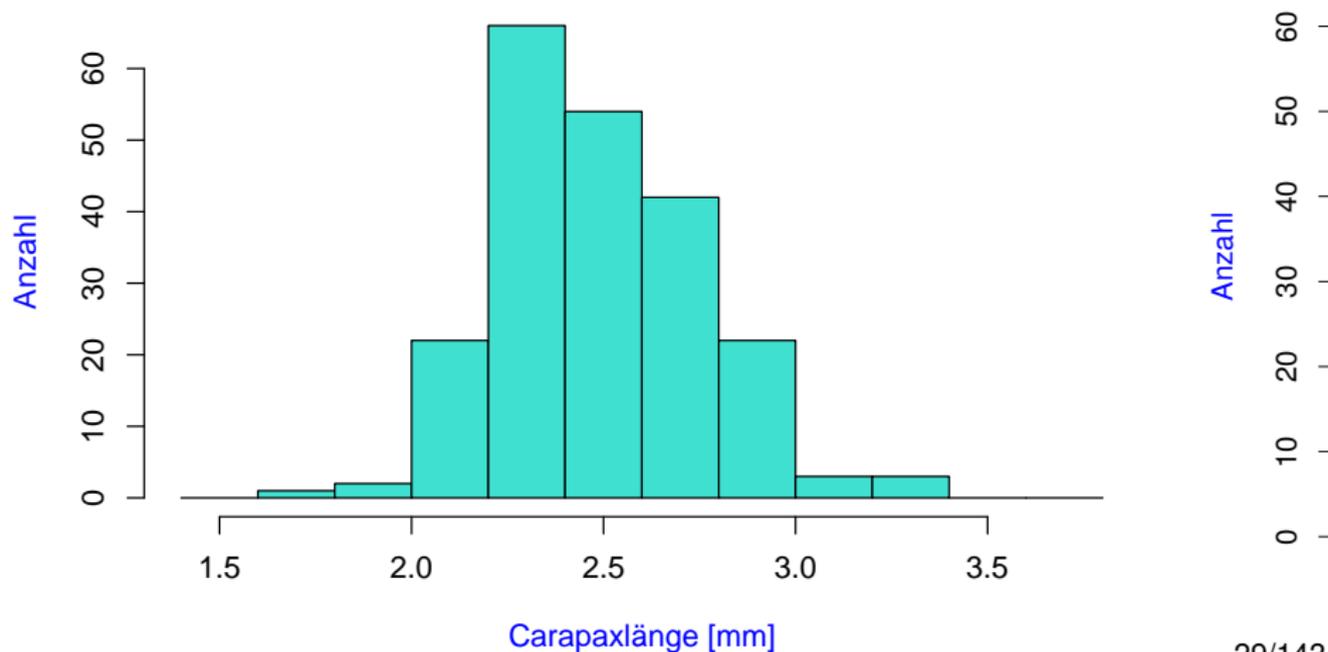
Insbesondere gilt für $X \sim \text{bin}(n, p)$:

$$\text{Var}X = n \cdot p \cdot (1 - p).$$

(Denn wir können $X = Y_1 + Y_2 + \dots + Y_n$ schreiben mit Y_1, \dots, Y_n unabhängig, jeweils Bernoulli(p)-verteilt.)

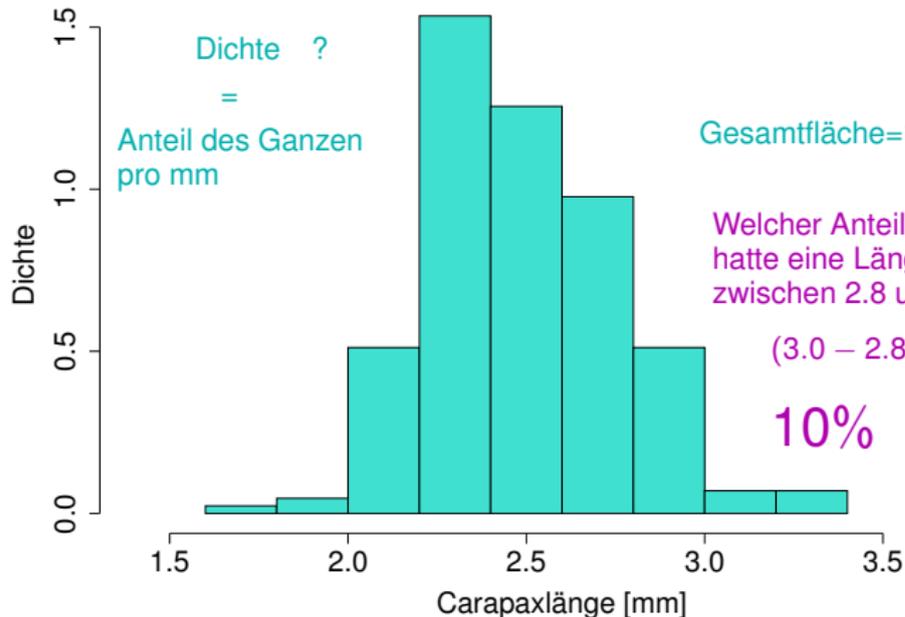
Histogramme

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



Histogramme

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

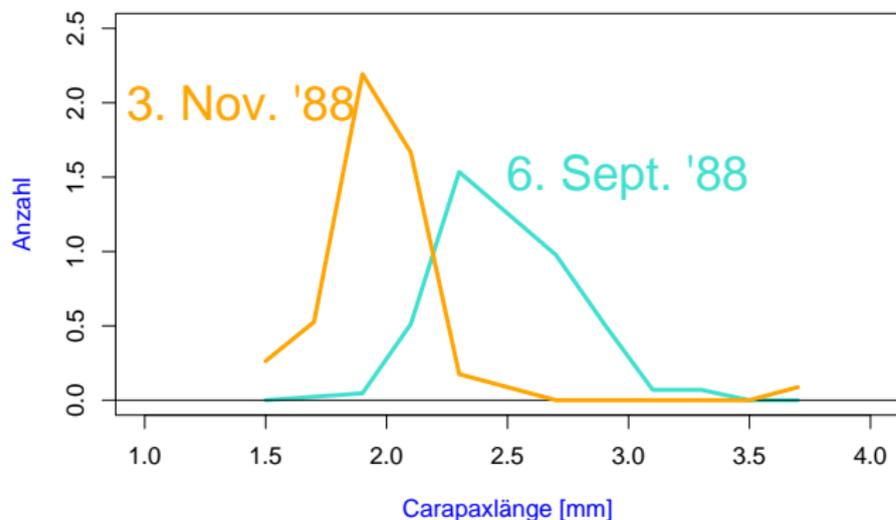


Nich

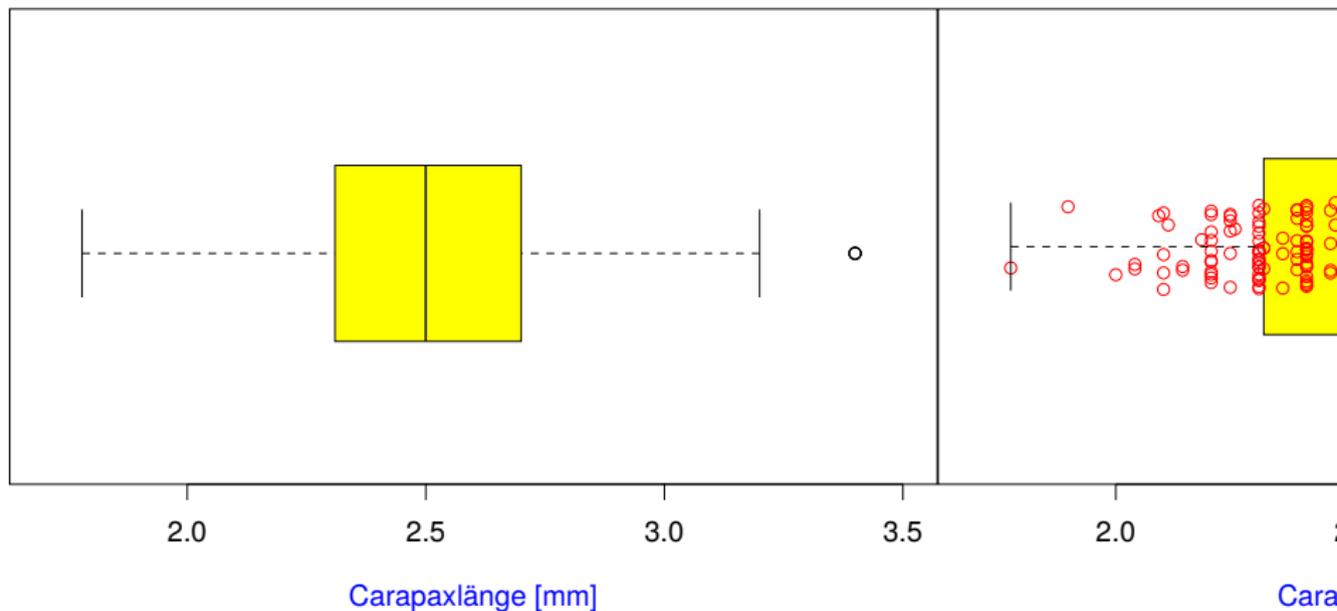


Zwei und mehr Dichtepolygone in einem Plot

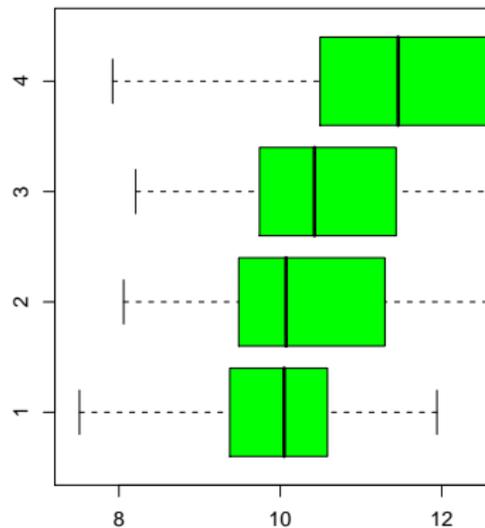
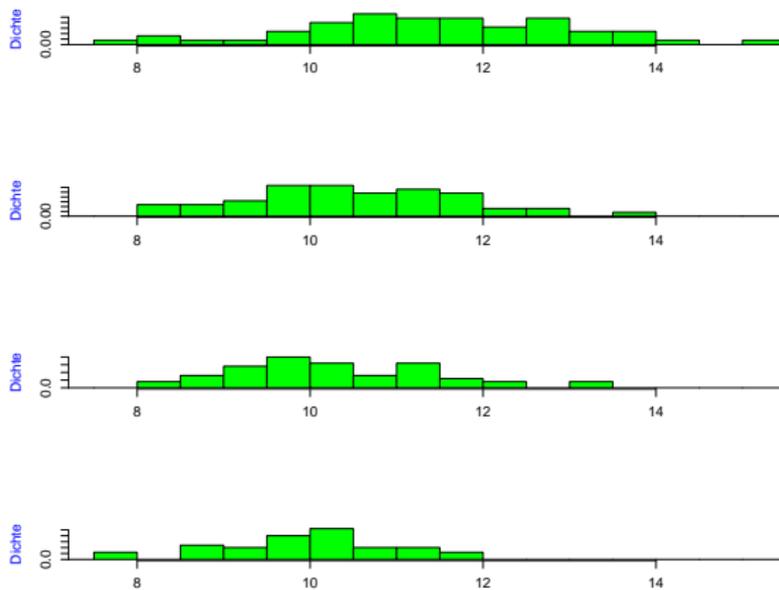
Nichteiertragende Weibchen



Der Boxplot



Beispiel: Vergleich von mehreren Gruppen



(ist mit Boxplots oft übersichtlicher)

Es ist oft möglich,
das Wesentliche
an einer Stichprobe

mit ein paar Zahlen
zusammenzufassen.

Wesentlich:

1. Wie groß?

Lageparameter

2. Wie variabel?

Streuungsparameter

Eine Möglichkeit
kennen wir schon
aus dem Boxplot:

Lageparameter

Der Median

Streuungsparameter

Der Quartilabstand ($Q_3 - Q_1$)

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 :
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 :
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}

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}

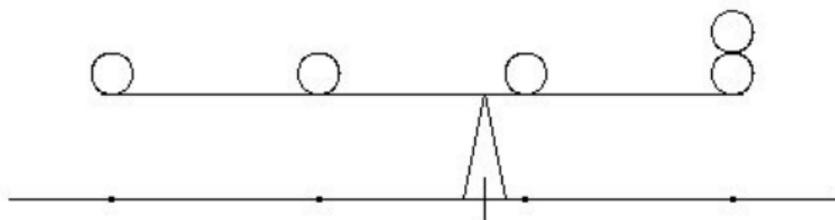
für den Mittelwert.

DEFINITION:

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

Wie weit weicht
eine typische Beobachtung
vom
Mittelwert
ab ?

Die **Standardabweichung** σ (“sigma”)

ist ein

etwas komisches

gewichtetes Mittel
der Abweichungsbeträge

und zwar

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

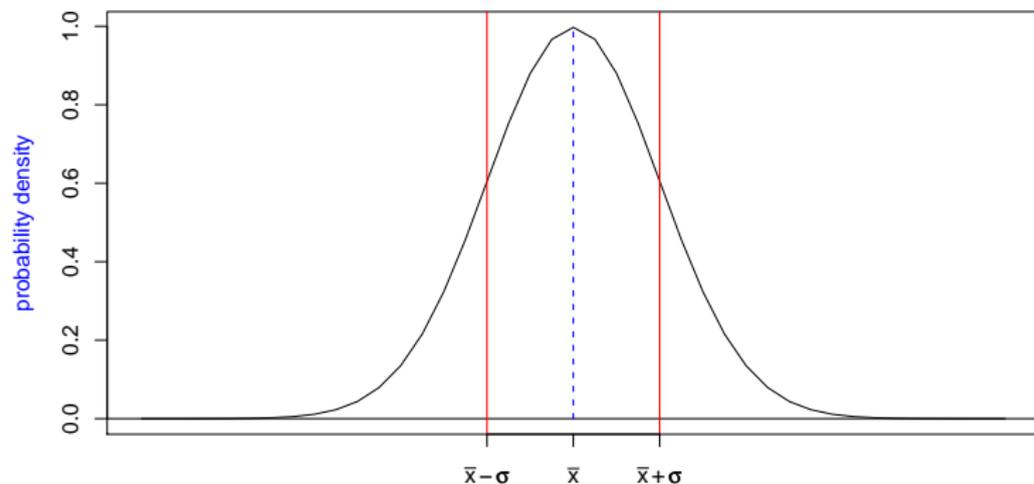
Die **Standardabweichung** von x_1, x_2, \dots, x_n
als Formel:

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

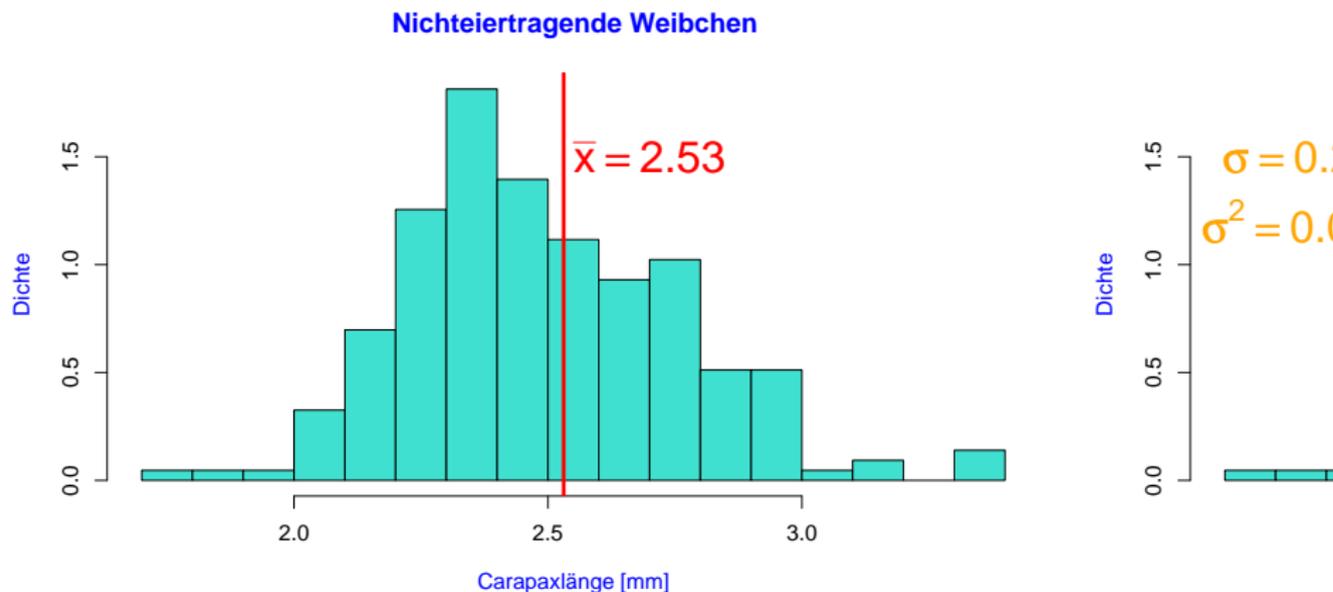
$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ heißt **Varianz**.

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



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

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

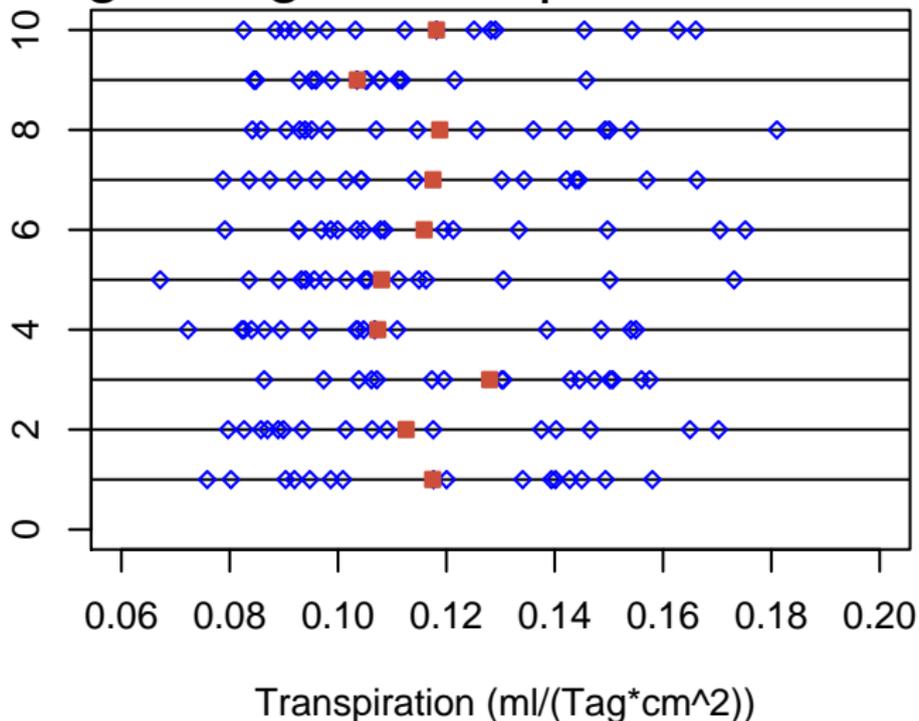
Stichprobenvarianz: $\sigma_S^2 = \frac{1}{n} \sum_{i=1}^n (S_i - \bar{S})^2$

korrigierte Stichprobenvarianz:

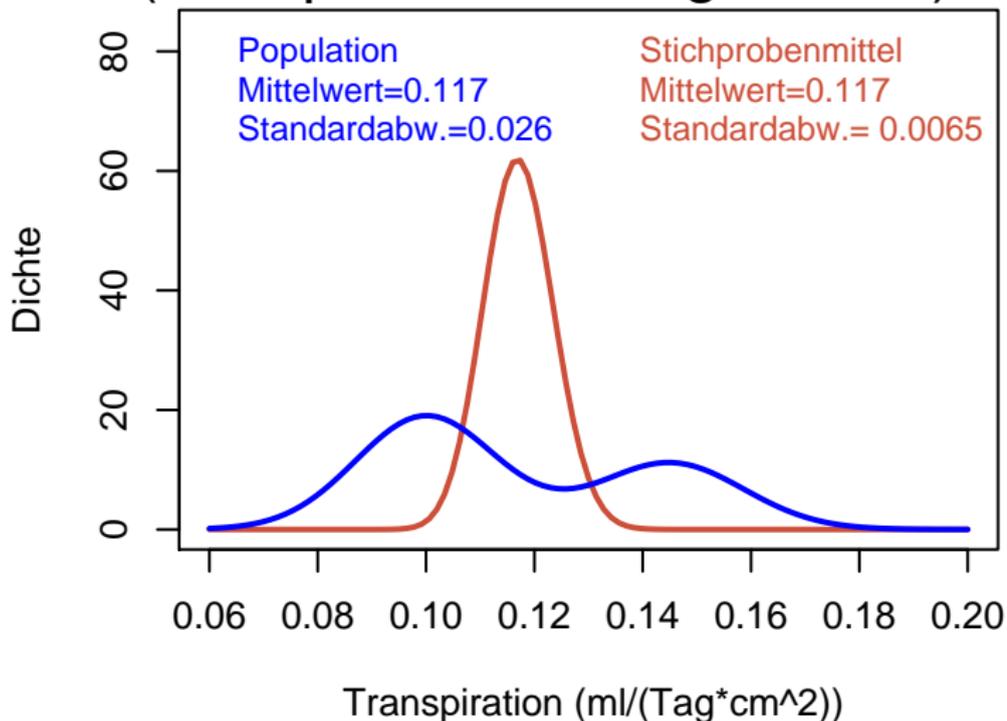
$$\begin{aligned} s^2 &= \frac{n}{n-1} \sigma_S^2 \\ &= \frac{n}{n-1} \cdot \frac{1}{n} \cdot \sum_{i=1}^n (S_i - \bar{S})^2 \\ &= \frac{1}{n-1} \cdot \sum_{i=1}^n (S_i - \bar{S})^2 \end{aligned}$$

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

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

ist

$$1/\sqrt{n}$$

mal

der Standardabweichung
der Population.

Die Standardabweichung der Population
bezeichnet man mit

σ
(sigma).

Die Regel schreibt man häufig so:

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

In der Praxis ist

σ

unbekannt.

Es wird durch

die Stichproben-Standardabweichung S

geschätzt:

$$\sigma \approx S$$

$$s/\sqrt{n}$$

(die geschätzte
Standardabweichung
von \bar{x})

nennt man den

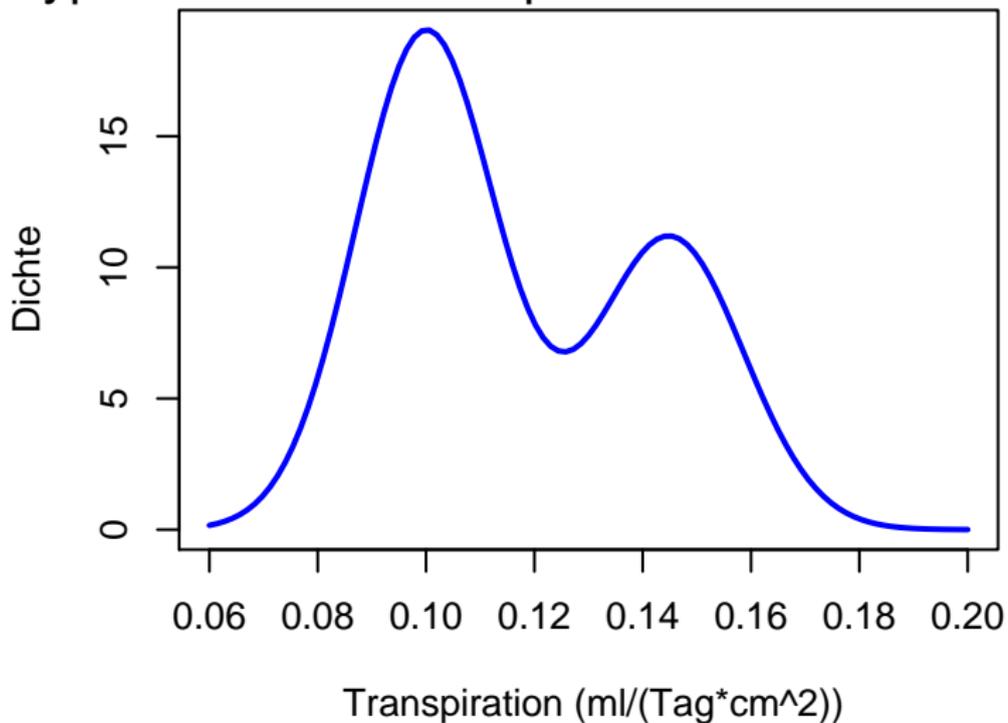
Standardfehler.

(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)

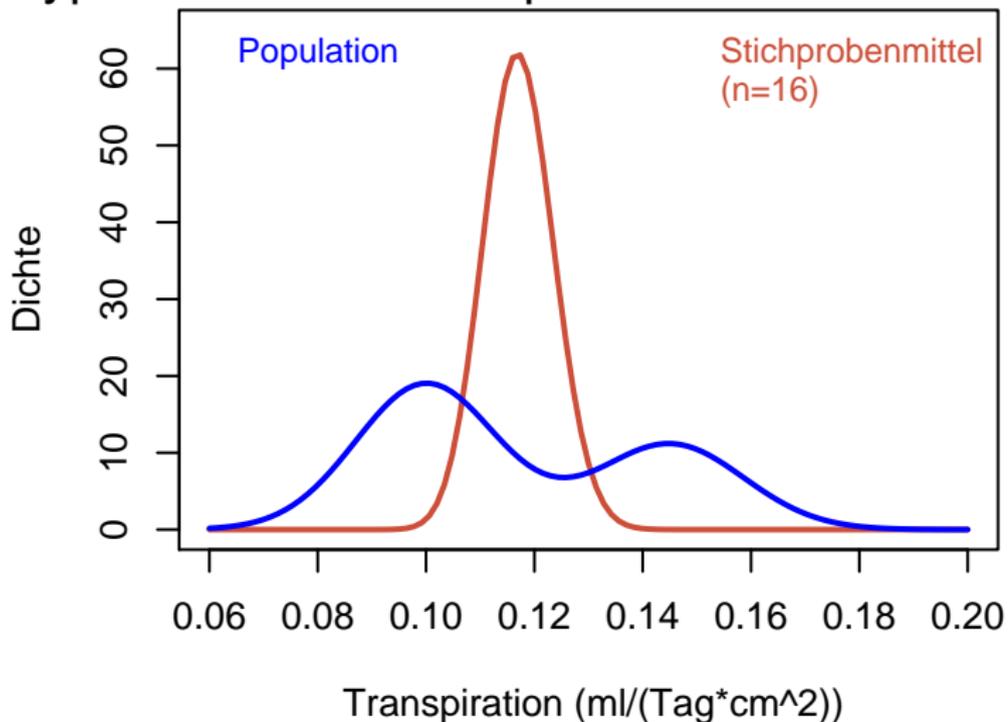
eingipfelig

&

symmetrisch

(wenn der Stichprobenumfang n nur groß genug ist)

Hypothetische Transpirationsratenverteilung



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

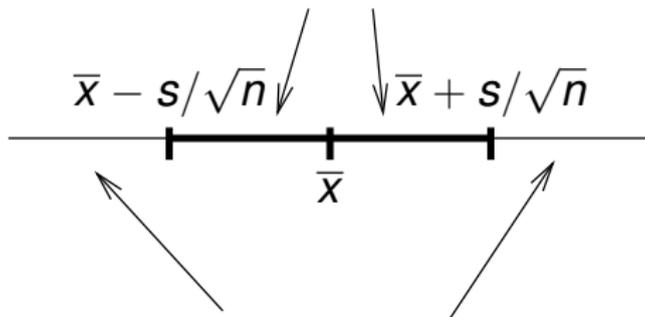
Dichte der Normalverteilung



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

Wichtige Folgerung

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



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

Demnach:

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

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

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

so ist

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

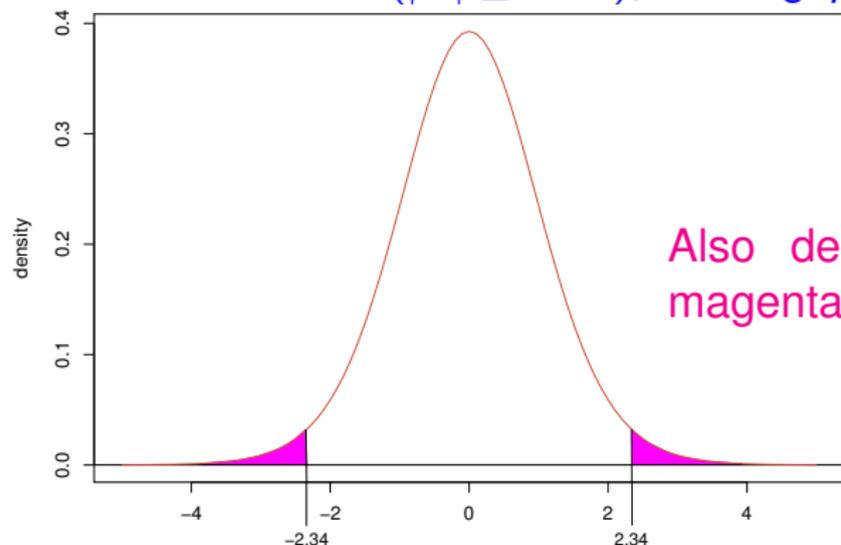
t-verteilt mit $n - 1$ Freiheitsgraden (df=*degrees of freedom*).

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

Wie (un)wahrscheinlich ist nun eine **mindestens** so große Abweichung wie 2.34 Standardfehler?

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

Zu berechnen ist $\mathbb{P}(|T| \geq 2.34)$, der sog. p -Wert.



Also der Gesamthalt der magentafarbenen Flächen.

Wir halten fest:

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

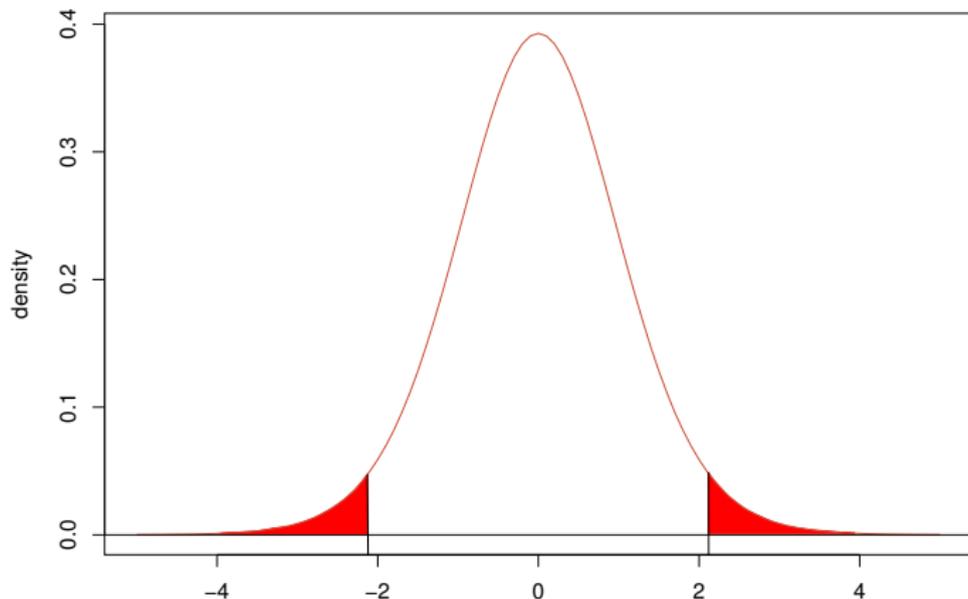
d.h.: Wenn die **Nullhypothese** “alles nur Zufall”, also in diesem Fall die Hypothese $\mu = 0$ gilt, dann ist eine mindestens so große Abweichung sehr unwahrscheinlich.

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)

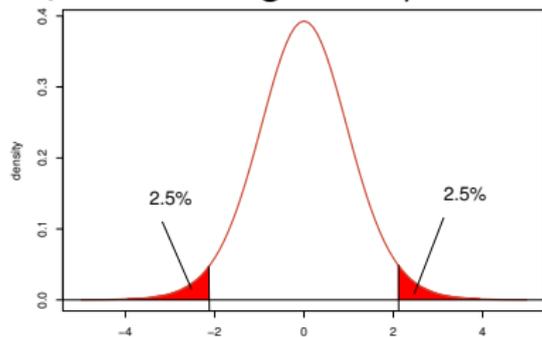
Welche t -Werte sind “auf dem 5%-Niveau” signifikant?

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

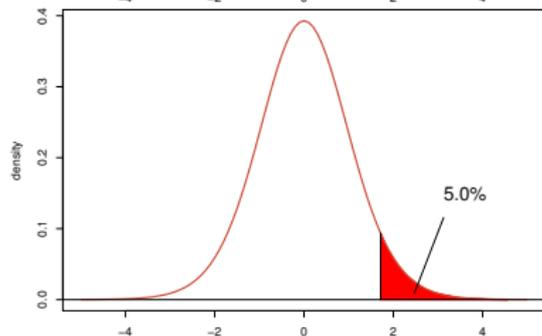
- Wir möchten belegen, dass eine Abweichung in den Daten vermutlich nicht allein auf Zufallsschwankung beruht.
- Dazu spezifizieren wir zunächst eine **Nullhypothese H_0** , d.h. wir konkretisieren, was “allein auf Zufall beruhen” bedeutet.
- Dann versuchen wir zu zeigen: Wenn H_0 gilt, dann ist eine **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 μ ist.



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



$$p\text{-Wert} = \mathbb{P}_{H_0}(X \geq x)$$

Reine Lehre des statistischen Testens

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

$$\mathbb{P}_{H_0}(\mathcal{A}) = \alpha$$

(oder zumindest $\mathbb{P}_{H_0}(\mathcal{A}) \leq \alpha$).

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

allgemein: $H_0 = \{\text{p-Wert} \leq \alpha\}$

- **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 α .

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 μ_{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.

Allgemeiner: Ist \mathcal{A} das Ereignis, dass zum Verwerfen von H_0 führt (falls es eintritt), so muss die Festlegung von H_0 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:

$$\mathbb{P}_{H_0}(\mathcal{A}) = \alpha$$

und

$$\mathbb{P}_{H_1}(\mathcal{A}) = \text{möglichst groß,}$$

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.
- 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.

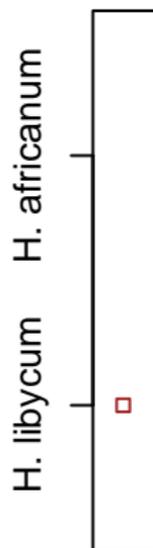
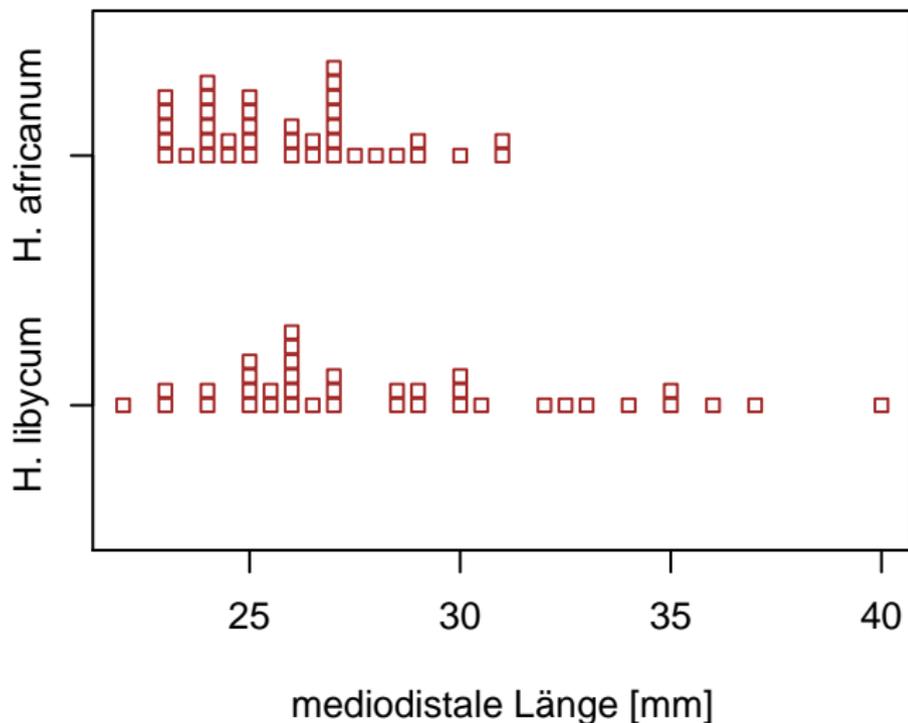
Angenommen, H_0 wird auf dem 5%-Niveau verworfen. Welche Aussage gilt dann?

- Die Nullhypothese ist falsch. ~~Die Nullhypothese ist falsch.~~
- H_0 ist mit 95%-iger Wahrscheinlichkeit 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. 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.
~~Wir müssen die Alternative H_1 verwerfen.~~
- H_0 ist wahr. ~~H_0 ist wahr.~~
- H_0 ist wahrscheinlich wahr. ~~H_0 ist wahrscheinlich wahr.~~
- Es ist ungefährlich, davon auszugehen, dass H_0 zutrifft.
~~Es ist ungefährlich, davon auszugehen, dass H_0 zutrifft.~~
- Auch wenn H_0 zutrifft, ist es nicht sehr unwahrscheinlich, dass unsere Teststatistik einen so extrem erscheinenden Wert annimmt. Auch wenn H_0 zutrifft, ist es nicht sehr unwahrscheinlich, dass unsere Teststatistik einen so extrem erscheinenden Wert annimmt. ✓
- Die Nullhypothese ist in dieser Hinsicht mit den Daten verträglich. Die Nullhypothese ist in dieser Hinsicht mit den Daten verträglich. ✓

Gegeben: Zwei Stichproben:



(Im Beispiel: Mesiodistale Längen fossiler Hipparion-Zähne von zwei

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
 $f_A = s_A/\sqrt{n_A} = 2,2/\sqrt{39} = 0,36$ (Standardfehler)),

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

(unser Schätzwert für die Streuung von \bar{x}_L ist also
 $f_L = s_L/\sqrt{n_L} = 4,3/\sqrt{38} = 0,70$).

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

Ungepaarter t-Test (mit Annahme gleicher Varianzen)

Die „Bilderbuchsituation“: Wir haben zwei unabhängige Stichproben

$x_{1,1}, \dots, x_{1,n_1}$ und $x_{2,1}, \dots, x_{2,n_2}$.

Die $x_{1,i}$ stammen aus einer Normalverteilung mit (unbekanntem) Mittelwert μ_1 und (unbekannter) Varianz $\sigma^2 > 0$, die $x_{2,i}$ aus einer Normalverteilung mit (unbekanntem) Mittelwert μ_2 und derselben Varianz σ^2 .

Seien

$$\bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1,i}, \quad \bar{x}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} x_{2,i}$$

die jeweiligen Stichprobenmittelwerte,

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{1,i} - \bar{x}_1)^2, \quad s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (x_{2,i} - \bar{x}_2)^2,$$

die (korrigierten) Stichprobenvarianzen.

Wir möchten die Hypothese $H_0 : \mu_1 = \mu_2$ prüfen.

Wenn $\mu_1 = \mu_2$ gilt, so sollte $\bar{x}_1 = \bar{x}_2$ „bis auf
Zufallsschwankungen“ gelten, denn $\mathbb{E}[\bar{X}_1] = \mu_1$, $\mathbb{E}[\bar{X}_2] = \mu_2$.

Was ist die Skala der typischen Schwankungen von $\bar{x}_1 - \bar{x}_2$?

$$\text{Var}(\bar{X}_1 - \bar{X}_2) = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

Problem (wie bereits im ein-Stichproben-Fall): Wir kennen σ^2 nicht.

Wir schätzen es im zwei-Stichproben-Fall durch die gepoolte
Stichprobenvarianz

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(= \frac{1}{n_1 + n_2 - 2} \left(\sum_{i=1}^{n_1} (x_{1,i} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{2,i} - \bar{x}_2)^2 \right) \right)$$

und bilden die Teststatistik

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ungepaarter zwei-Stichproben t-Test (mit der Annahme gleicher Varianzen)

Seien X_1, \dots, X_n und Y_1, \dots, Y_m unabhängige normalverteilte Zufallsvariablen mit der selben Varianz σ^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.

Im Hipparion-Beispiel war $\bar{x}_L = 28,4$, $\bar{x}_A = 25,9$, $s_A = 2,2$,
 $s_A = 4,3$.

Wir finden $t = 3,2$, das 99,5%-Quantil der Student-Vert. mit 75 Freiheitsgraden ist 2,64.

Wir können die Nullhypothese „die mittlere mesiodistale Länge bei *H. lybicum* und bei *H. africanum* sind gleich“ zum Signifikanzniveau 1% ablehnen.

Mögliche Formulierung:

„Die mittlere mesiodistale Länge war signifikant größer (28,4 mm) bei *H. libycum* als bei *H. africanum* (25,9 mm) (t -Test, $\alpha = 0,01$).“

Wann gepaarter t -Test und wann ungepaarter t -Test?

Wenn die **Stichprobenlänge unterschiedlich** ist, ergibt „gepaart“ keinen Sinn.

Wenn **die Stichprobenlänge gleich** ist:

- Sind die Stichproben unabhängig voneinander?
Falls ja, dann ungepaart testen. Ein gepaarter Test würde sinnlose Abhängigkeiten unterstellen und hätte wegen der geringeren Anzahl Freiheitsgrade auch eine geringere Macht.
- Sind die Stichproben voneinander abhängig?
(z.B. Messungen von denselben Individuen bzw. Objekten)
Falls ja, dann ist ein gepaarter Test sinnvoll. Bei starker Abhängigkeitsstruktur hat der gepaarte t -Test höhere Macht (da der Test von Variabilität zwischen den Individuen bereinigt ist)

Zwei-Stichproben-t-Test ohne Annahme gleicher Varianzen (Welch-Test) mit R

```
> A <- md[Art=="africanum"]  
> L <- md[Art=="libycum"]  
> t.test(L,A)
```

Welch Two Sample t-test

data: L and A

t = 3.2043, df = 54.975, p-value = 0.002255

alternative hypothesis: true difference in means
is not equal to 0

95 percent confidence interval:

0.9453745 4.1025338

sample estimates:

mean of x mean of y

28.43421 25.91026

Formulierung:

„Die mittlere mesiodistale Länge war signifikant größer (28,4 mm) bei *H. libycum* als bei *H. africanum* (25,9 mm) (*t*-Test, $p = 0,002$).“

Idee der Rangsummentests

Beobachtungen:

$X : x_1, x_2, \dots, x_m$

$Y : y_1, y_2, \dots, y_n$

- Sortiere alle Beobachtungen der Größe nach.
- Bestimme die Ränge der m X -Werte unter allen $m + n$ Beobachtungen.
- Wenn die Nullhypothese zutrifft, sind die m X -Ränge eine rein zufällige Wahl aus $\{1, 2, \dots, m + n\}$.
- Berechne die Summe der X -Ränge, prüfe, ob dieser Wert untypisch groß oder klein.

Wilcoxon's Rangsummenstatistik

Beobachtungen:

$$X : x_1, x_2, \dots, x_m$$

$$Y : y_1, y_2, \dots, y_n$$

$U =$ Summe der X -Ränge $- (1 + 2 + \dots + m)$
heißt

Wilcoxon's Rangsummenstatistik

(Die Normierung ist so gewählt, dass W Werte zwischen 0 und $m \cdot n$ annehmen kann.)

Beispiel:

- Beobachtungen:

X : 1,5; 5,6; 35,2

Y : 7,9; 38,1; 41,0; 56,7; 112,1; 197,4; 381,8

- Lege Beobachtungen zusammen und sortiere:

1,5; 5,6; 7,9; 35,2; 38,1; 41,0; 56,7; 112,1; 197,4; 381,8

- Bestimme Ränge:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

- Rangsummenstatistik hier: $U = 1 + 2 + 4 - (1 + 2 + 3) = 1$

Im Beispiel:

$m = 3$ beobachtete X -Werte, $n = 7$ beobachtete Y -Werte

Die Quantiltabelle der Rangsummen-Verteilungen (für $m = 3$, $n = 7$) zeigt

$$\mathbb{P}(U_{3,7} \leq 19) = 0,975,$$

d.h.

$$\mathbb{P}(U_{3,7} > 19) \leq 0,025$$

und aus Symmetrie ($m \cdot n = 21$, $21 - 19 = 2$) auch

$$\mathbb{P}(U_{3,7} < 2) \leq 0,025,$$

somit gilt unter der Nullhypothese

„die X - und die Y -Wert stammen aus derselben Verteilung“

$$\mathbb{P}(U_{3,7} \in \{0, 1, 20, 21\}) \leq 0,05.$$

Wir haben $U = 1$ gefunden und lehnen demnach die Nullhypothese zum Signifikanzniveau 5% ab (zweiseitiger Rangsummen-Test).

Wilcoxon-Test: Allgemeines Vorgehen

m x -Werte und n y -Werte beobachtet, der Wert von $U_{m,n} =$ Summe der x -Ränge $-\frac{m(m+1)}{2}$ sei u . (Falls Werte mehrfach vorkommen, vergibt man geteilte Ränge.)

- Wir lehnen die Nullhypothese

H_0 : „die x -Werte und die y -Werte stammen aus derselben Verteilung“

zum Signifikanzniveau α ab, wenn $u > q_{U_{m,n}, 1-\alpha/2}$ oder $u < q_{U_{m,n}, \alpha/2}$ gilt
($q_{U_{m,n}, a} = a$ -Quantil der (m, n) -Wilcoxon-Rangsummenverteilung).

Wilcoxon-Test: Allgemeines Vorgehen, 2

m x -Werte und n y -Werte beobachtet, der Wert von $U_{m,n} =$ Summe der x -Ränge $-\frac{m(m+1)}{2}$ sei u . (Falls Werte mehrfach vorkommen, vergibt man geteilte Ränge.)

- Einseitiger Test: Die Nullhypothese

H_0 : „die x -Verteilung ist gegenüber der y -Verteilung nach links verschoben“

lehnen wir zum Signifikanzniveau α ab, wenn der beobachtete Wert $u > q_{W_{m,n}, 1-\alpha}$ ist.

(analog, wenn wir eine Verschiebung nach rechts prüfen wollen, oder vertausche die Rollen von x und y)

Wilcoxon-Test: Anmerkung

- Die Quantile von $U_{m,n}$ entnimmt man einer Tabelle oder einem Computerprogramm, beachte:

$$q_{U_{m,n},\alpha/2} = n \cdot m - q_{U_{m,n},1-\alpha/2} \text{ wegen Symmetrie.}$$

(Für kleine m und n kann man die Verteilung auch durch Auszählen per Hand bestimmen, für den allgemeinen Fall gibt es eine Rekursionsformel;

$$\text{es gilt } \mathbb{E}_{H_0}[U_{m,n}] = \frac{mn}{2} \text{ und } \text{Var}_{H_0}[U_{m,n}] = \frac{mn(m+n+1)}{12},$$

für große m und n ist $U_{m,n}$ unter H_0 ungefähr normalverteilt.)

R kennt den Wilcoxon-Test mittels

`wilcox.test`

```
> x
[1] 1.5 5.6 35.2
> y
[1] 7.9 38.1 41.0 56.7 112.1 197.4 381.8
> wilcox.test(x,y)
```

Wilcoxon rank sum test

data: x and y

$W = 1$, p-value = 0.03333

alternative hypothesis: true location shift is
not equal to 0

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?

Vorgehen:

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

95%-Quantil der χ^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

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

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

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

Das 95%-Quantil der χ^2 -Verteilung mit 5 Freiheitsgraden ist $11,07 > 4,115$, wir lehnen H_0 nicht ab (zum Signifikanzniveau 5%).

95%-Quantil der χ^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: $\chi_5^2([4,115, \infty)) = 0,533$, d.h. wir finden einen p -Wert von 53%, der Test gibt keinen Anlass zu Zweifel an 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 Beförderungentscheidung 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

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

Unter H_0 ist X^2 (approximativ) χ^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 χ^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 $\chi_1^2([5.17, \infty)) = 0.023$, 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) χ^2 -verteilt mit $rs - 1 - (r - 1) - (s - 1) = (r - 1)(s - 1)$ Freiheitsgraden.

Student-Konfidenzintervall für den Mittelwert

Sei $\alpha \in (0, 1)$ (oft $\alpha = 0.05$), $t_{n-1, 1-\alpha/2}$ das $(1 - \alpha/2)$ -Quantil der Student-Verteilung mit $n - 1$ Freiheitsgraden (d.h. für Student- $(n - 1)$ -verteiltes T_{n-1} gilt $\mathbb{P}(T_{n-1} \leq t_{n-1, 1-\alpha/2}) = 1 - \alpha/2$). Dann ist die Wahrscheinlichkeit, dass der wahre Mittelwert μ von dem Intervall

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

überdeckt wird, (approximativ*) $1 - \alpha$.

I heißt ein **Konfidenzintervall für μ zum Niveau $1 - \alpha$** oder kurz ein $(1 - \alpha)$ -Konfidenzintervall.

* Die Aussage ist wörtlich korrekt, wenn die Daten als normalverteilt angenommen werden dürfen, die Näherung ist sehr gut und für die Praxis ausreichend, wenn die Daten ungefähr symmetrisch und glockenförmig verteilt sind oder n genügend groß.

Dualität von Konfidenzintervallen und zweiseitigen Tests

Beispiel:

Stichprobe der Größe $n = 29$, in der wir Stichprobenmittelwert $\bar{x} = 3.23$ und Stichprobenstreuung $s = 0.9$ beobachtet haben. Konfidenzintervall für den wahren Mittelwert μ zum Irrtumsniveau 5% ($t_{28,0.975} = 2.048$)

$$\left[\bar{x} - t_{n-1,0.975} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,0.975} \frac{s}{\sqrt{n}} \right] = [2.88, 3.58]$$

Nehmen wir an, wir wollten (anhand derselben Beobachtungen) die Nullhypothese „ $\mu = 3.2$ “ zum Signifikanzniveau 5% testen:

Verwende den t -Test: $t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = 0.18$,

$|t| = |0.18| \leq t_{28,0.975} = 2.048$, d.h. wir würden die Nullhypothese nicht verwerfen (der p -Wert ist $0.86 (= \mathbb{P}_{\mu=3.2}(|t| \geq 0.18))$).

Anteilsschätzung: Szenario (Beispiel)

Wir beobachten X Männchen in einer Stichprobe der Größe n und möchten den (unbekannten) Männchenanteil p in der Gesamtpopulation schätzen.

Der offensichtliche Schätzer ist

$$\text{die relative Häufigkeit } \hat{p} := \frac{X}{n}$$

in der Stichprobe.

Gewünscht: Ein in Abhängigkeit von den Beobachtungen konstruiertes (und möglichst kurzes) $(1 - \alpha)$ -Konfidenzintervall $[\hat{p}_u, \hat{p}_o]$, d.h. es soll gelten

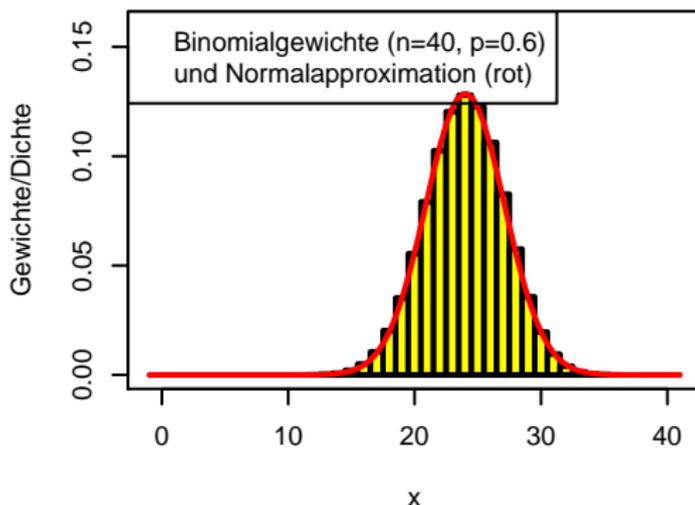
$$\mathbb{P}_p([\hat{p}_u, \hat{p}_o] \text{ überdeckt } p) \geq 1 - \alpha$$

für *jede Wahl* von p .

Für gegebenes p ist X Binomial(n,p)-verteilt,

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

Für (genügend) großes n ist X ungefähr normalverteilt
mit Mittelwert np und Varianz $np(1 - p)$
(zentraler Grenzwertsatz):



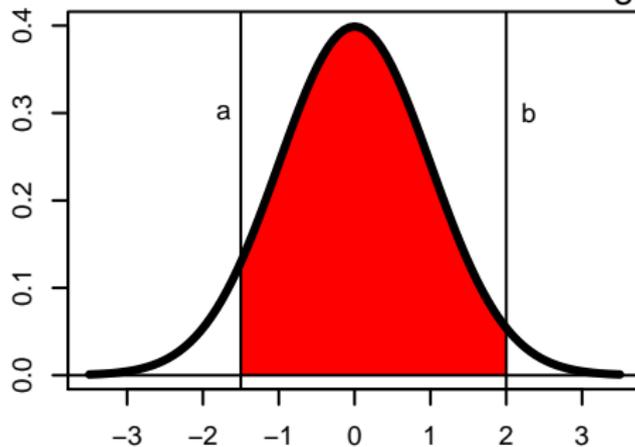
Also ist $\hat{p} = \frac{X}{n}$ (ungefähr) normalverteilt mit Mittelwert p und
Varianz $\frac{1}{n}p(1 - p)$.

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

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

(mit standard-normalverteiletem Z)

Dichte der Standard-Normalverteilung:



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

Wähle $z_{1-\alpha/2}$ so dass $\mathbb{P}(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha$, dann ist

$$\left[\hat{p} - z_{1-\alpha/2} \frac{\sqrt{\hat{p}(1 - \hat{p})}}{\sqrt{n}}, \hat{p} + z_{1-\alpha/2} \frac{\sqrt{\hat{p}(1 - \hat{p})}}{\sqrt{n}} \right]$$

ein (approximatives) Konfidenzintervall für p zum Niveau $1 - \alpha$.

$z_{1-\alpha/2}$ ist das $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung,
für die Praxis wichtig sind die Werte

$z_{0,975} \doteq 1,96$ (für $\alpha = 0,05$) und $z_{0,995} \doteq 2,58$ (für $\alpha = 0,01$).



Im Porzellankrebs-Beispiel war $n = 53$, $\hat{p} = \frac{30}{53} = 0,57$, also ist

$$\left[\frac{30}{53} - 1,96 \frac{\sqrt{\frac{30}{53} \cdot \frac{23}{53}}}{\sqrt{53}}, \frac{30}{53} + 1,96 \frac{\sqrt{\frac{30}{53} \cdot \frac{23}{53}}}{\sqrt{53}} \right] = [0,43, 0,70]$$

ein 95%-Konfidenzintervall für den Männchenanteil in dieser Population.

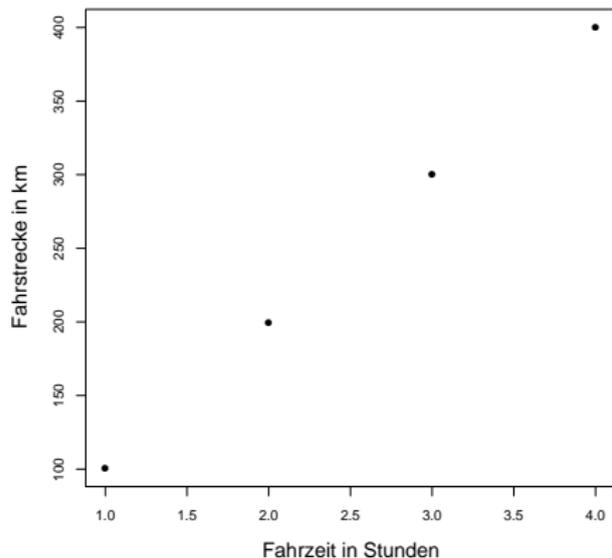
Anmerkungen

$\left[\hat{p} \pm z_{1-\alpha/2} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \right]$ ist ein (approximatives) Konfidenzintervall für p zum Niveau $1 - \alpha$.

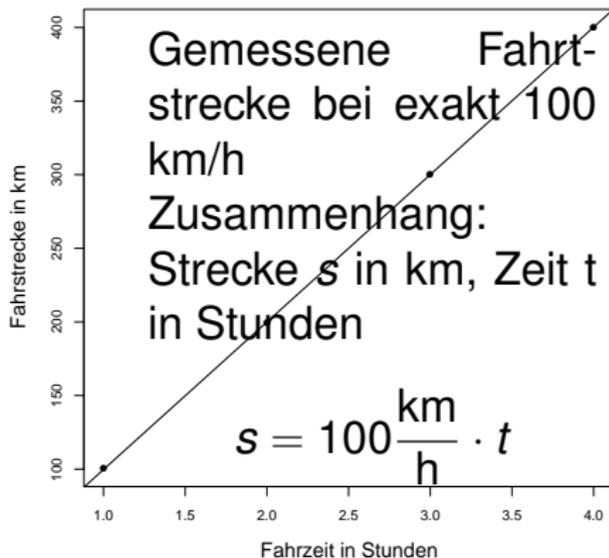
- Für die Gültigkeit der Approximation muss n genügend groß und p nicht zu nahe an 0 oder 1 sein. (Eine häufig zitierte „Faustregel“ ist „ $np \geq 9$, $n(1 - p) \geq 9$ “.)
- Die Philosophie der Konfidenzintervalle entstammt der *frequentistischen* Interpretation der Statistik: Für jede Wahl des Parameters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

Erst mal ohne Zufallsschwankungen

Gefahrene Strecke bei 100km/h



Gefahrene Strecke bei 100km/h



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

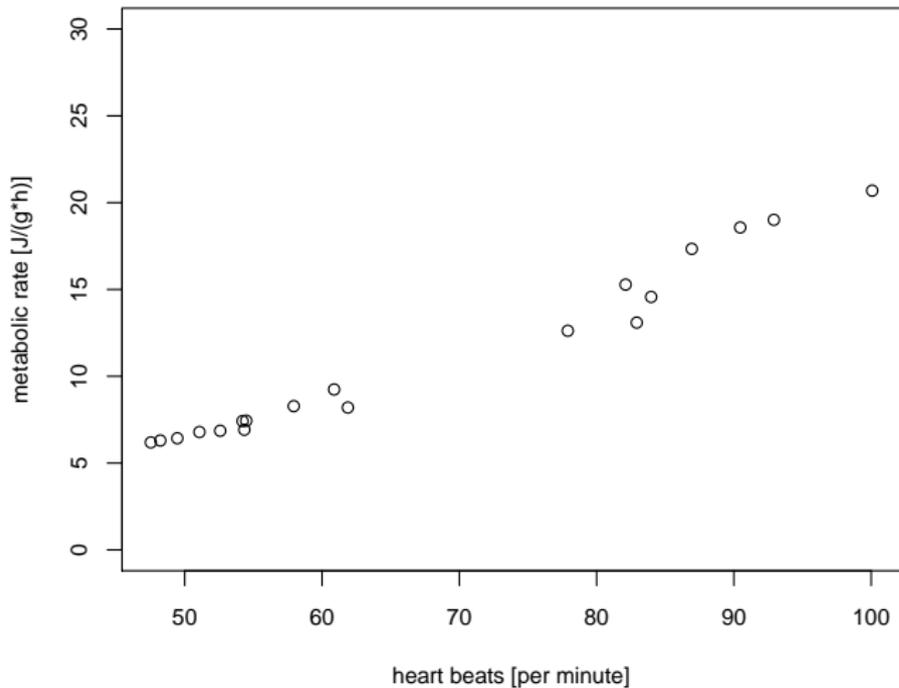
Englisch: Grif-
fon Vulture
Gypus fulvus
Gänsegeier

Beispiel: Herzfrequenz und Stoffwechselrate beim Gänsegeier

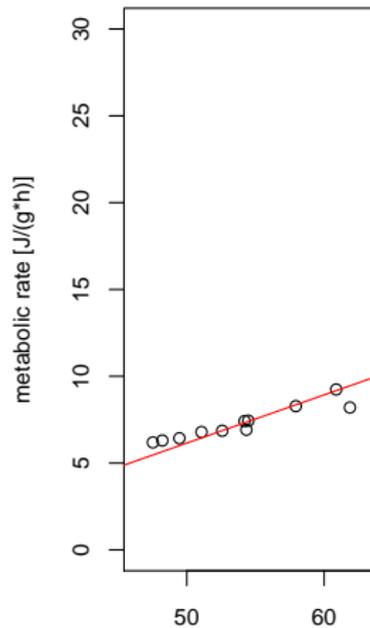
- Frage: Was ist die Stoffwechselrate bei Gänsegeiern im Alltag (z.B. im Flug)?
- Problem: Stoffwechselrate ist aufwändig zu messen (eigentlich nur im Labor)
- Beobachtung: Herzfrequenz ist leicht zu messen.
- Lösung: Nutze linearen Zusammenhang zwischen Stoffwechselrate und Herzfrequenz.
- Komplikation: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

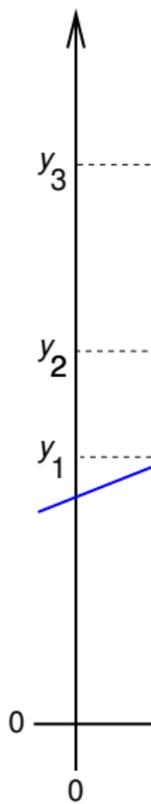
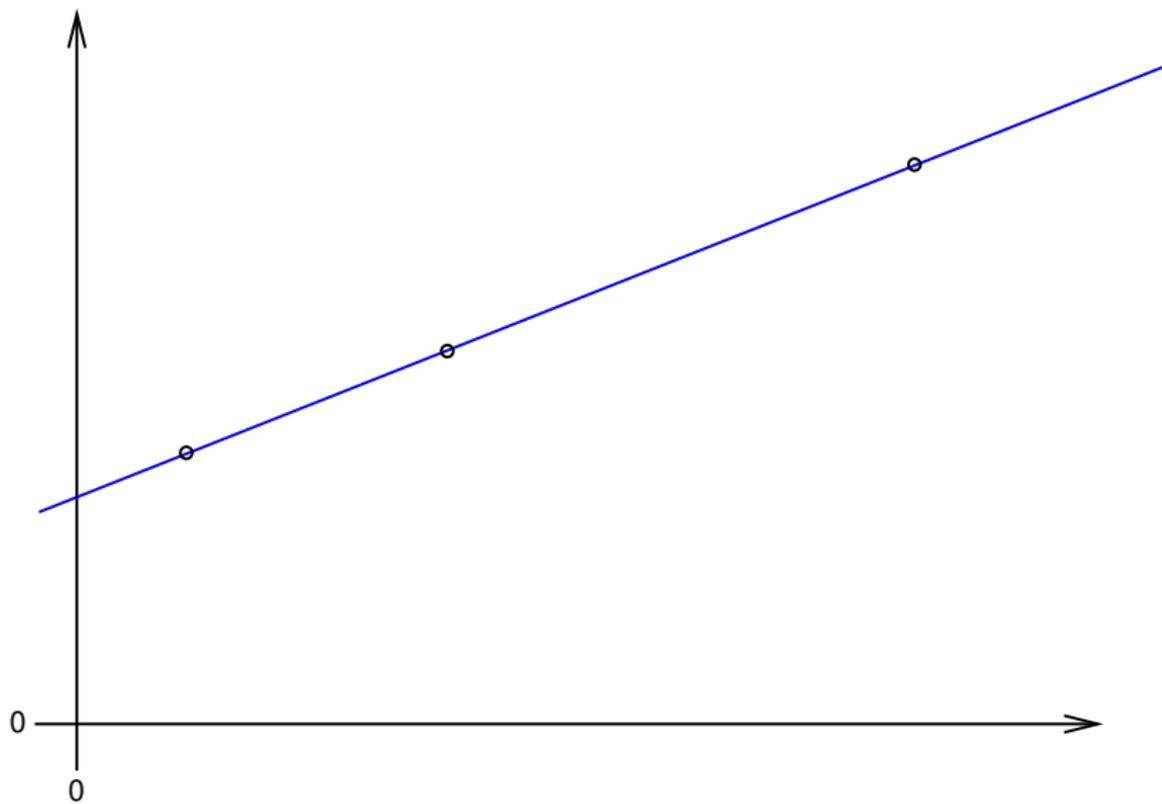
Beispiel: Herzfrequenz und Stoffwechselrate beim Gänsegeier

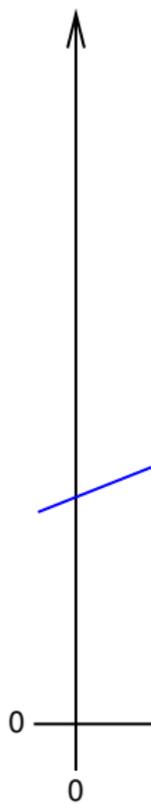
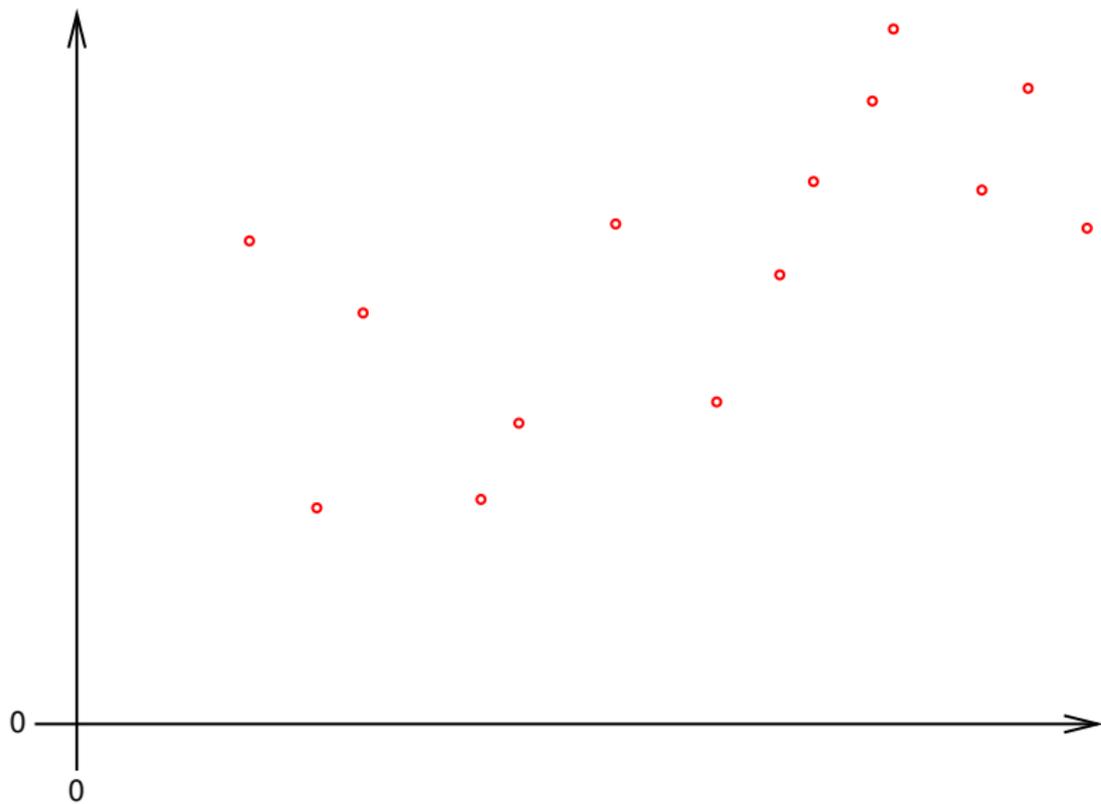
griffon vulture, 17.05.99, 16 degrees C



griffon v







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 ε_i unabhängig und normalverteilt sind und alle dieselbe Varianz σ^2 haben.

gegebene Daten:

Y	X
y_1	x_1
y_2	x_2
y_3	x_3
\vdots	\vdots
y_n	x_n

Modell: es gibt Zahlen
 a, b, σ^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 $y_i \sim \mathcal{N}(a + b \cdot x_i, \sigma^2)$.

a, b, σ^2 sind unbekannt, aber **nicht zufällig**.

Wir schätzen a und b , indem wir

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

Theorem

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

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

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 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

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, σ^2 zufällig: ε, 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

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

Problem: Wir kennen σ^2 nicht.

Wir schätzen σ^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 σ^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.

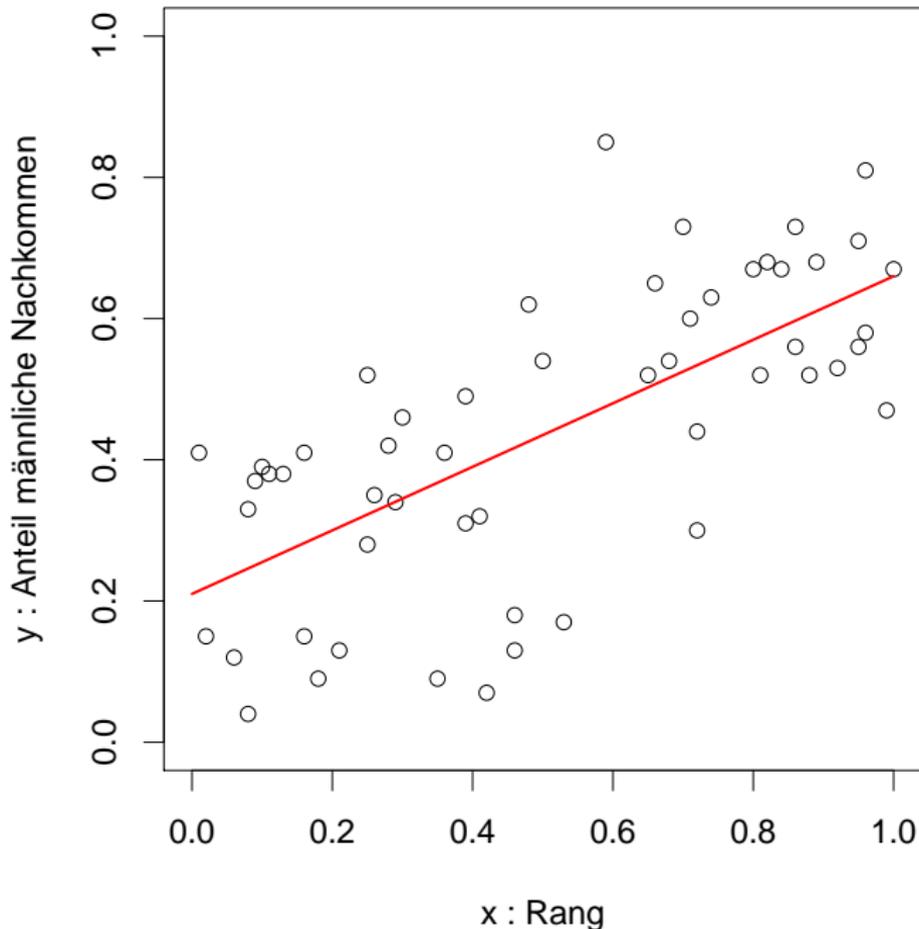
Verwerfe $H_0: „b = 0“$ zum Signifikanzniveau α , wenn

$\left| \frac{\hat{b}}{s / \sqrt{\sum_i (x_i - \bar{x})^2}} \right| \geq q_{1-\alpha/2}$, 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.

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



Es ist

$$\bar{x} = 0.51$$

(mittl. Rang)

$$\bar{y} = 0.44$$

(mittl. Ant. männl.
Nachk.)

$$\sigma_x^2 = 0.097$$

$$\text{cov}(x, y) = 0.044$$

$$\hat{b} = \frac{0.044}{0.097} \doteq 0.45$$

$$\hat{a}$$

$$= 0.44 - 0.51 \cdot 0.45$$

$$\doteq 0.21$$

Im Rothirschkühe-Beispiel ($n = 54$ Datenpunkte):

$$\hat{b} = 0.45, \frac{s_{\text{res}}}{\sqrt{\sum_i (x_i - \bar{x})^2}} = 0.0673,$$

also beobachten wir $t = \frac{\hat{b} - 0}{s_{\text{res}} / \sqrt{\sum_i (x_i - \bar{x})^2}} = 6.7$

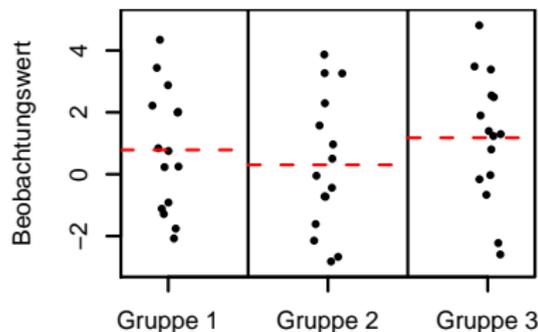
Einer Tabelle entnehmen wir: Das 99.95%-Quantil der Student-Verteilung mit 50 Freiheitsgraden ist 3.496 (und das der Student-Vert. mit 60 Freiheitsgraden ist 3.460).

Wir können also die Nullhypothese „das wahre $b = 0$ “ zum Signifikanzniveau 0.1% ablehnen.

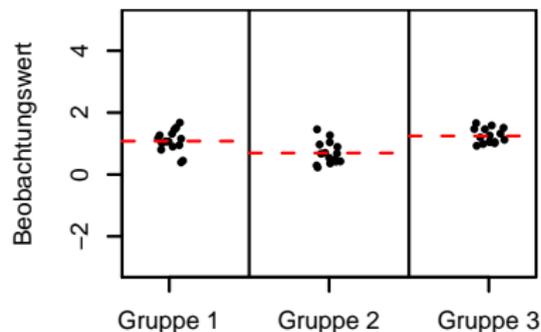
Bemerkung: Das beweist natürlich nicht, dass Hirschkühe das Geschlecht ihrer Nachkommen willentlich bestimmen können. Es scheint eher plausibel anzunehmen, dass es Faktoren gibt, die den Rang und die Geschlechterverteilung der Nachkommen zugleich beeinflussen, siehe die Diskussion in dem zitierten Artikel von T. H. Clutton-Brock et. al.

Grundidee der Varianzanalyse

Sind beobachtete Unterschiede von Gruppenmittelwerten ernst zu nehmen — oder könnte das alles Zufall sein?



Variabilität innerhalb
der Gruppen groß



Variabilität innerhalb
der Gruppen klein

Das hängt vom Verhältnis der Variabilität der Gruppenmittelwerte und der Variabilität der Beobachtungen innerhalb der Gruppen ab:

Quantifiziert im F -Test

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 ε_{ij} , $\mathbb{E}[\varepsilon_{ij}] = 0$, $\text{Var}[\varepsilon_{ij}] = \sigma^2$
 (μ_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, 2

X_{ij} = j -te Beobachtung in der i -ten Gruppe, $j = 1, \dots, n_i$,

Modellannahme: $X_{ij} = \mu_i + \varepsilon_{ij}$. $\mathbb{E}[\varepsilon_{ij}] = 0$, $\text{Var}[\varepsilon_{ij}] = \sigma^2$

$SS_{\text{innerh}} = \sum_{i=1}^I \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2$ Quadratsumme innerhalb d. Gruppen,
 $n - I$ Freiheitsgrade

$SS_{\text{zw}} = \sum_{i=1}^I n_i (\bar{X}_{i.} - \bar{X}_{..})^2$ 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 μ_i sind gleich“) ist F Fisher-verteilt mit $I - 1$ und $n - I$ Freiheitsgraden (unabhängig vom tatsächlichen gemeinsamen Wert der μ_i).

F-Test: Wir lehnen H_0 zum Signifikanzniveau α ab, wenn $F \geq q_\alpha$, wobei q_α das $(1 - \alpha)$ -Quantil der Fisher-Verteilung mit $I - 1$ und $n - I$ Freiheitsgraden ist.

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

Zusammengefasst in 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,

das 95%-Quantil der $F_{3,20}$ -Verteilung ist 3,098 ($< 13,57$).

Wir können demnach H_0 zum Signifikanzniveau 5% ablehnen.

Multipl. Testen und Bonferroni-Korrektur

Wenn m Tests zum *multiplen Signifikanzniveau* $\alpha \in (0, 1)$ durchgeführt werden sollen,

so führe jeden Test für sich zum *lokalen Signifikanzniveau* $\frac{\alpha}{m}$ durch.

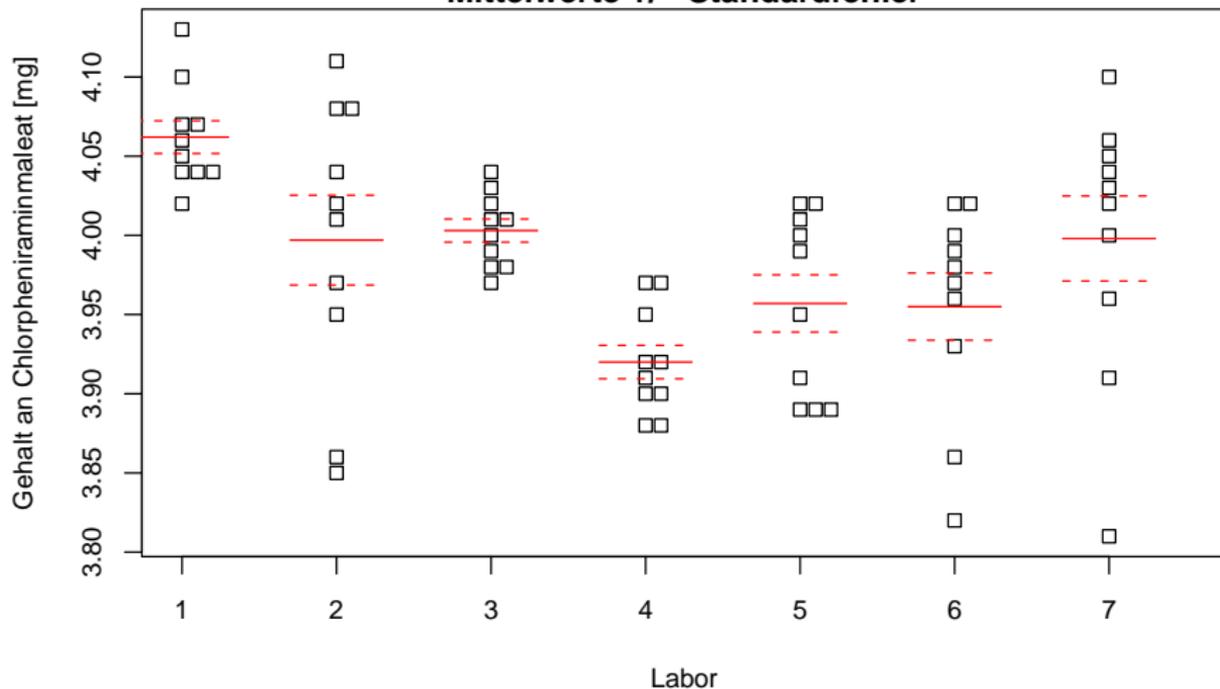
(Alternativ bedeutet dies: Multipliziere jeden (individuellen) p -Wert mit der Anzahl m der durchgeführten Tests.

[denn wenn die jeweilige Nullhypothese zutrifft, so ist der p -Wert uniform verteilt in $[0, 1]$)

Dann gilt: Die Wahrscheinlichkeit, dass *irgendeine zutreffende* Nullhypothese zu Unrecht ablehnt wird,
beträgt höchstens α .

Beispiel:

**7 verschiedene Labors haben jeweils 10 Messungen des Chlorpheniraminmaleat-Gehalts von Medikamentenproben vorgenommen:
Mittelwerte \pm Standardfehler**



Labor-Vergleichs-Beispiel mit Bonferroni-Korrektur

Wert der t -Statistik aus paarweisen Vergleichen mittels t -Tests:

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Betrachte $\alpha = 5\%$. Hier $m = 21$, das $(1 - \frac{1}{2} \frac{\alpha}{m})$ -Quantil

$(1 - \frac{1}{2} \frac{\alpha}{m} = 0.99881)$ der

t -Verteilung mit 18 Freiheitsgraden ist 3.532,

also können wir für die **rot markierten** Paare $H_{0,(i,j)}$ zum multiplen Signifikanzniveau 5% ablehnen.

Viel Erfolg beim Lernen!