

# Biostatistik, WS 2015/2016

## Der Standardfehler

Matthias Birkner

<http://www.staff.uni-mainz.de/birkner/Biostatistik1516/>

20.11.2015



JOHANNES GUTENBERG  
UNIVERSITÄT MAINZ

- 1 Ein kurze Wiederholung zur deskriptiven Statistik
  - Mittelwert und Standardabweichung
  - Ein Beispiel zur Warnung
  
- 2 Der Standardfehler
  - Ein Versuch
  - Ein allgemeiner Rahmen
  - Zur Verteilung von  $\bar{x}$
  - Anwendungen
  - Zusammenfassung

# Bachstelzen fressen Dungfliegen

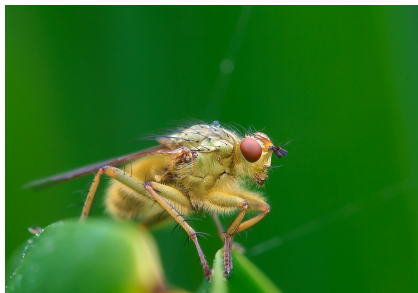
Räuber



Bachstelze (White Wagtail)  
*Motacilla alba alba*

image (c) by Artur Mikolajewski

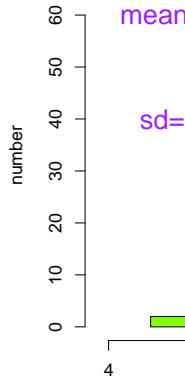
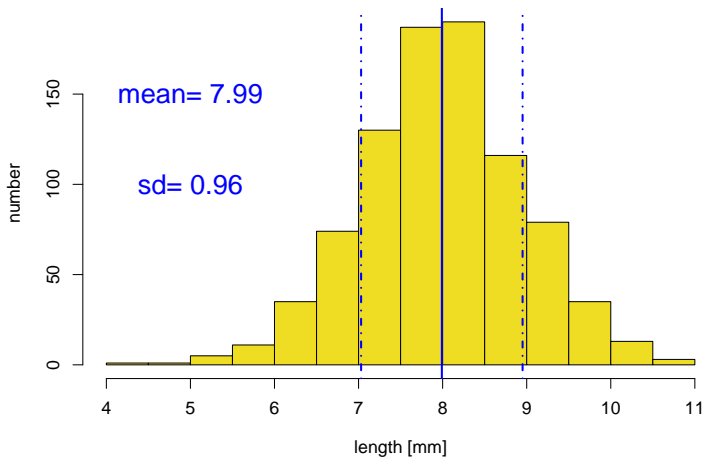
Beute



Gelbe Dungfliege  
*Scatophaga stercoraria*

image (c) by Viatour Luc

## available dung flies

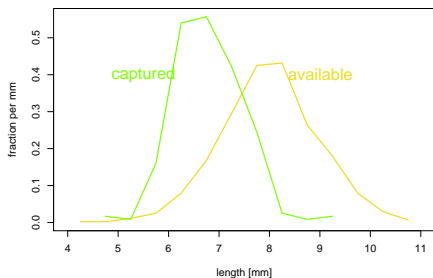


# Wählerische Bachstelzen

Länge der Dungfliegen:

	gefangen		verfügbar
Mittelwert	6.79	<	7.99
Standardabw.	0.69	<	0.96

dung flies: available, captured



Interpretation: Die Bachstelzen bevorzugen Dungfliegen aus einem relativ engen Größenbereich um 7mm herum.

# Beobachtung

In der Biologie sind viele Datenverteilungen annähernd glockenförmig und können durch den **Mittelwert** und die **Standardabweichung** angemessen beschrieben werden (z.B. die verfügbaren und die tatsächlich gefangenen Dungfliegen aus dem Bachstelzen-Beispiel).

Es gibt aber auch Ausnahmen.

# Kupfertolerantes Rotes Straußgras



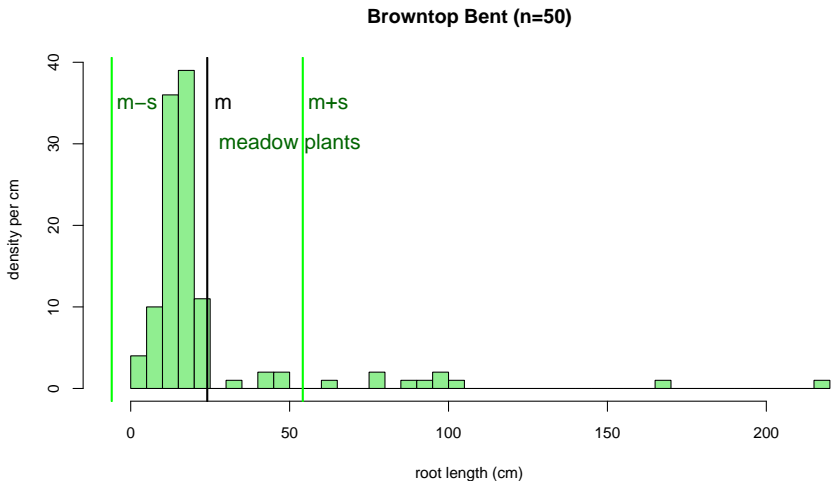
Rotes Straußgras  
*Agrostis tenuis*

image (c) Kristian Peters



Kupfer  
*Cuprum*

Hendrick met de Bles



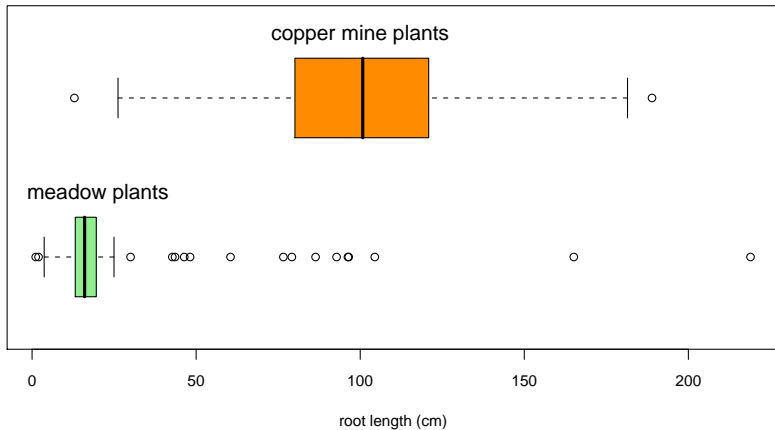
In etwa 2/3 der Wurzellängen innerhalb  $[m-s, m+s]$ ????

**Nein!** Deutlich mehr.



Manche Verteilungen können nur mit mehr als zwei Variablen angemessen beschrieben werden,

z.B. mit den fünf Werten der Boxplots:  
**min,  $Q_1$ , median,  $Q_3$ , max**

**Browntop Bent n=50+50**

# Schlussfolgerung

In der Biologie sind viele Datenverteilungen annähernd glockenförmig und können durch den **Mittelwert** und die **Standardabweichung** angemessen beschrieben werden.

Es gibt aber auch Ausnahmen. Also:  
**Immer** die Daten erst mal graphisch untersuchen!

Verlassen Sie sich **nicht** allein auf numerische Kenngrößen!



## 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

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

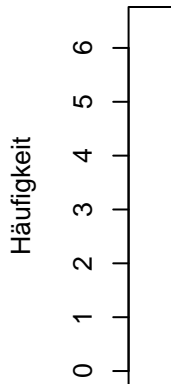
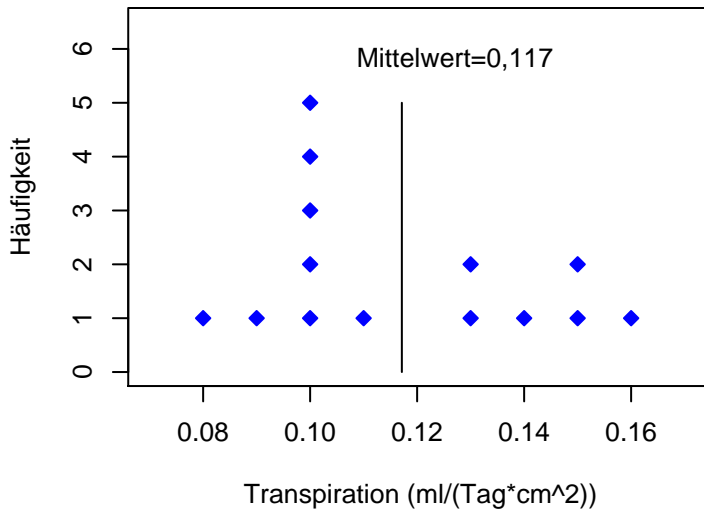
Ein Ziel des Versuchs: die mittlere  
Transpirationsrate  $\mu$   
(für diese Hirsesorte unter diesen  
Bedingungen)  
zu bestimmen.

In einem großen Versuch mit sehr vielen Pflanzen könnte man  $\mu$  beliebig genau bestimmen.

FRAGE:

Wie genau ist die Schätzung von  $\mu$  in diesem kleinen ( $n = 14$ ) Versuch?



Trockengestresste Hirse ( $n = 14$ )

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

$$\bar{x} = (x_1 + x_2 + \dots + x_{14})/14 = \frac{1}{14} \sum_{i=1}^{14} x_i$$

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

Unsere Schätzung:

$$\mu \approx 0,117$$

Wie genau ist diese Schätzung?

Wie weit weicht  $\bar{X}$  (unser Schätzwert)  
von  $\mu$  (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:  $N = \infty$ )

$\mu$

Stichprobe

$n = 14$

$\bar{x}$

Wir schätzen  
den Populationsmittelwert  
 $\mu$   
durch  
den Stichprobenmittelwert  
 $\bar{X}$ .

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

$\bar{x}$  hängt vom Zufall ab:  
eine *Zufallsgröße*

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

Genauer: Wie weit weicht  $\bar{x}$  typischerweise  
von  $\mu$  ab?



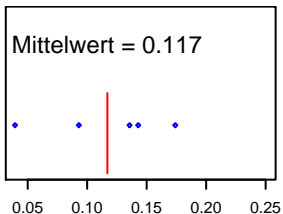
$$\bar{x} = (x_1 + x_2 + \cdots + x_n) / n$$

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

1.

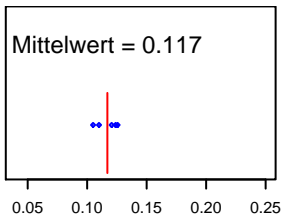
von der  
Variabilität  
der einzelnen  
Beobachtungen

$X_1, X_2, \dots, X_n$



$x$  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  $\bar{x}$ .

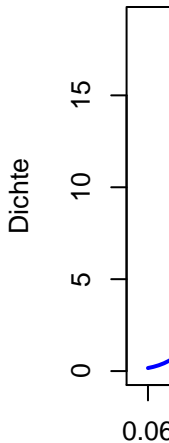
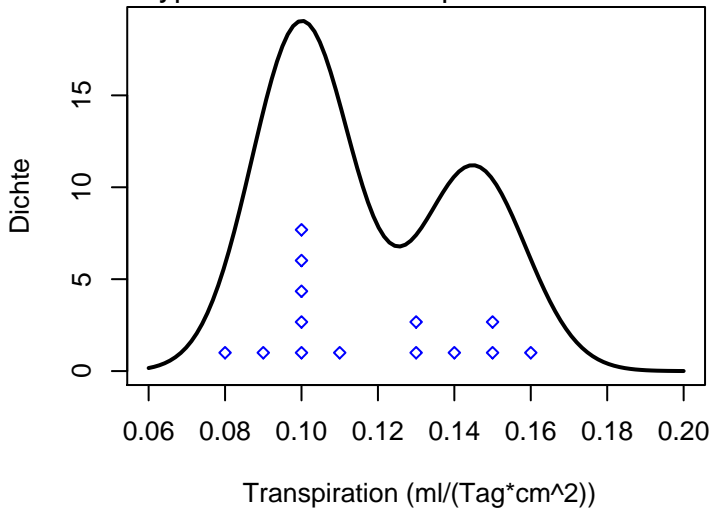
Um diese Abhängigkeit  
zu untersuchen,  
machen wir ein  
(Computer-)Experiment.

## Experiment:

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

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

## Hypothetische Transpirationsratenverteilung

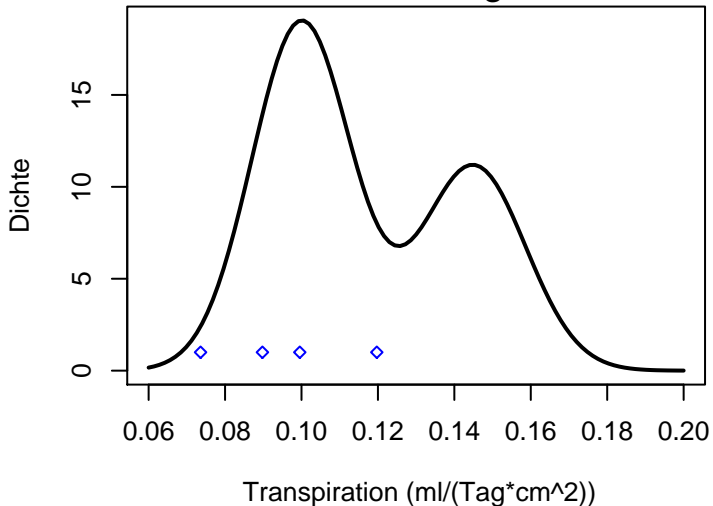




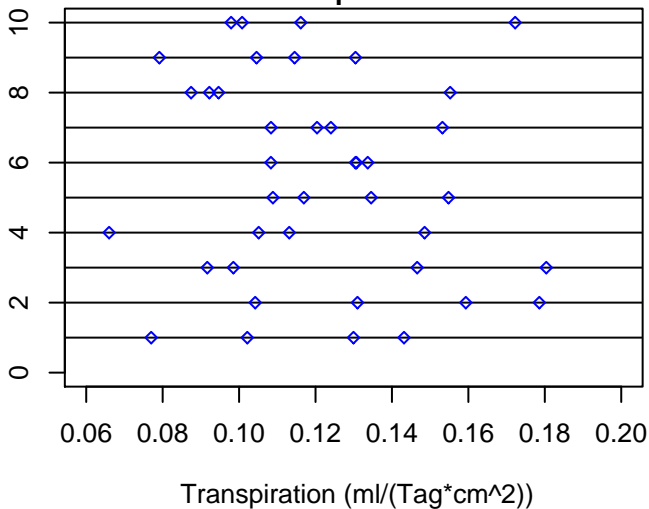
Wir beginnen mit kleinen Stichproben:

$$n = 4$$

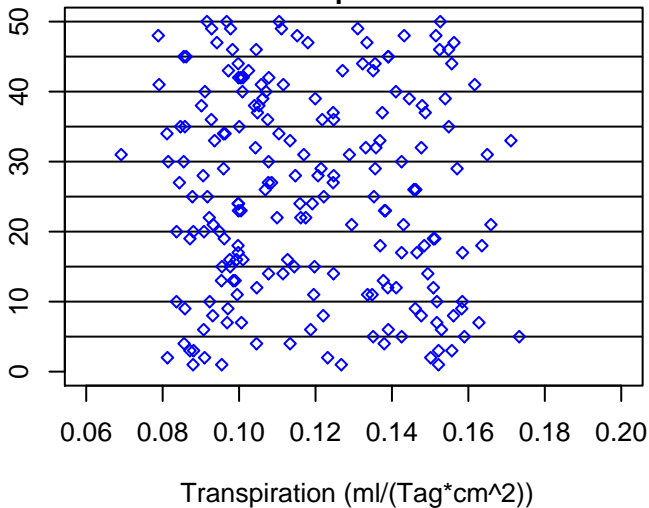
Eine Stichprobe vom Umfang 4 Eine zweite  
Stichprobe vom Umfang 4 Eine dritte Stichprobe vom  
Umfang 4



# 10 Stichproben

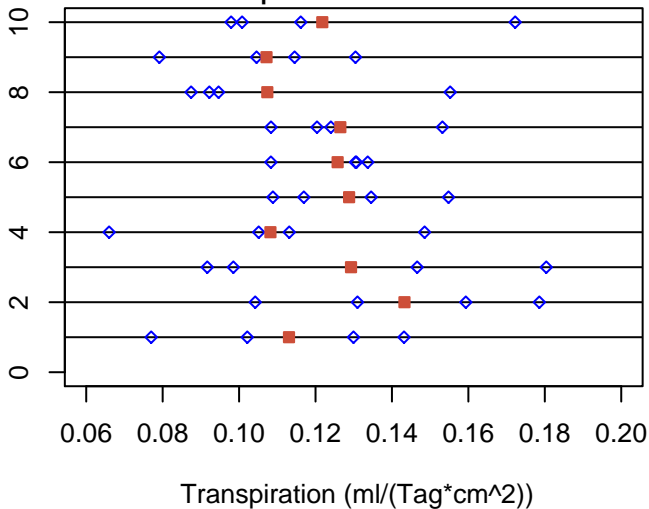


# 50 Stichproben

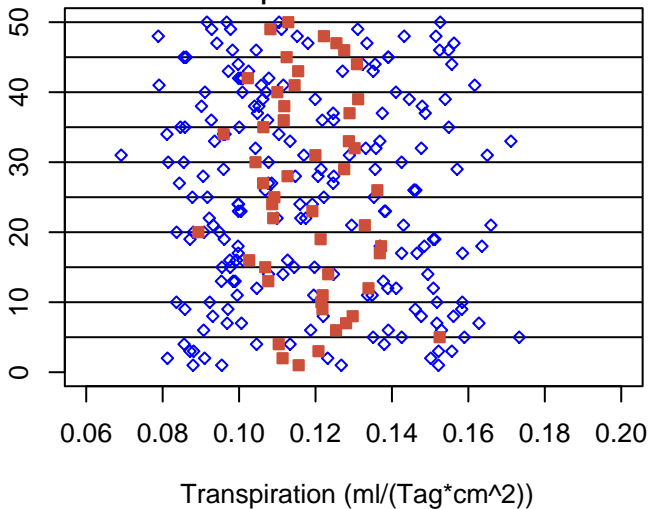


Wie variabel sind  
die Stichprobenmittelwerte?

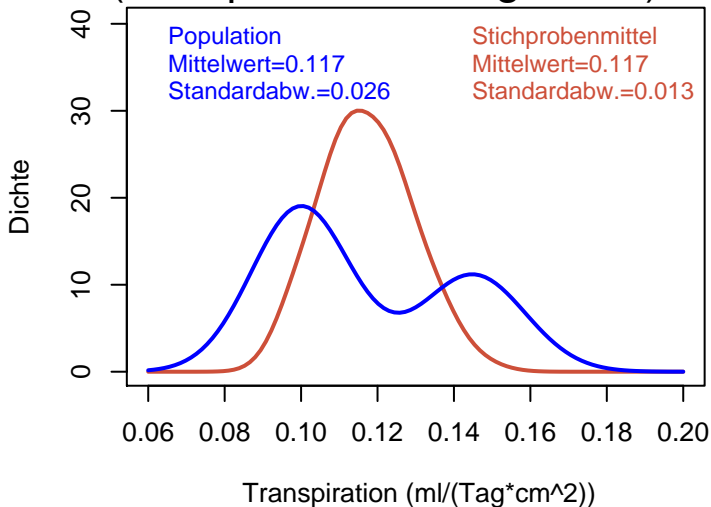
# 10 Stichproben vom Umfang 4 und die zugehörigen Stichprobenmittel



# 50 Stichproben vom Umfang 4 und die zugehörigen Stichprobenmittel



# Verteilung der Stichprobenmittelwerte (Stichprobenumfang $n = 4$ )





Population:

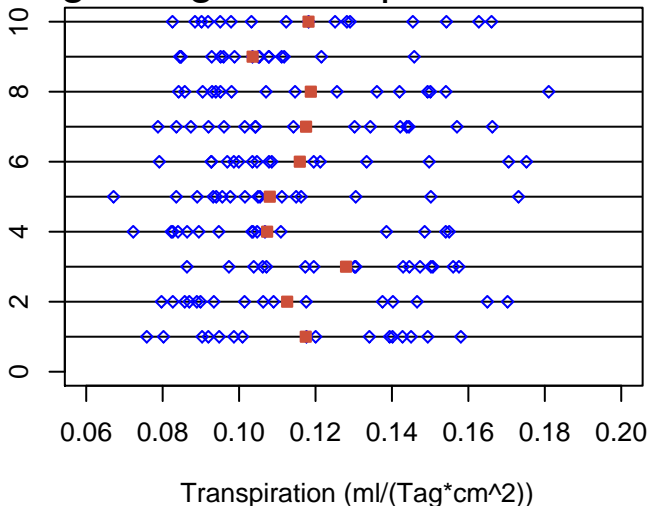
Standardabweichung = 0,026

Stichprobenmittelwerte ( $n = 4$ ):

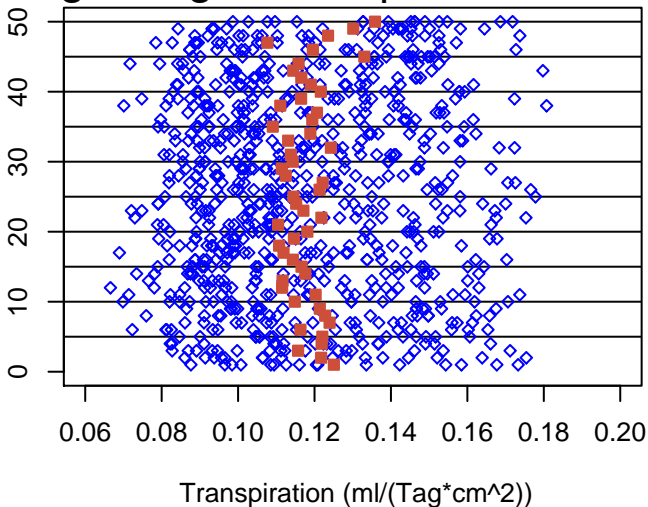
$$\begin{aligned}\text{Standardabweichung} &= 0,013 \\ &= 0,026 / \sqrt{4}\end{aligned}$$

Erhöhen wir  
den Stichprobenumfang  
von  
4  
auf  
16

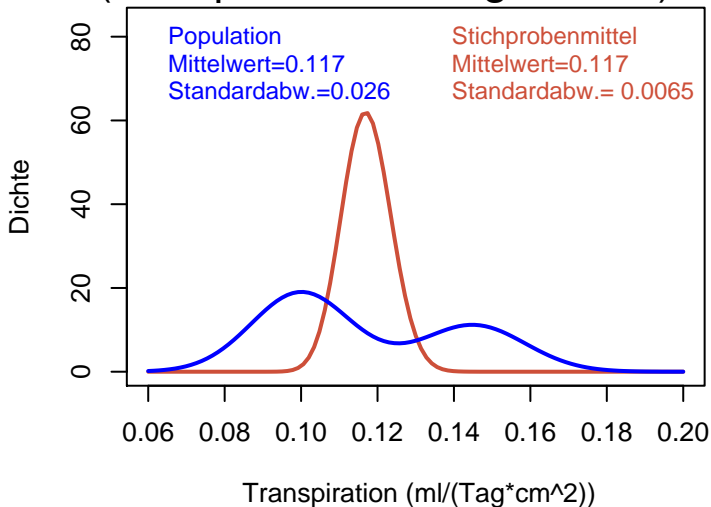
# 10 Stichproben vom Umfang 16 und die zugehörigen Stichprobenmittel



# 50 Stichproben vom Umfang 16 und die zugehörigen Stichprobenmittel



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



Population:

Standardabweichung = 0,026

Stichprobenmittelwerte ( $n = 16$ ):

$$\begin{aligned}\text{Standardabweichung} &= 0,0065 \\ &= 0,026/\sqrt{16}\end{aligned}$$

# 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:

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



In der Praxis ist  $\sigma$  unbekannt.

Es wird durch

die Stichproben-Standardabweichung  $S$   
geschätzt:

$$\sigma \approx S$$

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

$$s/\sqrt{n}$$

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

nennt man den

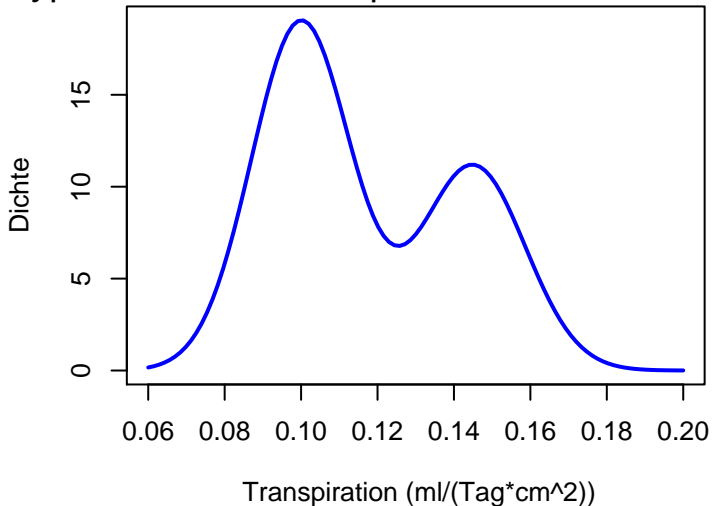
*Standardfehler.*

(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)

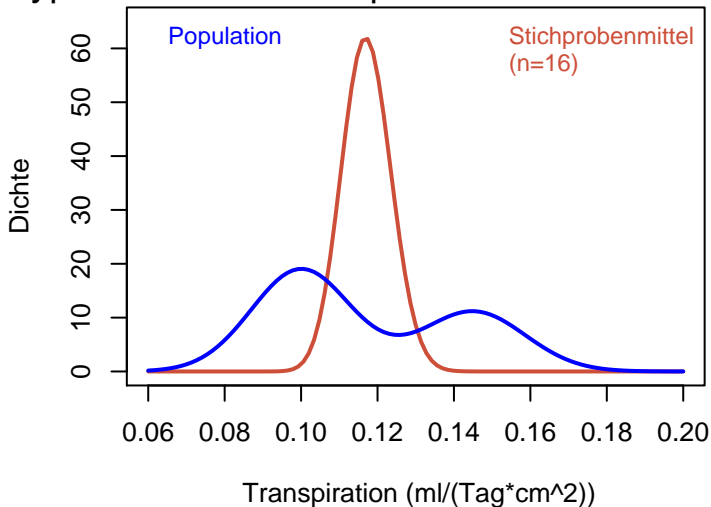
eingipfelig

&

symmetrisch

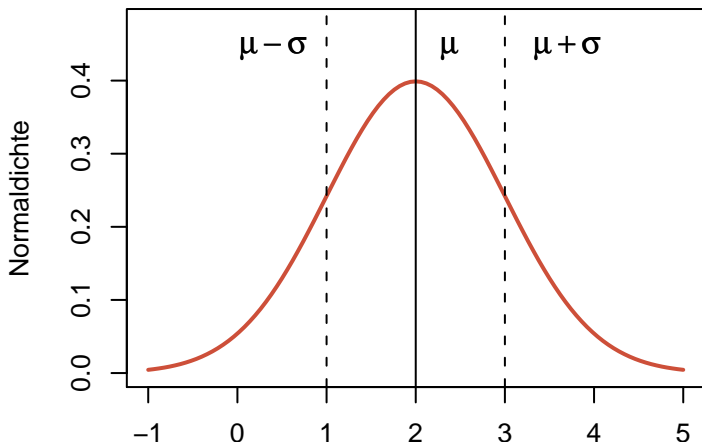
(wenn der Stichprobenumfang  $n$  nur groß genug ist)

# Hypothetische Transpirationsratenverteilung



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

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



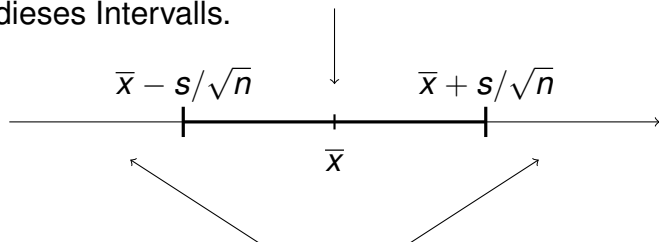
Die Normalverteilungsdichte heisst  
auch *Gauß'sche Glockenkurve*  
(nach Carl Friedrich Gauß, 1777-1855)



# Wichtige Folgerung

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

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



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

Beachte: Das wahre  $\mu$  ist unbekannt, aber nicht zufällig.

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

Demnach:

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

## Anwendung 1: Welche Werte von $\mu$ sind plausibel?

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

**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 Springkrebse



# Galathea: Carapaxlänge in einer Stichprobe

Männchen:

$$\bar{x}_1 = 3,04 \text{ mm}$$

$$s_1 = 0,9 \text{ mm}$$

$$n_1 = 25$$

Weibchen:

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

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

$$n_2 = 29$$

Die Weibchen  
scheinen größer zu sein.

Ist das ernst zu nehmen?

Oder könnte es nur **Zufall** sein?

# Wie genau sind die beiden Mittelwerte?

Männchen:

$$\bar{x}_1 = 3,04 \text{ mm}$$

$$s_1 = 0,9 \text{ mm}$$

$$n_1 = 25$$

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

Mit Schwankungen von

$$\pm 0,18 \text{ (mm) in } \bar{x}_1$$

müssen wir rechnen.



Wie genau sind die beiden Mittelwerte?

Weibchen:

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

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

$$n_2 = 29$$

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

Es ist nicht unwahrscheinlich, dass  $\bar{x}_2$  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

$$\bar{X}_2 > \bar{X}_1$$

Genauer formuliert:

Wenn in Wirklichkeit  
die Populationsmittelwerte  
gleich sind,

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

kann es trotzdem leicht passieren,  
dass die Stichprobenmittelwerte

$$\bar{X}_2 \text{ und } \bar{X}_1$$

so weit auseinander liegen.

Der Statistiker sagt:

Die Differenz  
der Mittelwerte  
ist

(statistisch)

nicht signifikant.

nicht signifikant

=

könnte Zufall sein

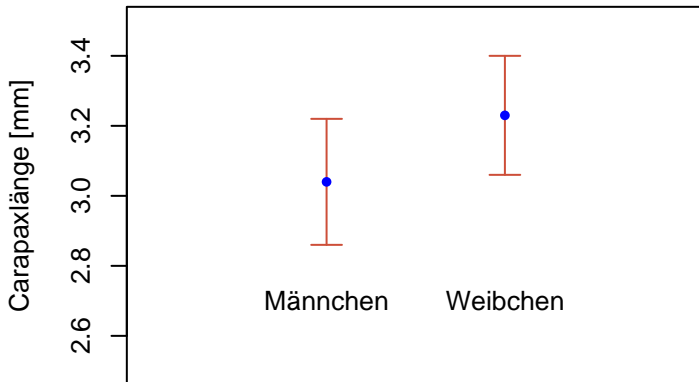
## Anwendung 3:

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

Also so:

Carapaxlängen:

Mittelwerte  $\pm$  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  $\bar{x}$  man erreichen will,

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

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

( $g$  = gewünschter Standardfehler)



Beispiel:  
Gestresste Transpirationswerte  
bei einer anderen Hirse-Sorte:

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

$$s = 0,06$$

$$n = 13$$

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:

$$s \approx 0,06,$$

$$\text{also } \sqrt{n} \approx 6$$

$$n \approx 36$$

## ZUSAMMENFASSUNG

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