

# Wahrscheinlichkeitsrechnung und Statistik für Studierende der Biologie **Einführung: Deskriptive Statistik**

Dirk Metzler, Noémie Becker

12. April 2020

## Inhaltsverzeichnis

<b>1 Einführung</b>	<b>1</b>
1.1 Zum Inhalt des ganzen Kurses . . . . .	1
1.2 Wie das alles lernen? . . . . .	4
<b>2 Ziele der deskriptiven (d.h. beschreibenden) Statistik</b>	<b>4</b>
<b>3 Graphische Darstellungen</b>	<b>5</b>
3.1 Histogramme und Dichtepolygone . . . . .	6
3.2 Scatterplots/Stripcharts . . . . .	11
3.3 Boxplots . . . . .	11
3.4 Beispiel: Ringeltaube . . . . .	13
3.5 Beispiel: Darwin-Finken . . . . .	15
<b>4 Statistische Kenngrößen</b>	<b>18</b>
4.1 Median und andere Quartile . . . . .	19
4.2 Mittelwert und Standardabweichung . . . . .	20
<b>5 Vom Sinn und Unsinn von Mittelwerten</b>	<b>29</b>
5.1 Beispiel: Wählerische Bachstelzen . . . . .	29
5.2 Beispiel: Spiderman & Spiderwoman . . . . .	30
5.3 Beispiel: Kupfertoleranz beim Roten Straußgras . . . . .	31

## 1 Einführung

### 1.1 Zum Inhalt des ganzen Kurses

*It is easy to lie with statistics. It is hard to tell the truth without it.*

Andrejs Dunkels

Die Natur ist voller Variabilität.

Wie geht man mit variablen Daten um?

Es gibt eine mathematische Theorie des Zufalls: die **Stochastik**.

IDEE DER STATISTIK:

Variabilität (Erscheinung der Natur) durch Zufall (mathematische Abstraktion) modellieren.

Also: Statistik ist Datenanalyse mit Hilfe stochastischer Modelle.

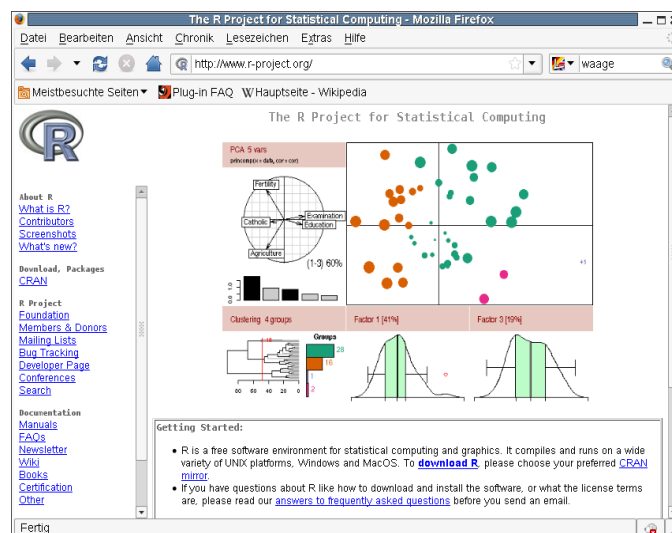
## Quellen

Wir danken Brooks Ferebee, Gaby Schneider und Anton Wakolbinger von der Uni Frankfurt für die Bereitstellung vieler Beispiele und Lehrmaterialien sowie Matthias Birkner (jetzt Uni Mainz) und Martin Hutzenthaler (jetzt Uni Essen) für die intensive Zusammenarbeit beim Erstellen früherer Version dieser Vorlesung.

## Themen der Vorlesung

1. Beschreibende Statistik
2. Der Standardfehler
3. Grundlagen aus der der Wahrscheinlichkeitstheorie
4. Statistische Tests (jeweils mehrere Varianten)
  - t-Test
  - Chi-Quadrat Test und Fishers exakter Test
  - nichtparametrische Tests
  - Varianzanalyse (ANOVA)
5. Lineare Regression
6. Parameterschätzung und Konfidenzintervalle
7. evtl. Diskriminanzanalyse
8. bei allen Themen: R

## Statistik-Software R



<http://www.r-project.org>

## Wieso R?

- R wird immer mehr zu DEM Standardprogramm für Datenanalysen aller Art
- R und sehr viele Zusatzpakete für spezielle Methoden frei erhältlich
- Befehle sollten in R-Skripten gespeichert werden, dann reproduzierbar
- Datenanalysen automatisierbar, da R auch Programmiersprache ist
- Zur Eingabe und Vorbehandlung der Daten kann ein Tabellenkalkulationsprogramm wie LibreOffice Calc oder Excel verwendet werden (bei komplexen Datensätzen besser eine Datenbank).

## R und RStudio

**RStudio:** Ein Programm, das aus Sicht von vielen Benutzer(inne)n die Verwendung von R einfacher, intuitiver und effizienter macht.

Falls Sie ein Laptop haben,

- installieren Sie bitte noch heute R und RStudio, z.B. von den Webseiten
  - <https://ftp.gwdg.de/pub/misc/cran/> und
  - <https://www.rstudio.com/products/rstudio/download/#download>,
- laden Sie die Datei [http://evol.bio.lmu.de/\\_statgen/StatBiol/R\\_intro.R](http://evol.bio.lmu.de/_statgen/StatBiol/R_intro.R) in R oder RStudio und probieren Sie die darin enthaltenen Befehlszeilen aus
- und bringen Sie, wenn möglich, Ihr Laptop in die erste Übung mit.

## Folien, R-Befehle, Quellen und Übungen

[http://evol.bio.lmu.de/\\_statgen/StatBiol/](http://evol.bio.lmu.de/_statgen/StatBiol/)

## Lernziele

- Prinzipien und Argumentationsweisen der Statistik verstehen
- Wahrscheinlichkeitstheoretische Grundlagen verstehen
- Wichtige Tests und Methoden verstehen
- Diese auf Daten anwenden können (mit R)
- Wann welche Methode und welche Voraussetzungen müssen erfüllt sein?
- Wann muss ich einen Experten zu Rate ziehen?
- Wissenschaftliche Literatur verstehen können, einschließlich darin verwendeter statistischer Methoden, die von Standardmethoden abweichen können
- Ein Gefühl für Zufallseffekte

## 1.2 Wie das alles lernen?

### Arbeitsaufwand

6 ECTS = 180 Stunden (h), d.h.  $\approx$  13 h pro Woche.

Im Durchschnitt über alle Wochen:

- 90+45+90 min  $\approx$  4 h in Vorlesungen und Übung
- 4 h Stoff aus der Vorlesung lernen
- 5 h Übungsaufgaben bearbeiten

### Mögliche Lernstrategie

- Finden Sie Erklärungen für “Was Sie u.a. erklären können sollten” (siehe Ende eines Themengebiets).
- Diskutieren Sie Ihre Erklärungen mit einigen Mitstudierenden
- Lernen Sie dann auch alle anderen Inhalte aus der Vorlesung (auch mit Büchern)
- Bearbeiten Sie die Übungsaufgaben
- Präsentieren und diskutieren Sie Ihre Lösungen in den Übungsgruppen
- Tun Sie das alles rechtzeitig, um, wenn nötig, in der nächsten Vorlesung Fragen stellen zu können

### Wie wird die Klausur aussehen?

- Anders als die aus den letzten Semestern!
- Sie dürfen mitbringen:
  - Taschenrechner
  - selbst-handgeschriebenes Formelblatt
- Was brauchen Sie sonst noch?
  - Verständnis der Konzepte aus der Vorlesung
  - Fähigkeit, Konzepte anzuwenden (aus den Übungen)
  - Denkvermögen während der Klausur (nicht nur Faktenwissen oder rezepthaftes Anwenden)
  - Kenntnisse in R

## 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik

### Beschreibende Statistik

Beschreibende Statistik: Ein erster Blick auf die Daten

### 3 Graphische Darstellungen

#### Beispiel

Daten aus einer Diplomarbeit aus 2001 am Forschungsinstitut Senckenberg,  
Frankfurt am Main

Crustaceensektion

*Leitung: Dr. Michael Türkay*

*Charybdis acutidens TÜRKAY 1985*

Der Springkrebs

*Galathea intermedia*

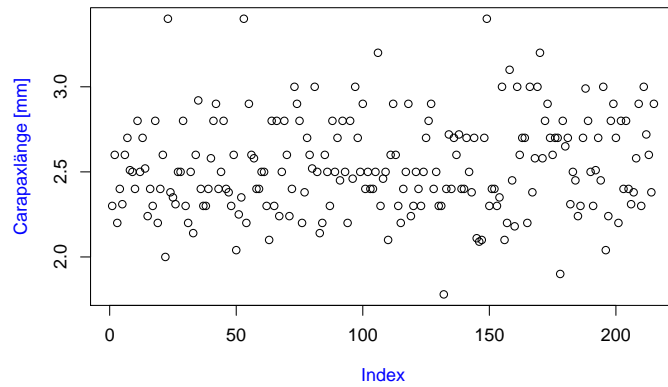


Helgoländer Tiefe Rinne, Fang vom 6.9.1988

Carapaxlänge (mm): Nichteiertragende Weibchen ( $n = 215$ )

2,9	3,0	2,9	2,5	2,7	2,9	2,9	3,0
3,0	2,9	3,4	2,8	2,9	2,8	2,8	2,4
2,8	2,5	2,7	3,0	2,9	3,2	3,1	3,0
2,7	2,5	3,0	2,8	2,8	2,8	2,7	3,0
2,6	3,0	2,9	2,8	2,9	2,9	2,3	2,7
2,6	2,7	2,5	.	.	.	.	.

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

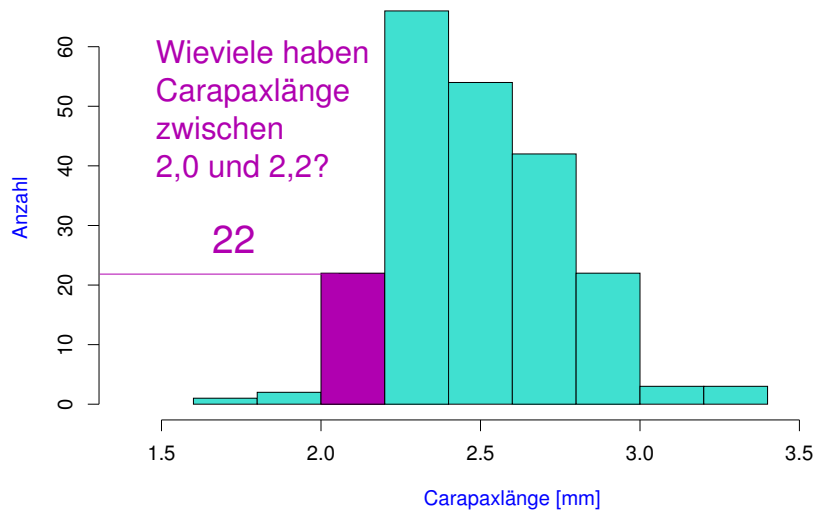


### 3.1 Histogramme und Dichtepolygone

Eine Möglichkeit der graphischen Darstellung:

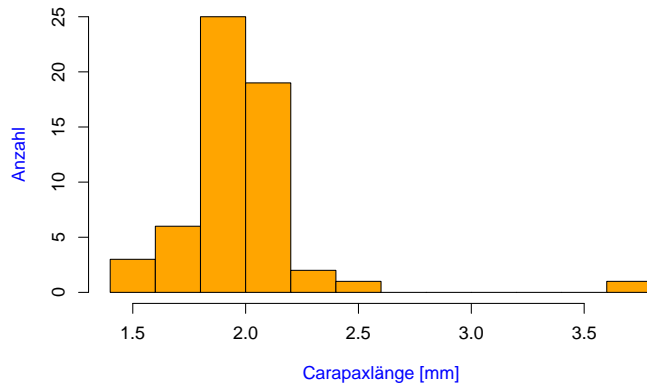
das Histogramm

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

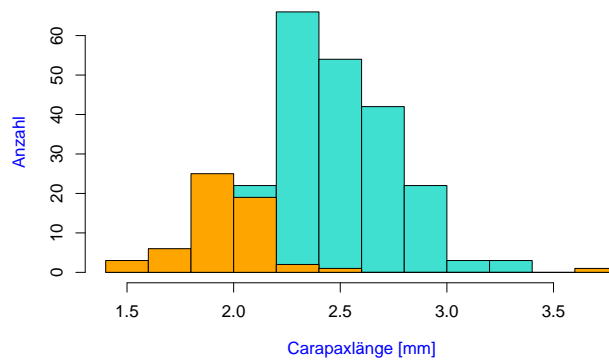


Analoge Daten zwei Monate später (3.11.88):

Nichteiertragende Weibchen am 3. Nov. '88, n=57



Nichteiertragende Weibchen



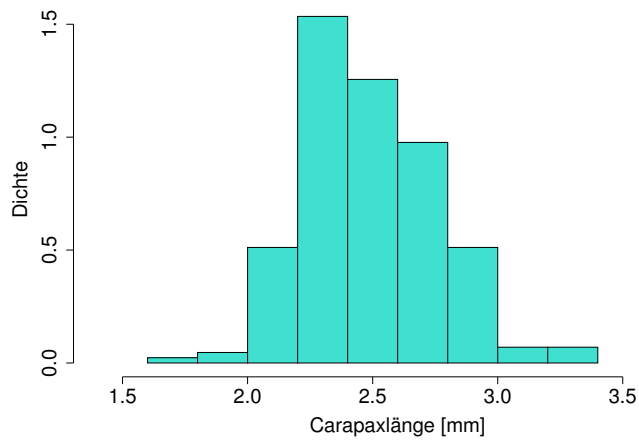
Vergleich der beiden Verteilungen

Problem: ungleiche Stichprobenumfänge: 6.Sept:  $n = 215$

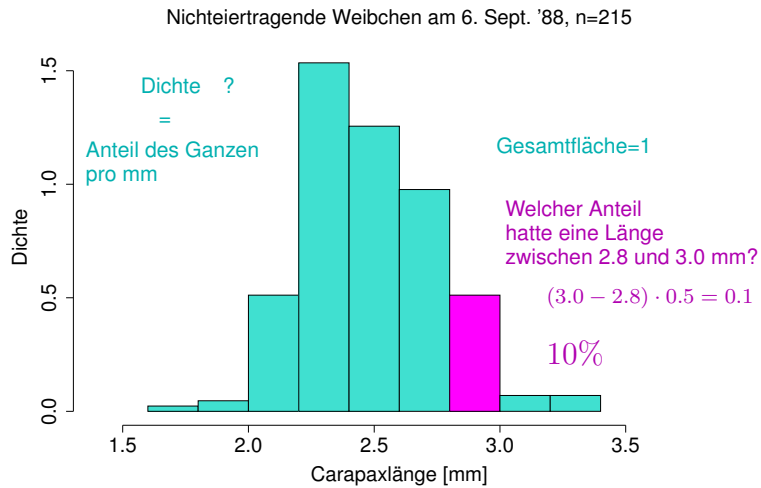
3.Nov :  $n = 57$

Idee: stauche vertikale Achse so, dass Gesamtfläche = 1.

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

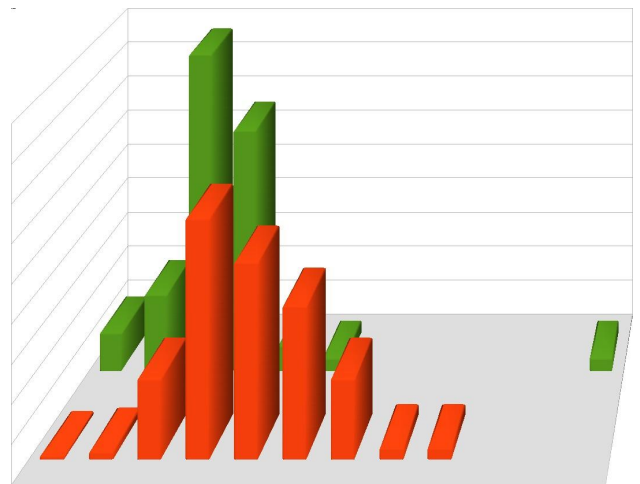
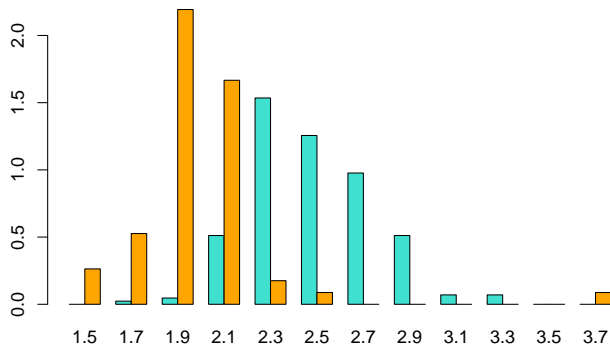
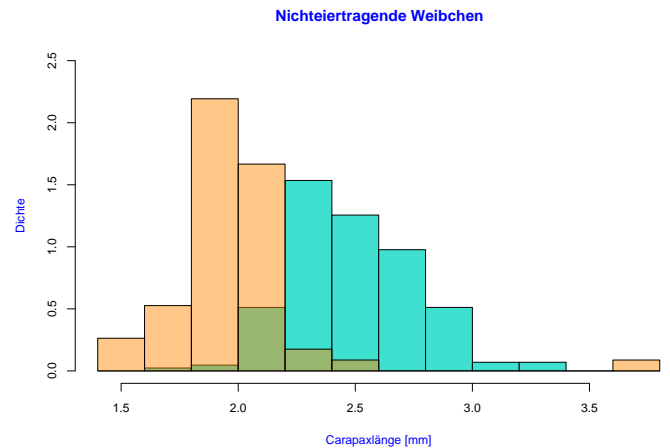
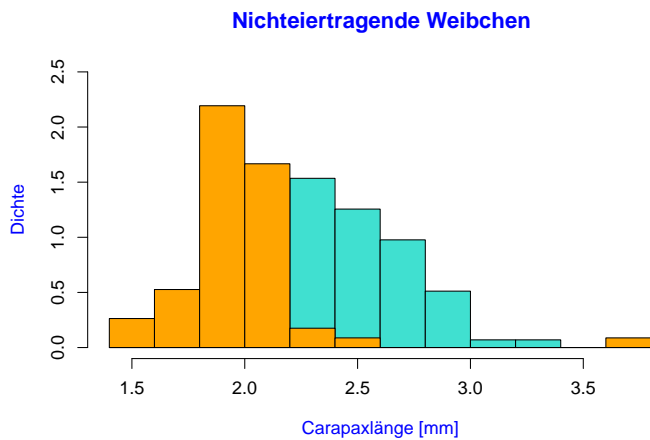


Die neue vertikale Koordinate ist jetzt eine Dichte (engl. *density*).



Die beiden Histogramme sind jetzt vergleichbar, denn sie haben dieselbe Gesamtfläche:

Versuche, die Histogramme zusammen zu zeigen:





**Unser Rat an Sie:**

Wenn Sie in der Werbebranche arbeiten:

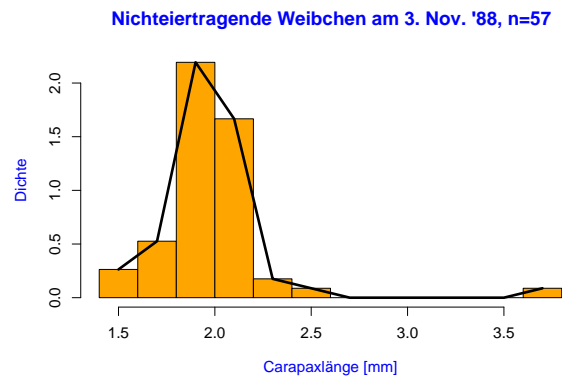
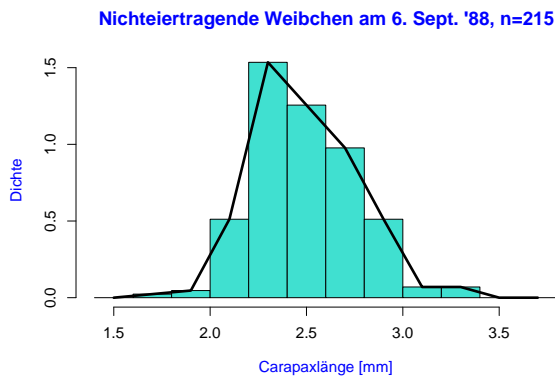
Beeindrucken Sie Jung und Alt mit total abgefahrenen 3D-Plots!

Wenn Sie Wissenschaftler(in) werden wollen:

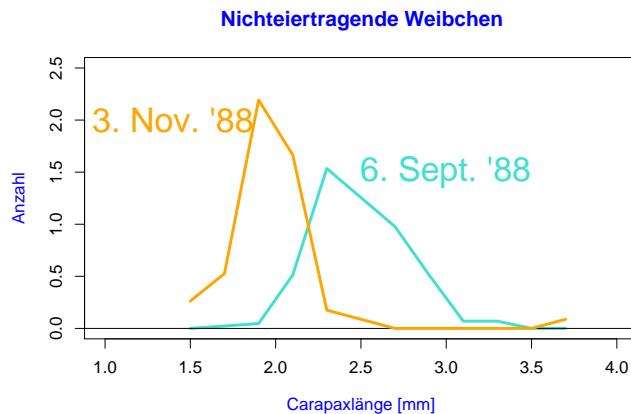
Bevorzugen Sie einfache und klare 2D-Darstellungen.

Problem: Histogramme kann man nicht ohne weiteres in demselben Graphen darstellen, weil sie einander überdecken würden.

**Einfache und klare Lösung: Dichtepolygone**

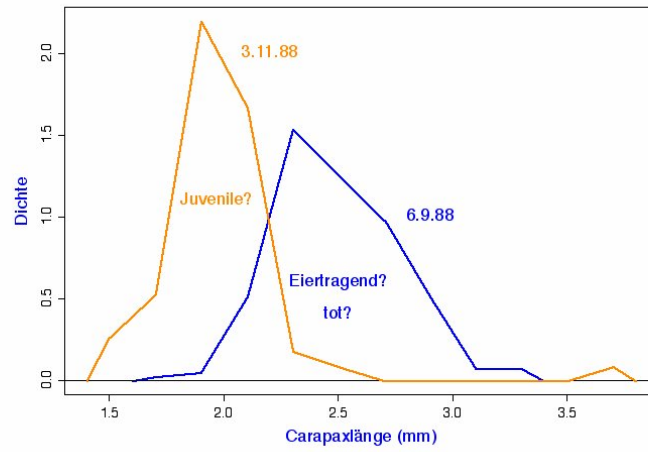


**Zwei und mehr Dichtepolygone in einem Plot**

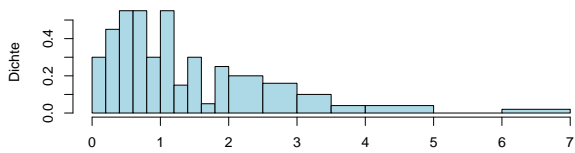
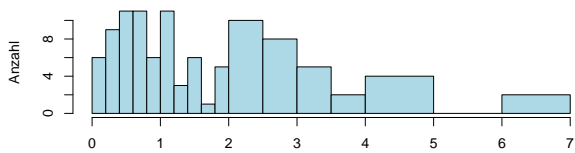
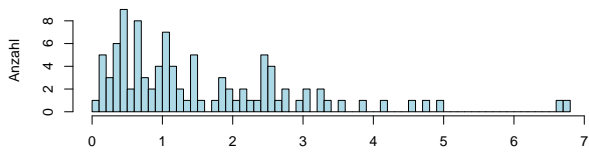


Biologische Interpretation der Verschiebung?

Nichteiertragende Weibchen 6.9.88 und 3.11.88

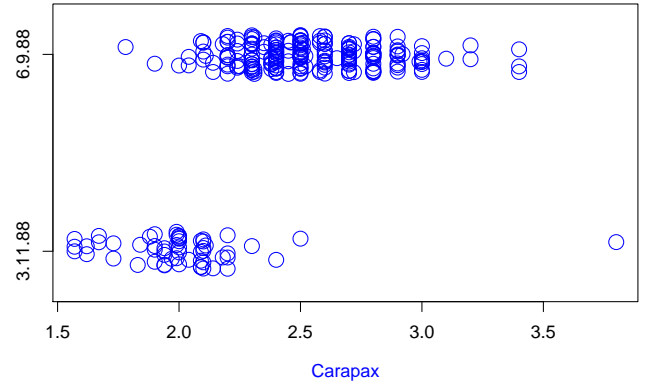
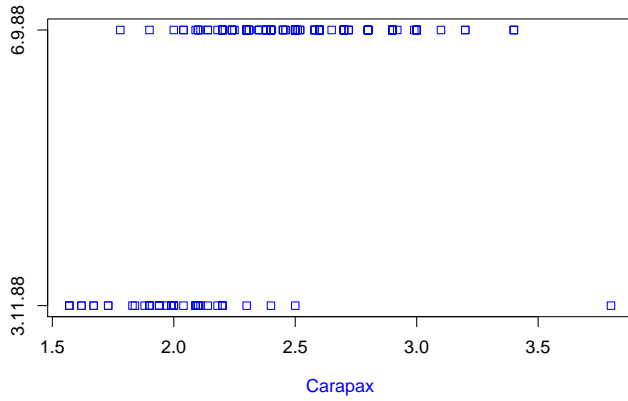


### Anzahl vs. Dichte

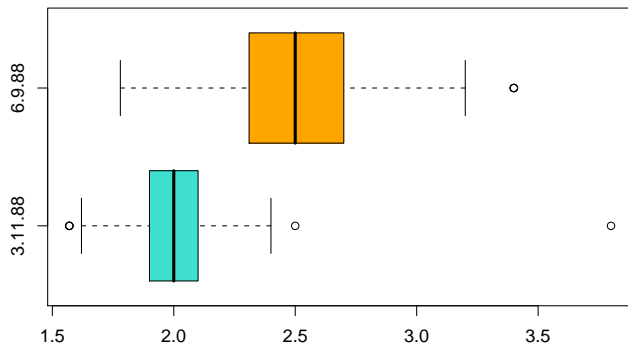


Also: Bei Histogrammen mit ungleichmäßiger Unterteilung immer Dichten verwenden!

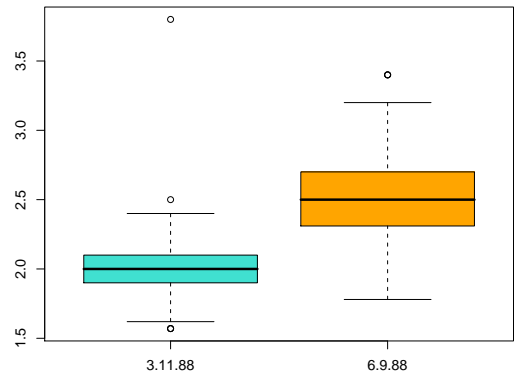
### 3.2 Scatterplots/Stripcharts



Boxplots, horizontal



Boxplots, vertikal



Histogramme und Dichtepolygone geben ein ausführliches Bild eines Datensatzes. Manchmal zu ausführlich.

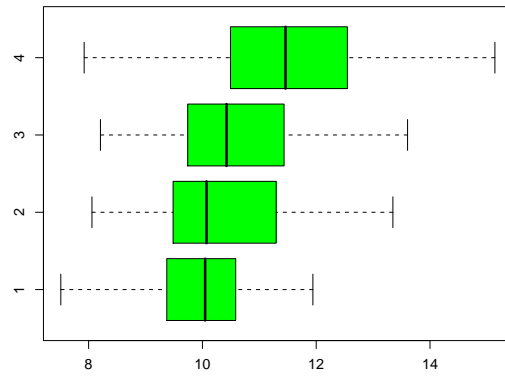
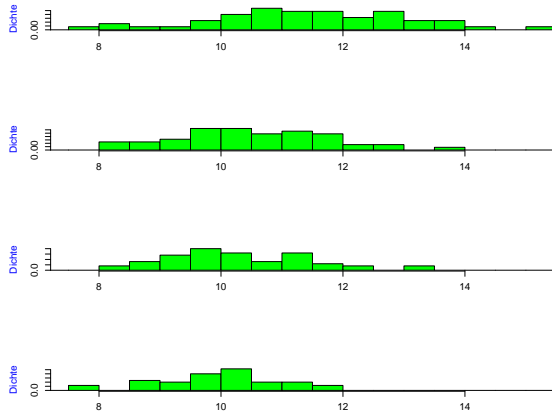
### 3.3 Boxplots

Zu viel Information erschwert den Überblick

Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum

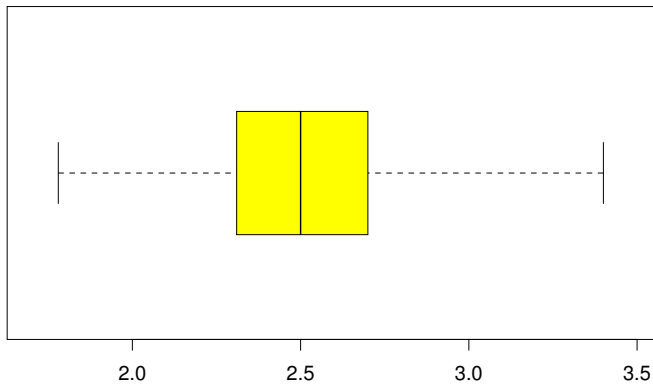
Wald?

Beispiel:  
Vergleich von mehreren Gruppen

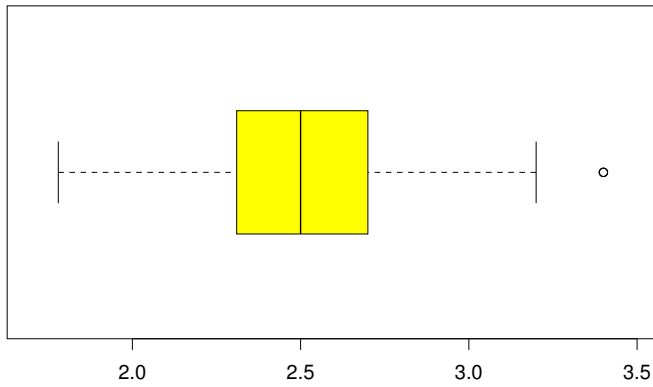


Der Boxplot

Boxplot, einfache Ausführung

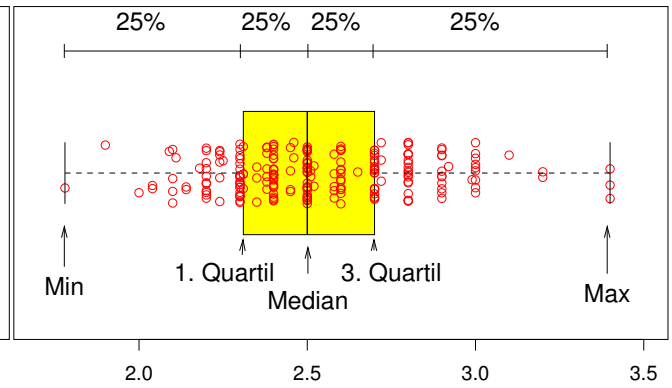


Carapaxlänge [mm]  
Boxplot, Standardausführung

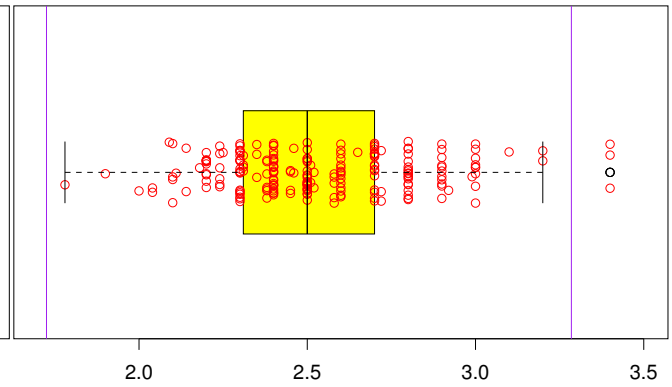


Carapaxlänge [mm]

Boxplot, einfache Ausführung

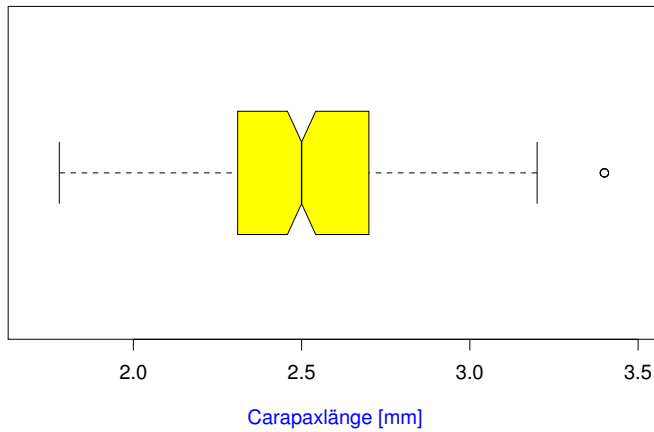


Carapaxlänge [mm]  
Boxplot, Standardausführung

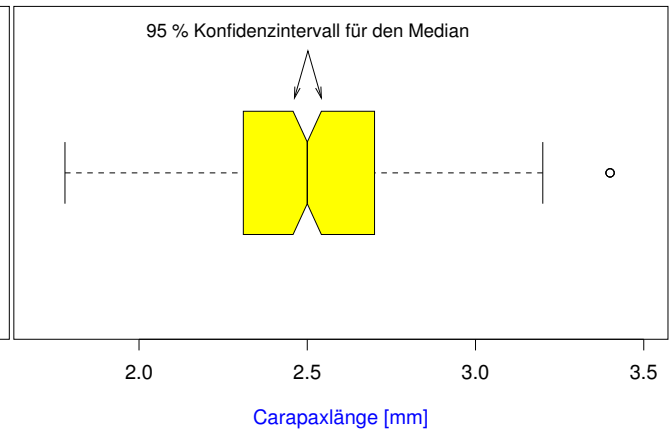


Carapaxlänge [mm]

Boxplot, Profiausstattung



Boxplot, Profiausstattung



### 3.4 Beispiel: Ringeltaube

Beispiel:

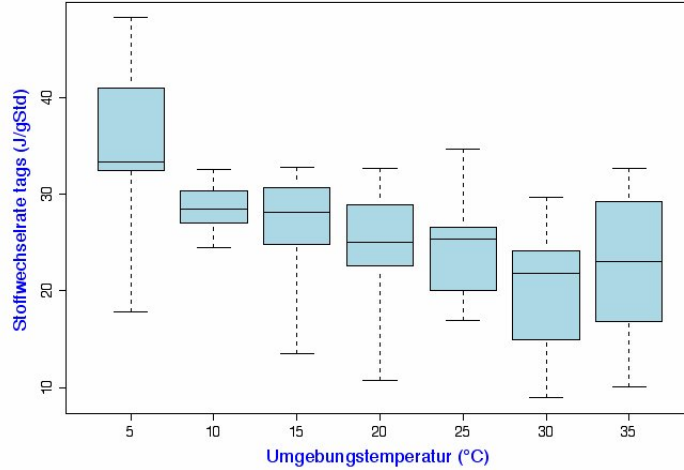
Die Ringeltaube

*Palumbus palumbus*

Wie hängt die Stoffwechselrate bei der Ringeltaube von der Umgebungstemperatur ab?

Daten aus dem AK Stoffwechselphysiologie  
Prof. Prinzing Universität Frankfurt

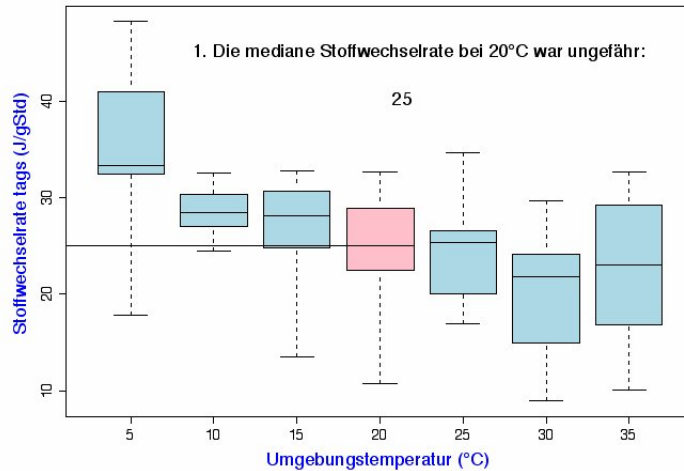
Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)



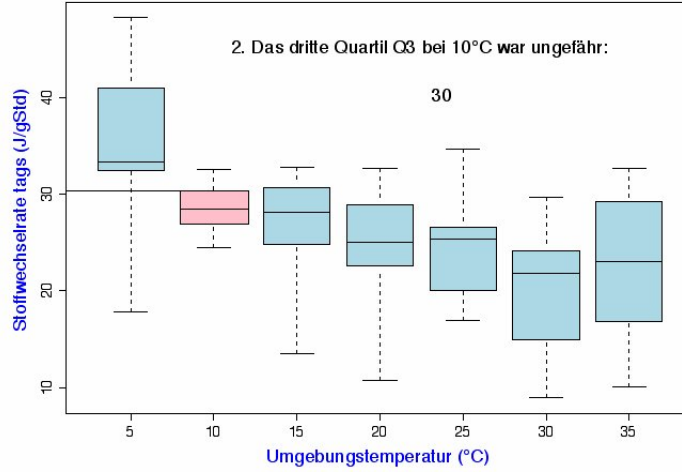
Klar: Stoffwechselrate *höher* bei *tiefen* Temperaturen

Vermutung: Bei *hohen* Temperaturen nimmt die Stoffwechselrate wieder zu (Hitzestress).

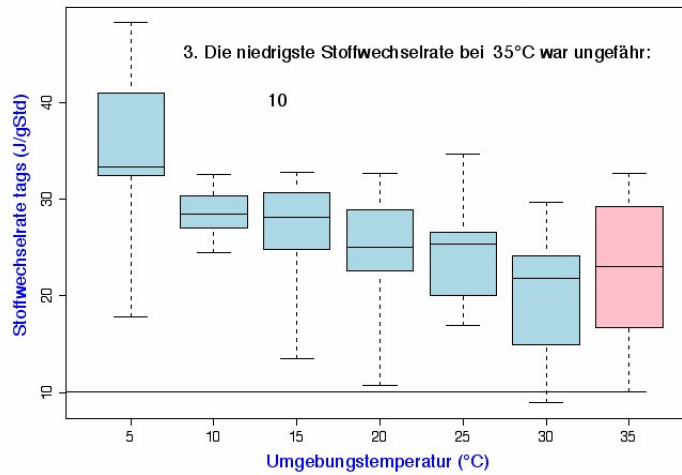
Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)



Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)



Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)

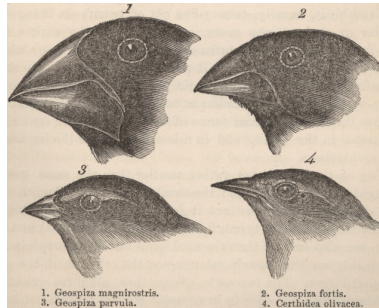


### 3.5 Beispiel: Darwin-Finken

Charles Robert Darwin (1809-1882)



## Darwin-Finken



[http://darwin-online.org.uk/graphics/Zoology\\_Illustrations.html](http://darwin-online.org.uk/graphics/Zoology_Illustrations.html)

## Darwins Finken-Sammlung

## Literatur

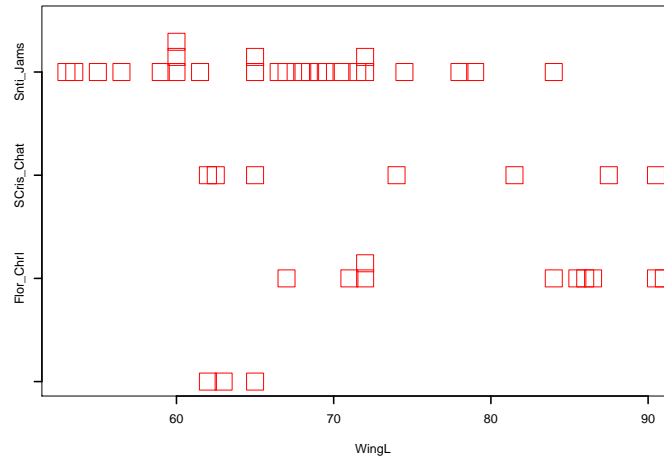
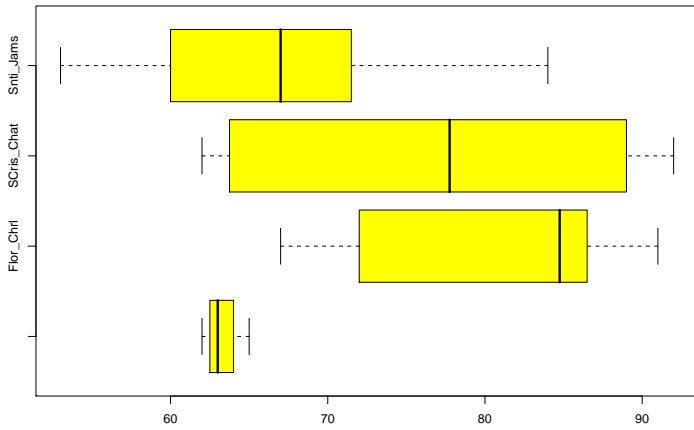
[1] Sulloway, F.J. (1982) The Beagle collections of Darwin's Finches (Geospizinae). *Bulletin of the British Museum (Natural History), Zoology series* **43**: 49-94.

[2] <http://datadryad.org/repo/handle/10255/dryad.154>

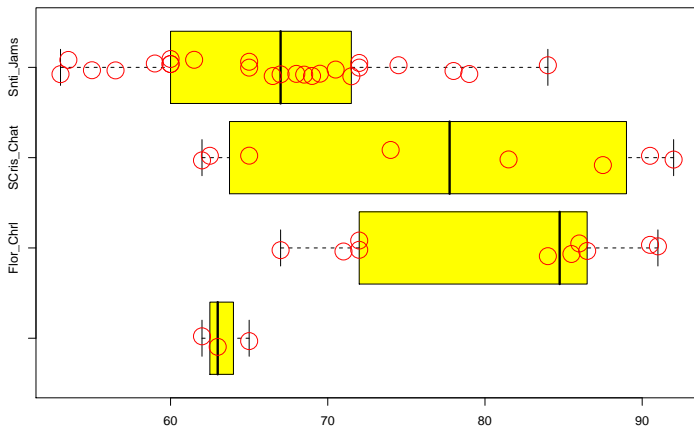
## Flügelängen der Darwin-Finken



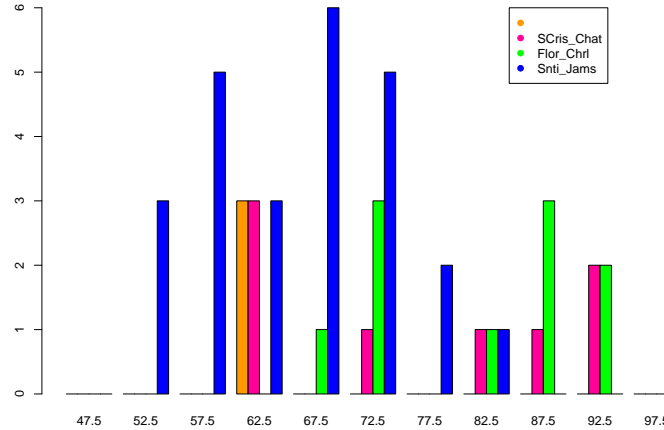
Flügelängen je nach Insel



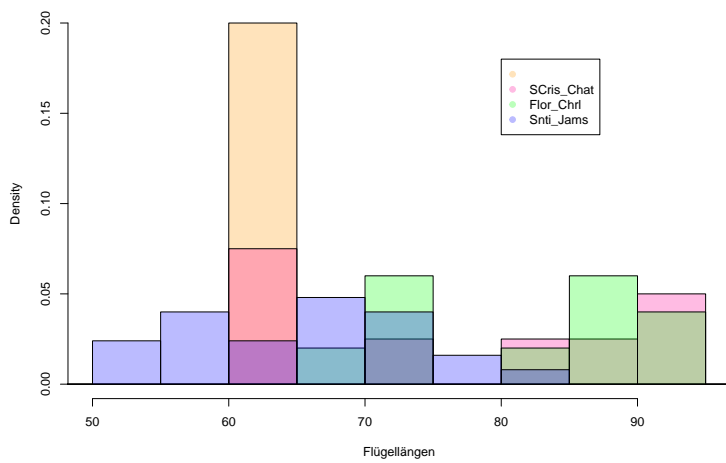
Flügelängen je nach Insel



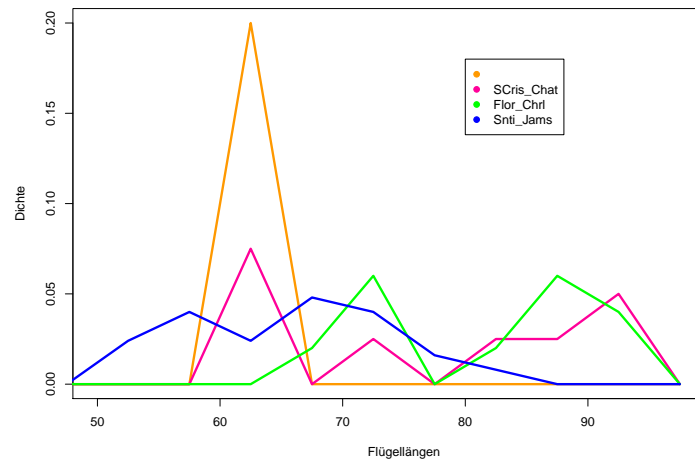
Barplot für Flügelängen (Anzahlen)

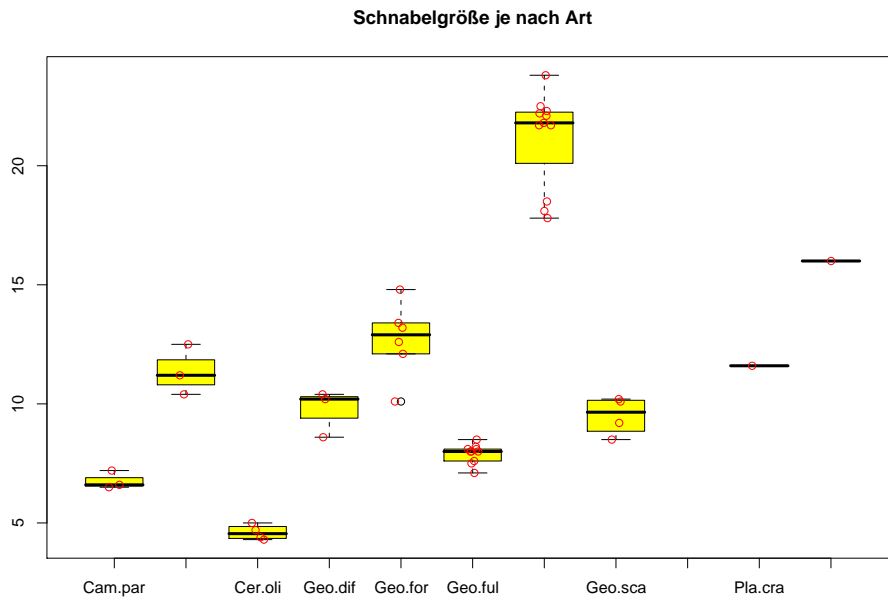


Histogramm (Dichten!) mit Transparenz



Dichteplot





## Fazit

1. Histogramme erlauben einen detaillierten Blick auf die Daten
2. Dichtepolygone erlauben Vergleiche zwischen vielen Verteilungen
3. Boxplot können große Datenmengen vereinfacht zusammenfassen
4. Bei kleinen Datenmengen eher Stripcharts verwenden
5. Vorsicht mit Tricks wie 3D oder halbtransparenten Farben
6. Jeder Datensatz ist anders; keine Patentrezepte

## Was Sie u.a. erklären können sollten

- Ihr Konzept, um die Inhalte dieses Kurses zu lernen
- Was ist ein Dichte?
- Wie man Histogramme und Dichte-Plots interpretiert
- Boxplots und Stripcharts/Scatterplots und wann man sie benutzt
- Quantile, Quartile und Median

## 4 Statistische Kenngrößen

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 ( $Q_3 - Q_1$ )*

#### 4.1 Median und andere Quartile

Der *Median*:

die Hälfte der Beobachtungen sind kleiner, die Hälfte sind größer. (grob gesagt.)

Genauer: (da auch Messwerte mit dem Median übereinstimmen können):  
Mindestens 50% der Beobachtungen sind kleiner oder gleich und mindestens 50% sind größer oder gleich.

Der Median ist  
das 50%-*Quantil*  
der Daten.

## Die Quartile

*Das erste Quartil,  $Q_1$ :* Grob gesagt: ein Viertel der Beobachtungen sind kleiner, drei Viertel sind größer.

Genauer (da auch Messwerte mit  $Q_1$  übereinstimmen können):  
Mindestens ein Viertel der Beobachtungen sind kleiner oder gleich und  
mindestens drei Viertel sind größer oder gleich.

$Q_1$  ist das 25%-*Quantil* der Daten.

## Die Quartile

*Das dritte Quartil,  $Q_3$ :* Mindestens drei Viertel der Beobachtungen sind kleiner oder gleich und mindestens ein Viertel der Werte sind größer oder gleich.

$Q_3$  ist das 75%-*Quantil* der Daten.

## 4.2 Mittelwert und Standardabweichung

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:**

$$\begin{aligned} \text{Mittelwert[lex]} &= [\text{lex}] \\ &= \frac{\text{Summe der Messwerte}}{\text{Anzahl der Messwerte}} \\ &= \frac{\text{Summe}}{\text{Anzahl}} \end{aligned}$$

Der Mittelwert von  $x_1, x_2, \dots, x_n$  als Formel:

$$\begin{aligned} \bar{x} &= (x_1 + x_2 + \dots + x_n)/n \\ &= \frac{1}{n} \sum_{i=1}^n x_i \end{aligned}$$

Beispiel:

$$x_1 = 3, x_2 = 0, x_3 = 2, x_4 = 3, x_5 = 1$$

$$\bar{x} = \text{Summe}/\text{Anzahl}$$

$$\bar{x} = (3 + 0 + 2 + 3 + 1)/5$$

$$\bar{x} = 9/5$$

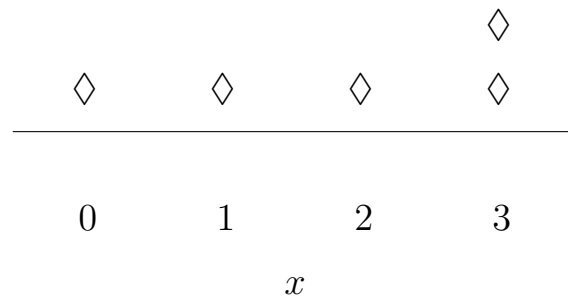
$$\bar{x} = 1,8$$

Geometrische Bedeutung des Mittelwerts:

Der Schwerpunkt

Wir stellen uns die Beobachtungen als gleich schwere Gewichte auf einer Waage vor:

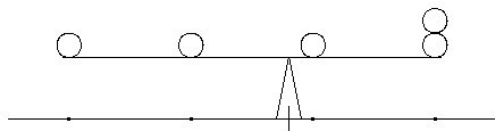
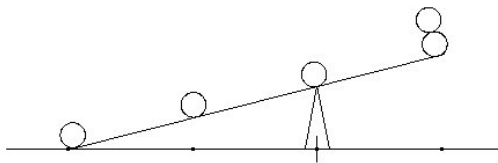
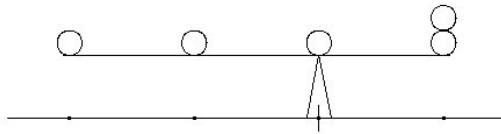
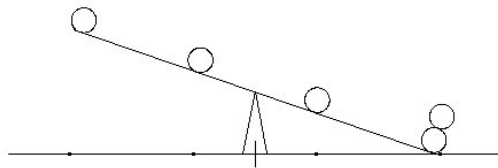
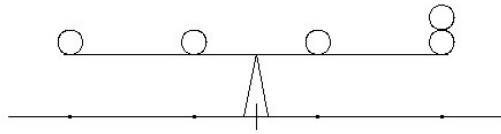
Wo muß der Drehpunkt sein, damit die Waage im Gleichgewicht ist?



$m = 1,5 ?$

$m = 2 ?$

$m = 1,8 ?$



zu klein

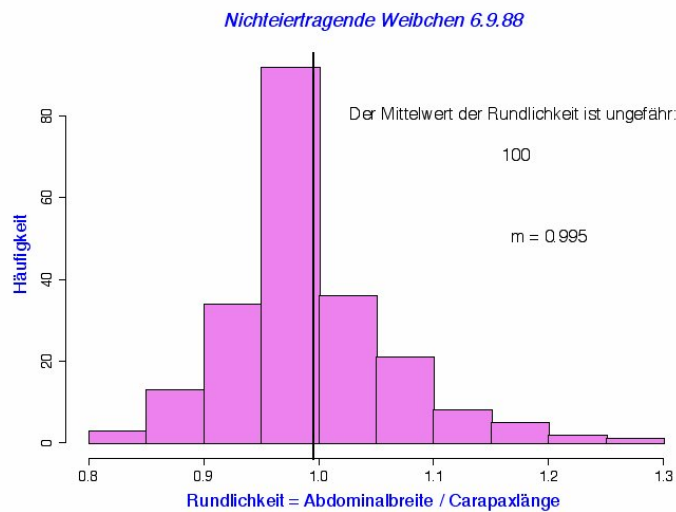
zu groß

richtig

Beispiel: *Galathea intermedia*

„Rundlichkeit“ :=  $\text{Abdominalbreite} / \text{Carapaxlänge}$

Vermutung: Rundlichkeit nimmt bei Geschlechtsreife zu

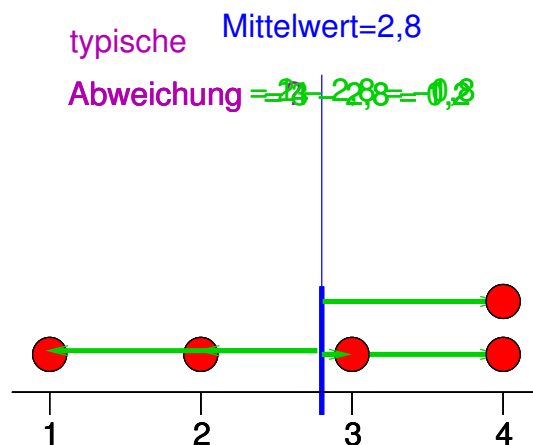


Beispiel:

3.11.88

Die Standardabweichung

Wie weit weicht eine typische Beobachtung vom Mittelwert ab ?



Die *Standardabweichung*  $\sigma$  (“sigma”) [auch *SD* von engl. *standard deviation*] 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:

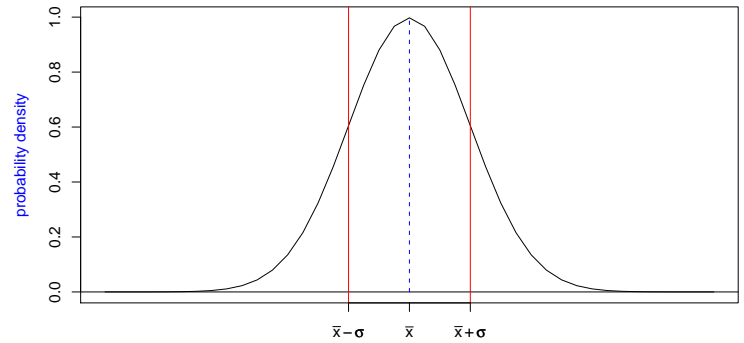
$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

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

**Faustregel für die Standardabweichung**

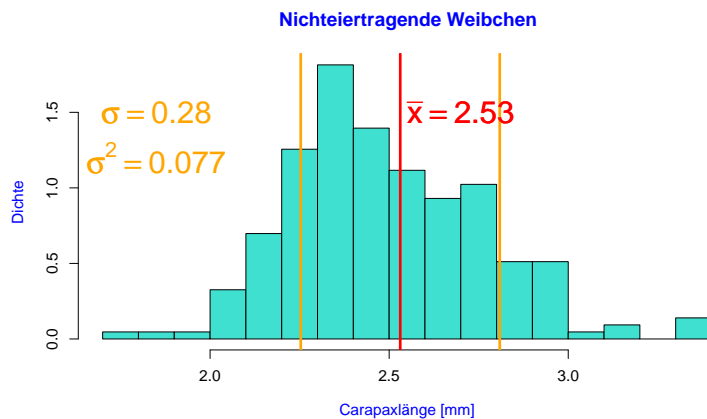
Bei ungefähr glockenförmigen (also eingipfligen und symmetrischen) Verteilungen liegen ca. 2/3 der Ver-





teilung zwischen  $\bar{x}-\sigma$  und  $\bar{x}+\sigma$ .

### Standardabweichung der Carapaxlängen nichteiertragender Weibchen vom 6.9.88



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

### Varianz der Carapaxlängen nichteiertragender 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  $\sigma_S^2$  im Durchschnitt um den Faktor  $\frac{n-1}{n}$  ( $= 214/215 \approx 0,995$ ) kleiner als  $\sigma_{\mathcal{X}}^2$

### Parameter vs. Statistik

**Parameter:** Größe in einer mathematischen Beschreibung, also Modellierung, des realen Systems

- Beispiel: Varianz in der Gesamtpopulation
- Genauer Wert i.d.R. unbekannt, evtl. wird ein hypothetischer Wert angenommen

**Statistik:** Wert, der aus Daten berechnet werden kann (also Funktion der Daten)

- Beispiel: Stichprobenvarianz

- Ist zufällig, da die Daten als zufällig angenommen werden

**Schätzer:** Statistik, die zum Schätzen des Werts eines Parameters verwendet wird

### Varianzbegriffe

Varianz in der Population (ein Parameter!):  $\sigma_X^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2$

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

korrigierte Stichprobenvarianz (auch eine Statistik):

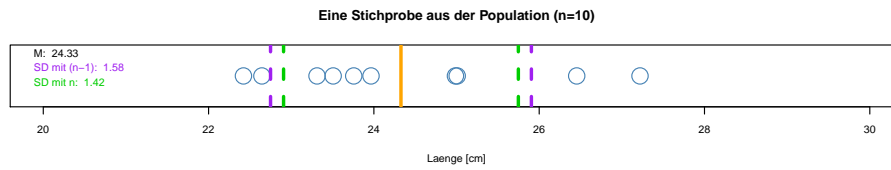
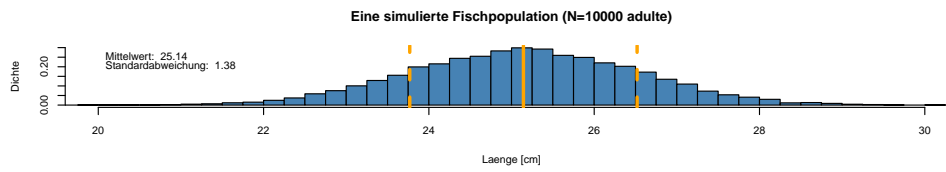
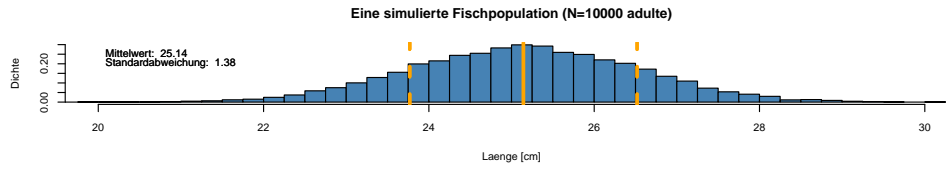
$$\begin{aligned} s^2 &= \frac{n}{n-1} \sigma_S^2 \\ &= \frac{n}{n-1} \cdot \frac{1}{n} \cdot \sum_{i=1}^n (S_i - \bar{S})^2 \\ &= \frac{1}{n-1} \cdot \sum_{i=1}^n (S_i - \bar{S})^2 \end{aligned}$$

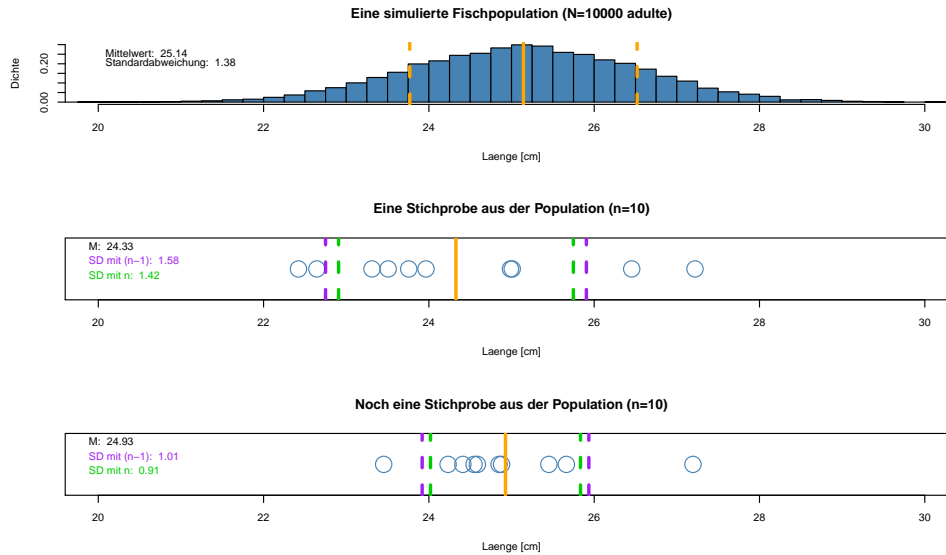
Mit "Standardabweichung von  $S$ " ist meistens das korrigierte  $s$  gemeint. Die beiden Statistiken  $\sigma_S^2$  und  $s^2$  sind Schätzer für den Parameter  $\sigma_X^2$ .

Beispiel Die Daten  $\bar{x} = ?$   $\bar{x} = 10/5 = 2$   
*Summe*

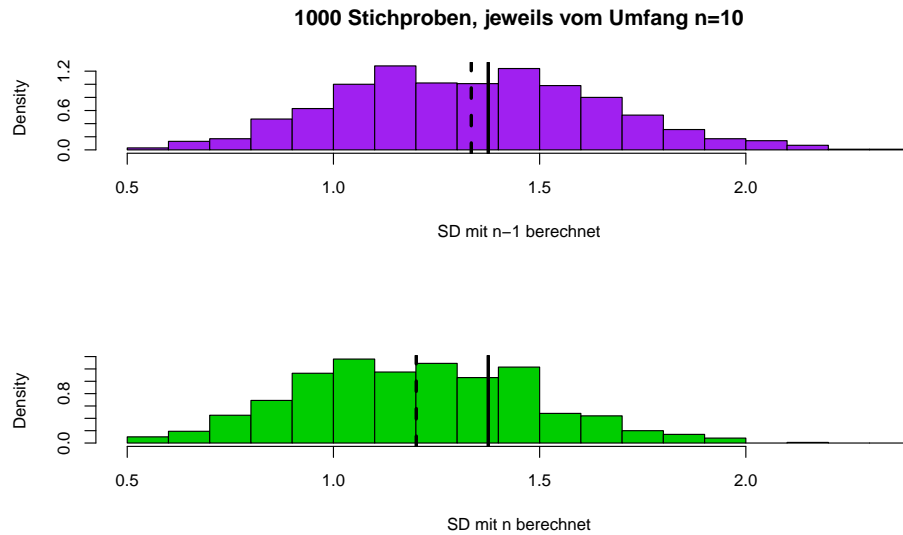
$x$	1	3	0	5	1	10
$x - \bar{x}$	-1	1	-2	3	-1	0
$(x - \bar{x})^2$	1	1	4	9	1	16

$$\begin{aligned} s^2 &= \text{Summe}((x - \bar{x})^2) / (n - 1) \\ &= 16 / (5 - 1) = 4 \\ s &= 2 \end{aligned}$$





Die folgenden Histogramme zeigen die Standardabweichungen, die aus 1000 verschiedenen Stichproben aus der selben Verteilung geschätzt wurden. Die durchgezogenen Linien stellen die tatsächliche Standardabweichung der Verteilung dar, die gestrichelten Linien die Mittelwerte der geschätzten Standardabweichungen.



### $\sigma$ mit $n$ oder $n - 1$ berechnen?

Die Standardabweichung  $\sigma$  eines Zufallsexperiments mit  $n$  gleichwahrscheinlichen Ausgängen  $x_1, \dots, x_n$  (z.B. Würfelwurf) ist ein **Parameter** und klar definiert durch

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

Wenn es sich bei  $x_1, \dots, x_n$  um eine Stichprobe handelt (wie meistens in der Statistik), wird für die

daraus berechnete Standardabweichung, also die **Statistik**, in der Regel

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

verwendet.

## 5 Vom Sinn und Unsinn von Mittelwerten

Mittelwert und Standardabweichung...

- charakterisieren die Daten gut, falls deren Verteilung glockenförmig ist
- und müssen andernfalls mit Vorsicht interpretiert werden.

Wir betrachten dazu einige Lehrbuch-Beispiele aus der Ökologie, siehe z.B.

### Literatur

[BTH08] M. Begon, C. R. Townsend, and J. L. Harper. *Ecology: From Individuals to Ecosystems*. Blackell Publishing, 4 edition, 2008.

Im Folgenden verwenden wir zum Teil simulierte Daten, wenn die Originaldaten nicht verfügbar waren. Glauben Sie uns also nicht alle Datenpunkte.

### 5.1 Beispiel: Wählerische Bachstelzen

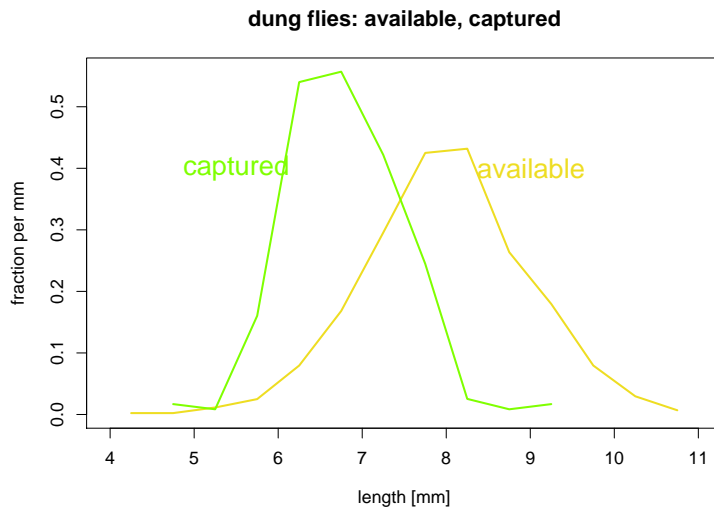
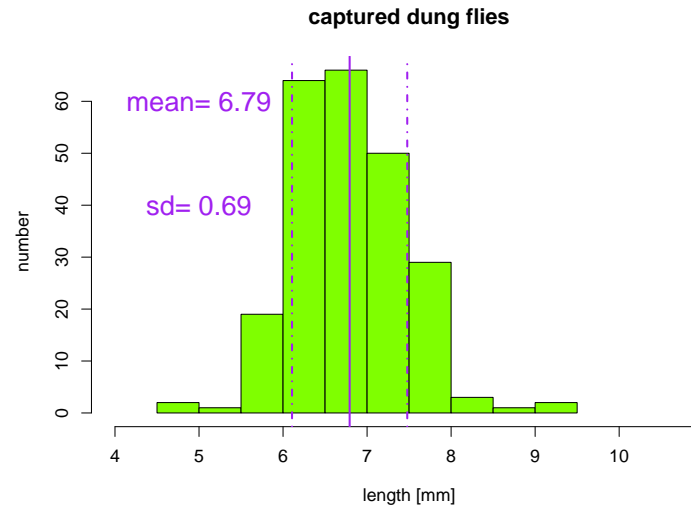
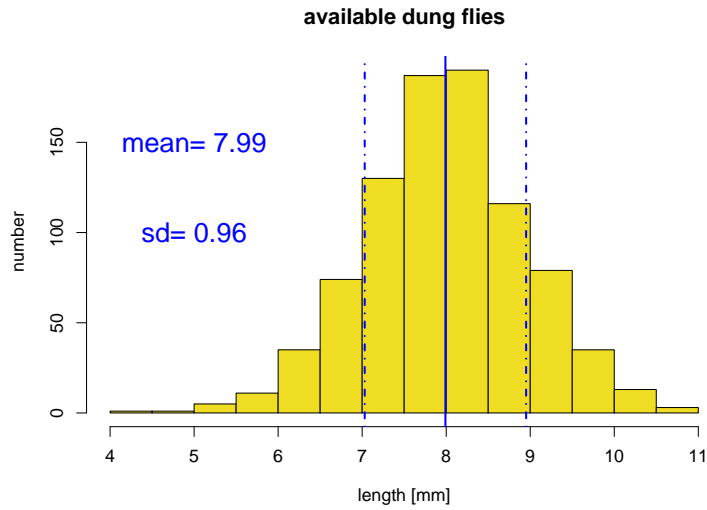
#### Bachstelzen fressen Dungfliegen

##### Vermutung

- Die Fliegen sind unterschiedlich groß
- Effizienz für die Bachstelze = Energiegewinn / Zeit zum Fangen und fressen
- Laborexperimente lassen vermuten, dass die Effizienz bei 7mm großen Fliegen maximal ist.

### Literatur

[Dav77] N.B. Davies. Prey selection and social behaviour in wagtails (Aves: Motacillidae). *J. Anim. Ecol.*, 46:37–57, 1977.



### Vergleich der Größenverteilungen

	captured		available
Mittelwert	6.29	<	7.99
Standardabweichung	0.69	<	0.96

### Interpretation

Die Bachstelzen bevorzugen Dungfliegen, die etwa 7mm groß sind.

Hier waren die Verteilungen glockenförmig und es genügten 4 Werte (die beiden Mittelwerte und die beiden Standardabweichungen), um die Daten adäquat zu beschreiben.

## 5.2 Beispiel: Spiderman & Spiderwoman

### Simulated Data:

Eine Stichprobe von 70 Spinnen

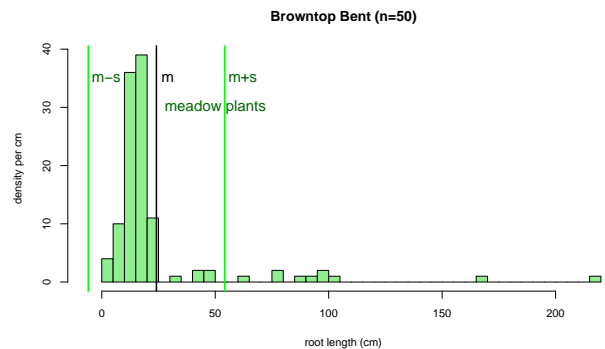
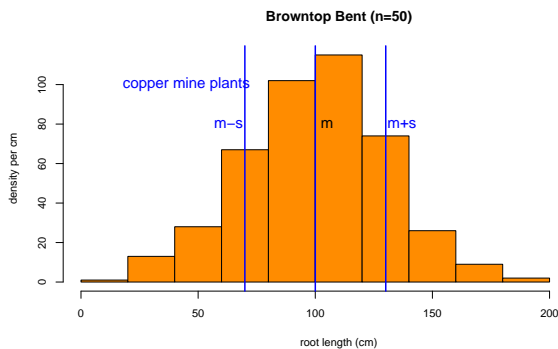
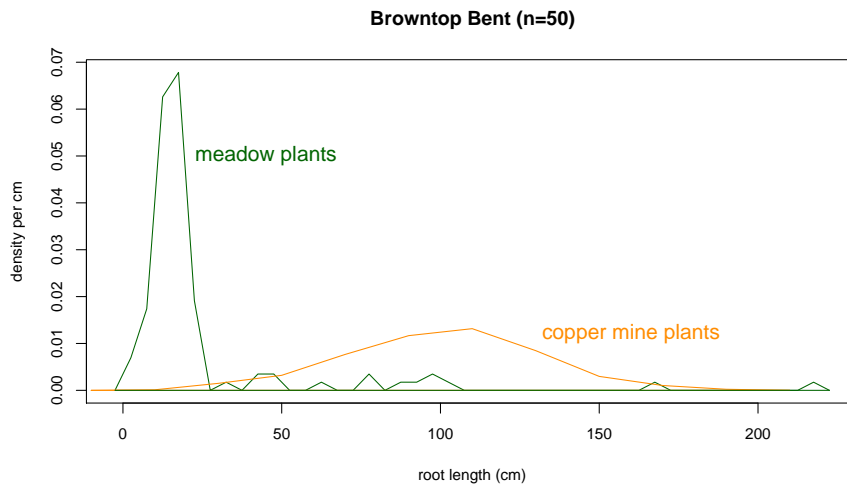


[MB68] T. McNeilly and A.D Bradshaw. Evolutionary Processes in Populations of Copper Tolerant *Agrostis tenuis* Sibth. *Evolution*, 22:108–118, 1968.

Wir verwenden hier wieder simulierte Daten, da die Originaldaten nicht zur Verfügung stehen.

### Anpassung an Kupfer?

- Pflanzen, denen das Kupfer schadet, haben kürzere Wurzeln.
- Die Wurzellängen von Pflanzen aus der Umgebung von Kupferminen wird gemessen.
- Samen von unbelasteten Wiesen werden bei Kupferminen eingesät.
- Die Wurzellängen dieser “Wiesepflanzen” werden gemessen.



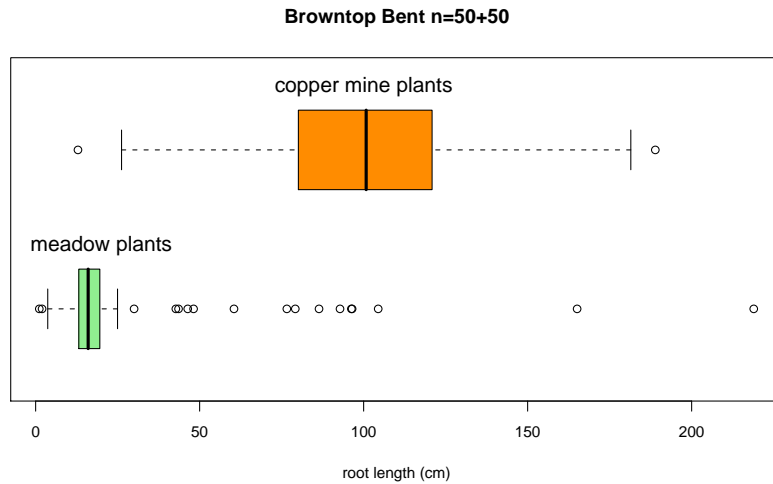
2/3 der Wurzellängen innerhalb  $[m-sd, m+sd]$ ???? **Nein!**

### Fazit des Straußgras-Beispiels

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





### Schlussfolgerung

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

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

Verlassen sie sich **niemals** allein auf numerische Kenngrößen!

### Copyright-Informationen zu einigen in der Präsentation verwendeten Bildern

[http://en.wikipedia.org/wiki/File:Motacilla\\_alba\\_alba.JPG](http://en.wikipedia.org/wiki/File:Motacilla_alba_alba.JPG) [Artur Mikołajewski] [http://commons.wikimedia.org/wiki/File:Scatophaga\\_stercoraria\\_1\\_Luc\\_Viatour.jpg](http://commons.wikimedia.org/wiki/File:Scatophaga_stercoraria_1_Luc_Viatour.jpg) [Viatour Luc] [http://commons.wikimedia.org/wiki/File:Nphila\\_inaurata\\_Madagascar\\_02.jpg](http://commons.wikimedia.org/wiki/File:Nphila_inaurata_Madagascar_02.jpg) [Bernard Gagnon] [http://de.wikipedia.org/w/index.php?title=Datei:Agrostis\\_capillaris.jpeg](http://de.wikipedia.org/w/index.php?title=Datei:Agrostis_capillaris.jpeg) [Kri-stian Peters] [http://de.wikipedia.org/w/index.php?title=Datei:Hendrick\\_met\\_de\\_Bles\\_001.jpg](http://de.wikipedia.org/w/index.php?title=Datei:Hendrick_met_de_Bles_001.jpg) [Hendrick met de Bles]

### Was Sie u.a. erklären können sollten

- Mittelwerte und Standardabweichung und wie man sie aus Histogrammen, Dichteplots und Strip-charts/Scatterplots ungefähr ablesen kann
- Wann man bei der Berechnung von Standardabweichungen und Varianzen durch  $n$  teilt und wann durch  $n - 1$  und wieso
- Konzept: Parameter, Statistik, Schätzer
- Wieso visualisieren?
- Wann können Mittelwerte etc. irreführend sein?

(Bitte beachten Sie auch die Punkte auf S. 18)