Statistik für Informatiker, SS 2021

2. Ideen aus der Statistik

2.2 Schätzer und Tests für (Populations-)Mittelwerte

Matthias Birkner

http://www.staff.uni-mainz.de/birkner/StatInfo21/

Eine „Standardsituation“ in der Statistik:

Wir haben n Beobachtungswerte $x_1, \ldots, x_n$ (z.B. die Carapaxlängen der gefangenen Springkrebse aus dem Einstiegsbeispiel zur deskriptiven Statistik)

Eine „Standardsituation“ in der Statistik:

Wir haben n Beobachtungswerte $x_1, \ldots, x_n$ (z.B. die Carapaxlängen der gefangenen Springkrebse aus dem Einstiegsbeispiel zur deskriptiven Statistik)
und möchten anhand dessen etwas über die „Gesamtpopulation“, aus der die Daten stammen, aussagen.

Eine „Standardsituation“ in der Statistik:

Wir haben n Beobachtungswerte $x_1, \ldots, x_n$ (z.B. die Carapaxlängen der gefangenen Springkrebse aus dem Einstiegsbeispiel zur deskriptiven Statistik)
und möchten anhand dessen etwas über die „Gesamtpopulation“, aus der die Daten stammen, aussagen.

Die grundlegende Idee: Die Daten werden als Realisierungen von Zufallsvariablen aufgefasst, die in einem stochastischen Modell spezifiert werden.

Man versucht dann, anhand der Daten Rückschlüsse auf Parameter des Modells zu ziehen, und so systematische Effekte von Zufälligem zu trennen.

Wir formulieren dies folgendermaßen („abstrakt“):

Definition 2.1

Ein *statistisches Modell* ist ein Tripel $(\Omega, \mathcal{F}, (P_\vartheta)_{\vartheta\in\Theta})$, wobei Ω der „Stichprobenraum“, $\mathcal{F} \subset 2^\Omega$ eine σ-Algebra über Ω, Θ eine (möglicherweise recht allgemeine) Indexmenge und jedes P_ϑ ein Wahrscheinlichkeitsmaß (auf $(\Omega, \mathcal{F})$), (vgl. Def. 1.1)

$\vartheta \in \Theta$

Wir formulieren dies folgendermaßen („abstrakt“):

Definition 2.1

Ein *statistisches Modell* ist ein Tripel $(\Omega, \mathcal{F}, (P_\vartheta)_{\vartheta\in\Theta})$, wobei Ω der „Stichprobenraum“, $\mathcal{F} \subset 2^\Omega$ eine σ-Algebra über Ω, Θ eine (möglicherweise recht allgemeine) Indexmenge und jedes P_ϑ ein Wahrscheinlichkeitsmaß (auf $(\Omega, \mathcal{F})$), (vgl. Def. 1.1)

Die verschiedenen P_ϑ beschreiben unterschiedliche Arten, „wie der Zufall wirken kann“ (d.h. unterschiedliche Verteilungen der Zufallsvariablen auf Ω). Im Kontext des Modells nehmen wir an, dass die Daten gemäß einem gewissen P_ϑ generiert wurden, wir kennen aber das „wahre“ ϑ nicht.

Wir formulieren dies folgendermaßen („abstrakt“):

Definition 2.1

Ein *statistisches Modell* ist ein Tripel $(\Omega, \mathcal{F}, (P_\vartheta)_{\vartheta\in\Theta})$, wobei Ω der „Stichprobenraum“, $\mathcal{F} \subset 2^\Omega$ eine σ-Algebra über Ω, Θ eine (möglicherweise recht allgemeine) Indexmenge und jedes P_ϑ ein Wahrscheinlichkeitsmaß (auf $(\Omega, \mathcal{F})$), (vgl. Def. 1.1)

Die verschiedenen P_ϑ beschreiben unterschiedliche Arten, „wie der Zufall wirken kann“ (d.h. unterschiedliche Verteilungen der Zufallsvariablen auf Ω). Im Kontext des Modells nehmen wir an, dass die Daten gemäß einem gewissen P_ϑ generiert wurden, wir kennen aber das „wahre“ ϑ nicht.

Die Aufgabe der (schließenden) Statistik besteht grob gesprochen darin, anhand der Beobachtungen Informationen über das „zugrundeliegende“ ϑ zu gewinnen.

Inhalt

Inhalt

Punktschätzer (für die Lageschätzung)

Modell: $X_1, \ldots, X_n$ unabhängige ZVn mit derselben Verteilung ϑ (die wir möglicherweise nicht kennen), mit Mittelwert μ ($= \mu(\vartheta)$) und Varianz σ^2 ($= \sigma^2(\vartheta)$)

$$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

$\overline{X}$ ist ein *Schätzer* für μ.

Punktschätzer (für die Lageschätzung)

Modell: $X_1, \ldots, X_n$ unabhängige ZVn mit derselben Verteilung ϑ (die wir möglicherweise nicht kennen), mit Mittelwert μ ($= \mu(\vartheta)$) und Varianz σ^2 ($= \sigma^2(\vartheta)$)

$$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

$\overline{X}$ ist ein *Schätzer* für μ.

$\overline{X}$ ist *erwartungstreu*:

$$\mathbb{E}_\vartheta\left[\overline{X}\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_\vartheta[X_i] = \mathbb{E}_\vartheta[X_1] = \mu = \mu(\vartheta) \quad \text{unter } P_\vartheta,$$

egal, was der tatsächliche Wert von μ ist

Lageschätzung

$X_1, \ldots, X_n$ unabhängige ZVn mit derselben Verteilung, mit Mittelwert μ ($= \mu(\vartheta)$) und Varianz σ^2 ($= \sigma^2(\vartheta)$)

$\overline{X}$ ist *konsistent*:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\text{stoch.}} \mu = \mu(\vartheta) \quad \text{unter } P_\vartheta$$

(gemäß dem Gesetz der großen Zahlen).

Lageschätzung

$X_1, \ldots, X_n$ unabhängige ZVn mit derselben Verteilung, mit Mittelwert μ (= $\mu(\vartheta)$) und Varianz σ^2 (= $\sigma^2(\vartheta)$)

$\overline{X}$ ist *konsistent*:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\text{stoch.}} \mu = \mu(\vartheta) \quad \text{unter } P_\vartheta$$

(gemäß dem Gesetz der großen Zahlen).

Allgemein:

Definition 2.2

Wir interessieren uns für eine gewisse (numerische) Eigenschaft $h(\vartheta)$ der Verteilung ϑ der X_i, eine Funktion Y der $X_1, \ldots, X_n$ ist ein erwartungstreuer Schätzer für $h(\vartheta)$, wenn stets gilt

$$\mathbb{E}_\vartheta[Y] = h(\vartheta) \quad \text{für } \vartheta \in \Theta$$

$$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

(ist ein „Punktschätzer“ für $\mu = \mathbb{E}_\vartheta[X_i]$ mit „guten“ Eigenschaften)

$$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

(ist ein „Punktschätzer“ für $\mu = \mathbb{E}_\vartheta[X_i]$ mit „guten“ Eigenschaften)

Soweit, so schön, aber: wie genau ist die Schätzung?

Es ist $\mathrm{Var}_\vartheta[\overline{X}] = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}_\vartheta[X_i] = \frac{\sigma^2}{n}$.

$$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

(ist ein „Punktschätzer“ für $\mu = \mathbb{E}_\vartheta[X_i]$ mit „guten“ Eigenschaften)

Soweit, so schön, aber: wie genau ist die Schätzung?

Es ist $\mathrm{Var}_\vartheta[\overline{X}] = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}_\vartheta[X_i] = \frac{\sigma^2}{n}$.

Wir machen zunächst „unser Leben leicht“:

Nehmen wir an, dass wir σ kennen, und dass die Daten aus einer Normalverteilung stammen
(im Modell: X_i u.i.v., $\sim \mathcal{N}_{\mu,\sigma^2}$).

Inhalt

Konfidenzintervall für den Mittelwert im normalen Modell bei bekannter Varianz

$\bar{X} \pm \ldots ?$

Beobachtung 2.3

$X_1, X_2, \ldots, X_n$ u.i.v. $\sim \vartheta = \mathcal{N}_{\mu,\sigma^2}$ mit (uns unbekanntem) $\mu \in \mathbb{R}$, $\sigma^2 > 0$ sei bekannt (und fest).

Sei $\alpha \in (0,1)$ [z.B. $\alpha = 0.05$ oder $\alpha = 0.01$], $q := \Phi^{-1}(1 - \frac{\alpha}{2})$ das $(1-\alpha/2)$-Quantil der Standardnormalverteilung [mit R : `qnorm(1 - α/2)`].

Konfidenzintervall für den Mittelwert im normalen Modell bei bekannter Varianz

Beobachtung 2.3

$X_1, X_2, \ldots, X_n$ u.i.v. $\sim \vartheta = \mathcal{N}_{\mu,\sigma^2}$ mit (uns unbekanntem) $\mu \in \mathbb{R}$, $\sigma^2 > 0$ sei bekannt (und fest).

Sei $\alpha \in (0,1)$ [z.B. $\alpha = 0.05$ oder $\alpha = 0.01$], $q := \Phi^{-1}(1 - \frac{\alpha}{2})$ das $(1-\alpha/2)$-Quantil der Standardnormalverteilung [mit R : `qnorm(1 - α/2)`].

Das (anhand der Daten konstruierte) Intervall

$$I := \left[\overline{X} - q \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + q \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Konfidenzintervall für den Mittelwert im normalen Modell bei bekannter Varianz

Beobachtung 2.3

$X_1, X_2, \ldots, X_n$ u.i.v. $\sim \vartheta = \mathcal{N}_{\mu,\sigma^2}$ mit (uns unbekanntem) $\mu \in \mathbb{R}$, $\sigma^2 > 0$ sei bekannt (und fest).

Sei $\alpha \in (0,1)$ [z.B. $\alpha = 0.05$ oder $\alpha = 0.01$], $q := \Phi^{-1}(1 - \frac{\alpha}{2})$ das $(1-\alpha/2)$-Quantil der Standardnormalverteilung [mit R : `qnorm(1 − α/2)`].

Das (anhand der Daten konstruierte) Intervall

$$I := \left[\overline{X} - q \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + q \cdot \frac{\sigma}{\sqrt{n}}\right]$$

hat die Eigenschaft

$$P_\vartheta(\mu \in I) = 1 - \alpha$$

(egal, was μ ist)

$\vartheta = (\mu, \sigma^2)$ bekannt

$$I := \left[\overline{X} - q \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + q \cdot \frac{\sigma}{\sqrt{n}}\right]$$

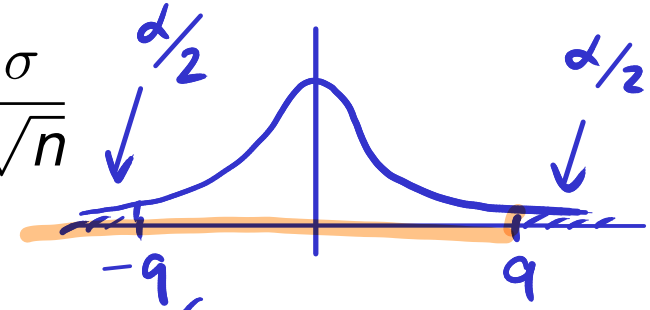
$$P_\vartheta\left(\mu \in I\right) = 1 - \alpha$$

denn

$$\overline{X} - q \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + q \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Longleftrightarrow$$

$$-q \leq \sqrt{n}\frac{\overline{X} - \mu}{\sigma} \leq q$$



(Erinnerung $\mathrm{Var}_\vartheta[\overline{X}] = \frac{\sigma^2}{n}$)

und (egal was μ und σ^2 sind) stets ist unter P_ϑ

$$\sqrt{n}\frac{\overline{X} - \mu}{\sigma} \sim \mathcal{N}_{0,1}$$

Konfidenzintervall

Definition 2.4

Allgemein heißt ein (anhand der Beobachtungswerte zu konstruierendes) Intervall I ein *Konfidenzintervall* (manchmal auch „Vertrauensintervall“) für μ zum (Sicherheits-)Niveau $1 - \alpha$ (bzw. Irrtumsniveau α), wenn gilt

$$\forall\, \vartheta \in \Theta \;:\quad P_\vartheta\big(\mu \in I\big) \geq 1 - \alpha.$$

(mit Notation $\mu = \mu(\vartheta)$).

Konfidenzintervall

Definition 2.4

Allgemein heißt ein (anhand der Beobachtungswerte zu konstruierendes) Intervall I ein *Konfidenzintervall* (manchmal auch „Vertrauensintervall“) für μ zum (Sicherheits-)Niveau $1-\alpha$ (bzw. Irrtumsniveau α), wenn gilt

$$\forall\, \vartheta \in \Theta \;:\quad P_\vartheta\big(\mu \in I\big) \geq 1-\alpha.$$

(mit Notation $\mu = \mu(\vartheta)$).

Beachte: I ist zufällig, nicht aber ϑ (zumindest in unserer (sogenannten frequentistischen) Interpretation).

Konfidenzintervall

Definition 2.4

Allgemein heißt ein (anhand der Beobachtungswerte zu konstruierendes) Intervall I ein *Konfidenzintervall* (manchmal auch „Vertrauensintervall“) für μ zum (Sicherheits-)Niveau $1-\alpha$ (bzw. Irrtumsniveau α), wenn gilt

$$\forall\, \vartheta \in \Theta : \quad P_\vartheta\big(\mu \in I\big) \geq 1-\alpha.$$

(mit Notation $\mu = \mu(\vartheta)$).

Beachte: I ist zufällig, nicht aber ϑ (zumindest in unserer (sogenannten frequentistischen) Interpretation).

Offenbar möchte man i.A. I so kurz wie möglich wählen (soweit verträglich mit dem geforderten Niveau).

Bemerkung. Auch wenn die X_i nicht wörtlich normalverteilt sind (sondern unabhängig und identisch verteilt mit einer gewissen Verteilung ϑ, die Mittelwert μ und Varianz σ^2 hat), so ist

$$\sqrt{n}\frac{\overline{X}-\mu}{\sigma} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \overset{d}{\approx} \mathcal{N}_{0,1}$$

gemäß dem zentralen Grenzwertsatz.

Bemerkung. Auch wenn die X_i nicht wörtlich normalverteilt sind (sondern unabhängig und identisch verteilt mit einer gewissen Verteilung ϑ, die Mittelwert μ und Varianz σ^2 hat), so ist

$$\sqrt{n}\frac{\overline{X}-\mu}{\sigma} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \overset{d}{\approx} \mathcal{N}_{0,1}$$

gemäß dem zentralen Grenzwertsatz.

Daher ist auch für allgemeine Verteilung das auf Normalität fußende Konfidenzintervall zumindest für große n „approximativ korrekt“

Bemerkung. Auch wenn die X_i nicht wörtlich normalverteilt sind (sondern unabhängig und identisch verteilt mit einer gewissen Verteilung ϑ, die Mittelwert μ und Varianz σ^2 hat), so ist

$$\sqrt{n}\frac{\overline{X}-\mu}{\sigma} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \overset{d}{\approx} \mathcal{N}_{0,1}$$

gemäß dem zentralen Grenzwertsatz.

Daher ist auch für allgemeine Verteilung das auf Normalität fußende Konfidenzintervall zumindest für große n „approximativ korrekt“

(man muss aber σ kennen, um es zu konstruieren).

Inhalt

Wenn man die Varianz nicht kennt ...

Wenn man die Varianz nicht kennt ...

... kann man sie immerhin schätzen:

$$S^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

[im Sinne der deskriptiven Statistik wäre dies die „korrigierte Stichprobenvarianz“]

Wenn man die Varianz nicht kennt ...

... kann man sie immerhin schätzen:

$$S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

[im Sinne der deskriptiven Statistik wäre dies die „korrigierte Stichprobenvarianz“]
ist erwartungstreuer Schätzer für σ^2:

$$\mathbb{E}_\vartheta\Big[\sum_{i=1}^{n}(X_i - \overline{X})^2\Big] = n\mathbb{E}_\vartheta\big[(X_i - \overline{X})^2\big] = n\mathrm{Var}_\vartheta\big[X_i - \overline{X}\big]$$

$$= n\mathrm{Var}_\vartheta\Big[\frac{n-1}{n}X_1 - \frac{1}{n}\sum_{i=2}^{n}X_i\Big]$$

$$= n\Big(\big(\tfrac{n-1}{n}\big)^2\mathrm{Var}_\vartheta[X_1] + \tfrac{n-1}{n^2}\mathrm{Var}_\vartheta[X_1]\Big) = (n-1)\mathrm{Var}_\vartheta[X_1] = (n-1)\sigma^2$$

$\frac{1}{n}\sum_{j=1}^{n}X_j = \frac{1}{n}X_1 + \text{Rest}$

S^2 ist konsistent (als Folge in n betrachtet):

$$\frac{n-1}{n} S^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2\right) - 2\left(\frac{1}{n} \sum_{i=1}^{n} X_i \overline{X}\right) + \left(\overline{X}\right)^2$$

$$= \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2\right) - \left(\overline{X}\right)^2 \overset{\text{stoch.}}{\underset{n\to\infty}{\longrightarrow}} \mathbb{E}_\vartheta\left[X_1^2\right] - \left(\mathbb{E}_\vartheta\left[X_1\right]\right)^2 = \sigma^2$$

(verwende jeweils das Gesetz der großen Zahlen).

S^2 ist konsistent (als Folge in n betrachtet):

$$\frac{n-1}{n} S^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2\right) - 2\left(\frac{1}{n} \sum_{i=1}^{n} X_i \overline{X}\right) + \left(\overline{X}\right)^2$$

$$= \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2\right) - \left(\overline{X}\right)^2 \overset{\text{stoch.}}{\underset{n\to\infty}{\longrightarrow}} \mathbb{E}_\vartheta\left[X_1^2\right] - \left(\mathbb{E}_\vartheta\left[X_1\right]\right)^2 = \sigma^2$$

(verwende jeweils das Gesetz der großen Zahlen).

Wir wissen bereits: Die Standardabweichung von $\overline{X}$ ist $\frac{\sigma}{\sqrt{n}}$.

S^2 ist konsistent (als Folge in n betrachtet):

$$\frac{n-1}{n}S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right) - 2\left(\frac{1}{n}\sum_{i=1}^{n}X_i\overline{X}\right) + \left(\overline{X}\right)^2$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right) - \left(\overline{X}\right)^2 \overset{\text{stoch.}}{\underset{n\to\infty}{\longrightarrow}} \mathbb{E}_\vartheta\left[X_1^2\right] - \left(\mathbb{E}_\vartheta\left[X_1\right]\right)^2 = \sigma^2$$

(verwende jeweils das Gesetz der großen Zahlen).

Wir wissen bereits: Die Standardabweichung von $\overline{X}$ ist $\frac{\sigma}{\sqrt{n}}$.
Somit ist

$$\frac{S}{\sqrt{n}}$$

ein (naheliegender) Schätzer für die (unbekannte) Standardabweichung von $\overline{X}$

S^2 ist konsistent (als Folge in n betrachtet):

$$\frac{n-1}{n}S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right) - 2\left(\frac{1}{n}\sum_{i=1}^{n}X_i\overline{X}\right) + \left(\overline{X}\right)^2$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right) - \left(\overline{X}\right)^2 \overset{\text{stoch.}}{\underset{n\to\infty}{\longrightarrow}} \mathbb{E}_\vartheta\left[X_1^2\right] - \left(\mathbb{E}_\vartheta\left[X_1\right]\right)^2 = \sigma^2$$

(verwende jeweils das Gesetz der großen Zahlen).

Wir wissen bereits: Die Standardabweichung von $\overline{X}$ ist $\frac{\sigma}{\sqrt{n}}$.
Somit ist

$$\frac{S}{\sqrt{n}}$$

ein (naheliegender) Schätzer für die (unbekannte) Standardabweichung von $\overline{X}$,

man nennt es auch den *Standardfehler*.

Satz 2.5 (Student (=William Gosset), 1908))

$X_1, \ldots, X_n$ u.i.v. $\sim \mathcal{N}_{\mu,\sigma^2}$ mit $\mu \in \mathbb{R}$, $\sigma > 0$,

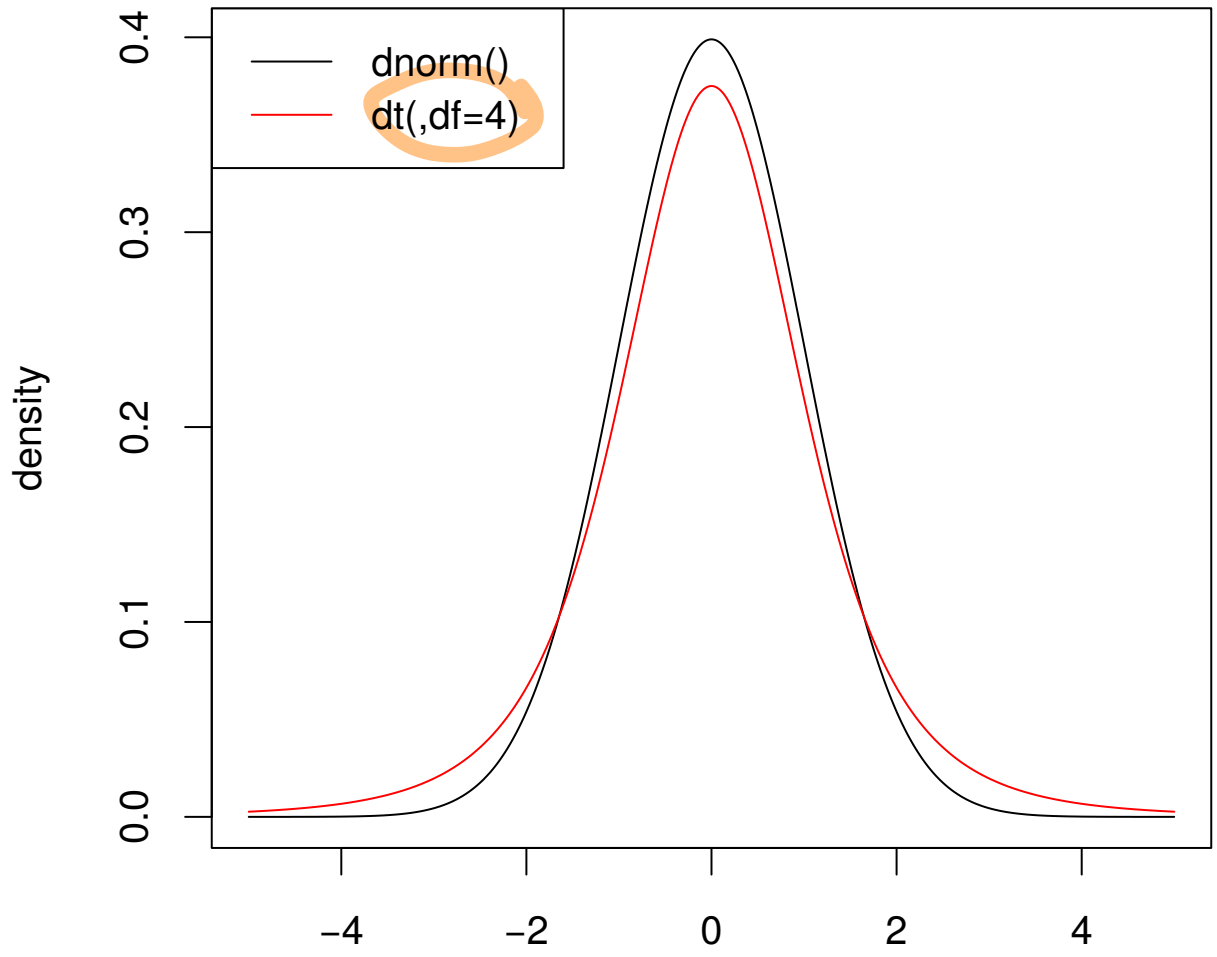
$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Dann besitzt $T := \dfrac{\sqrt{n}(\overline{X} - \mu)}{\sqrt{S^2}}$ *die Dichte* $t_{n-1}(x)$*, wobei*

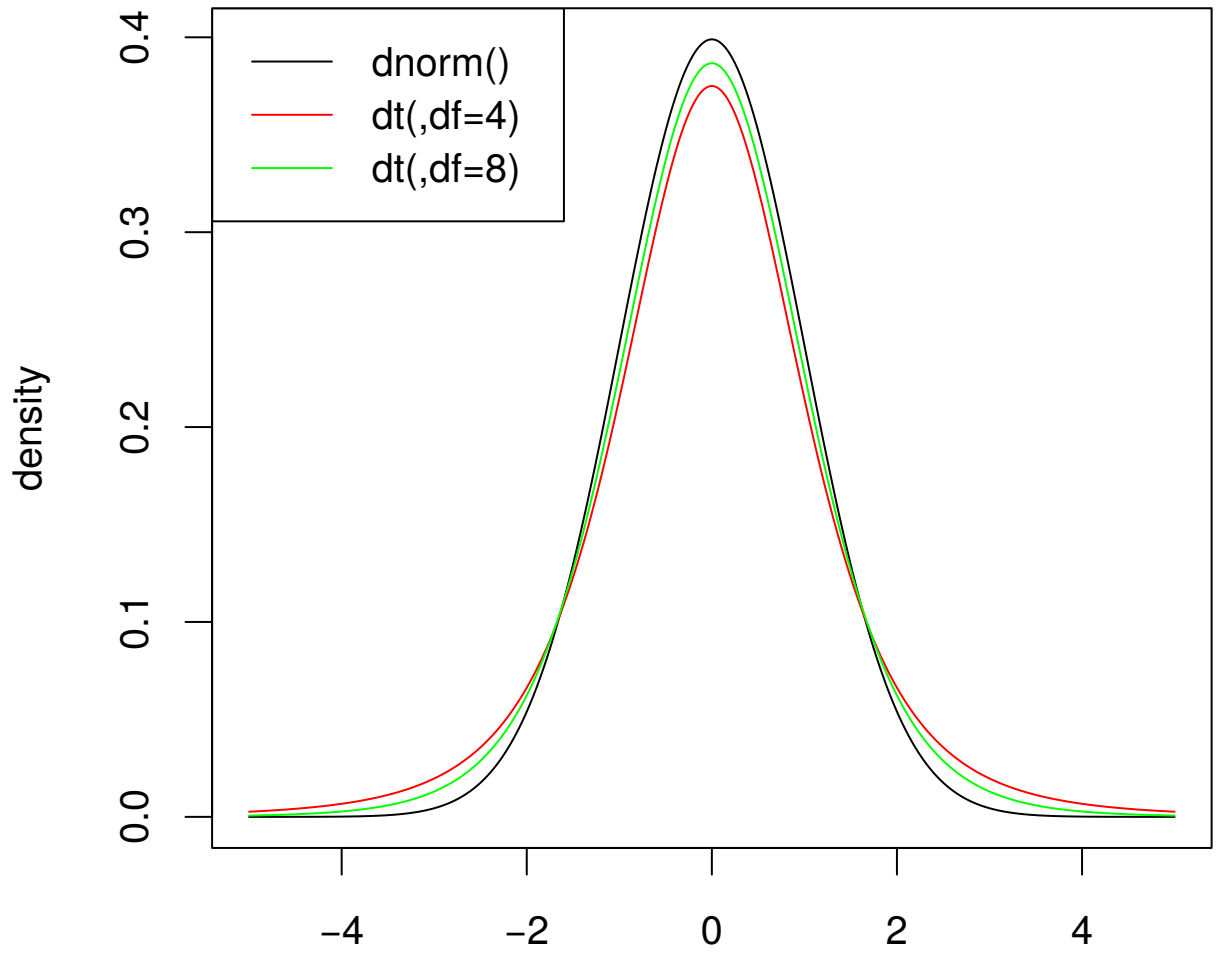
$$t_m(x) = \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{m}{2})}\frac{1}{\sqrt{m}}\Big(1 + \frac{x^2}{m}\Big)^{-\frac{m+1}{2}}$$

t_m ist die Dichte der Student-t-Verteilung mit m Freiheitsgraden (manchmal auch Student-Verteilung oder t-Verteilung genannt).

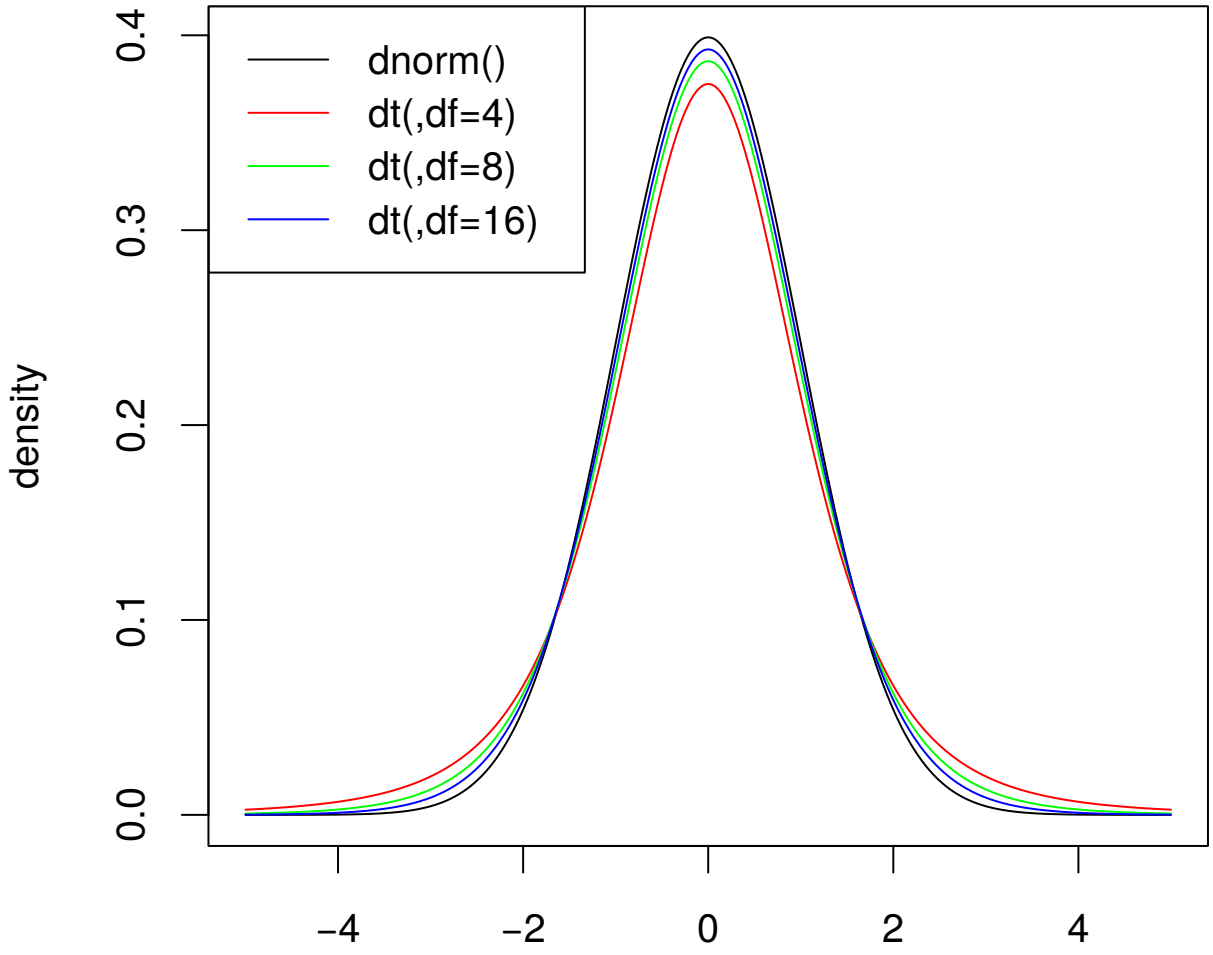
Dichte der t-Verteilung



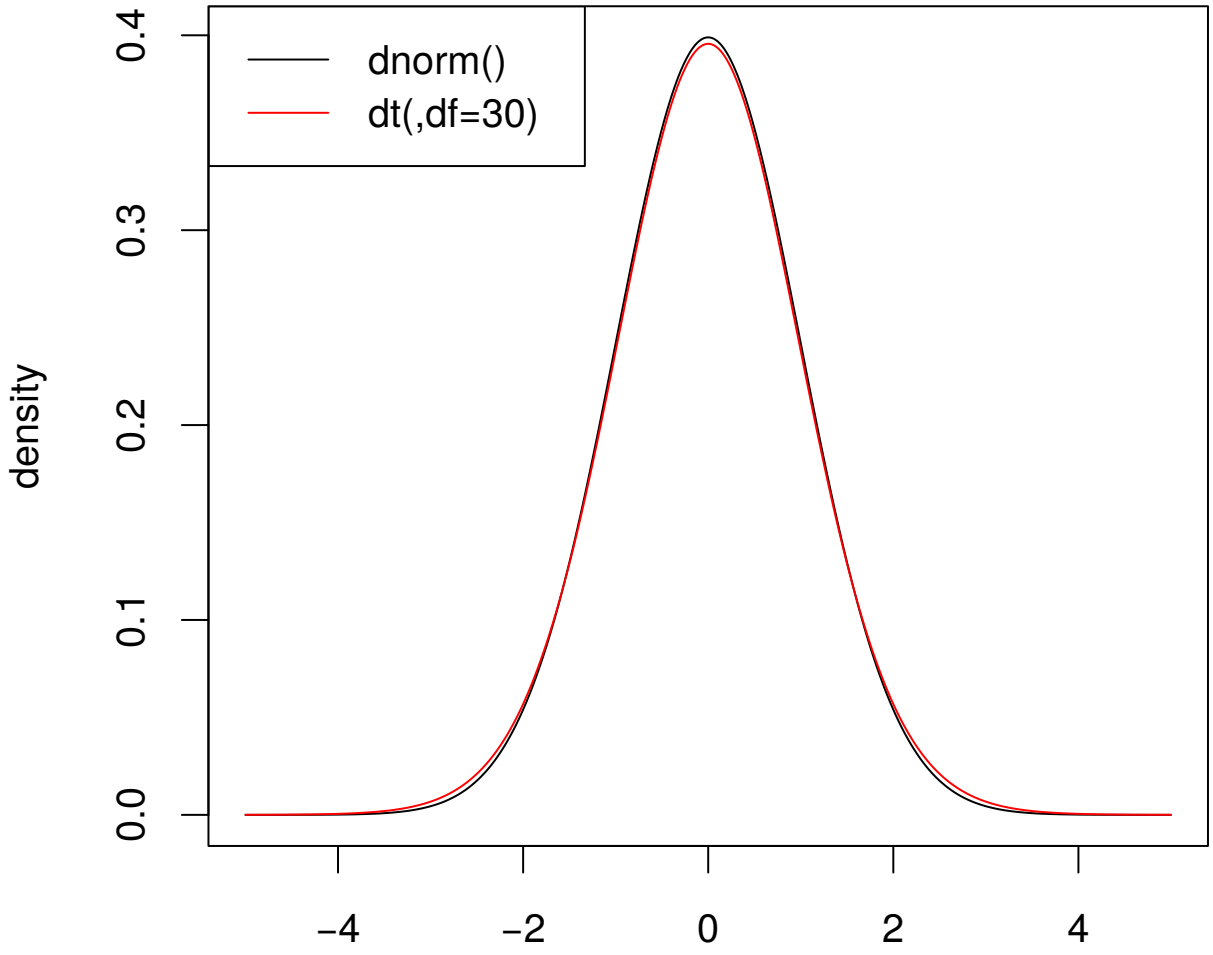
Dichte der t-Verteilung



Dichte der t-Verteilung



Dichte der t-Verteilung



Bemerkung.
Die t-Verteilung wurde von William Gosset erforscht und von ihm 1908 veröffentlicht, während er in einer Guinness-Brauerei arbeitete. Da sein Arbeitgeber die Veröffentlichung nicht gestattete, veröffentlichte Gosset sie unter dem Pseudonym *Student*.

Bemerkung.
Die *t*-Verteilung wurde von William Gosset erforscht und von ihm 1908 veröffentlicht, während er in einer Guinness-Brauerei arbeitete. Da sein Arbeitgeber die Veröffentlichung nicht gestattete, veröffentlichte Gosset sie unter dem Pseudonym *Student*.

Beweise finden sich in der Literatur, siehe z.B. Kersting & Wakolbinger, Kap.22 oder Georgii, Satz 9.17

(Es gibt dazu ein schönes geometrisches Argument, das die Rotationsinvarianz der multivariaten Normalverteilung ausnutzt.)

Freiheitsgrade?

Beispiel: Es gibt 5 Freiheitsgrade im Vektor

$$x = (x_1, x_2, x_3, x_4, x_5)$$

da 5 Werte frei wählbar sind. Der Vektor

$$v := x - \overline{x} := (x_1 - \overline{x}, x_2 - \overline{x}, x_3 - \overline{x}, x_4 - \overline{x}, x_5 - \overline{x})$$

hat 4 Freiheitsgrade, denn nach Wahl von v_1, v_2, v_3, v_4 ist v_5 festgelegt wegen $v_1 + \cdots + v_4 + v_5 = 0$.

Freiheitsgrade?

Beispiel: Es gibt 5 Freiheitsgrade im Vektor

$$x = (x_1, x_2, x_3, x_4, x_5)$$

da 5 Werte frei wählbar sind. Der Vektor

$$v := x - \overline{x} := (x_1 - \overline{x}, x_2 - \overline{x}, x_3 - \overline{x}, x_4 - \overline{x}, x_5 - \overline{x})$$

hat 4 Freiheitsgrade, denn nach Wahl von v_1, v_2, v_3, v_4 ist v_5 festgelegt wegen $v_1 + \cdots + v_4 + v_5 = 0$.

Die Bezeichnung „Student-Verteilung mit $n-1$ Freiheitsgraden“ ist motiviert durch die Tatsache, dass $t = (\overline{x} - \mu)/(s/\sqrt{n})$, wo $s = \sqrt{s^2} = \frac{1}{\sqrt{n-1}}\Big((x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2\Big)^{1/2}$ (bis auf Normierung) die Länge eines „Vektors mit $n-1$ Freiheitsgraden“ ist.

Student-Konfidenzintervall für den Erwartungswert im normalen Modell

Korollar 2.6

Unter P_ϑ, $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$ *seien*

$$X_1, X_2, \ldots, X_n \text{ u.i.v. } \sim \mathcal{N}_{\mu,\sigma^2}.$$

Sei $\alpha \in (0,1)$, $q = q_{n-1,1-\alpha/2}$ *das* $1 - \frac{\alpha}{2}$*-Quantil der Student-Verteilung mit* $n-1$ *Freiheitsgraden [mit* `R`*:* `qt(1-α/2, df=n-1)`*],*

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Dann ist

$$I := \left[\overline{X} - q\sqrt{\frac{S^2}{n}}, \overline{X} + q\sqrt{\frac{S^2}{n}}\right]$$

ein Konfidenzintervall für μ *zum Irrtumsniveau* α.

$$I := \left[\overline{X} - q\sqrt{\frac{S^2}{n}}, \overline{X} + q\sqrt{\frac{S^2}{n}} \right]$$

ist ein Konfidenzintervall für μ zum Irrtumsniveau α

$$I := \left[\overline{X} - q\sqrt{\frac{S^2}{n}}, \overline{X} + q\sqrt{\frac{S^2}{n}} \right]$$

ist ein Konfidenzintervall für μ zum Irrtumsniveau α,

denn $T := \frac{\sqrt{n}(\overline{X}-\mu)}{\sqrt{S^2}}$ ist Student-verteilt mit $n-1$ Freiheitsgraden (für jede Wahl von μ und σ^2), also

$$P_{(\mu,\sigma^2)}\left(\overline{X} - q\sqrt{\frac{S^2}{n}} \le \mu \le \overline{X} + q\sqrt{\frac{S^2}{n}}\right)$$
$$= P(-q \le T \le q) = P(T \le q) - P(T \le -q) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha.$$

Beispiel. 10 Patienten erhielten in aufeinanderfolgenden Nächten Schlafmittel A, dann Mittel B.

Die Daten[1] (x_i = Anz. Stunden Schlaf mit Mittel A - Anz. Stunden Schlaf mit Mittel B bei Patient Nr. i):

i	1	2	3	4	5	6	7	8	9	10
x_i	1,2	2,4	1,3	1,3	0,0	1,0	1,8	0,8	4,6	1,4

[1] Aus Student (William S. Gosset), The Probable Error of a Mean, Biometrika 6:1–25 (1908)

Beispiel. 10 Patienten erhielten in aufeinanderfolgenden Nächten Schlafmittel A, dann Mittel B.

Die Daten[1] (x_i = Anz. Stunden Schlaf mit Mittel A - Anz. Stunden Schlaf mit Mittel B bei Patient Nr. i):

i	1	2	3	4	5	6	7	8	9	10
x_i	1,2	2,4	1,3	1,3	0,0	1,0	1,8	0,8	4,6	1,4

Es ist

$$\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i \approxeq 1{,}58, \quad s = \left(\frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x})^2\right)^{1/2} \approxeq 1{,}23.$$

[1] Aus Student (William S. Gosset), The Probable Error of a Mean, Biometrika 6:1–25 (1908)

$$\overline{x} \approxeq 1{,}58, \quad s \approxeq 1{,}23$$

Es ist $q_{9,\,0{,}995} \approxeq 3{,}25$ (aus einer Quantiltabelle oder beispielsweise mit R berechnet), demnach ist

$$\left[\overline{x} \pm q\frac{s}{\sqrt{n}}\right] \approxeq [0{,}31,\ 2{,}85]$$

ein Konfidenzintervall für μ (die mittlere zusätzliche Anzahl Stunden Schlaf, die Medikament A mehr bringt als Medikament B) zum Sicherheitsniveau $0{,}99 = 1 - 0{,}01$.

$$\overline{x} \approxeq 1{,}58, \quad s \approxeq 1{,}23$$

Es ist $q_{9,\,0{,}995} \approxeq 3{,}25$ (aus einer Quantiltabelle oder beispielsweise mit R berechnet), demnach ist

$$\left[\overline{x} \pm q\frac{s}{\sqrt{n}}\right] \approxeq [0{,}31,\ 2{,}85]$$

ein Konfidenzintervall für μ (die mittlere zusätzliche Anzahl Stunden Schlaf, die Medikament A mehr bringt als Medikament B) zum Sicherheitsniveau $0{,}99 = 1 - 0{,}01$.

Beachte: (Sinnlos) genaue Werte mit Rechnergenauigkeit sind $\overline{x} - q\frac{s}{\sqrt{n}} \approxeq 0{,}3159481$, $\overline{x} + q\frac{s}{\sqrt{n}} \approxeq 2{,}8440519$, man sollte allerdings die Grenzen eines Konfidenzintervalls stets „konservativ“, d.h. nach außen, runden.

Inhalt

Inhalt

Die Fragestellung statistischer Tests

Sei $(\Omega, \mathcal{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell, $\Theta = \Theta_0 \dot\cup \Theta_1$ disjunkte Zerlegung in „Nullhypothese“ und „Alternative“ (auch „Gegenhypothese“).

Die Fragestellung statistischer Tests

Sei $(\Omega, \mathcal{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell, $\Theta = \Theta_0 \dot\cup \Theta_1$ disjunkte Zerlegung in „Nullhypothese“ und „Alternative“ (auch „Gegenhypothese“).

Wir suchen ein Verfahren, um anhand von Beobachtungswerten $x_1, \ldots, x_n$ zu entscheiden, ob wir

$$H_0 : \vartheta \in \Theta_0 \quad \text{(die Nullhypothese) oder}$$
$$H_1 : \vartheta \in \Theta_1 \quad \text{(die Alternative)}$$

für plausibler halten (und zugleich eine geeignete Art, „plausibler“ zu quantifizieren).

Die Fragestellung statistischer Tests

Abstrakt ist ein statistischer Test von Θ_0 gegen Θ_1 eine Funktion $\varphi(x_1, \ldots, x_n)$ der Beobachtungen mit Werten $\{0, 1\}$ und der Entscheidungsregel:

$\varphi(x_1, \ldots, x_n) = 0$ dann	entscheide für H_0 („behalte die Nullhypothese bei“)
$\varphi(x_1, \ldots, x_n) = 1$ dann	entscheide für H_1 („verwirf die Nullhypothese zugunsten der Alternative“)

Die Fragestellung statistischer Tests

Abstrakt ist ein statistischer Test von Θ_0 gegen Θ_1 eine Funktion $\varphi(x_1, \ldots, x_n)$ der Beobachtungen mit Werten $\{0, 1\}$ und der Entscheidungsregel:

$\varphi(x_1, \ldots, x_n) = 0$ dann	entscheide für H_0 („behalte die Nullhypothese bei“)
$\varphi(x_1, \ldots, x_n) = 1$ dann	entscheide für H_1 („verwirf die Nullhypothese zugunsten der Alternative“)

Sei $\alpha \in (0, 1)$. Der Test φ hat *Niveau* α, wenn gilt

$$\sup_{\vartheta \in \Theta_0} P_\vartheta\big(\varphi(X_1, \ldots, X_n) = 1\big) \leq \alpha$$

(d.h. die W’keit für einen „Fehler 1. Art“ (die Nullhypothese fälschlicherweise zu verwerfen) ist $\leq \alpha$)

Die Fragestellung statistischer Tests

Zugleich ist wünschenswert:

$$P_\vartheta\big(\varphi(X_1,\ldots,X_n)=0\big) \stackrel{!}{=} \text{möglichst klein} \quad \text{für } \vartheta \in \Theta_1,$$

die W'keit für einen „Fehler 2. Art" (die Nullhypothese fälschlicherweise zu akzeptieren) sollte klein sein.

Dann hat φ große „Schärfe" oder „Macht".

Statistischer Tests: Niveau vs. p-Wert

In der Praxis haben Tests meist die folgende Form:

Berechne eine gewisse (Test-)Statistik $Y = Y(x_1, \ldots, x_n)$, setze $\varphi(x_1, \ldots, x_n) = \mathbf{1}\big(Y(x_1, \ldots, x_n) > q\big)$ für einen gewissen Wert $q = q(\alpha)$, der in Abhängigkeit von den Parametern des Tests gewählt wird.

Statistischer Tests: Niveau vs. *p*-Wert

In der Praxis haben Tests meist die folgende Form:

Berechne eine gewisse (Test-)Statistik $Y = Y(x_1, \ldots, x_n)$, setze $\varphi(x_1, \ldots, x_n) = \mathbf{1}\big(Y(x_1, \ldots, x_n) > q\big)$ für einen gewissen Wert $q = q(\alpha)$, der in Abhängigkeit von den Parametern des Tests gewählt wird.

Sei $y = Y(x_1, \ldots, x_n)$ beobachtet worden. Dann nennt man (speziell wenn $\Theta_0 = \{\vartheta_0\}$)

$$P_{\vartheta_0}\big(Y(X_1, \ldots, X_n) > y\big)$$

den *p-Wert* des Test(ergebnisses).

Statistischer Tests: Niveau vs. *p*-Wert

In der Praxis haben Tests meist die folgende Form:

Berechne eine gewisse (Test-)Statistik $Y = Y(x_1, \ldots, x_n)$, setze $\varphi(x_1, \ldots, x_n) = \mathbf{1}\big(Y(x_1, \ldots, x_n) > q\big)$ für einen gewissen Wert $q = q(\alpha)$, der in Abhängigkeit von den Parametern des Tests gewählt wird.

Sei $y = Y(x_1, \ldots, x_n)$ beobachtet worden. Dann nennt man (speziell wenn $\Theta_0 = \{\vartheta_0\}$)

$$P_{\vartheta_0}\big(Y(X_1, \ldots, X_n) > y\big)$$

den *p-Wert* des Test(ergebnisses).

Dies ist die W'keit, bei Gültigkeit der Nullhypothese einen mindestens so „extremen" Wert der Teststatistik zu finden wie den tatsächlich anhand der Daten beobachteten.

Statistischer Tests: Niveau vs. *p*-Wert

In der Praxis haben Tests meist die folgende Form:

Berechne eine gewisse (Test-)Statistik $Y = Y(x_1, \ldots, x_n)$, setze $\varphi(x_1, \ldots, x_n) = \mathbf{1}\big(Y(x_1, \ldots, x_n) > q\big)$ für einen gewissen Wert $q = q(\alpha)$, der in Abhängigkeit von den Parametern des Tests gewählt wird.

Sei $y = Y(x_1, \ldots, x_n)$ beobachtet worden. Dann nennt man (speziell wenn $\Theta_0 = \{\vartheta_0\}$)

$$P_{\vartheta_0}\big(Y(X_1, \ldots, X_n) > y\big)$$

den *p-Wert* des Test(ergebnisses).

Dies ist die W'keit, bei Gültigkeit der Nullhypothese einen mindestens so „extremen" Wert der Teststatistik zu finden wie den tatsächlich anhand der Daten beobachteten.

(Es gilt: Test lehnt H_0 zum Niveau α ab gd.w. *p*-Wert $\leq \alpha$.)

Inhalt

ein Stichproben-t-Test

Modell: n u.i.v. Beobachtungen $X_1, \ldots, X_n$, $X_i \sim \mathcal{N}_{\mu,\sigma^2}$, $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ unbekannt.

$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$, $S^2 := \frac{1}{n-1} \sum_{i=1}^{n-1} (X_i - \overline{X})^2$

Wir wissen: Für jedes $\mu_0 \in \mathbb{R}$, $\sigma^2 > 0$ ist unter $P_{(\mu_0,\sigma^2)}$

$$T := \sqrt{n} \frac{\overline{X} - \mu_0}{\sqrt{S^2}} \sim \text{Student-}(n-1)$$

(vgl. Satz 2.5)

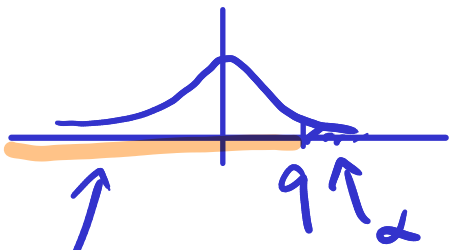
ein Stichproben-t-Test

$\Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma > 0\}$

Modell: n u.i.v. Beobachtungen $X_1, \ldots, X_n$, $X_i \sim \mathcal{N}_{\mu,\sigma^2}$, $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ unbekannt.

$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$, $S^2 := \frac{1}{n-1}\sum_{i=1}^{n-1}(X_i - \overline{X})^2$

Wir wissen: Für jedes $\mu_0 \in \mathbb{R}$, $\sigma^2 > 0$ ist unter $P_{(\mu_0,\sigma^2)}$



$$T := \sqrt{n}\frac{\overline{X} - \mu_0}{\sqrt{S^2}} \sim \text{Student-}(n-1)$$

(vgl. Satz 2.5

$= \{(\mu_0, \sigma^2) : \sigma > 0\}$

Signifikanzniveau $\alpha \in (0,1)$:

Zweiseitiger Test von $H_0 : \{\mu = \mu_0\}$ gegen $H_1 : \{\mu \neq \mu_0\}$:
Lehne H_0 ab, wenn $|T| > q_{n-1,1-\alpha/2}$, wobei $q_{n-1,1-\alpha/2}$ = $(1-\alpha/2)$-Quantil der Student-$(n-1)$-Vert.

$\{(\mu, \sigma^2) : \sigma > 0, \mu \neq \mu_0\}$

Einseitiger Test von $H_0 : \{\mu \leq \mu_0\}$ gegen $H_1 : \{\mu > \mu_0\}$:
Lehne H_0 ab, wenn $T > q_{n-1,1-\alpha}$, wobei $q_{n-1,1-\alpha}$ = $(1-\alpha)$-Quantil der Student-$(n-1)$-Vert.

Beispiel:
Die Wirksamkeit eines gewissen Schlafmittels soll geprüft werden.

10 Patienten erhalten das Schlafmittel, die Anzahl zusätzlicher Stunden Schlaf wird in einer Nacht beobachtet.

Wir nehmen an, die Beob. sind u.i.v. $\sim \mathcal{N}_{\mu,\sigma^2}$ und wir möchten die Nullhypothese $\mu = 0$, sagen wir, zum Niveau $\alpha = 0.05$ testen.

Die Daten*

Patient i	1	2	3	4	5	6	7	8	9	10
zus. Schl.	0.7	-1.6	-0.2	-1.2	-0.1	3.4	3.7	0.8	0.0	2.0

Es ist $n = 10$, $\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 0.75$, $s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x})^2 \approx 1.79$, $t = \frac{\overline{x}-0}{s/\sqrt{10}} \approx 1.326$

Das 0.975-Quantil der Student-9-Verteilung ist ≈ 2.262, demnach können wir die Nullhypothese nicht ablehnen.

(Für ein Student-9-verteiltes T ist $P(|T| \geq 1.326) \approx 0.2176$, dies ist der p-Wert des Tests.)

---

* Aus Student (William S. Gosset), The Probable Error of a Mean, Biometrika 6:1–25 (1908)

Man kann unseren Befund folgendermaßen formulieren:

„Die Beobachtungen sind mit der Nullhypothese $\mu = 0$ (im statistischen Sinne) verträglich.“

oder

„Die beobachtete Abweichung $\overline{x} = 0.75$ ist nicht signifikant von 0 verschieden (t-Test, $\alpha = 0.05$).“

Das Beispiel in R:

```
> schlaf <- c(0.7, -1.6, -0.2, -1.2, -0.1, 3.4,
              3.7, 0.8, 0.0, 2.0)
> t.test(schlaf)

One Sample t-test

data:  schlaf
t = 1.3257, df = 9, p-value = 0.2176
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.5297804  2.0297804
sample estimates:
mean of x
     0.75
```

Beispiel:
Die Wirksamkeit eines Schlafmittels soll mit der eines anderen verglichen werden

10 Patienten erhalten Schlafmittel *A*, die Anzahl zusätzlicher Stunden Schlaf wird in einer Nacht beobachtet.
Dann erhalten dieselben 10 Patienten Schlafmittel *B*, wieder wird die Anzahl zusätzlicher Stunden Schlaf wird in einer Nacht beobachtet.

Da dieselben Patienten untersucht werden, können (und sollten) wir die Messungen paaren:
Wir interessieren uns bei jedem Patienten für die Differenz des (zusätzlichen) Schlafs bei Mittel *B* und bei Mittel *A*.

Wir nehmen an, die beobachteten Differenzen sind Realisierungen von u.i.v. ZVn mit Vert. $\mathcal{N}_{\mu,\sigma^2}$ und wir möchten die Nullhypothese $\mu \leq 0$ gegen die Alternative $\mu > 0$, sagen wir, zum Niveau $\alpha = 0.05$ testen.

Wir nehmen an, die beobachteten Differenzen sind Realisierungen von u.i.v. ZVn mit Vert. $\mathcal{N}_{\mu,\sigma^2}$ und wir möchten die Nullhypothese $\mu \leq 0$ gegen die Alternative $\mu > 0$, sagen wir, zum Niveau $\alpha = 0.05$ testen.

Dies wäre beispielsweise in folgender Situation angemessen: Wir möchten darlegen, dass Mittel *B* wirksamer ist als Mittel *A*, indem wir die Nullhypothese „$\mu \leq 0$“ entkräften.

Wir nehmen an, die beobachteten Differenzen sind Realisierungen von u.i.v. ZVn mit Vert. $\mathcal{N}_{\mu,\sigma^2}$ und wir möchten die Nullhypothese $\mu \leq 0$ gegen die Alternative $\mu > 0$, sagen wir, zum Niveau $\alpha = 0.05$ testen.

Dies wäre beispielsweise in folgender Situation angemessen: Wir möchten darlegen, dass Mittel *B* wirksamer ist als Mittel *A*, indem wir die Nullhypothese „$\mu \leq 0$“ entkräften.

Die Daten*

Patient *i*	1	2	3	4	5	6	7	8	9	10
Mittel *A*	0.7	-1.6	-0.2	-1.2	-0.1	3.4	3.7	0.8	0.0	2.0
Mittel *B*	1.9	0.8	1.1	0.1	-0.1	4.4	5.5	1.6	4.6	3.4
Diff.	1.2	2.4	1.3	1.3	0.0	1.0	1.8	0.8	4.6	1.4

* Aus Student (William S. Gosset), The Probable Error of a Mean, Biometrika 6:1–25 (1908)

Patient i	1	2	3	4	5	6	7	8	9	10
Diff.	1.2	2.4	1.3	1.3	0.0	1.0	1.8	0.8	4.6	1.4

Es ist $n = 10$, $\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 1.58$, $s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x})^2 \approx 1.23$, $t = \frac{\overline{x}-0}{s/\sqrt{10}} \approx 4.062$

Das 0.95-Quantil der Student-9-Verteilung ist ≈ 1.833, demnach können wir die Nullhypothese ablehnen.

(Für ein Student-9-verteiltes T ist $P(T > 4.062) \approx 0.0014$, dies ist der p-Wert des Tests.)

Patient i	1	2	3	4	5	6	7	8	9	10
Diff.	1.2	2.4	1.3	1.3	0.0	1.0	1.8	0.8	4.6	1.4

Es ist $n = 10$, $\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 1.58$, $s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x})^2 \approx 1.23$, $t = \frac{\overline{x}-0}{s/\sqrt{10}} \approx 4.062$

Das 0.95-Quantil der Student-9-Verteilung ist ≈ 1.833, demnach können wir die Nullhypothese ablehnen.

(Für ein Student-9-verteiltes T ist $P(T > 4.062) \approx 0.0014$, dies ist der p-Wert des Tests.)

Mögliche knappe Formulierung:

„Die beobachtete Differenz $\overline{x} = 1.58$ ist signifikant größer als 0 (einseitiger t-Test, $\alpha = 0.05$).“

Das Beispiel in R:

```
> diff <- c(1.2, 2.4, 1.3, 1.3, 0.0, 1.0, 1.8,
            0.8, 4.6, 1.4)
> t.test(diff, alternative="greater")

One Sample t-test

data:  diff
t = 4.0621, df = 9, p-value = 0.001416
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 0.8669947       Inf
sample estimates:
mean of x
     1.58
```

Inhalt

ungepaarter t-Test: Modell

m u.i.v. Beobachtungen $X_1, \ldots, X_m$ und davon unabhängig n u.i.v. Beobachtungen $Y_1, \ldots, Y_n$, unter P_ϑ
$X_i \sim \mathcal{N}_{\mu_1,\sigma^2}$, $Y_j \sim \mathcal{N}_{\mu_2,\sigma^2}$ mit $\vartheta = (\mu_1, \mu_2, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times (0, \infty)$

Seien
$$\overline{X} := \frac{1}{m}\sum_{i=1}^{m} X_i, \ \overline{Y} := \frac{1}{n}\sum_{j=1}^{n} Y_j$$

$$S_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X})^2, \ S_Y^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \overline{Y})^2,$$

ungepaarter t-Test: Modell

m u.i.v. Beobachtungen $X_1, \dots, X_m$ und davon unabhängig n u.i.v. Beobachtungen $Y_1, \dots, Y_n$, unter P_ϑ
$X_i \sim \mathcal{N}_{\mu_1,\sigma^2}$, $Y_j \sim \mathcal{N}_{\mu_2,\sigma^2}$ mit $\vartheta = (\mu_1, \mu_2, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times (0, \infty)$

Seien
$$\overline{X} := \frac{1}{m}\sum_{i=1}^{m} X_i, \ \overline{Y} := \frac{1}{n}\sum_{j=1}^{n} Y_j$$

$$S_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X})^2, \ S_Y^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \overline{Y})^2,$$

$$S^2 := \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2} \left(= \frac{1}{m+n-2}\left(\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2 \right) \right),$$

(bem.: $\mathbb{E}_{(\mu_1,\mu_2,\sigma^2)}[S^2] = \sigma^2$)

ungepaarter t-Test: Modell

m u.i.v. Beobachtungen $X_1, \ldots, X_m$ und davon unabhängig n u.i.v. Beobachtungen $Y_1, \ldots, Y_n$, unter P_ϑ
$X_i \sim \mathcal{N}_{\mu_1,\sigma^2}$, $Y_j \sim \mathcal{N}_{\mu_2,\sigma^2}$ mit $\vartheta = (\mu_1, \mu_2, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times (0, \infty)$

Seien
$$\overline{X} := \frac{1}{m}\sum_{i=1}^{m} X_i, \ \overline{Y} := \frac{1}{n}\sum_{j=1}^{n} Y_j$$

$$S_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X})^2, \ S_Y^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \overline{Y})^2,$$

$$S^2 := \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2} \left(= \frac{1}{m+n-2}\left(\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2 \right) \right),$$

(bem.: $\mathbb{E}_{(\mu_1,\mu_2,\sigma^2)}[S^2] = \sigma^2$)

$$T = \frac{\overline{X} - \overline{Y}}{S\sqrt{\frac{1}{m} + \frac{1}{n}}}.$$

Für $\mu_0 \in \mathbb{R}$, $\sigma^2 > 0$ ist unter $P_{(\mu_0,\mu_0,\sigma^2)}$

$$T = \frac{\overline{X} - \overline{Y}}{S\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim \text{Student-}(m+n-2)$$

Signifikanzniveau $\alpha \in (0,1)$:

Zweiseitiger Test von $H_0 : \{\mu_1 = \mu_2\}$ gegen $H_1 : \{\mu_1 \neq \mu_2\}$:
Lehne H_0 ab, wenn $|T| > q_{m+n-2,1-\alpha/2}$, wobei $q_{m+n-2,1-\alpha/2}$ = $(1-\alpha/2)$-Quantil der Student-$(m+n-2)$-Vert.

Einseitiger Test von $H_0 : \{\mu_1 \leq \mu_2\}$ gegen $H_1 : \{\mu_1 > \mu_2\}$:
Lehne H_0 ab, wenn $T > q_{m+n-2,1-\alpha}$, wobei $q_{m+n-2,1-\alpha}$ = $(1-\alpha)$-Quantil der Student-$(m+n-2)$-Vert.

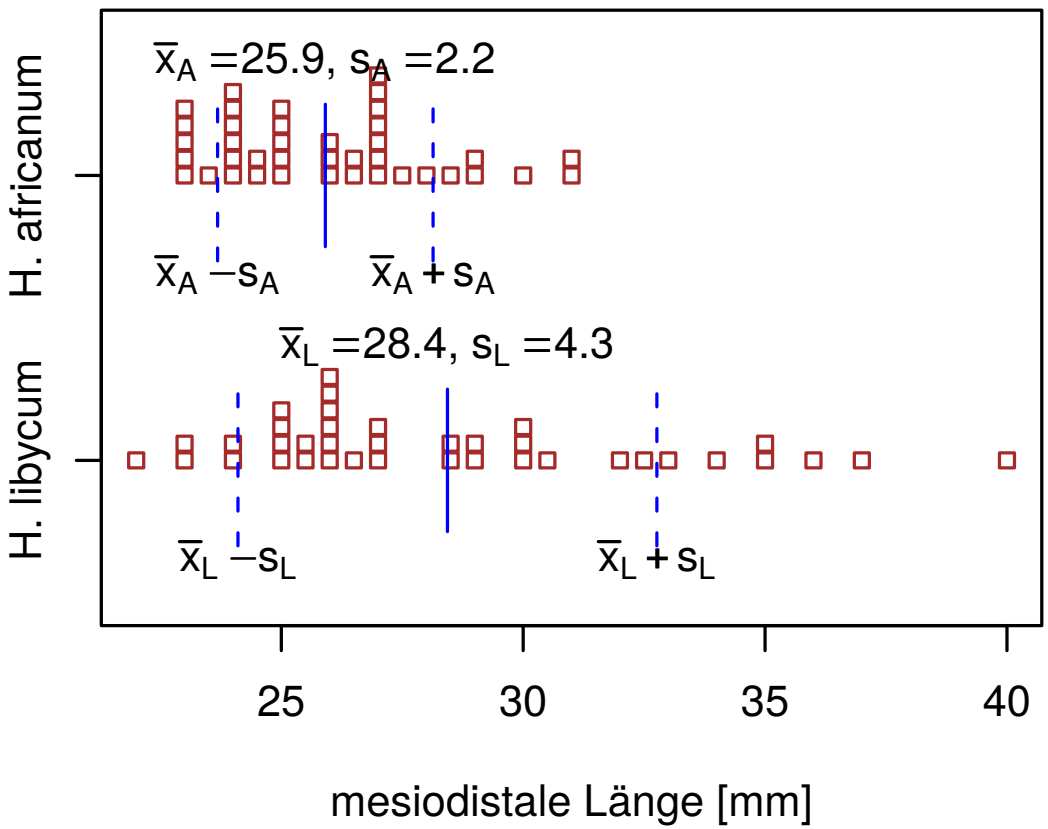
Beispiel: Es wurden fossile Backenzähne gefunden, die zwei Arten von Urpferden zugeordnet wurden, und jeweils die („mesiodistale“) Länge bestimmt.

Wir möchten die (Null-)Hypothese prüfen, ob die mittlere Zahnlänge bei den beiden Arten gleich ist.

Beispiel: Es wurden fossile Backenzähne gefunden, die zwei Arten von Urpferden zugeordnet wurden, und jeweils die („mesiodistale“) Länge bestimmt.

Wir möchten die (Null-)Hypothese prüfen, ob die mittlere Zahnlänge bei den beiden Arten gleich ist.

Die Daten



Hipparion africanum

$n_A = 39$

$\bar{x}_A = 25.9$

$s_A = 2.2$

Hipparion lybicum

$n_L = 38$

$\bar{x}_L = 28.4$

$s_L = 4.3$

Wir verwenden Signifikanzniveau $\alpha = 0.01$, das 99,5%-Quantil der Student-Vert. mit 75 Freiheitsgraden ist ≈ 2.64.
Es ist

$$s^2 = \frac{(n_A - 1)s_A^2 + (n_L - 1)s_L^2}{n_A + n_L - 1} \approx 11.94, \quad t = \frac{\overline{x}_A - \overline{x}_L}{s\sqrt{\frac{1}{n_A} + \frac{1}{n_L}}} \approx -3.229,$$

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

(Für ein Student-75-verteiltes T ist $P(|T| > 3.229) \approx 0.0018$, dies ist der p-Wert des Tests.)

Wir verwenden Signifikanzniveau $\alpha = 0.01$, das 99,5%-Quantil der Student-Vert. mit 75 Freiheitsgraden ist ≈ 2.64.
Es ist

$$s^2 = \frac{(n_A - 1)s_A^2 + (n_L - 1)s_L^2}{n_A + n_L - 1} \approx 11.94, \quad t = \frac{\overline{x}_A - \overline{x}_L}{s\sqrt{\frac{1}{n_A} + \frac{1}{n_L}}} \approx -3.229,$$

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

(Für ein Student-75-verteiltes T ist $P(|T| > 3.229) \approx 0.0018$, dies ist der p-Wert des Tests.)

Mögliche Formulierung unseres Befunds:

„Die mittlere mesiodistale Länge war signifikant größer (28.4 mm) bei *H. libycum* als bei *H. africanum* (25.9 mm) (t-Test, $\alpha = 0{,}01$).“

Das Beispiel in R:

```
> t.test(md[Art=="africanum"],md[Art=="libycum"],
         var.equal=TRUE)

Two Sample t-test

data:  md[Art == "africanum"] and md[Art == "libycu
t = -3.2289, df = 75, p-value = 0.001845
alternative hypothesis:
true difference in means is not equal to 0
95 percent confidence interval:
 -4.0811448 -0.9667634
sample estimates:
mean of x mean of y
 25.91026  28.43421
```