Biostatistik, SS 2016

Der Standardfehler

Matthias Birkner

http://www.staff.uni-mainz.de/birkner/Biostatistik16/

20.5.2016

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Hirse

Bild: *Panicum miliaceum*

Ein Versuch

Versuchsaufbau:

14 Hirse-Pflanzen von einer Sorte wurden 7 Tage lang nicht mehr gegossen („trockengestresst“).

An den letzten drei Tagen wurde die Wasserabgabe der Pflanzen durch Wägung ermittelt und ein Mittelwert über drei Tage errechnet.

Zum Schluß des Versuchs wurden die Pflanzen abgeschnitten und die Blattfläche bestimmt.

Transpirationsrate

=

(Wasserabgabe pro Tag)/Blattfläche

$$\left[\frac{\text{ml}}{\text{cm}^2 \cdot \text{Tag}}\right]$$

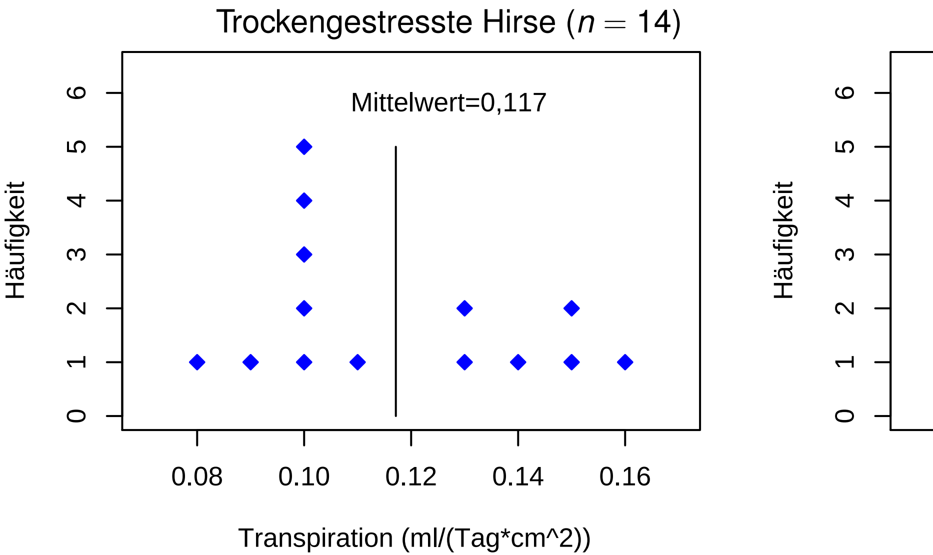
Ein Ziel des Versuchs: die mittlere Transpirationsrate μ

(für diese Hirsesorte unter diesen Bedingungen)

zu bestimmen.

In einem großen Versuch mit sehr vielen Pflanzen könnte man μ beliebig genau bestimmen.

FRAGE:
Wie genau ist die Schätzung von μ in diesem kleinen ($n = 14$) Versuch?


Trockengestresste Hirse ($n = 14$)
Mittelwert=0,117
Häufigkeit
0
1
2
3
4
5
6
0.08
0.10
0.12
0.14
0.16
Transpiration (ml/(Tag*cm^2))
Häufigkeit
0
1
2
3
4
5
6

Transpirationsdaten: $x_1, x_2, \ldots, x_{14}$

$$\overline{x} = \left(x_1 + x_2 + \cdots + x_{14}\right)/14 \;= \frac{1}{14}\sum_{i=1}^{14} x_i$$

$$\overline{x} = 0{,}117$$

Unsere Schätzung:

$\mu \approx 0{,}117$

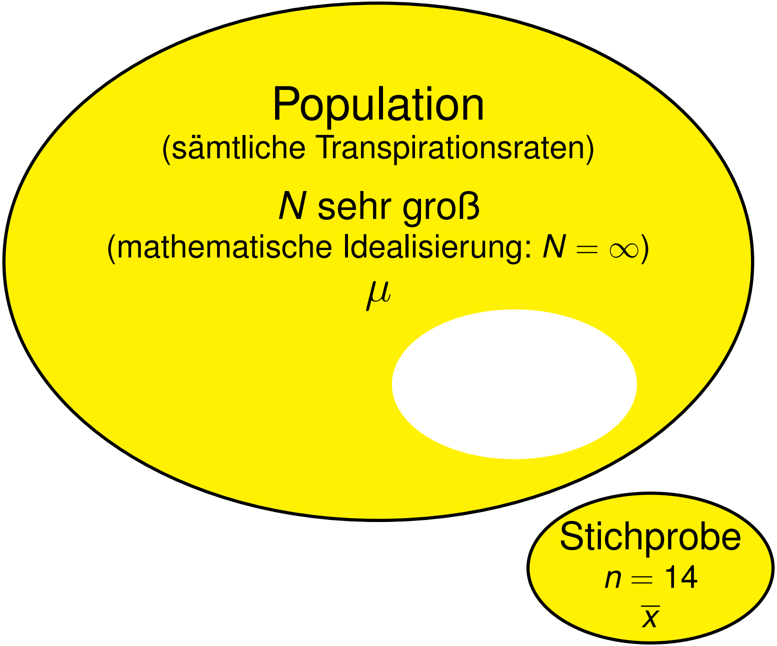
Wie genau ist diese Schätzung?

Wie weit weicht $\overline{x}$ (unser Schätzwert) von μ (dem wahren Mittelwert) ab?

Ein allgemeiner Rahmen

Wir stellen uns vor,
wir hätten den Versuch nicht 14 mal,
sondern 100 mal,
1.000 mal,
1.000.000 mal
wiederholt.

Unsere 14 Transpirationswerte
betrachten wir als
zufällige Stichprobe
aus dieser großen Population
von möglichen Werten.


Population
(sämtliche Transpirationsraten)
N sehr groß
(mathematische Idealisierung: $N = \infty$)
μ
Stichprobe
$n = 14$
$\bar{x}$

Wir schätzen
den Populationsmittelwert
μ
durch
den Stichprobenmittelwert
$\overline{x}$.

Jede neue Stichprobe liefert einen neuen Wert von $\overline{x}$.

$\overline{x}$ hängt vom Zufall ab:

eine *Zufallsgröße*

FRAGE: Wie variabel ist $\overline{x}$?

Genauer: Wie weit weicht $\overline{x}$ typischerweise von μ ab?

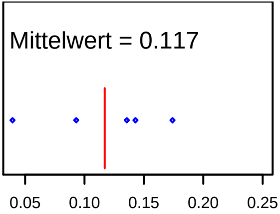
$$\overline{x} = \left(x_1 + x_2 + \cdots + x_n\right) / n$$

Wovon hängt die Variabilität von $\overline{x}$ ab?

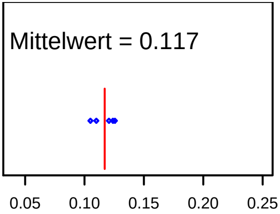
1.

von der
Variabilität
der einzelnen
Beobachtungen

$$x_1, x_2, \ldots, x_n$$



x variiert viel
$\Rightarrow \bar{x}$ variiert viel



x variiert wenig
$\Rightarrow \bar{x}$ variiert wenig

2.

vom
Stichprobenumfang
n

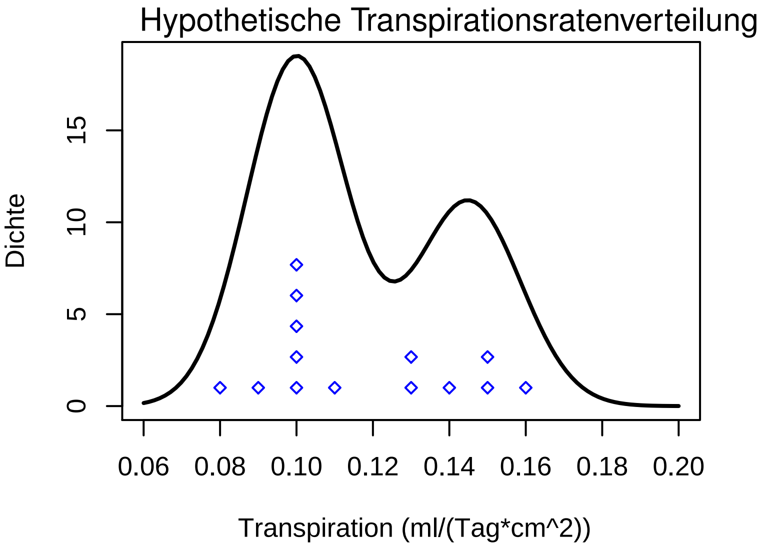
Je größer n,
desto kleiner
die Variabilität von $\bar{x}$.

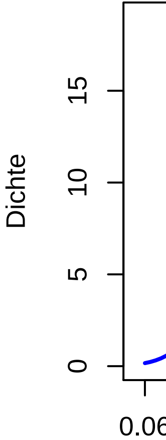
Um diese Abhängigkeit
zu untersuchen,
machen wir ein
(Computer-)Experiment.

Experiment:

Wir nehmen eine Population,
ziehen Stichproben,
und schauen,
wie $\overline{x}$ variiert.

Nehmen wir an,
die Verteilung
aller möglichen Transpirationswerte
sieht folgendermaßen aus:

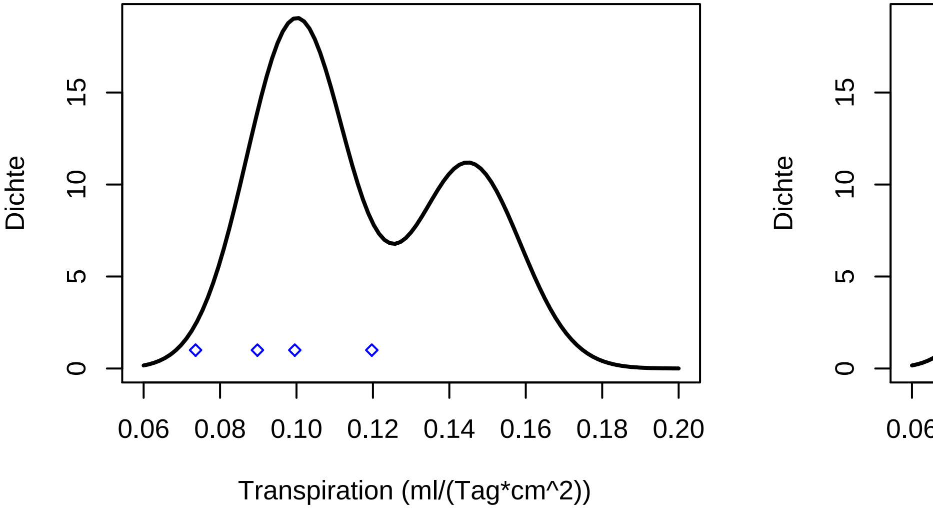

Hypothetische Transpirationsratenverteilung
Dichte
0
5
10
15
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
Transpiration (ml/(Tag*cm^2))

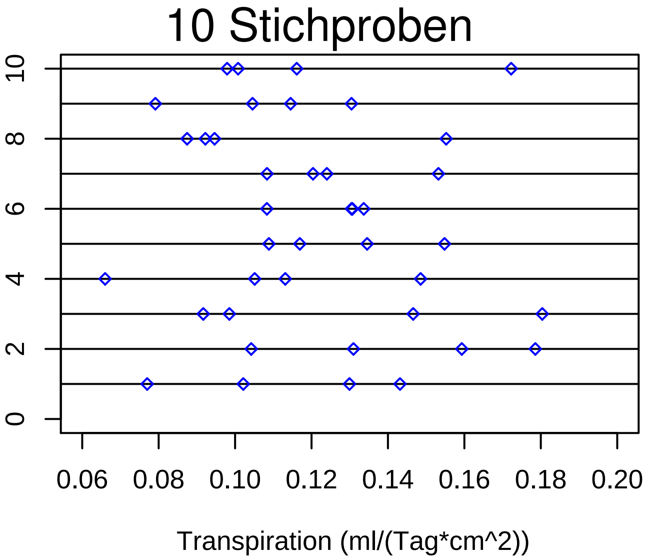

Dichte
0
5
10
15
0.06

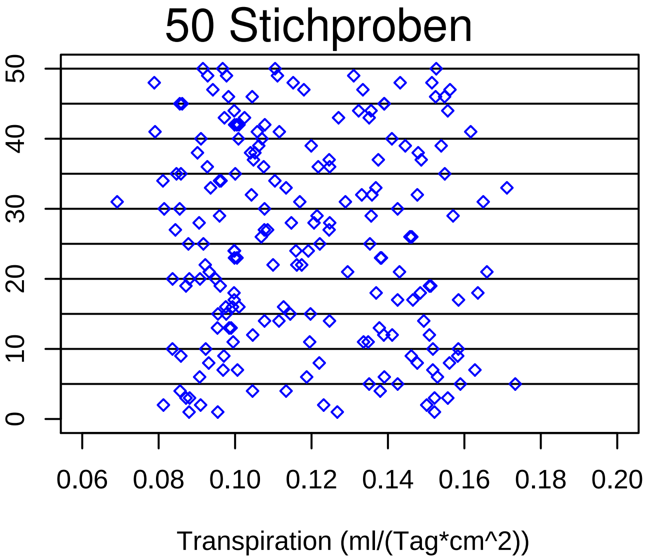
Wir beginnen mit kleinen Stichproben:

$$n = 4$$

Eine Stichprobe vom Umfang 4 Eine zweite Stichprobe vom Umfang 4 Eine dritte Stichprobe vom Umfang 4




10 Stichproben
0
2
4
6
8
10
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
Transpiration (ml/(Tag*cm^2))

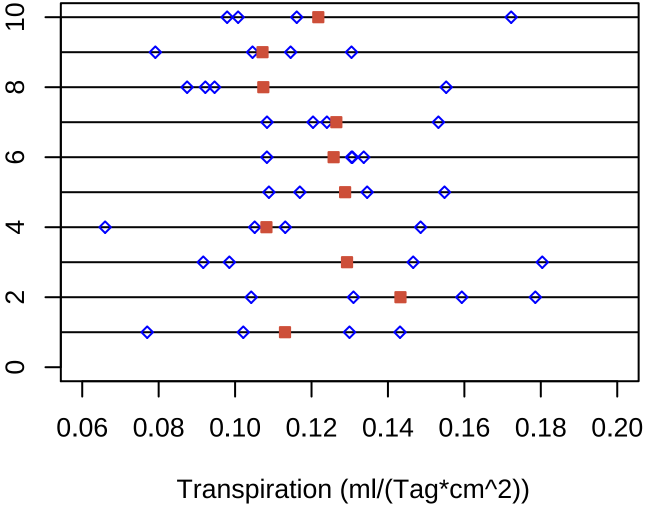


50 Stichproben
0
10
20
30
40
50
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
Transpiration (ml/(Tag*cm^2))

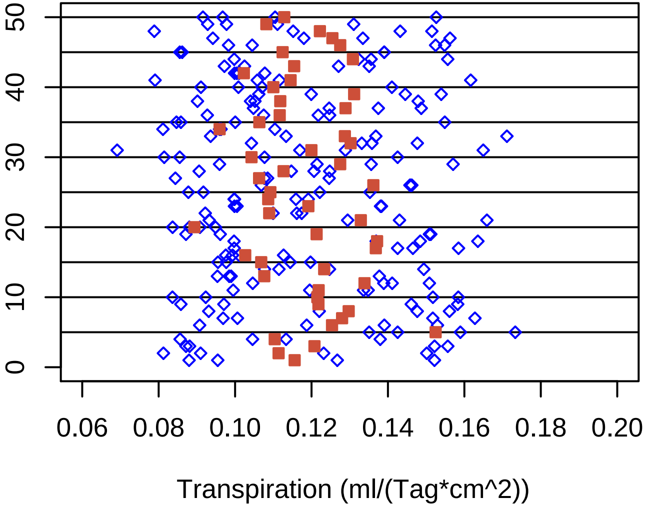
Wie variabel sind

die Stichprobenmittelwerte?

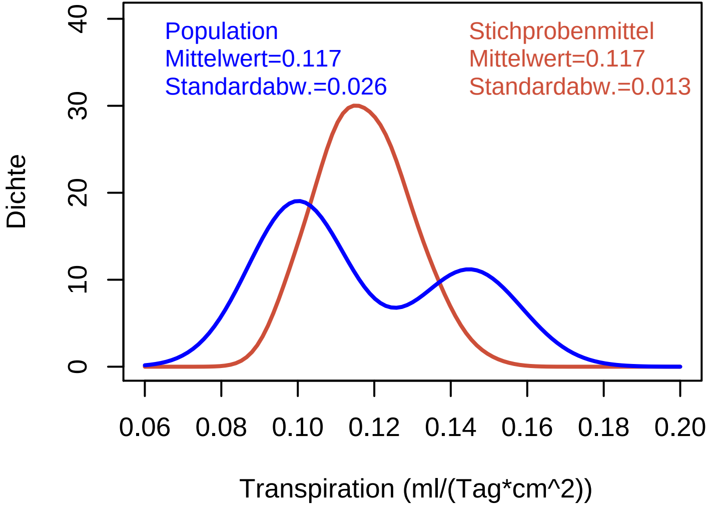
10 Stichproben vom Umfang 4 und die zugehörigen Stichprobenmittel



50 Stichproben vom Umfang 4 und die zugehörigen Stichprobenmittel



Verteilung der Stichprobenmittelwerte (Stichprobenumfang $n = 4$)



Population:

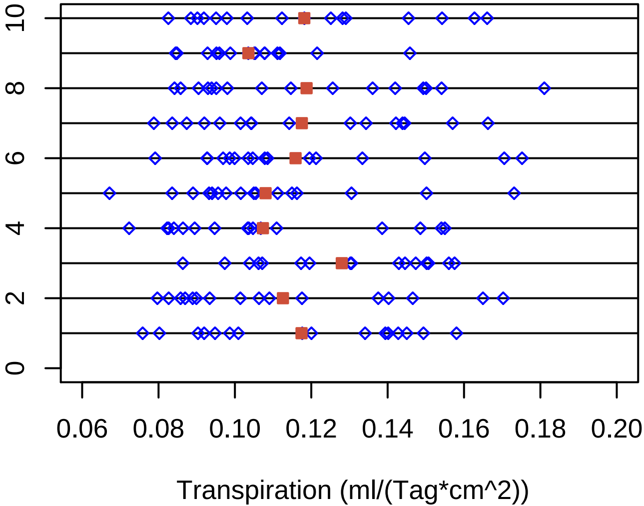
Standardabweichung = 0,026

Stichprobenmittelwerte ($n = 4$):

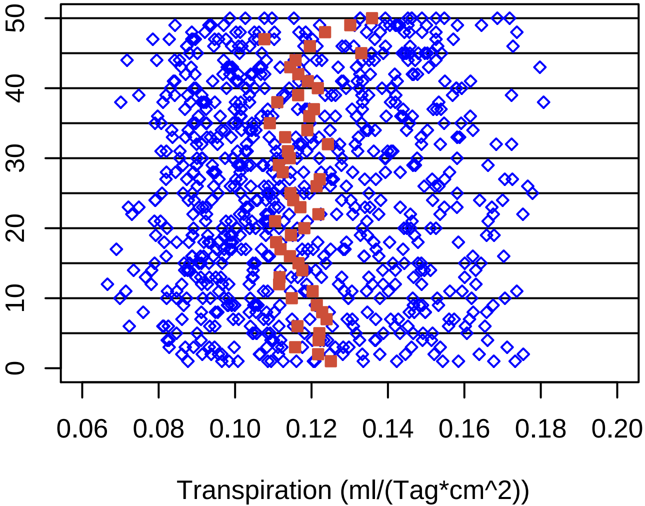
$$\begin{aligned} \text{Standardabweichung} &= 0{,}013 \\ &= 0{,}026/\sqrt{4} \end{aligned}$$

Erhöhen wir
den Stichprobenumfang
von
4
auf
16

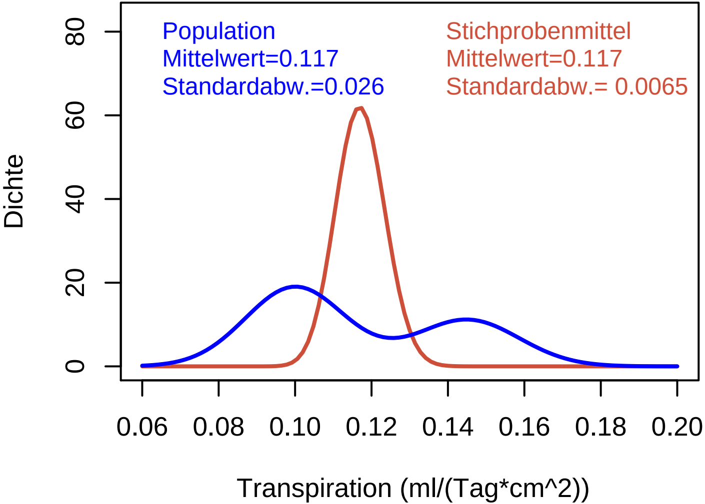
10 Stichproben vom Umfang 16 und die zugehörigen Stichprobenmittel



50 Stichproben vom Umfang 16 und die zugehörigen Stichprobenmittel



Verteilung der Stichprobenmittelwerte (Stichprobenumfang $n = 16$)



Population:

Standardabweichung = 0,026

Stichprobenmittelwerte ($n = 16$):

$$\text{Standardabweichung} = 0{,}0065 = 0{,}026/\sqrt{16}$$

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

$$\sigma$$

(sigma).

Die Regel schreibt man häufig so:

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

In der Praxis ist σ unbekannt.

Es wird durch

die Stichproben-Standardabweichung $\boldsymbol{s}$

geschätzt:

$$\sigma \approx \boldsymbol{s}$$

(Erinnerung: $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2}$)

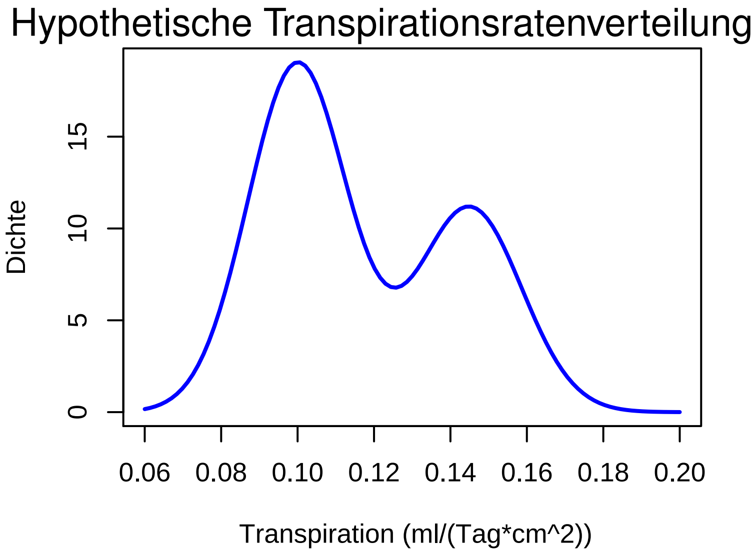
$s/\sqrt{n}$

(die geschätzte Standardabweichung von $\overline{x}$) 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
Dichte
0
5
10
15
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
Transpiration (ml/(Tag*cm^2))

ist die Verteilung von

$\overline{x}$

trotzdem

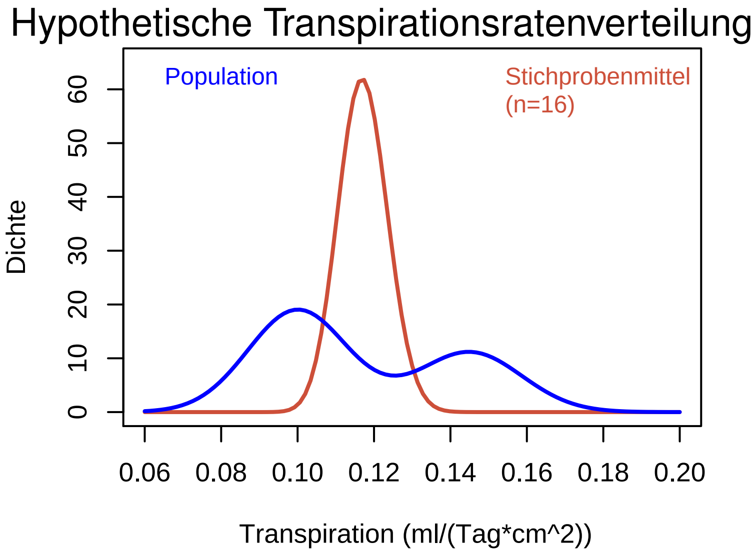
(annähernd)

eingipfelig

&

symmetrisch

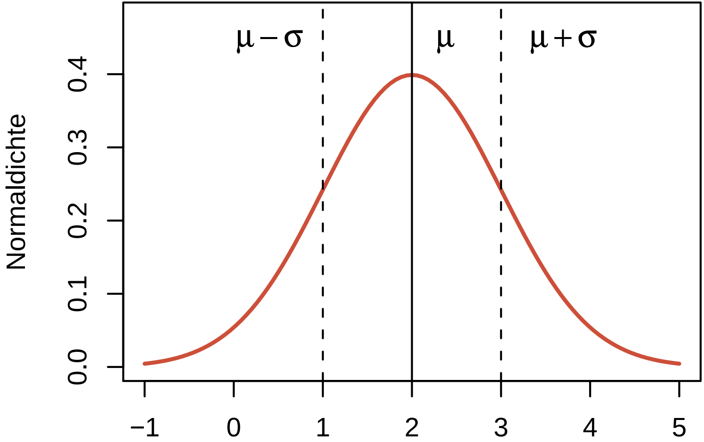
(wenn der Stichprobenumfang n nur groß genug ist)


Hypothetische Transpirationsratenverteilung
Population
Stichprobenmittel
(n=16)
Dichte
0
10
20
30
40
50
60
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
Transpiration (ml/(Tag*cm^2))

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

die Normalverteilung.

Dichte der Normalverteilung: $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

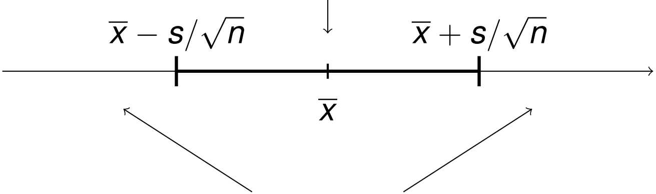


Die Normalverteilungsdichte heisst
auch *Gauß'sche Glockenkurve*
(nach Carl Friedrich Gauß, 1777-1855)

Wichtige Folgerung

Wir betrachten das Intervall $\left[\overline{x} - s/\sqrt{n}, \overline{x} + s/\sqrt{n}\right]$

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



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

Beachte: Das wahre μ ist unbekannt, aber nicht zufällig.

Die Wahrscheinlichkeit bezieht sich auf das Verhalten von $\overline{x}$ und s, die wir anhand der zufälligen Stichprobe berechnet haben.

Demnach:

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

Anwendung 1:
Welche Werte von μ sind plausibel?

$$\overline{x} = 0{,}12$$
$$s/\sqrt{n} = 0{,}007$$

Frage: Könnte es sein, dass

$$\mu = 0{,}115?$$

Antwort: Es ist gut möglich.

Abweichung

$$\bar{x} - \mu = 0{,}120 - 0{,}115 = 0{,}005.$$

Standardfehler

$$s/\sqrt{n} = 0{,}007$$

Abweichungen dieser Größe kommen häufig vor.

(Die Frage, welche Abweichungen *nicht* mehr plausibel sind, untersuchen wir im nächsten Kapitel.)

Anwendung 2: Vergleich von Mittelwerten

Beispiel: Eine Stichprobe von Springkrebsen

Galathea: Carapaxlänge in einer Stichprobe

Männchen:

$$\overline{x}_1 = 3{,}04 \text{ mm}$$
$$s_1 = 0{,}9 \text{ mm}$$
$$n_1 = 25$$

Weibchen:

$$\overline{x}_2 = 3{,}23 \text{ mm}$$
$$s_2 = 0{,}9 \text{ mm}$$
$$n_2 = 29$$

Die Weibchen
scheinen größer zu sein.

Ist das ernst zu nehmen?

Oder könnte es nur Zufall sein?

Wie genau sind die beiden Mittelwerte?

Männchen:

$$\overline{x}_1 = 3{,}04 \text{ mm}$$
$$s_1 = 0{,}9 \text{ mm}$$
$$n_1 = 25$$

$$s_1/\sqrt{n_1} = 0{,}18 \text{ [mm]}$$

Mit Schwankungen von
$\pm 0{,}18$ (mm) in $\overline{x}_1$
müssen wir rechnen.

Wie genau sind die beiden Mittelwerte?

Weibchen:

$$\overline{x}_2 = 3{,}23 \text{ mm}$$

$$s_2 = 0{,}9 \text{ mm}$$

$$n_2 = 29$$

$$s_2/\sqrt{n_2} = 0{,}17 \text{ [mm]}$$

Es ist nicht unwahrscheinlich, dass $\overline{x}_2$ um mehr als $\pm 0{,}17$ (mm) vom wahren Mittelwert abweicht.

Die Differenz der Mittelwerte

$$3{,}23 - 3{,}04 = 0{,}19 \text{ [mm]}$$

ist kaum größer als
die zu erwartenden Schwankungen.

Es könnte also einfach Zufall sein,
dass

$$\overline{x}_2 > \overline{x}_1$$

Genauer formuliert:

Wenn in Wirklichkeit
die Populationsmittelwerte
gleich sind,

$$\mu_{\text{Weibchen}} = \mu_{\text{Männchen}}$$

kann es trotzdem leicht passieren,
dass die Stichprobenmittelwerte

$\overline{x}_2$ und $\overline{x}_1$

so weit auseinander liegen.

Der Statistiker sagt:

Die Differenz
der Mittelwerte
ist
(statistisch)
nicht signifikant.

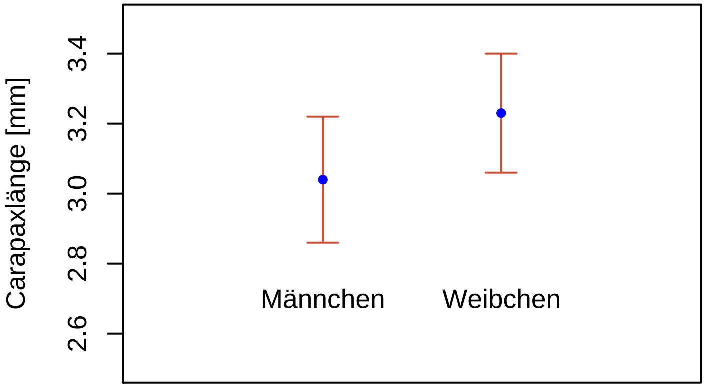
nicht signifikant
=
könnte Zufall sein

Anwendung 3:

Wenn man Mittelwerte
graphisch darstellt,
sollte man auch
ihre Genauigkeit
$(\pm s/\sqrt{n})$
anzeigen.

Also so:

Carapaxlängen:
Mittelwerte ± Standardfehler nach Geschlecht



Anwendung 4:

Bei der Versuchsplanung:

Wieviele Beobachtungen brauche ich?

(Wie groß sollte n sein?)

Wenn man weiß, welche Genauigkeit
(Standardfehler $s/\sqrt{n}$)
für $\overline{x}$ man erreichen will,

und wenn man (aus Erfahrung oder aus einem Vorversuch) s ungefähr kennt,
dann kann man
das notwendige n ungefähr abschätzen:

$$s/\sqrt{n} = g$$

(g = gewüschter Standardfehler)

Beispiel:
Gestresste Transpirationswerte
bei einer anderen Hirse-Sorte:

$$\overline{x} = 0{,}18$$

$$s = 0{,}06$$

$$n = 13$$

Nehmen wir an, der Versuch soll wiederholt werden und man will Standardfehler $\approx 0{,}01$ erreichen.

Wie groß sollte n sein?

Lösung:

gewünscht:

$$s/\sqrt{n} \approx 0{,}01$$

aus dem Vorversuch wissen wir:

$$s \approx 0{,}06,$$

also $\sqrt{n} \approx 6$

$$n \approx 36$$

ZUSAMMENFASSUNG

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

  Solche Schwankungen sind „nicht signifikant“: sie könnten Zufall sein.