

Biostatistik

Konfidenzintervalle

Matthias Birkner



JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

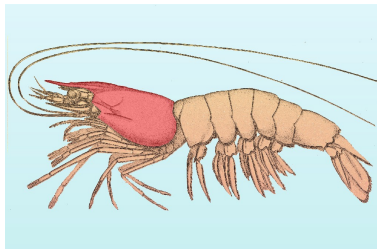
Beispiel: Springkrebs



Galathea squamifera

image (c) by Matthias Buschmann

Carapaxlänge:



(c): public domain

Wie groß ist die mittlere Carapaxlänge des weiblichen Springkrebses?

Alle weiblichen Springkrebse (also die Grundgesamtheit)
zu vermessen ist zu aufwändig.

Idee: Aus einer Stichprobe läßt sich
die mittlere Carapaxlänge schätzen.

Wie genau ist diese Schätzung?

Ziel: Ein (möglichst kurzes) Intervall, das den Mittelwert der
Carapaxlängen aller weiblichen Springkrebse mit hoher
Wahrscheinlichkeit überdeckt.

Ein solches Intervall nennen wir
Konfidenzintervall (oder auch Vertrauensbereich)
für den wahren Wert.

Galathea: Carapaxlänge
in einer Stichprobe von $n = 29$ Weibchen

Es wurde beobachtet

Stichprobenmittelwert: $\bar{x} = 3.23$ [mm]

Stichprobenstreuung: $s = 0.9$ [mm], also

$$\text{Standardfehler } \frac{s}{\sqrt{n}} = \frac{0.9}{\sqrt{29}} = 0.17 \text{ [mm]}$$

(Schätzwert für die Streuung von $\bar{x}$)

Erinnerung: In der Sitzung „5. Standardfehler“ haben wir bereits eine Faustformel kennengelernt:

Der wahre Mittelwert liegt im Intervall

$$\left[\bar{x} - \frac{s}{\sqrt{n}}, \bar{x} + \frac{s}{\sqrt{n}} \right]$$

(im Bsp.: [3.06, 3.4]) mit Wahrscheinlichkeit ungefähr 2/3.

Student-Konfidenzintervall für den Mittelwert

Allgemeiner:

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- $(n - 1)$ -verteilt 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.

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

Im Beispiel: $\bar{x} = 3.23$, $s = 0.9$, $n = 29$.

Sei $\alpha = 0.05$. Es ist $t_{28, 0.975} = 2.048$, also erhält man das Konfidenzintervall

$$[2.88, 3.58]$$

für den wahren Mittelwert der Carapaxlänge der Galathea-Weibchen zum Niveau 95%.

Konfidenzintervall für den wahren Mittelwert

Ziel: Bestimme ein Konfidenzintervall für den wahren Mittelwert zum Niveau $1 - \alpha$

Ein Konfidenzintervall für den wahren Mittelwert zum Irrtumsniveau α ist ein aus den Daten $x = (x_1, \dots, x_n)$ geschätztes (zufälliges) Intervall

$$I = [\underline{I}(x), \bar{I}(x)]$$

mit folgender Eigenschaft: Ist der wahre Mittelwert gleich μ und ist $(x_1, \dots, x_n)$ eine Stichprobe aus der Grundgesamtheit (mit Mittelwert μ), so gilt

$$\mathbb{P}(\mu \in [\underline{I}(x), \bar{I}(x)]) \geq 1 - \alpha$$

Selbstverständlich wollen wir das Konfidenzintervall möglichst klein wählen.

Konfidenzintervall für den wahren Mittelwert

Lösung: Wir wissen bereits ($\rightarrow$ Normalapproximation und die Sitzung zum t -Test),
dass die t -Statistik

$$t := \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

annähernd Student-verteilt ist mit $n - 1$ Freiheitsgraden (wenn n groß genug ist).

Sei $t_{n-1, 1-\alpha/2}$ das $(1 - \alpha/2)$ -Quantil der Student-Verteilung mit $n - 1$ Freiheitsgraden. Dann ist

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

ein Konfidenzintervall zum Niveau $1 - \alpha$, denn ...

Begründung:

$$\begin{aligned} & \mathbb{P}\left(\mu \in \left[\bar{X} - t_{n-1,1-\alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + t_{n-1,1-\alpha/2} \frac{s}{\sqrt{n}}\right]\right) \\ &= \mathbb{P}\left(\mu - \bar{X} \in \left[-t_{n-1,1-\alpha/2} \frac{s}{\sqrt{n}}, t_{n-1,1-\alpha/2} \frac{s}{\sqrt{n}}\right]\right) \\ &= \mathbb{P}\left(\frac{\mu - \bar{X}}{s/\sqrt{n}} \in \left[-t_{n-1,1-\alpha/2}, t_{n-1,1-\alpha/2}\right]\right) \\ &= \mathbb{P}\left(\left|\frac{\mu - \bar{X}}{s/\sqrt{n}}\right| \leq t_{n-1,1-\alpha/2}\right) \\ &= \mathbb{P}(|t| \leq t_{n-1,1-\alpha/2}) \\ &= 1 - \mathbb{P}(t < -t_{n-1,1-\alpha/2}) - \mathbb{P}(t > t_{n-1,1-\alpha/2}) \\ &= 1 - \alpha/2 - \alpha/2 \\ &= 1 - \alpha \end{aligned}$$

Beachte: $t_{n-1,1-\alpha/2}$ ist gerade so gewählt, dass die vorletzte Gleichung richtig ist (t ist symmetrisch verteilt).

Bemerkung Konfidenzintervall abstrakt

Sei θ ein Parameter der zu Grunde liegenden Verteilung.

Ein $(1 - \alpha)$ -Konfidenzintervall für den Parameter θ ist ein aus den Daten $x = (x_1, \dots, x_n)$ geschätztes (zufälliges) Intervall

$$[\underline{I}(x), \bar{I}(x)]$$

mit folgender Eigenschaft: Ist der wahre Parameter gleich θ und ist $(x_1, \dots, x_n)$ eine Stichprobe aus der Grundgesamtheit (mit Parameter θ), so gilt

$$\mathbb{P}_\theta(\theta \in [\underline{I}(x), \bar{I}(x)]) \geq 1 - \alpha$$

Im Allgemeinen kann ein Konfidenzintervall sehr schwer (wenn überhaupt) exakt zu bestimmen sein.

Im 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 $0.86 (= \mathbb{P}_{\mu=3.2}(|t| \geq 0.18))$).

$n = 29, \bar{x} = 3.23, s = 0.9$, Konfidenzintervall für den wahren Mittelwert μ : $I = \left[\bar{x} - t_{28,0.975} \frac{s}{\sqrt{n}}, \bar{x} + t_{28,0.975} \frac{s}{\sqrt{n}} \right] = [2.88, 3.58]$

Variieren wir nun μ in der Nullhypothese im t -Test:

$H_0 : \mu = \dots$	$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$	Wird H_0 verworfen? ($t_{28,0.975} = 2.048$)	p -Wert
$\mu = 3.2$	0.180	nein	0.859
$\mu = 3.0$	1.376	nein	0.180
$\mu = 2.89$	2.034	nein	0.051
$\mu = 2.6$	3.77	ja	0.0008

Der (zweiseitige) t -Test zum Signifikanzniveau 5% verwirft die Nullhypothese für ein gegebenes μ (anhand dieser Beobachtungen) genau dann nicht, wenn μ im Konfidenzintervall I liegt.

Der (zweiseitige) t -Test zum Signifikanzniveau α verwirft die Nullhypothese für ein gegebenes μ (anhand vorgegebener Beobachtungen) genau dann nicht, wenn μ im Konfidenzintervall I zum Irrtumsniveau α liegt, denn

$$\begin{aligned}\bar{X} - t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}} &\leq \mu \leq \bar{X} + t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}} \\ \Leftrightarrow -t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}} &\leq \bar{X} - \mu \leq t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}} \\ \Leftrightarrow -t_{n-1, 1-\alpha/2} &\leq \frac{\bar{X} - \mu}{s/\sqrt{n}} \leq t_{n-1, 1-\alpha/2} \\ \Leftrightarrow |t| &= \left| \frac{\bar{X} - \mu}{s/\sqrt{n}} \right| \leq t_{n-1, 1-\alpha/2}\end{aligned}$$

Dualität Tests $\leftrightarrow$ Konfidenzintervalle: allgemeines Prinzip

Gegeben Daten $\mathbf{x} = (x_1, \dots, x_n)$

Sei $[a, b]$ mit $a = a(\mathbf{x})$, $b = b(\mathbf{x})$ ein Konfidenzintervall zum Irrtumsniveau α für einen Parameter θ , betrachte folgenden Test:
Verwirf die Nullhypothese $\theta = \theta_0$ genau dann, wenn $\theta_0 \notin [a, b]$.

Dieser Test hat Signifikanzniveau α .

Sei umgekehrt für jedes θ_0 ein Test T_{θ_0} der Nullhypothese $\theta = \theta_0$ zum Signifikanzniveau α gegeben, so bildet die Menge aller Werte θ_0 , für die die Nullhypothese $\theta = \theta_0$ *nicht* verworfen wird, einen Konfidenzbereich für θ zum Irrtumsniveau α .

Bemerkung

Konfidenzintervalle sind auch und gerade dann hilfreich, wenn ein Test *keine* Signifikanz anzeigt.

Beispiel: Gibt es bei Steinläusen
(*Petrophaga lorioti*)
geschlechtsspezifische Unterschiede
in der Körperlänge?

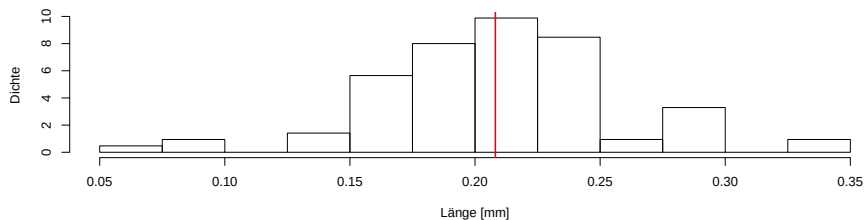


weibliche Steinlaus
(c) User Anton, CC-by-sa 3.0

Wir möchten die Nullhypothese
„die mittlere Körperlänge ist bei beiden Geschlechtern gleich“
zum Signifikanzniveau 5% testen.

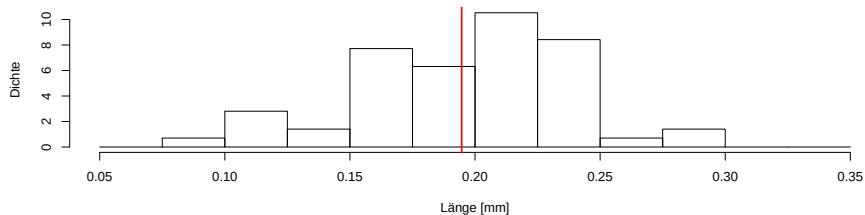
Datenlage: die Längen von 85 weiblichen und 57 männlichen
Steinläusen.

85 weibliche Steinläuse



$$n_w = 85, \bar{x}_w = 0.208, s_w = 0.047$$

57 männliche Steinläuse



$$n_m = 57, \bar{x}_m = 0.194, s_m = 0.041$$

Steinlaus-Längen: Analyse

Aus den Daten: $n_w = 85$, $\bar{x}_w = 0.208$, $s_w = 0.047$
 $n_m = 57$, $\bar{x}_m = 0.194$, $s_m = 0.041$

Demnach beobachten wir in der Stichprobe eine Differenz $d = \bar{x}_w - \bar{x}_m = 0.014$ zwischen der mittleren Länge weiblicher und männlicher Steinläuse.

Ist dieser Unterschied signifikant?

Konfidenzintervall für den wahren Wert?

Steinlaus-Längen: Analyse (Test)

$$n_w = 85, \bar{x}_w = 0.208, s_w = 0.047, \quad n_m = 57, \bar{x}_m = 0.194, s_m = 0.041$$

Wir bilden

$$s = \sqrt{\frac{n_w - 1}{n_w + n_m - 2} s_w^2 + \frac{n_m - 1}{n_w + n_m - 2} s_m^2} = 0.045$$

(unser Schätzwert für die unbekannte Streuung der individuellen Längen) und

$$t = \frac{\bar{x}_w - \bar{x}_m}{s \sqrt{\frac{1}{n_w} + \frac{1}{n_m}}} = 1.83$$

Das 97,5%-Quantil der Student-Verteilung mit $n_w + n_m - 2 = 140$ Freiheitsgraden ist $t_{140,0.975} = 1.977$, der beobachtete t -Wert erfüllt $|t| = 1.83 \leq 1.977$, demnach lehnt der zwei-Stichproben- t -Test die Nullhypothese „die mittleren Längen sind bei Weibchen und Männchen gleich“ nicht ab.

Steinlaus-Längen: Analyse (Konfidenzintervall)

$$n_w = 85, \bar{x}_w = 0.208, s_w = 0.047, \quad n_m = 57, \bar{x}_m = 0.194, s_m = 0.041$$

Wir haben berechnet $d = \bar{x}_w - \bar{x}_m = 0.014$, $s = 0.045$

Wir schätzen die Streuung von d durch $s\sqrt{\frac{1}{n_w} + \frac{1}{n_m}}$ und

$d/(s\sqrt{\frac{1}{n_w} + \frac{1}{n_m}})$ ist approximativ Student-verteilt mit
 $n_w + n_m - 2 = 140$ Freiheitsgraden, also ist

$$\left[d - t_{140,0.975} s \sqrt{\frac{1}{n_w} + \frac{1}{n_m}}, d + t_{140,0.975} s \sqrt{\frac{1}{n_w} + \frac{1}{n_m}} \right] = [-0.001, 0.029]$$

ein (approximatives) 95%-Konfidenzintervall für die Differenz der mittleren Länge von Weibchen und von Männchen in der Population.

Wie würden wir das Ergebnis berichten?

- ~~Es gibt keinen Unterschied zwischen männlichen und weiblichen Steinläusen.~~
- ~~Männliche und weibliche Steinläuse sind im Mittel gleich lang.~~
- Die Daten zeigen keine signifikanten Unterschiede zwischen den mittleren Längen männlicher und weiblicher Steinläuse. ✓
- Ein 95%-Konfidenzintervall für die Differenz zwischen der mittleren Länge der Weibchen und der Männchen (in mm) ist $[-0.001, 0.029]$. ✓



(c): Lamiot CC-by-sa 3.0

Pisidia longicornis

In einem Fang vom 21.02.1992 in der Helgoländer Tiefe Rinne waren 23 Weibchen und 30 Männchen (*Pisidia longicornis*), d.h. der Männchenanteil in der Stichprobe war $30/53 = 0,57$.

Was sagt uns dies über den
Männchenanteil in der Population?

Was ist ein 95%-Konfidenzintervall
für den Männchenanteil in der Population? ($0,57 \pm ??$)

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

Der offensichtliche Schätzer ist

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

in der Stichprobe.

Frage: Wie verlässlich ist die Schätzung?

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

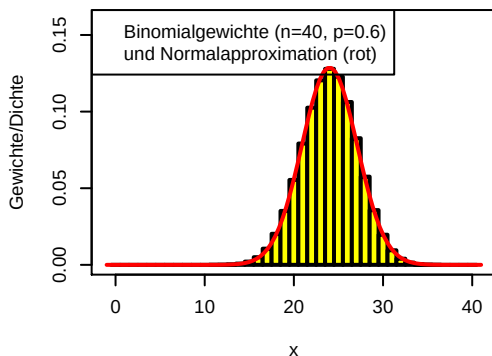
$$\mathbb{P}_p([\hat{p}_u, \hat{p}_o] \text{ überdeckt } p) \geq 1 - \alpha$$

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):



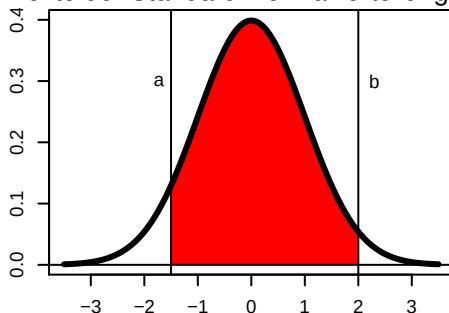
Also ist $\hat{p} = \frac{X}{n}$ (ungefähr) normalverteilt mit Mittelwert p und
Varianz $\frac{1}{n}p(1 - p)$.

$\hat{p} = \frac{X}{n}$ 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.

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

Bemerkung

Das auf den vorherigen Folien vorgestellte (auf approximativer Normalität beruhende) Konfidenzintervall für den Parameter p einer Binomialverteilung heißt in der Literatur auch das „Standard-Intervall“ oder das „Wald-Intervall“ (nach Abraham Wald, 1902–1950). Es gibt verschiedene andere Konstruktionen von Konfidenzintervallen für p , z.B.

$$\left[\tilde{p} \pm z_{1-\alpha/2} \frac{\sqrt{\tilde{p}(1-\tilde{p})}}{\sqrt{n}} \right] \quad \text{mit } \tilde{p} := \frac{X+1}{n+2}$$

(auf dem Bayes-Schätzer $\tilde{p} := \frac{X+1}{n+2}$ fußendes Konfidenzintervall) oder

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

(Student-basiertes Konfidenzintervall, hier wird ähnlich dem t -Test die Tatsache benutzt, dass $\frac{\hat{p}(1-\hat{p})}{n-1}$ ein erwartungstreuer Schätzer für $\text{Var}_p(\hat{p})$ ist.)

Wir werden im Zusammenhang dieser Vorlesung nur das Standard-Intervall für p verwenden.



Bild (c) Andreas Trepte CC-by-sa 2.5

Anas platyrhynchos

Stockente (engl. mallard)

Daten: In einer Stichprobe von $n = 2200$ Stockenten wurden 1240 Männchen gezählt, d.h. die beobachtete relative Häufigkeit der Männchen war 0.564.

Daten aus:



D.H. Johnson, A.B. Sargeant (1977) Impact of red fox predation on the sex ratio of prairie mallards
United States fish & wild life service res. rep. 6.

Zusammenhang: Füchse jagen Stockenten.

Hat dies einen Einfluss auf das Geschlechterverhältnis bei amerikanischen Stockenten?

Setzt man die Zahlenwerte $n = 2200$, $\hat{p} = 0,564$, $z_{0.975} = 1,96$
und $\sqrt{\hat{p}(1 - \hat{p})/n} = 0,011$ in

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

ein, so erhält man das Konfidenzintervall

$$[0,542, 0,585]$$

für den wahren Männchenanteil zum Niveau 95%.