

# Wahrscheinlichkeitsrechnung und Statistik für Biologen

## **Einführung: Deskriptive Statistik**

Martin Hutzenthaler & Dirk Metzler

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

3. Mai 2011

- 1 Einführung
  - Konzept und Quellen
  - Plan
- 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik
- 3 Graphische Darstellungen
  - Histogramme und Dichtepolygone
  - Stripcharts
  - Boxplots
  - Beispiel: Ringeltaube
  - Beispiel: Darwin-Finken
- 4 Statistische Kenngrößen
  - Median und andere Quartile
  - Mittelwert und Standardabweichung
- 5 Vom Sinn und Unsinn von Mittelwerten
  - Beispiel: Wählerische Bachstelzen
  - Beispiel: Spiderman & Spiderwoman
  - Beispiel: Kupfertoleranz beim Roten Straußgras

- 1 Einführung
  - Konzept und Quellen
  - Plan
- 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik
- 3 Graphische Darstellungen
  - Histogramme und Dichtepolygone
  - Stripcharts
  - Boxplots
  - Beispiel: Ringeltaube
  - Beispiel: Darwin-Finken
- 4 Statistische Kenngrößen
  - Median und andere Quartile
  - Mittelwert und Standardabweichung
- 5 Vom Sinn und Unsinn von Mittelwerten
  - Beispiel: Wählerische Bachstelzen
  - Beispiel: Spiderman & Spiderwoman
  - Beispiel: Kupfertoleranz beim Roten Straußgras

# 1 Einführung

- **Konzept und Quellen**

- Plan

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

# 3 Graphische Darstellungen

- Histogramme und Dichtepolygone
- Stripcharts
- Boxplots
- Beispiel: Ringeltaube
- Beispiel: Darwin-Finken

# 4 Statistische Kenngrößen

- Median und andere Quartile
- Mittelwert und Standardabweichung

# 5 Vom Sinn und Unsinn von Mittelwerten

- Beispiel: Wählerische Bachstelzen
- Beispiel: Spiderman & Spiderwoman
- Beispiel: Kupfertoleranz beim Roten Straußgras

*It is easy to lie with statistics.*

Andrejs Dunkels

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

Andrejs Dunkels

# Was ist Statistik?

Die Natur ist voller Variabilität.

# Was ist Statistik?

Die Natur ist voller Variabilität.

Wie geht man mit variablen Daten um?



# Was ist Statistik?

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

durch

Zufall

modellieren.

# IDEE DER STATISTIK

Variabilität

(Erscheinung der Natur)

durch

Zufall

modellieren.

# IDEE DER STATISTIK

Variabilität

(Erscheinung der Natur)

durch

Zufall

(mathematische Abstraktion)

modellieren.

# Statistik

=

## Datenanalyse

mit Hilfe

## stochastischer Modelle

# Quellen

Wir danken Matthias Birkner für die intensive Zusammenarbeit beim Erstellen der ersten Version dieser Vorlesung sowie Brooks Ferebee, Gaby Schneider und Anton Wakolbinger für die Bereitstellung vieler Beispiele und Lehrmaterialien.

<http://joguinf.informatik.uni-mainz.de/~birkner/>

<http://www.math.uni-frankfurt.de/~wakolbin/statbio/>

<http://ismi.math.uni-frankfurt.de/schneider/statbio0708.html>

# 1 Einführung

- Konzept und Quellen
- **Plan**

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

## 3 Graphische Darstellungen

- Histogramme und Dichtepolygone
- Stripcharts
- Boxplots
- Beispiel: Ringeltaube
- Beispiel: Darwin-Finken

## 4 Statistische Kenngrößen

- Median und andere Quartile
- Mittelwert und Standardabweichung

## 5 Vom Sinn und Unsinn von Mittelwerten

- Beispiel: Wählerische Bachstelzen
- Beispiel: Spiderman & Spiderwoman
- Beispiel: Kupfertoleranz beim Roten Straußgras

# Plan der Vorlesung

## Klassische Statistik



# Plan der Vorlesung

## Klassische Statistik

- 1 Beschreibende Statistik

# Plan der Vorlesung

## Klassische Statistik

- 1 Beschreibende Statistik
- 2 Der Standardfehler

# Plan der Vorlesung

## Klassische Statistik

- 1 Beschreibende Statistik
- 2 Der Standardfehler
- 3 Der t-Test für gepaarte Stichproben

# Plan der Vorlesung

## Klassische Statistik

- 1 Beschreibende Statistik
- 2 Der Standardfehler
- 3 Der t-Test für gepaarte Stichproben
- 4 Der t-Test für unabhängige Stichproben

# Plan der Vorlesung

## Klassische Statistik

- 1 Beschreibende Statistik
- 2 Der Standardfehler
- 3 Der t-Test für gepaarte Stichproben
- 4 Der t-Test für unabhängige Stichproben
- 5 Häufigkeiten

# Plan der Vorlesung

## Klassische Statistik

- 1 Beschreibende Statistik
- 2 Der Standardfehler
- 3 Der t-Test für gepaarte Stichproben
- 4 Der t-Test für unabhängige Stichproben
- 5 Häufigkeiten
- 6 Der Chi-Quadrat Test

# Plan der Vorlesung

## Klassische Statistik

- 1 Beschreibende Statistik
- 2 Der Standardfehler
- 3 Der t-Test für gepaarte Stichproben
- 4 Der t-Test für unabhängige Stichproben
- 5 Häufigkeiten
- 6 Der Chi-Quadrat Test
- 7 Lineare Regression

# Plan der Vorlesung

## Klassische Statistik

- 1 Beschreibende Statistik
- 2 Der Standardfehler
- 3 Der t-Test für gepaarte Stichproben
- 4 Der t-Test für unabhängige Stichproben
- 5 Häufigkeiten
- 6 Der Chi-Quadrat Test
- 7 Lineare Regression
- 8 Korrelation



# Plan der Vorlesung

## Klassische Statistik

- 1 Beschreibende Statistik
- 2 Der Standardfehler
- 3 Der t-Test für gepaarte Stichproben
- 4 Der t-Test für unabhängige Stichproben
- 5 Häufigkeiten
- 6 Der Chi-Quadrat Test
- 7 Lineare Regression
- 8 Korrelation
- 9 Varianzanalyse (ANOVA)

# Plan der Vorlesung

## Weitere Themen

- Nichtparametrische Tests

# Plan der Vorlesung

## Weitere Themen

- Nichtparametrische Tests
- Diskriminanzanalyse

# Plan der Vorlesung

## Weitere Themen

- Nichtparametrische Tests
- Diskriminanzanalyse
- Grundbegriffe der  
Wahrscheinlichkeitstheorie

# Plan der Vorlesung

## Weitere Themen

- Nichtparametrische Tests
- Diskriminanzanalyse
- Grundbegriffe der  
Wahrscheinlichkeitstheorie
- Parameterschätzung

# Plan der Vorlesung

## Weitere Themen

- Nichtparametrische Tests
- Diskriminanzanalyse
- Grundbegriffe der  
Wahrscheinlichkeitstheorie
- Parameterschätzung
- Moderne Anwendung: Analyse von  
Genexpressionsdaten (vielleicht)

# Plan der Vorlesung

## Weitere Themen

- Nichtparametrische Tests
- Diskriminanzanalyse
- Grundbegriffe der Wahrscheinlichkeitstheorie
- Parameterschätzung
- Moderne Anwendung: Analyse von Genexpressionsdaten (vielleicht)
- R

# Statistik-Software R

The screenshot shows the R Project for Statistical Computing website in a Mozilla Firefox browser window. The browser title is "The R Project for Statistical Computing - Mozilla Firefox". The address bar shows the URL "http://www.r-project.org/". The page content includes the R logo, navigation links, and several data visualization plots.

**Navigation Links:**

- About R:** [What is R?](#), [Contributors](#), [Screenshots](#), [What's new?](#)
- Download, Packages:** [CRAN](#)
- R Project Foundation:** [Members & Donors](#), [Mailing Lists](#), [Bug Tracking](#), [Developer Page](#), [Conferences](#), [Search](#)
- Documentation:** [Manuals](#), [FAQs](#), [Newsletter](#), [Wiki](#), [Books](#), [Certification](#), [Other](#)

**Visualizations:**

- PCA 5 vars:** A biplot showing variables (Fertility, Catholic, Agriculture, Emigration, Education) on a grid. A label "(1-3) 60%" is present.
- Clustering 4 groups:** A dendrogram and a bar chart showing the distribution of data points into four groups.
- Factor 1 [41%]:** A histogram and a normal distribution curve for the first principal component.
- Factor 3 [19%]:** A histogram and a normal distribution curve for the third principal component.

**Getting Started:**

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

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



# Folien, R-Befehle, Quellen und Übungen

<http://evol.bio.lmu.de/statgen/StatBiol/10SS>

- 1 Einführung
  - Konzept und Quellen
  - Plan

- 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik**

- 3 Graphische Darstellungen
  - Histogramme und Dichtepolygone
  - Stripcharts
  - Boxplots
  - Beispiel: Ringeltaube
  - Beispiel: Darwin-Finken

- 4 Statistische Kenngrößen
  - Median und andere Quartile
  - Mittelwert und Standardabweichung

- 5 Vom Sinn und Unsinn von Mittelwerten
  - Beispiel: Wählerische Bachstelzen
  - Beispiel: Spiderman & Spiderwoman
  - Beispiel: Kupfertoleranz beim Roten Straußgras

# Beschreibende Statistik



# Beschreibende Statistik



Beschreibende Statistik:  
Ein erster Blick auf die Daten

- 1 Einführung
  - Konzept und Quellen
  - Plan

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

- 3 Graphische Darstellungen**

- Histogramme und Dichtepolygone
- Stripcharts
- Boxplots
- Beispiel: Ringeltaube
- Beispiel: Darwin-Finken

- 4 Statistische Kenngrößen

- Median und andere Quartile
- Mittelwert und Standardabweichung

- 5 Vom Sinn und Unsinn von Mittelwerten

- Beispiel: Wählerische Bachstelzen
- Beispiel: Spiderman & Spiderwoman
- Beispiel: Kupfertoleranz beim Roten Straußgras

# Beispiel

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

# 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*



# Der Springkrebs

## *Galathea intermedia*

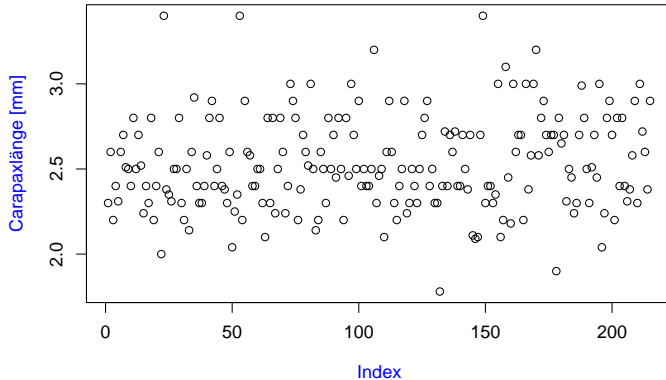


# Helgoländer Tiefe Rinne, Fang vom 6.9.1988

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**

- 1 Einführung
  - Konzept und Quellen
  - Plan

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

- 3 Graphische Darstellungen**

- **Histogramme und Dichtepolygone**
- Stripcharts
- Boxplots
- Beispiel: Ringeltaube
- Beispiel: Darwin-Finken

- 4 Statistische Kenngrößen

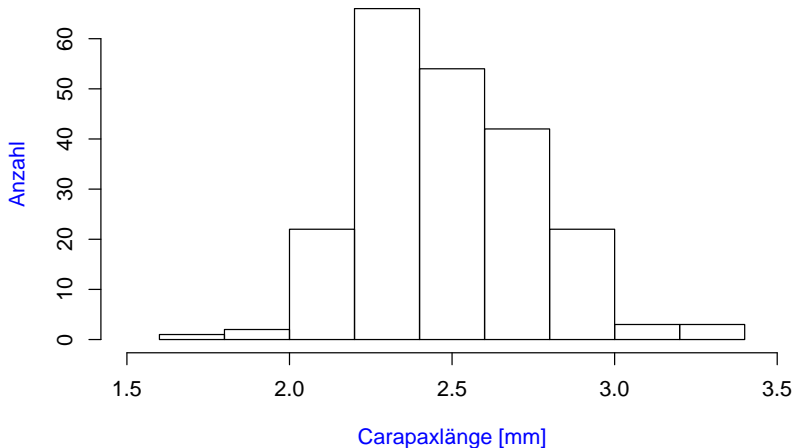
- Median und andere Quartile
- Mittelwert und Standardabweichung

- 5 Vom Sinn und Unsinn von Mittelwerten

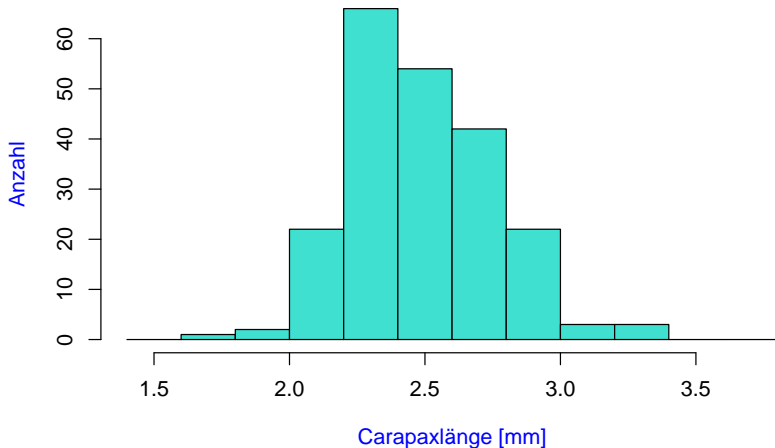
- Beispiel: Wählerische Bachstelzen
- Beispiel: Spiderman & Spiderwoman
- Beispiel: Kupfertoleranz beim Roten Straußgras

Eine Möglichkeit der graphischen  
Darstellung:  
das Histogramm

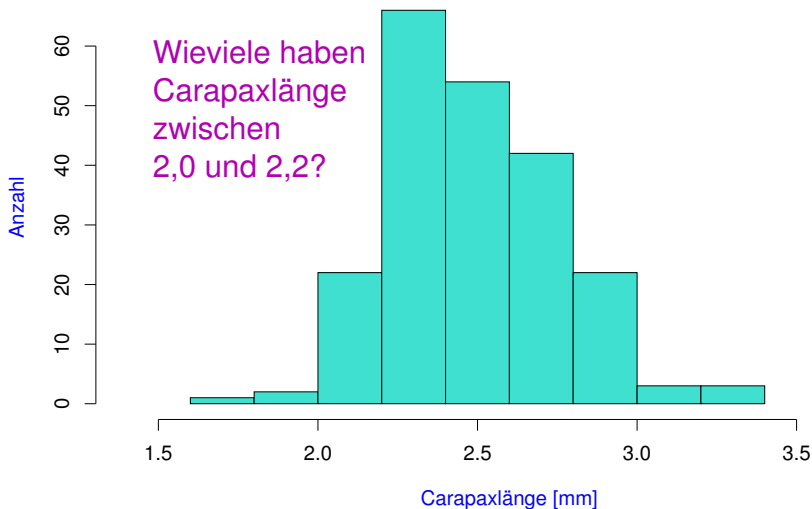
## Nichteiertragende Weibchen am 6. Sept. '88, n=215



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

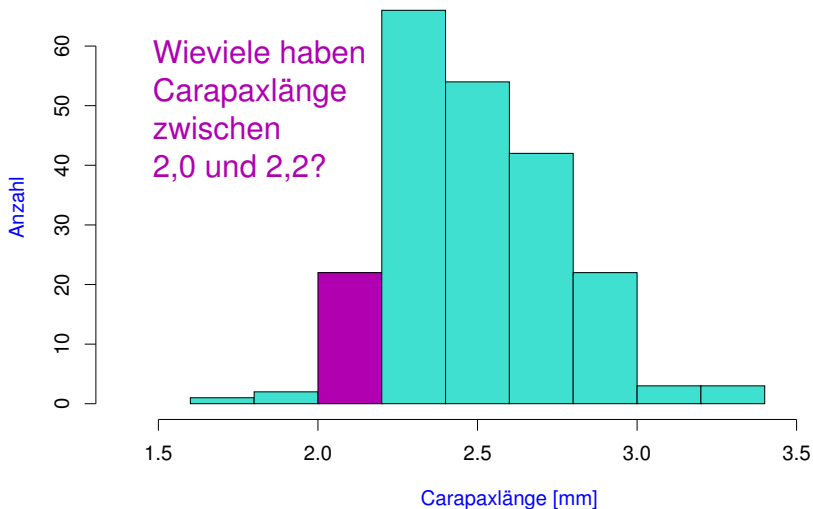


## Nichteiertragende Weibchen am 6. Sept. '88, n=215

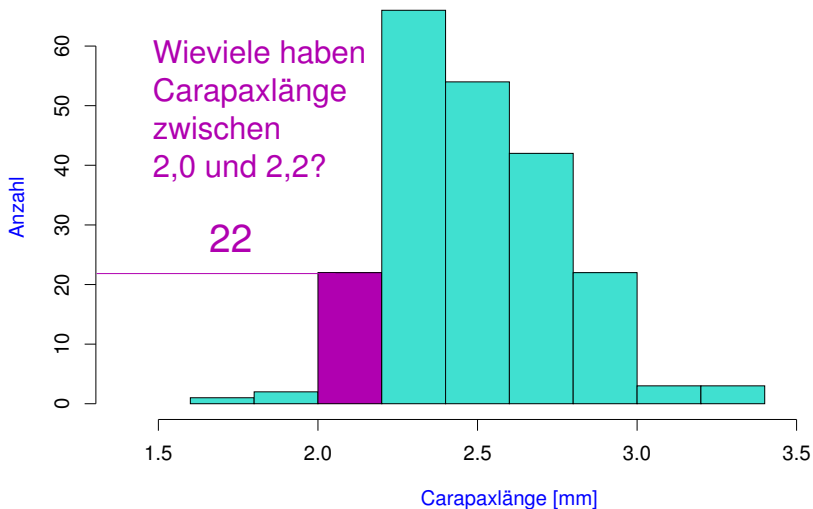




## Nichteiertragende Weibchen am 6. Sept. '88, n=215

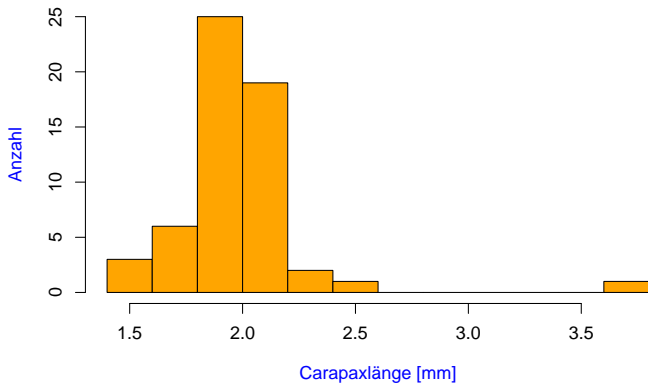


## Nichteiertragende Weibchen am 6. Sept. '88, n=215



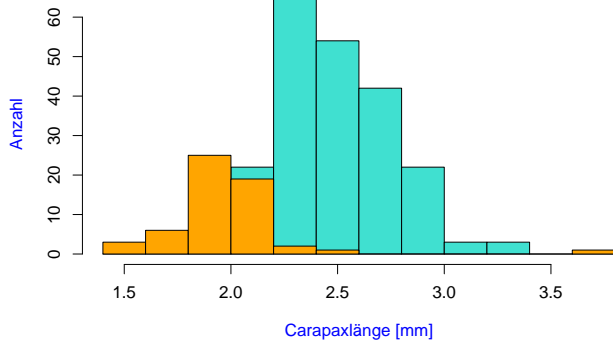
# Analoge Daten zwei Monate später (3.11.88):

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



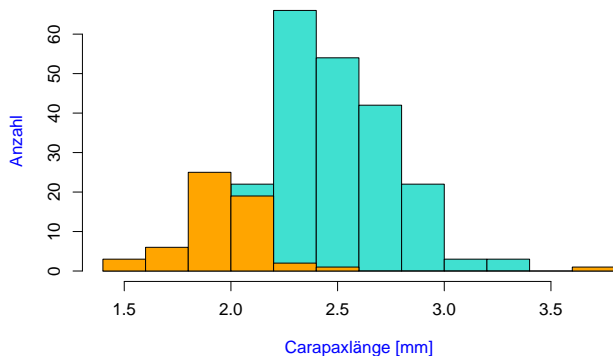
# Vergleich der beiden Verteilungen

## Nichteiertragende Weibchen



# Vergleich der beiden Verteilungen

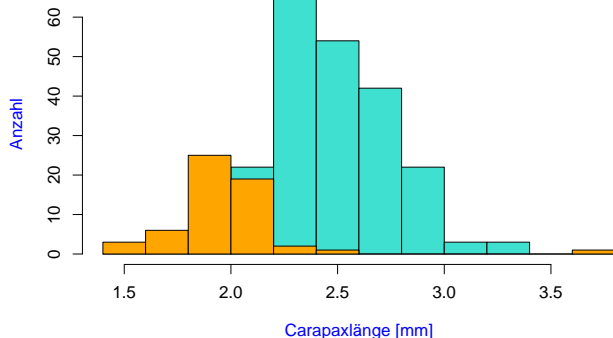
## Nichteiertragende Weibchen



Problem: ungleiche Stichprobenumfänge: 6.Sept:  $n = 215$   
3.Nov:  $n = 57$

# Vergleich der beiden Verteilungen

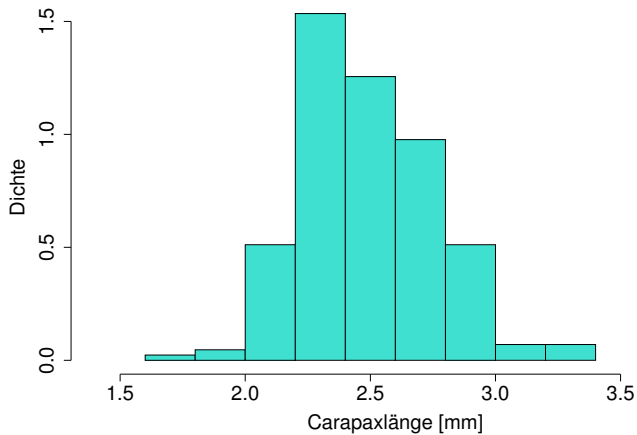
## Nichteiertragende Weibchen



Problem: ungleiche Stichprobenumfänge: 6.Sept:  $n = 215$   
 3.Nov:  $n = 57$

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

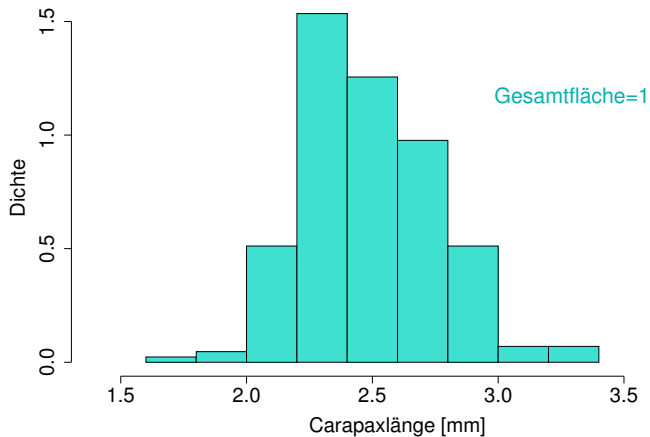
Nichteiertragende Weibchen am 6. Sept. '88, n=215



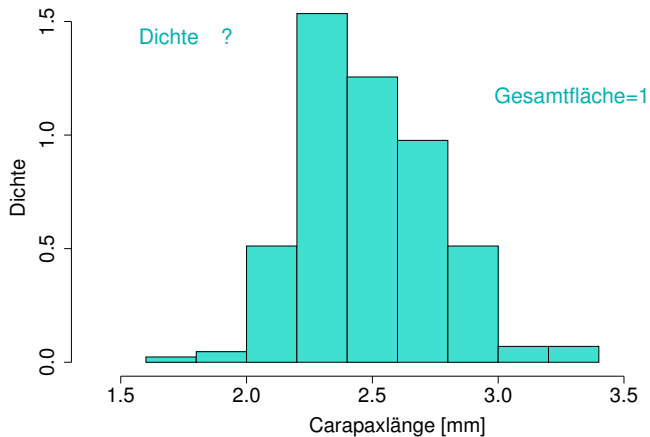


Die neue  
vertikale Koordinate  
ist jetzt eine  
**Dichte**  
(engl. **density**).

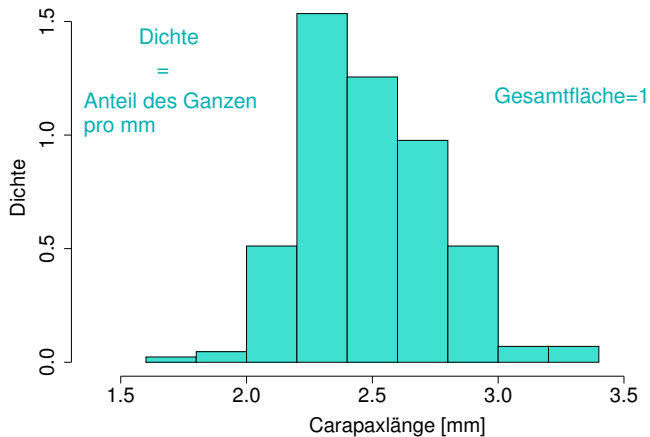
Nichteiertragende Weibchen am 6. Sept. '88, n=215



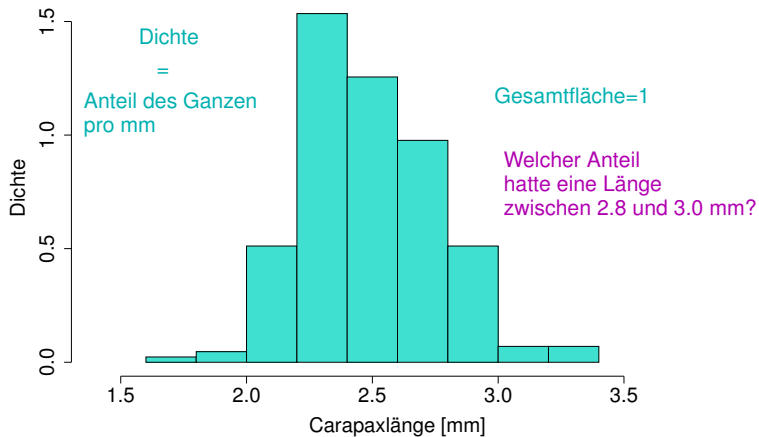
Nichteiertragende Weibchen am 6. Sept. '88, n=215



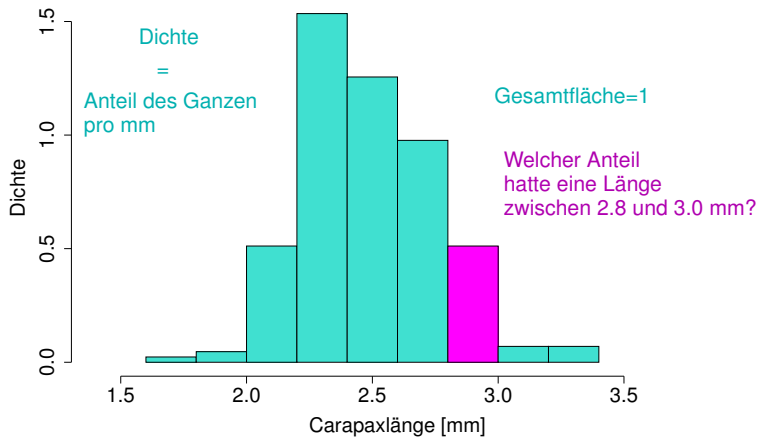
Nichteiertragende Weibchen am 6. Sept. '88, n=215



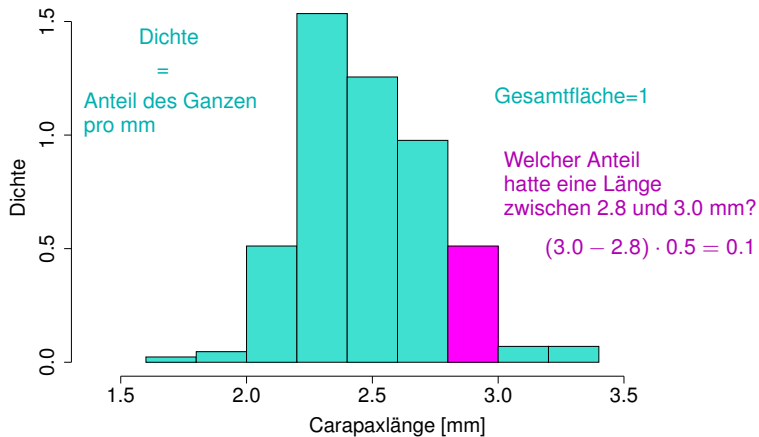
Nichteiertragende Weibchen am 6. Sept. '88, n=215



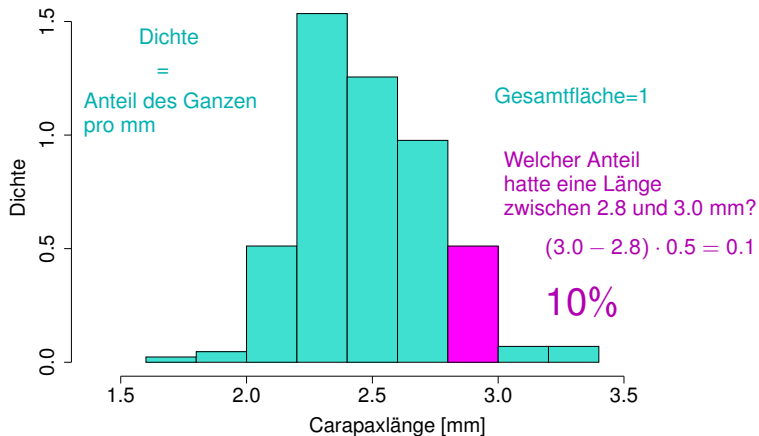
Nichteiertragende Weibchen am 6. Sept. '88, n=215



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



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

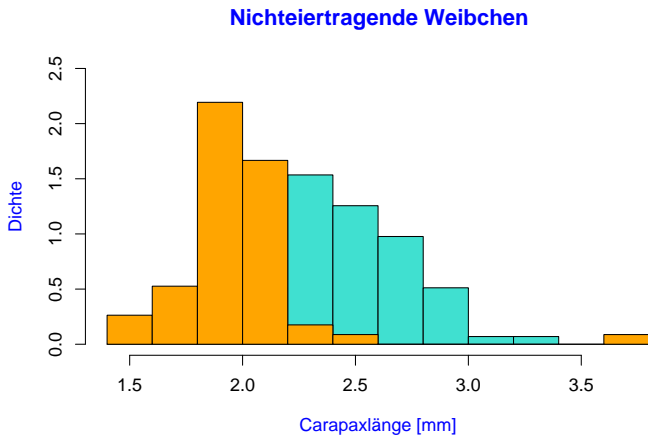




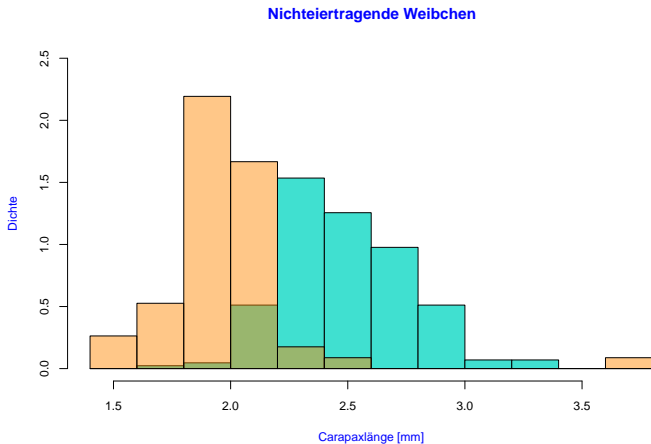
Die beiden Histogramme sind jetzt  
vergleichbar

Die beiden Histogramme sind jetzt  
vergleichbar  
(sie haben dieselbe Gesamtfläche).

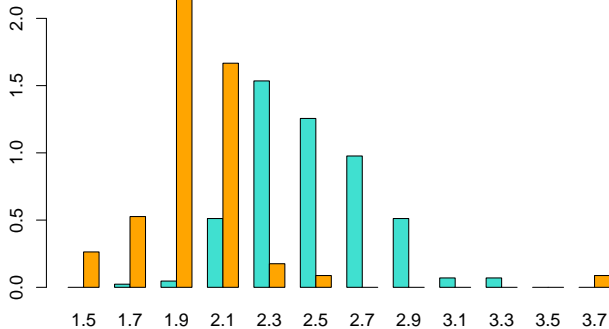
# Versuche, die Histogramme zusammen zu zeigen:



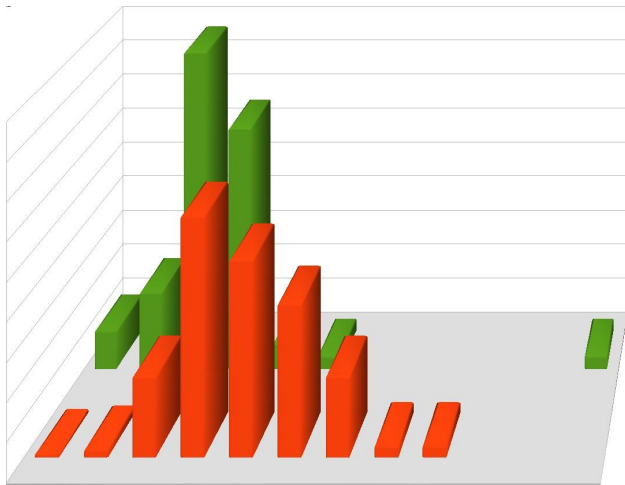
# Versuche, die Histogramme zusammen zu zeigen:



# Versuche, die Histogramme zusammen zu zeigen:



# Versuche, die Histogramme zusammen zu zeigen:



# Unser Rat an Sie:

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

# Unser Rat an Sie:

Wenn Sie Schauwerbegestalter(in) sind:

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



# Unser Rat an Sie:

Wenn Sie Schauwerbegestalter(in) sind:

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

Wenn Sie Wissenschaftler(in) werden wollen:

# Unser Rat an Sie:

Wenn Sie Schauwerbegestalter(in) sind:

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

Wenn Sie Wissenschaftler(in) werden wollen:

Bevorzugen Sie  
einfache und klare 2D-Darstellungen.

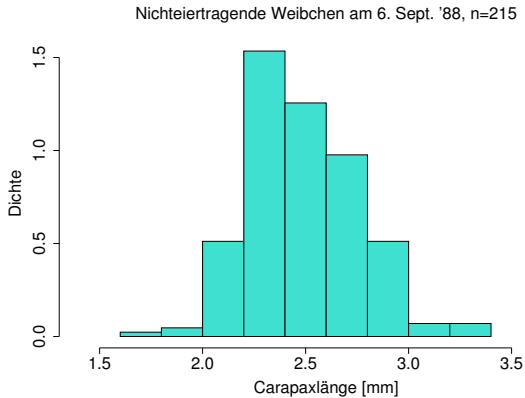
# Problem

Histogramme kann man nicht ohne weiteres  
in demselben Graphen  
darstellen,

# Problem

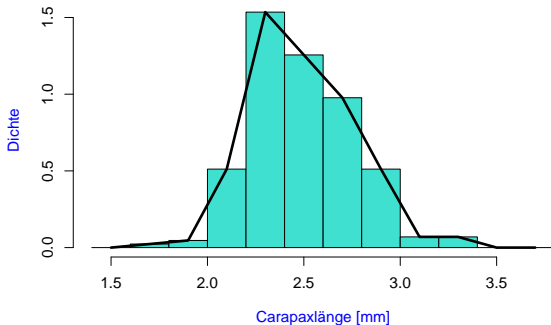
Histogramme kann man nicht ohne weiteres  
in demselben Graphen  
darstellen,  
weil sie einander  
überdecken würden.

# Einfache und klare Lösung: Dichtepolygone



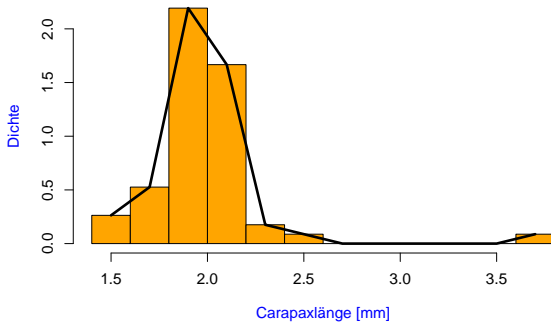
# Einfache und klare Lösung: Dichtepolygone

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



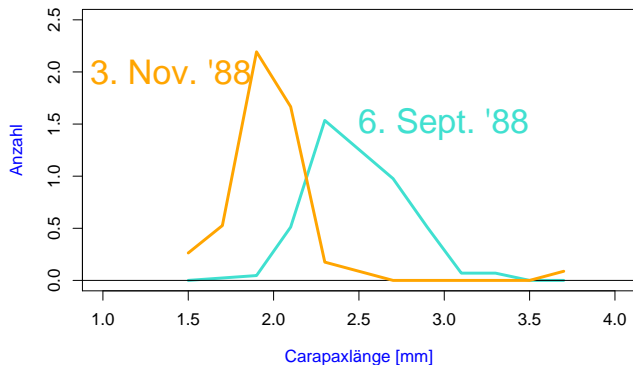
# Einfache und klare Lösung: Dichtepolygone

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



# Zwei und mehr Dichtepolygone in einem Plot

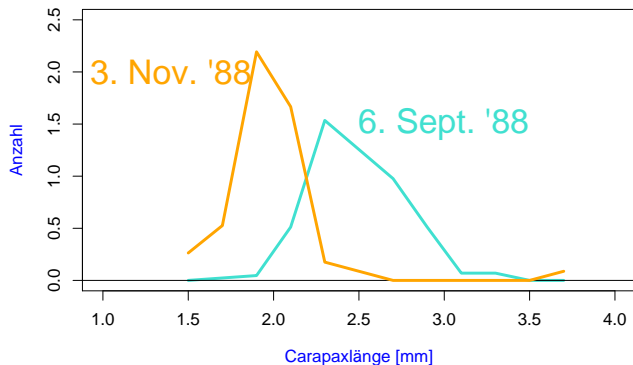
## Nichteiertragende Weibchen



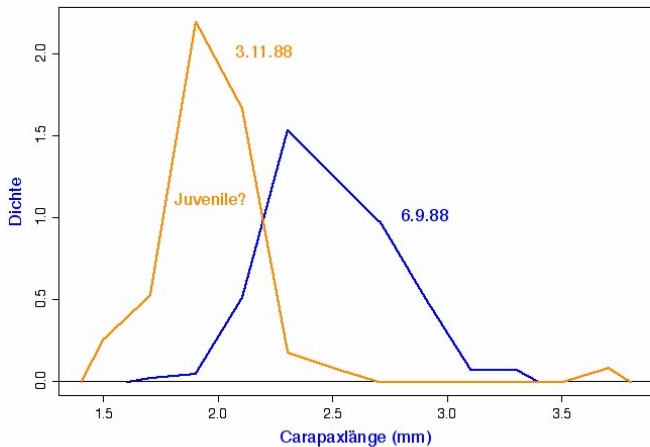


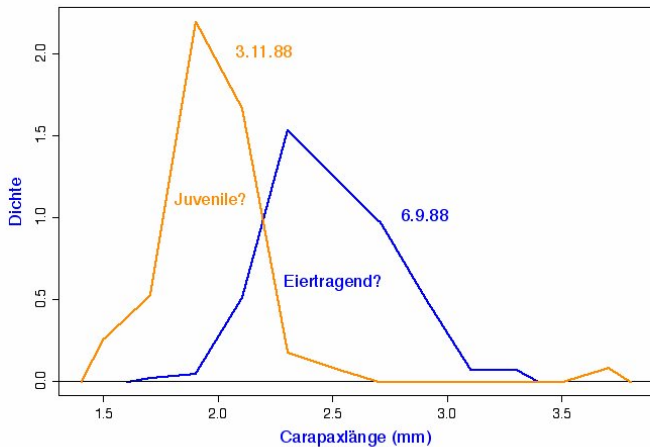
# Zwei und mehr Dichtepolygone in einem Plot

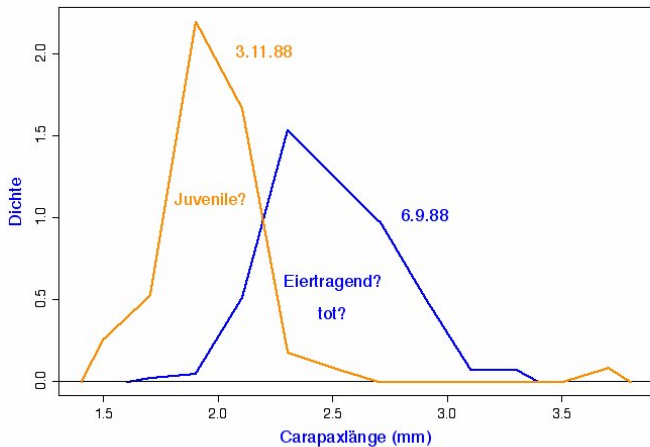
## Nichteiertragende Weibchen



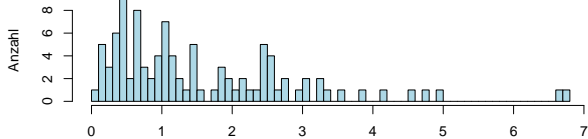
Biologische Interpretation der Verschiebung?

*Nichteiertragende Weibchen 6.9.88 und 3.11.88*

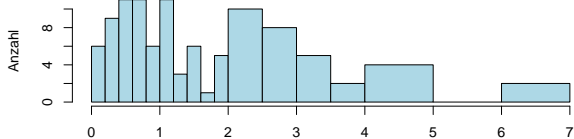
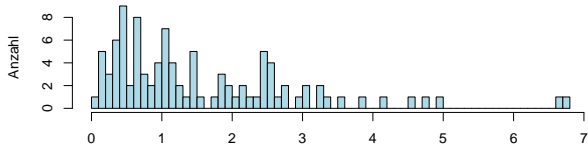
*Nichteiertragende Weibchen 6.9.88 und 3.11.88*

*Nichteiertragende Weibchen 6.9.88 und 3.11.88*

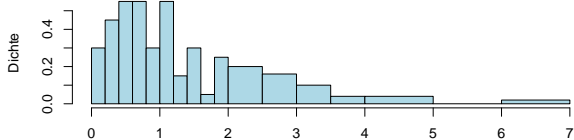
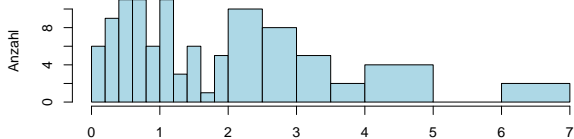
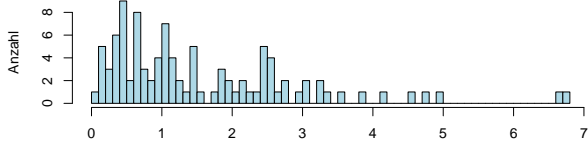
# Anzahl vs. Dichte



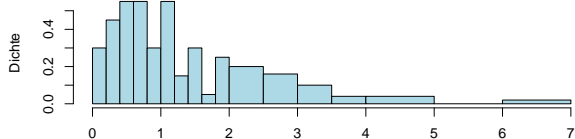
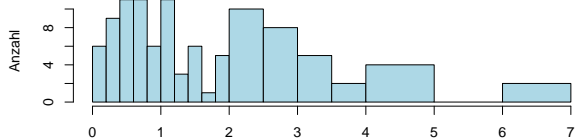
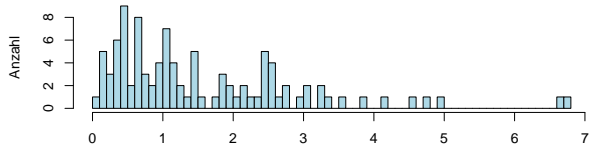
# Anzahl vs. Dichte



# Anzahl vs. Dichte



# Anzahl vs. Dichte



Also: Bei Histogrammen mit ungleichmäßiger Unterteilung immer Dichten verwenden!



- 1 Einführung
  - Konzept und Quellen
  - Plan

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

- 3 **Graphische Darstellungen**

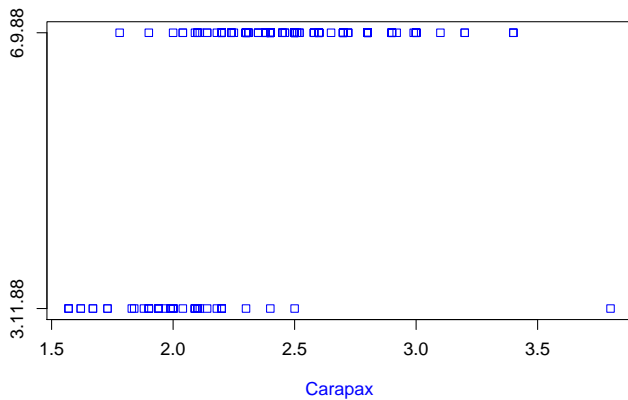
- Histogramme und Dichtepolygone
- **Stripcharts**
- Boxplots
- Beispiel: Ringeltaube
- Beispiel: Darwin-Finken

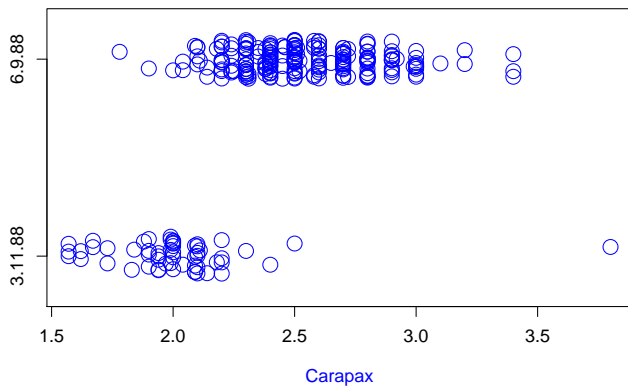
- 4 Statistische Kenngrößen

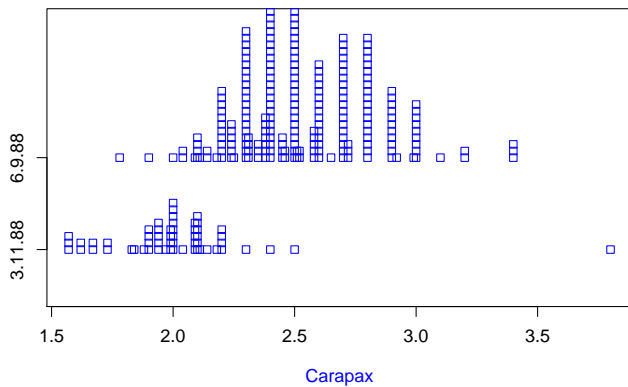
- Median und andere Quartile
- Mittelwert und Standardabweichung

- 5 Vom Sinn und Unsinn von Mittelwerten

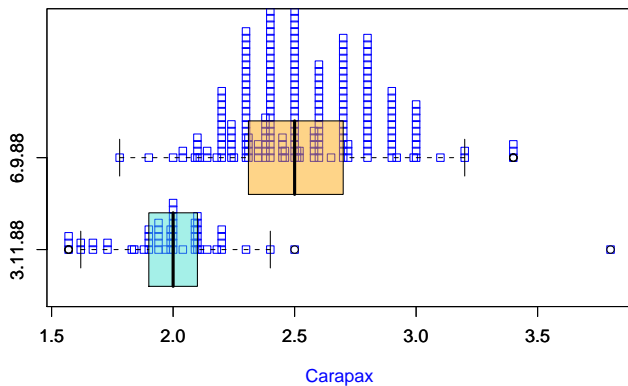
- Beispiel: Wählerische Bachstelzen
- Beispiel: Spiderman & Spiderwoman
- Beispiel: Kupfertoleranz beim Roten Straußgras



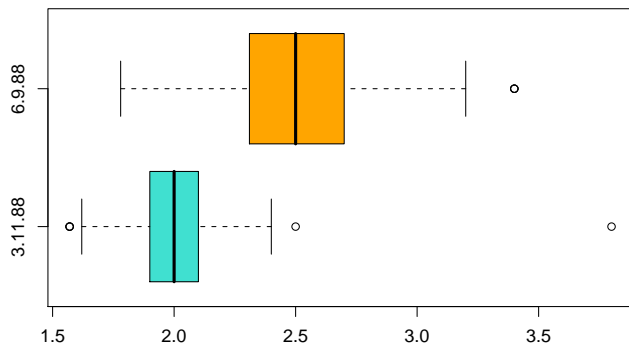




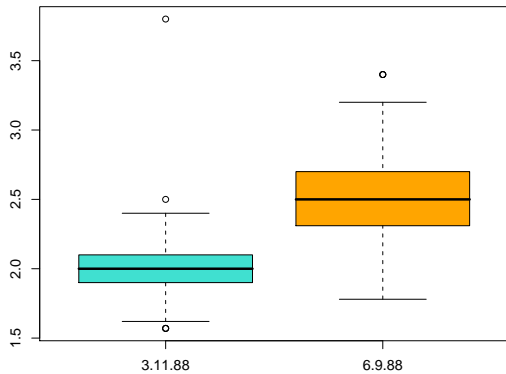
## Stripchart + Boxplots, horizontal



## Boxplots, horizontal



## Boxplots, vertikal



Histogramme und Dichtepolygone  
geben  
ein ausführliches Bild  
eines Datensatzes.



Histogramme und Dichtepolygone  
geben  
ein ausführliches Bild  
eines Datensatzes.

Manchmal zu ausführlich.

- 1 Einführung
  - Konzept und Quellen
  - Plan
- 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik
- 3 **Graphische Darstellungen**
  - Histogramme und Dichtepolygone
  - Stripcharts
  - **Boxplots**
  - Beispiel: Ringeltaube
  - Beispiel: Darwin-Finken
- 4 Statistische Kenngrößen
  - Median und andere Quartile
  - Mittelwert und Standardabweichung
- 5 Vom Sinn und Unsinn von Mittelwerten
  - Beispiel: Wählerische Bachstelzen
  - Beispiel: Spiderman & Spiderwoman
  - Beispiel: Kupfertoleranz beim Roten Straußgras

# Zu viel Information erschwert den Überblick



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

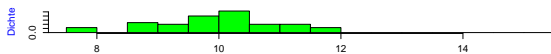
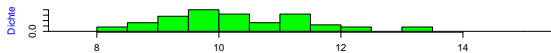
# Zu viel Information erschwert den Überblick

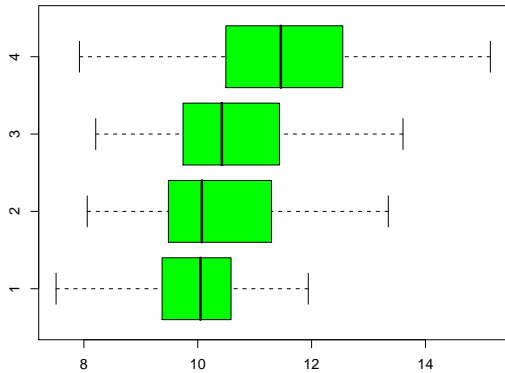


Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum

Wald?

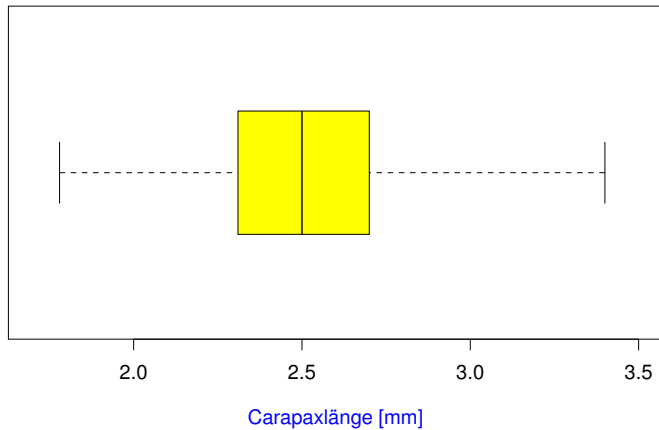
# Beispiel: Vergleich von mehreren Gruppen





# Der Boxplot

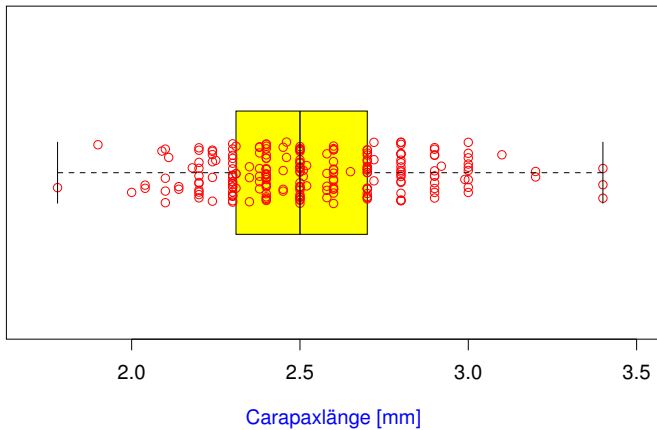
## Boxplot, einfache Ausführung





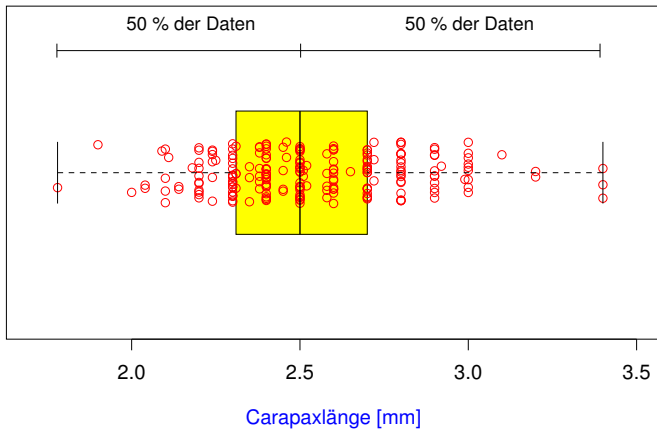
# Der Boxplot

## Boxplot, einfache Ausführung



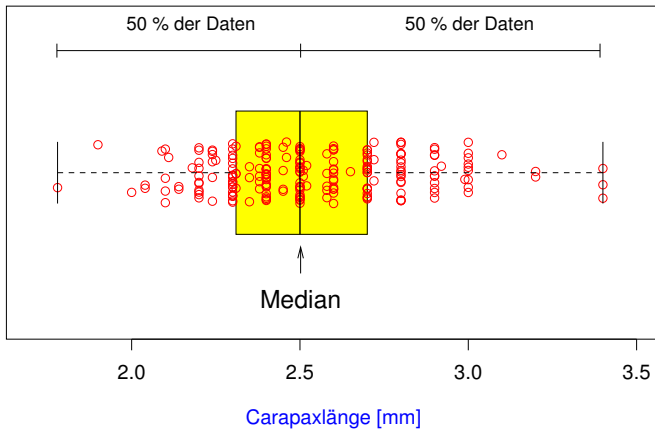
# Der Boxplot

## Boxplot, einfache Ausführung



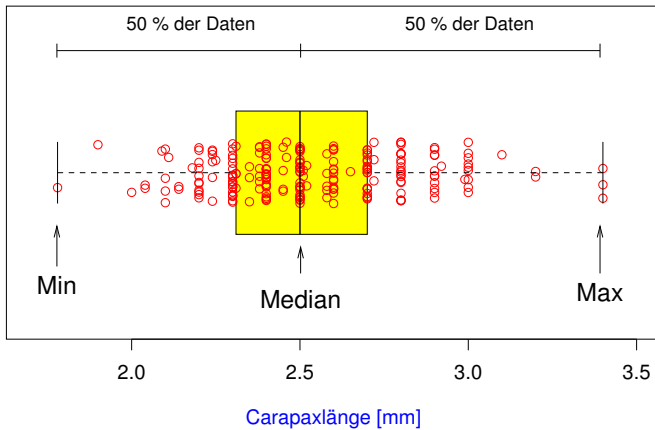
# Der Boxplot

## Boxplot, einfache Ausführung



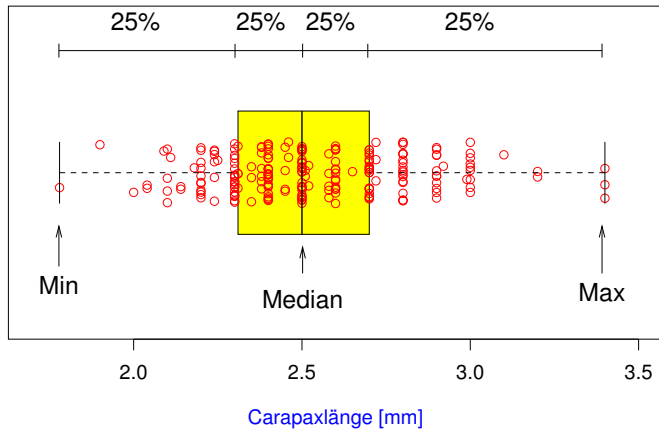
# Der Boxplot

## Boxplot, einfache Ausführung



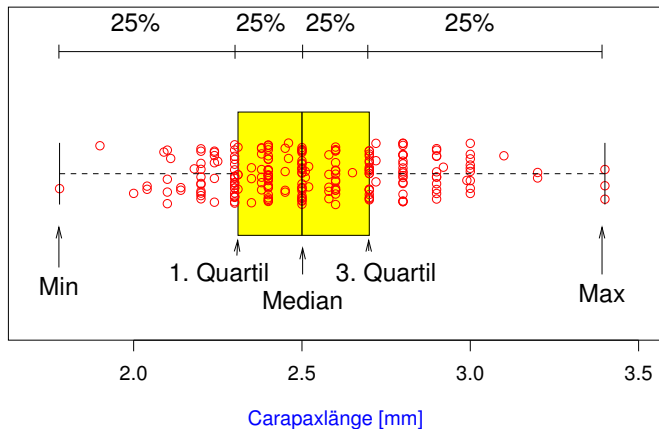
# Der Boxplot

## Boxplot, einfache Ausführung



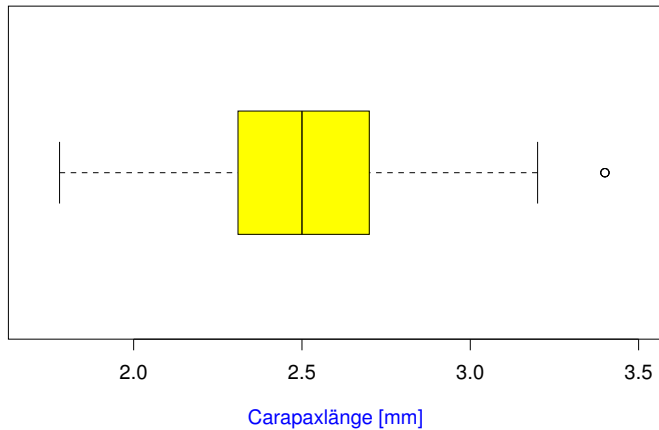
# Der Boxplot

## Boxplot, einfache Ausführung



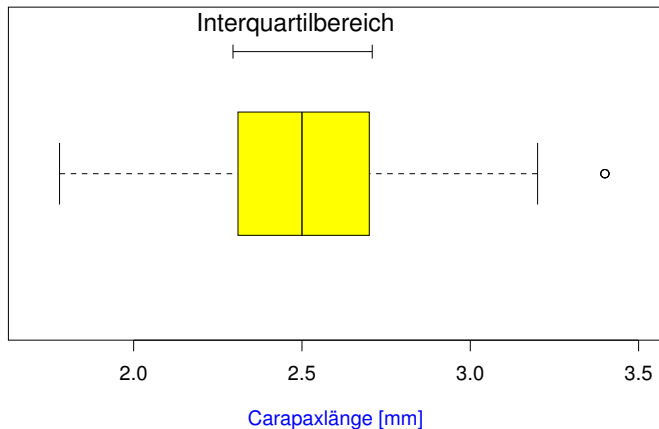
# Der Boxplot

## Boxplot, Standardausführung



# Der Boxplot

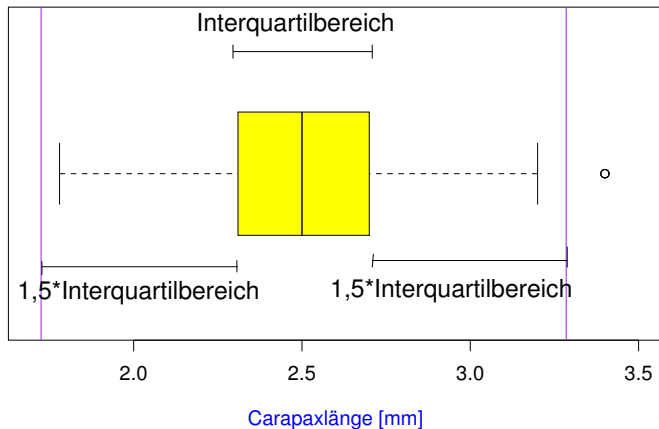
## Boxplot, Standardausführung





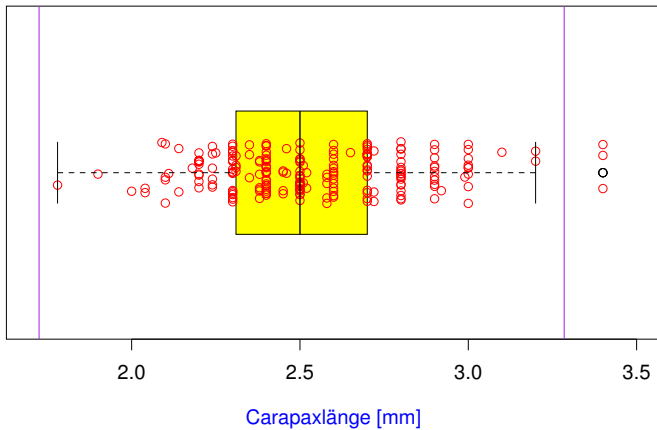
# Der Boxplot

## Boxplot, Standardausführung



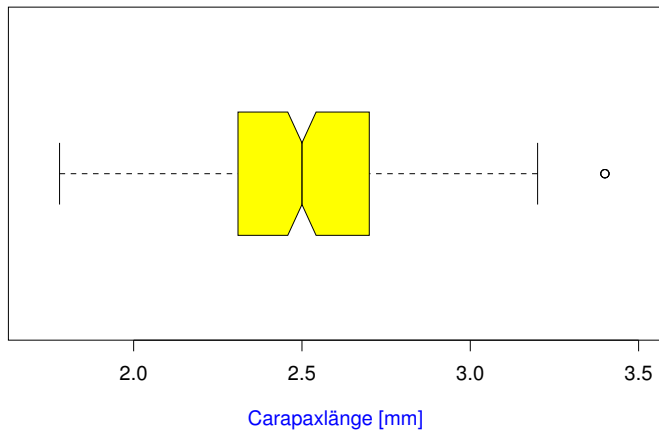
# Der Boxplot

## Boxplot, Standardausführung



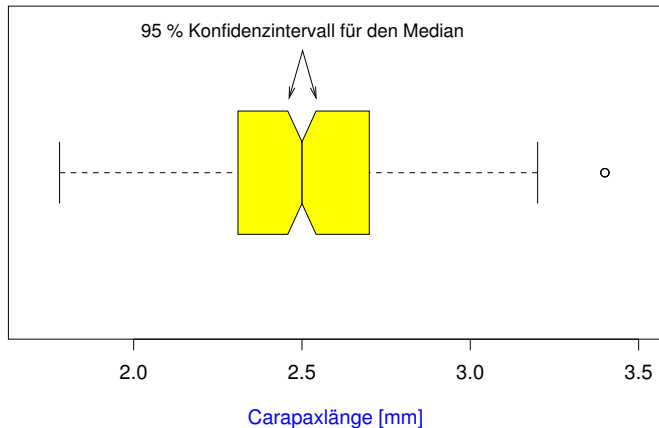
# Der Boxplot

## Boxplot, Profiausstattung



# Der Boxplot

## Boxplot, Profiausstattung



- 1 Einführung
  - Konzept und Quellen
  - Plan
- 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik
- 3 **Graphische Darstellungen**
  - Histogramme und Dichtepolygone
  - Stripcharts
  - Boxplots
  - **Beispiel: Ringeltaube**
  - Beispiel: Darwin-Finken
- 4 Statistische Kenngrößen
  - Median und andere Quartile
  - Mittelwert und Standardabweichung
- 5 Vom Sinn und Unsinn von Mittelwerten
  - Beispiel: Wählerische Bachstelzen
  - Beispiel: Spiderman & Spiderwoman
  - Beispiel: Kupfertoleranz beim Roten Straußgras

Beispiel:

Die Ringeltaube

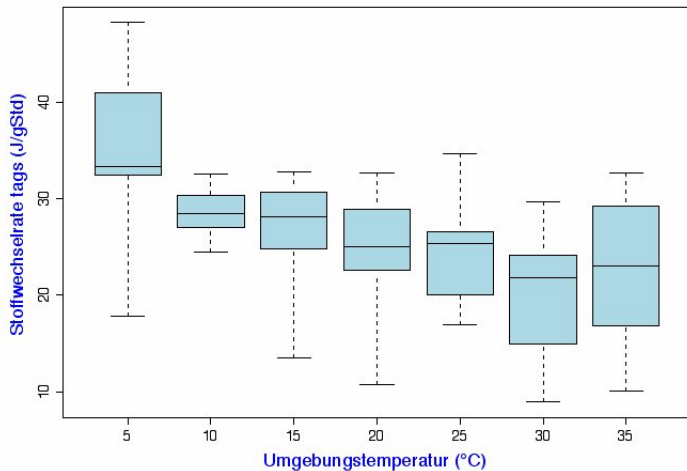
*Palumbus palumbus*



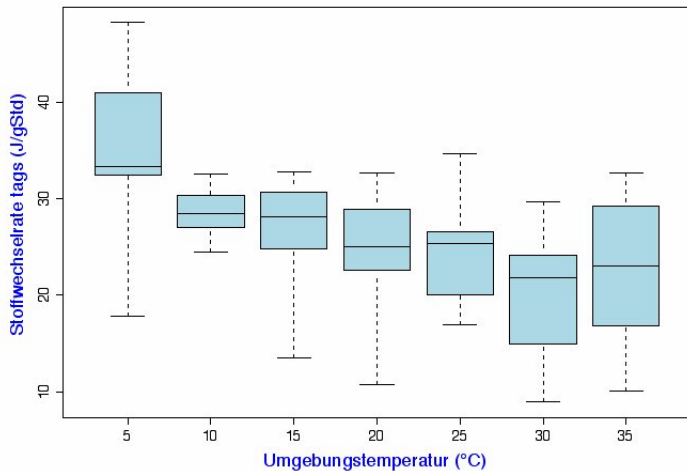
Wie hängt die Stoffwechselrate bei der Ringeltaube von der Umgebungstemperatur ab?



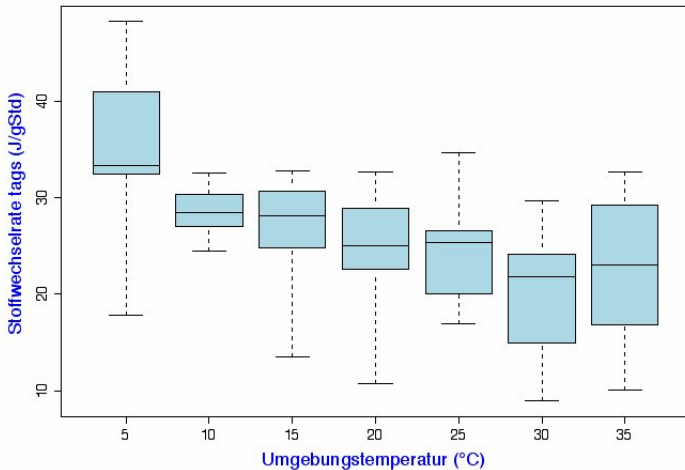
Daten  
aus dem  
AK Stoffwechselphysiologie  
Prof. Prinzinger  
Universität Frankfurt

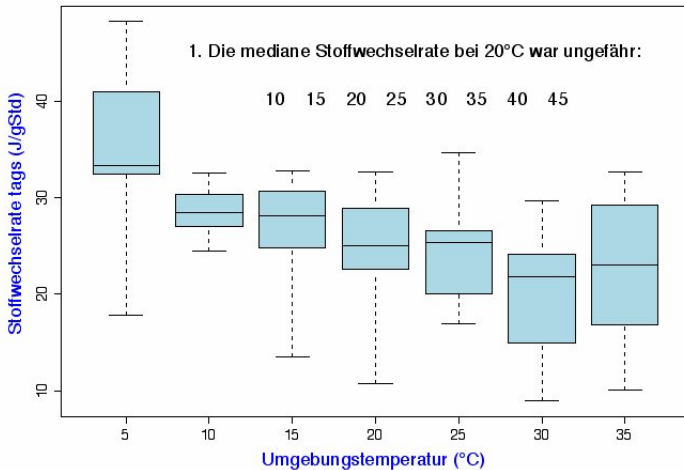
*Stoffwechselrate tags (J/gStd) und Umgebungstemperatur bei Ringeltauben (n=90)*

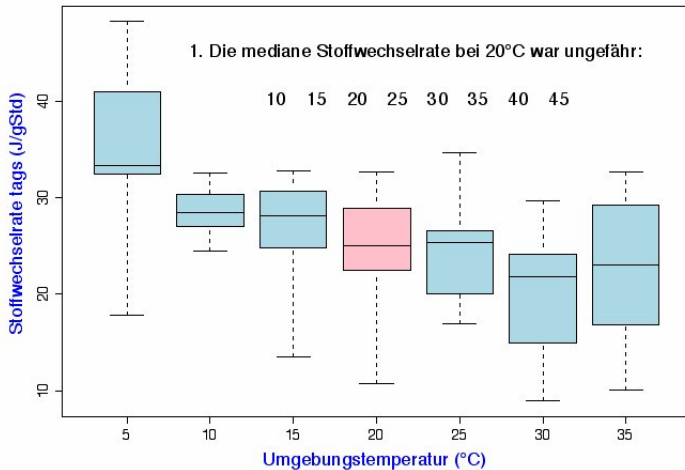
Klar:  
Stoffwechselrate  
höher  
bei  
tiefen Temperaturen

*Stoffwechselrate tags (J/gStd) und Umgebungstemperatur bei Ringeltauben (n=90)*

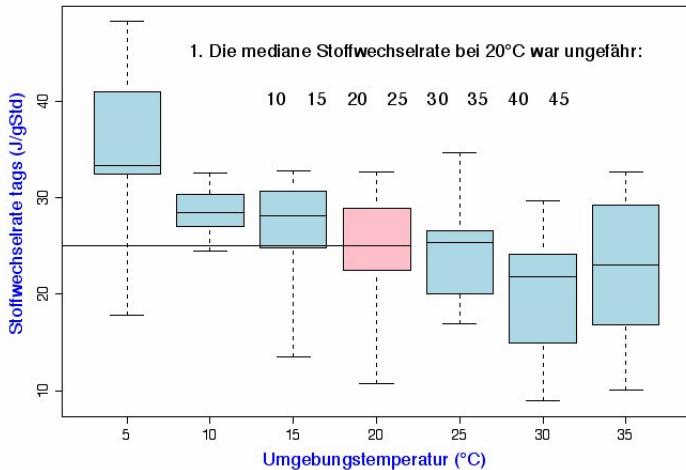
Vermutung:  
Bei **hohen** Temperaturen  
nimmt die Stoffwechselrate  
wieder zu  
(Hitzestress).

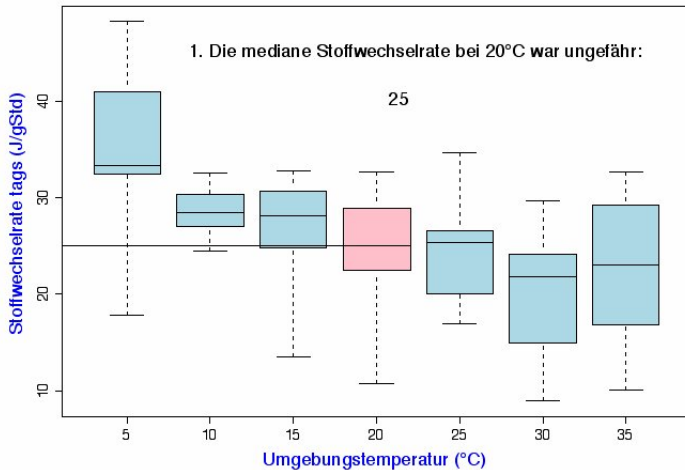
*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*

*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*

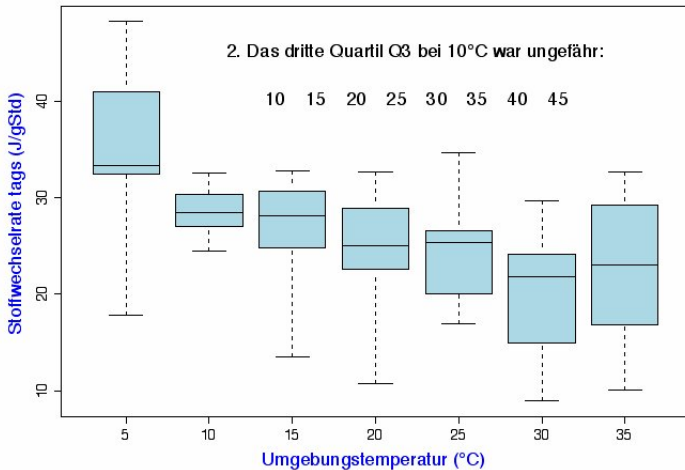
*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*

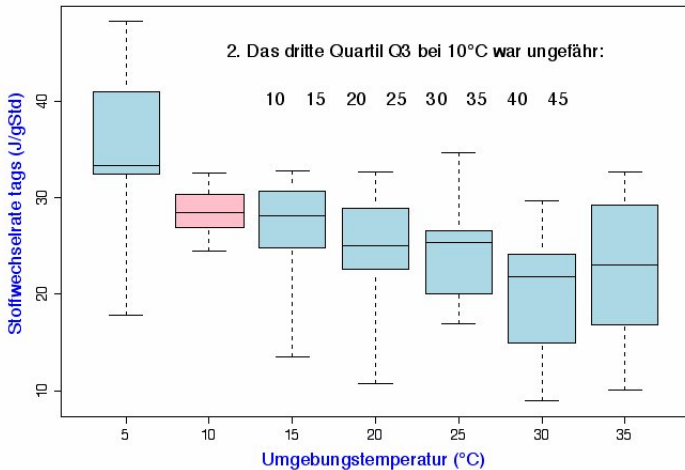


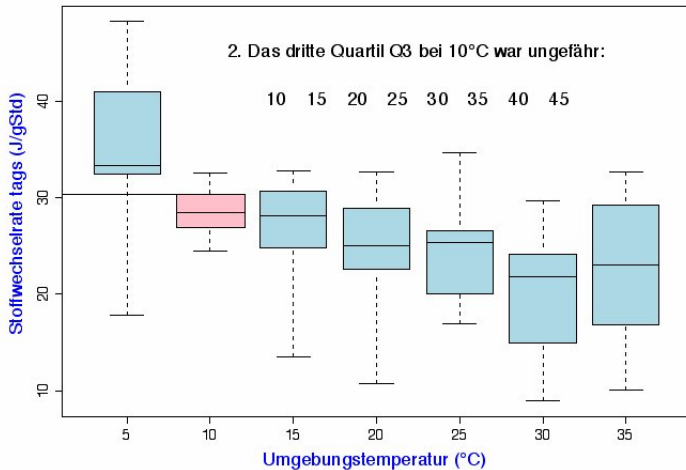
*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*

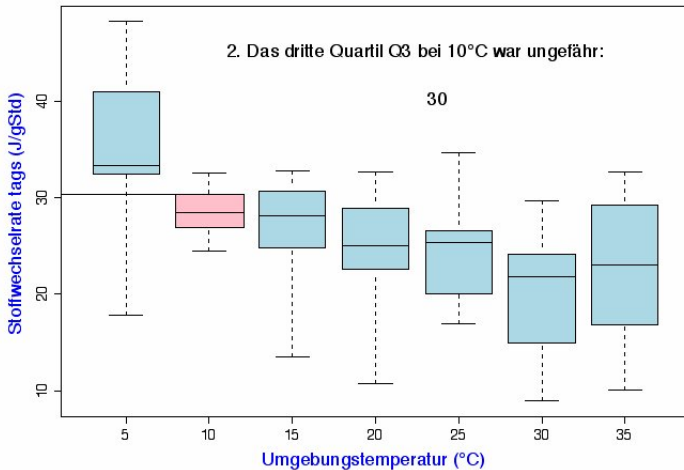
*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*



*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*

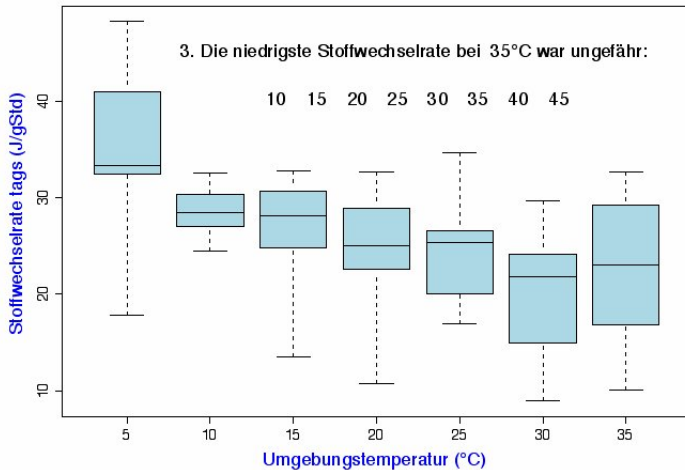
*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*

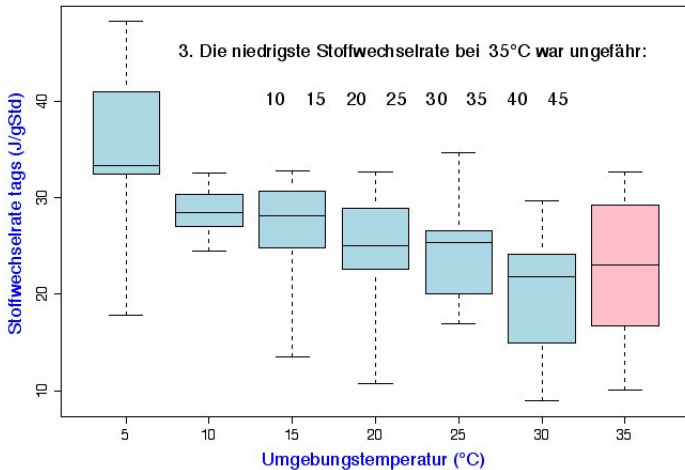
*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*

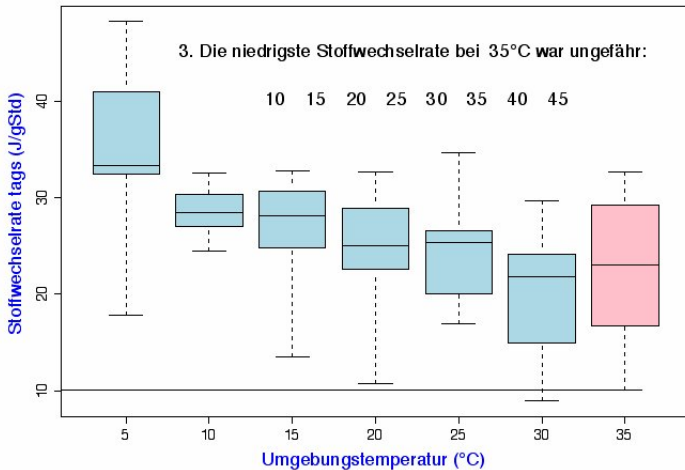
*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*

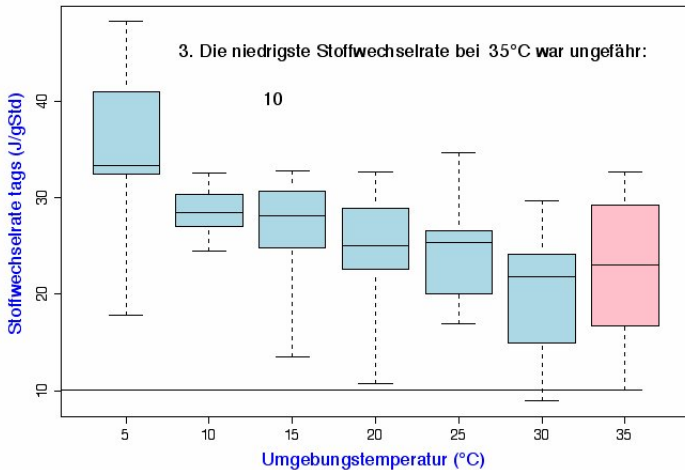




*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*

*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*

*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*

*Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)*



- 1 Einführung
  - Konzept und Quellen
  - Plan
- 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik
- 3 Graphische Darstellungen**
  - Histogramme und Dichtepolygone
  - Stripcharts
  - Boxplots
  - Beispiel: Ringeltaube
  - Beispiel: Darwin-Finken**
- 4 Statistische Kenngrößen
  - Median und andere Quartile
  - Mittelwert und Standardabweichung
- 5 Vom Sinn und Unsinn von Mittelwerten
  - Beispiel: Wählerische Bachstelzen
  - Beispiel: Spiderman & Spiderwoman
  - Beispiel: Kupfertoleranz beim Roten Straußgras

# Charles Robert Darwin (1809-1882)

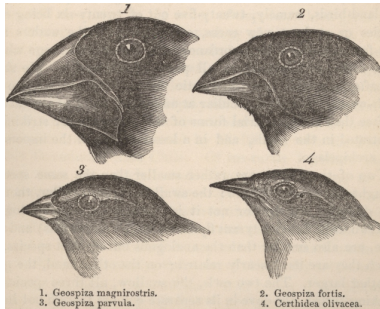


# Charles Robert Darwin (1809-1882)






# Darwin-Finken



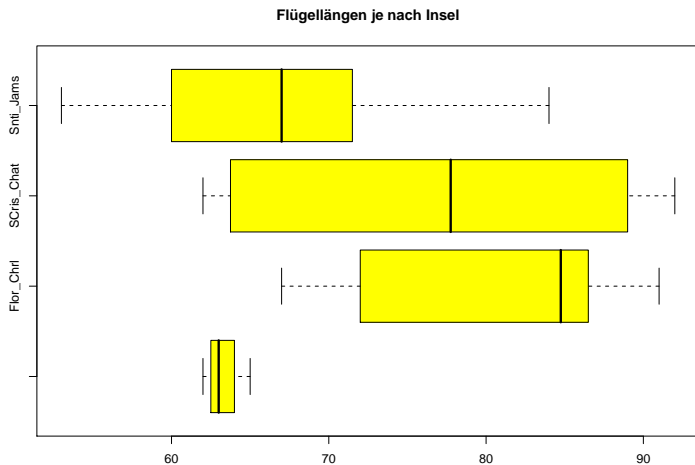
[http:](http://darwin-online.org.uk/graphics/Zoology_Illustrations.html)

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

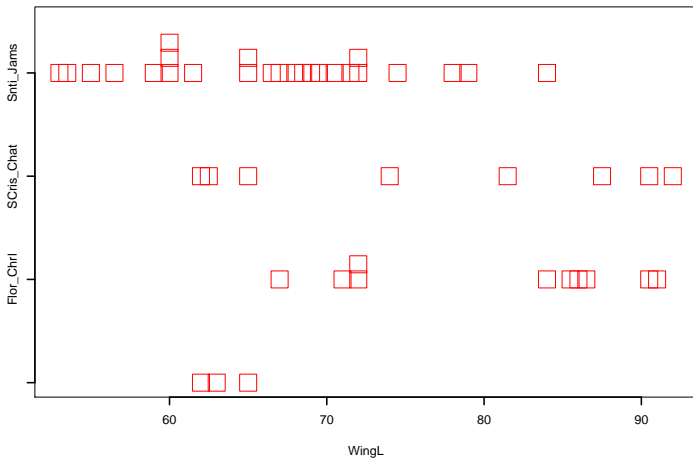
# Darwins Finken-Sammlung

-  Sulloway, F.J. (1982) The Beagle collections of Darwin's Finches (Geospizinae). *Bulletin of the British Museum (Natural History), Zoology series* **43**: 49-94.
- ▶ <http://datadryad.org/repo/handle/10255/dryad.154>

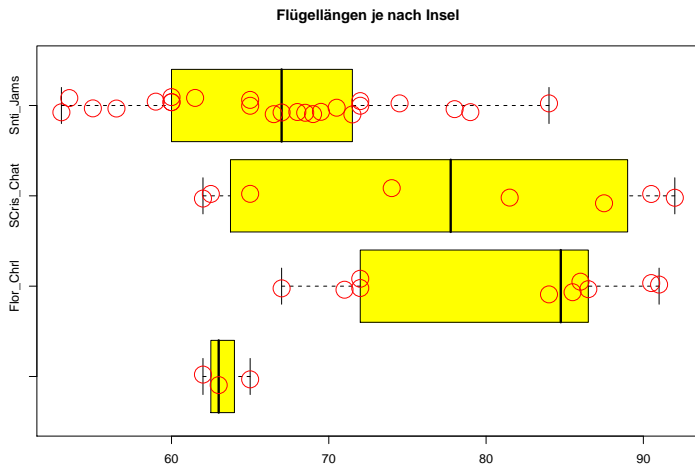
# Flügelängen der Darwin-Finken



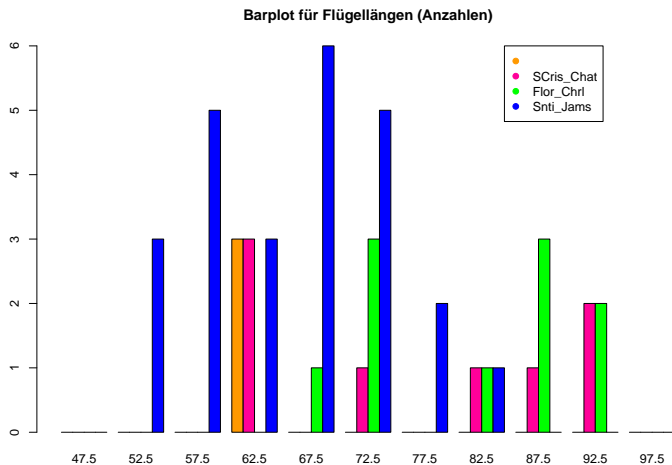
# Flügelängen der Darwin-Finken



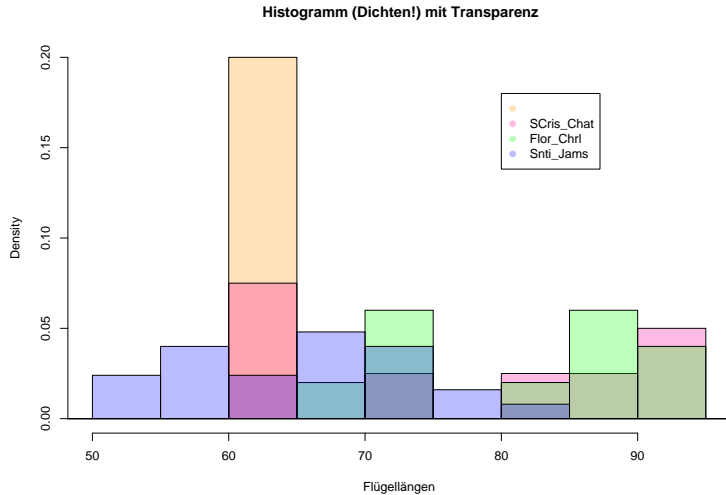
# Flügelängen der Darwin-Finken



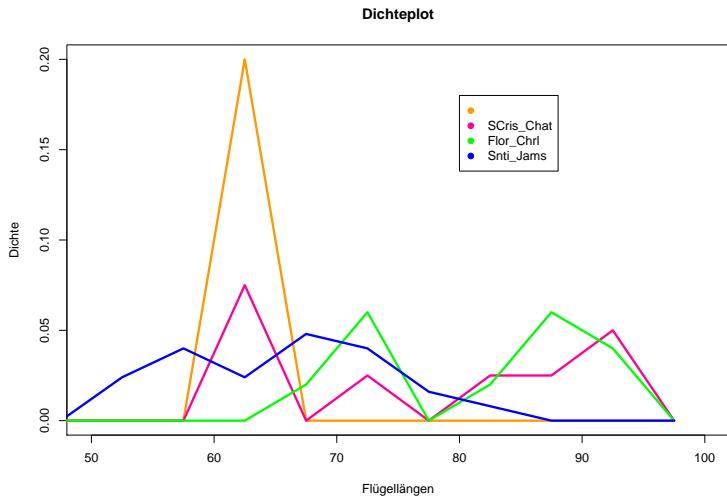
# Flügelängen der Darwin-Finken



# Flügelängen der Darwin-Finken

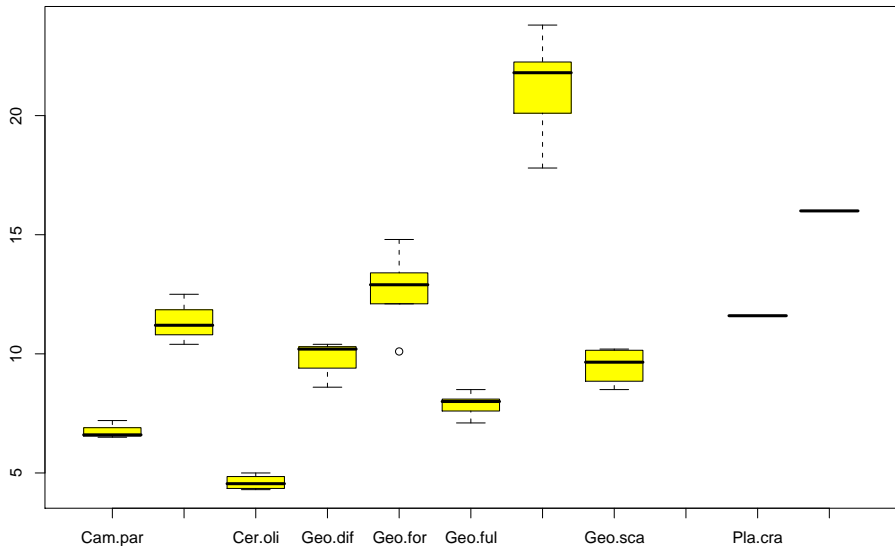


# Flügelängen der Darwin-Finken

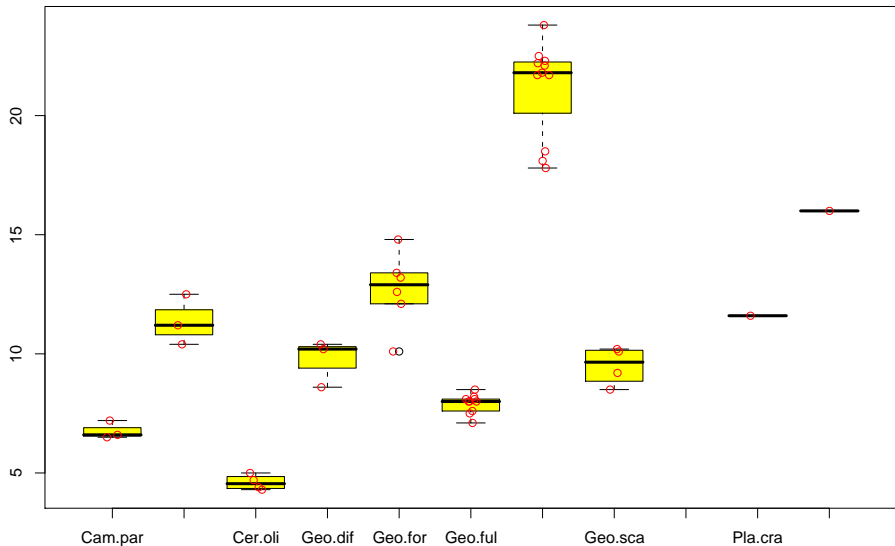




## Schnabelgröße je nach Art



## Schnabelgröße je nach Art



# Fazit

- 1 Histogramme erlauben einen detaillierten Blick auf die Daten

# Fazit

- 1 Histogramme erlauben einen detaillierten Blick auf die Daten
- 2 Dichtepolygone erlauben Vergleiche zwischen vielen Verteilungen

# Fazit

- 1 Histogramme erlauben einen detaillierten Blick auf die Daten
- 2 Dichtepolygone erlauben Vergleiche zwischen vielen Verteilungen
- 3 Boxplot können große Datenmengen vereinfacht zusammenfassen

# 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

# 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

# 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



- 1 Einführung
  - Konzept und Quellen
  - Plan
- 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik
- 3 Graphische Darstellungen
  - Histogramme und Dichtepolygone
  - Stripcharts
  - Boxplots
  - Beispiel: Ringeltaube
  - Beispiel: Darwin-Finken
- 4 Statistische Kenngrößen**
  - Median und andere Quartile
  - Mittelwert und Standardabweichung
- 5 Vom Sinn und Unsinn von Mittelwerten
  - Beispiel: Wählerische Bachstelzen
  - Beispiel: Spiderman & Spiderwoman
  - Beispiel: Kupfertoleranz beim Roten Straußgras

Es ist oft möglich,  
das Wesentliche  
an einer Stichprobe  
  
mit ein paar Zahlen  
zusammenzufassen.

Wesentlich:

1. Wie groß?

2. Wie variabel?

Wesentlich:

1. Wie groß?

Lageparameter

2. Wie variabel?

Wesentlich:

1. Wie groß?

Lageparameter

2. Wie variabel?

Streuungsparameter

Eine Möglichkeit  
kennen wir schon  
aus dem Boxplot:

# Lageparameter

## Der Median

Lageparameter

Der Median

Streuungsparameter



Lageparameter

Der Median

Streuungsparameter

Der Quartilabstand ( $Q_3 - Q_1$ )

- 1 Einführung
  - Konzept und Quellen
  - Plan
- 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik
- 3 Graphische Darstellungen
  - Histogramme und Dichtepolygone
  - Stripcharts
  - Boxplots
  - Beispiel: Ringeltaube
  - Beispiel: Darwin-Finken
- 4 Statistische Kenngrößen**
  - Median und andere Quartile**
  - Mittelwert und Standardabweichung
- 5 Vom Sinn und Unsinn von Mittelwerten
  - Beispiel: Wählerische Bachstelzen
  - Beispiel: Spiderman & Spiderwoman
  - Beispiel: Kupfertoleranz beim Roten Straußgras

## Der **Median**:

die Hälfte der Beobachtungen sind kleiner,  
die Hälfte sind größer.

Der **Median**:

die Hälfte der Beobachtungen sind kleiner,  
die Hälfte sind größer.

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

# Die Quartile

Das erste Quartil,  $Q_1$ :

# Die Quartile

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

# Die Quartile

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

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

# Die Quartile

Das dritte Quartil,  $Q_3$ :



# Die Quartile

Das dritte Quartil,  $Q_3$ :  
drei Viertel der Beobachtungen  
sind kleiner,  
ein Viertel sind größer.

# Die Quartile

Das dritte Quartil,  $Q_3$ :  
drei Viertel der Beobachtungen  
sind kleiner,  
ein Viertel sind größer.

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

- 1 Einführung
  - Konzept und Quellen
  - Plan
- 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik
- 3 Graphische Darstellungen
  - Histogramme und Dichtepolygone
  - Stripcharts
  - Boxplots
  - Beispiel: Ringeltaube
  - Beispiel: Darwin-Finken
- 4 Statistische Kenngrößen**
  - Median und andere Quartile
  - Mittelwert und Standardabweichung**
- 5 Vom Sinn und Unsinn von Mittelwerten
  - Beispiel: Wählerische Bachstelzen
  - Beispiel: Spiderman & Spiderwoman
  - Beispiel: Kupfertoleranz beim Roten Straußgras

Am häufigsten werden benutzt:

Lageparameter

Der Mittelwert  $\bar{x}$

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:

Mittelwert

=

$$\frac{\text{Summe der Messwerte}}{\text{Anzahl der Messwerte}}$$



# DEFINITION:

Mittelwert

=

$$\frac{\text{Summe}}{\text{Anzahl}}$$

# DEFINITION:

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

# DEFINITION:

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

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

# DEFINITION:

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:

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

## Beispiel:

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

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

## Beispiel:

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

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

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

## Beispiel:

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

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

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

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



## Beispiel:

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

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

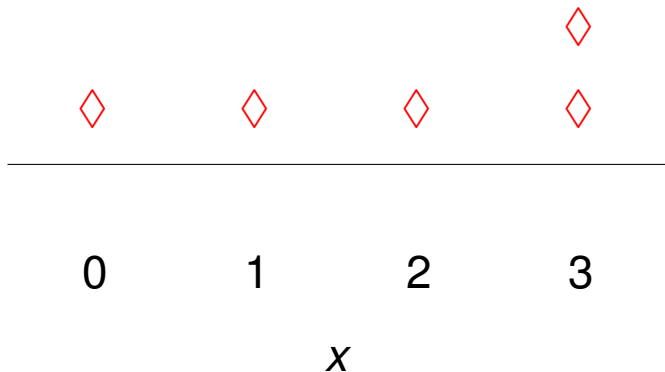
$$\bar{x} = (3 + 0 + 2 + 3 + 1)/5$$

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

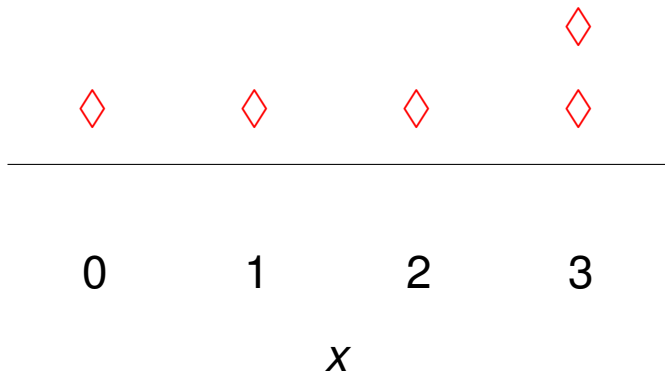
$$\bar{x} = 1,8$$

# Geometrische Bedeutung des Mittelwerts: Der Schwerpunkt

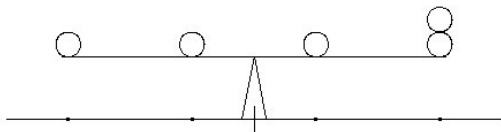
Wir stellen uns die Beobachtungen als gleich schwere Gewichte auf einer Waage vor:



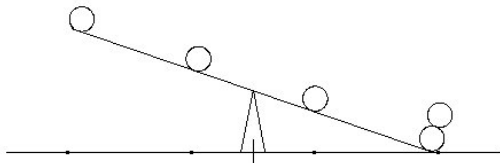
Wo muß der Drehpunkt sein, damit die Waage im Gleichgewicht ist?



$$m = 1,5 ?$$

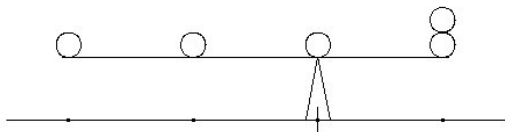


$$m = 1,5 ?$$

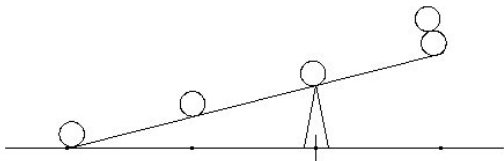


zu klein

$$m = 2 ?$$



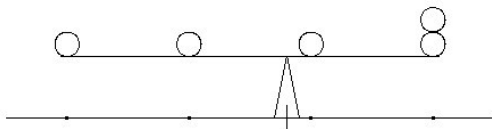
$$m = 2 ?$$



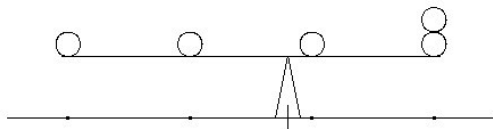
zu groß



$$m = 1,8 ?$$



$$m = 1,8 ?$$



richtig

Beispiel: *Galathea intermedia*

„Rundlichkeit“

:=

Abdominalbreite / Carapaxlänge

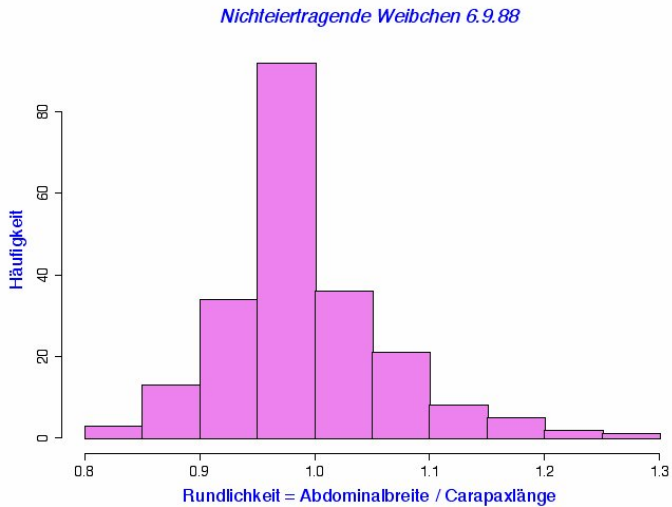
Beispiel: *Galathea intermedia*

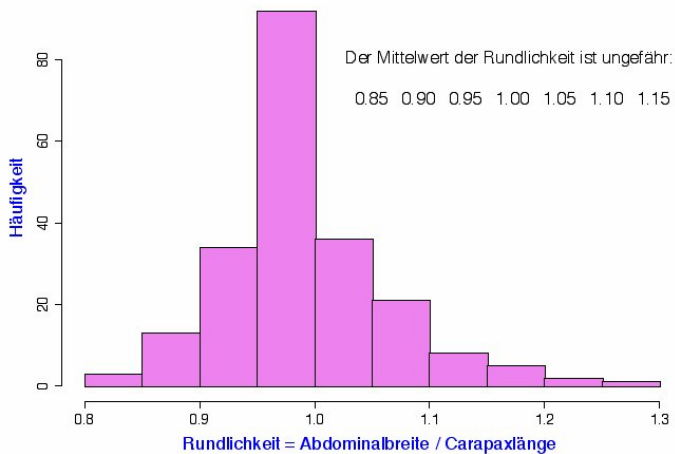
„Rundlichkeit“

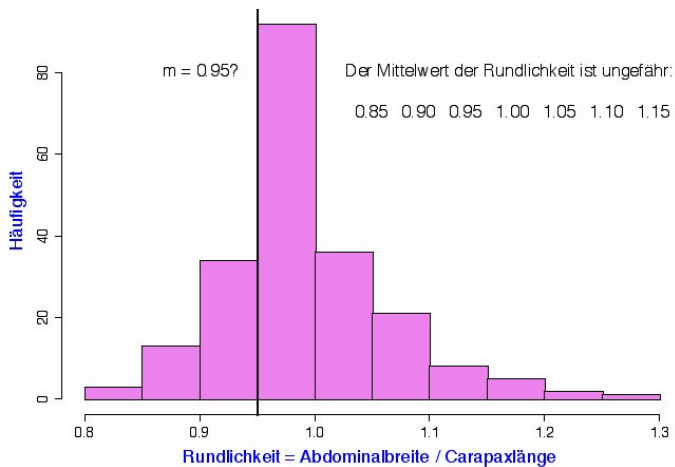
:=

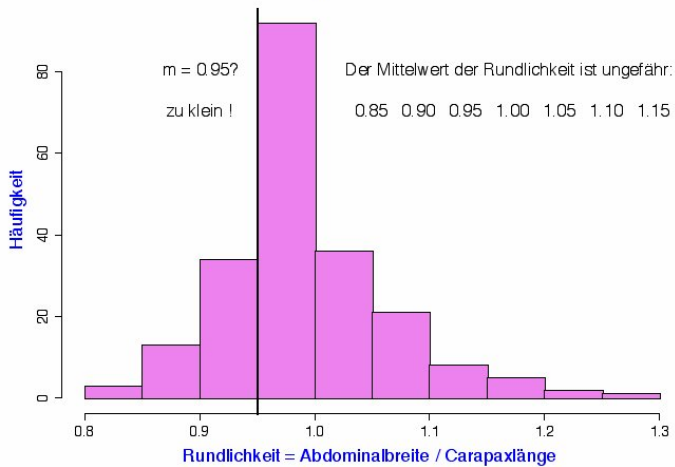
Abdominalbreite / Carapaxlänge

Vermutung:  
Rundlichkeit nimmt  
bei Geschlechtsreife zu

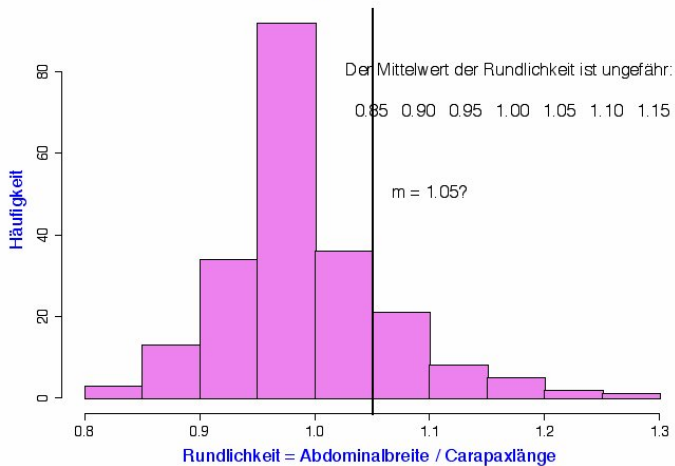


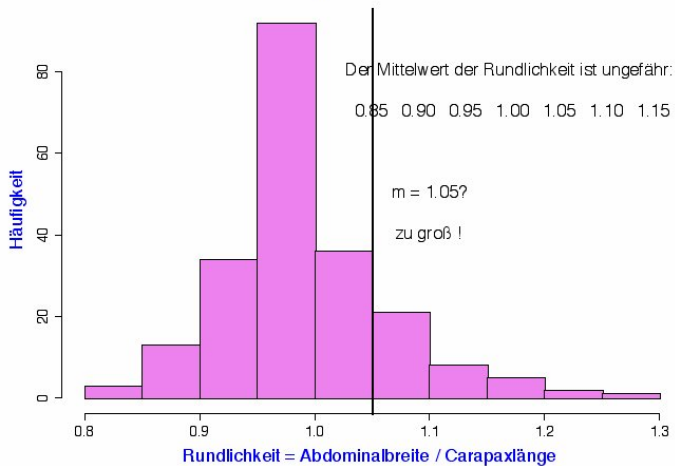
*Nichteiertragende Weibchen 6.9.88*

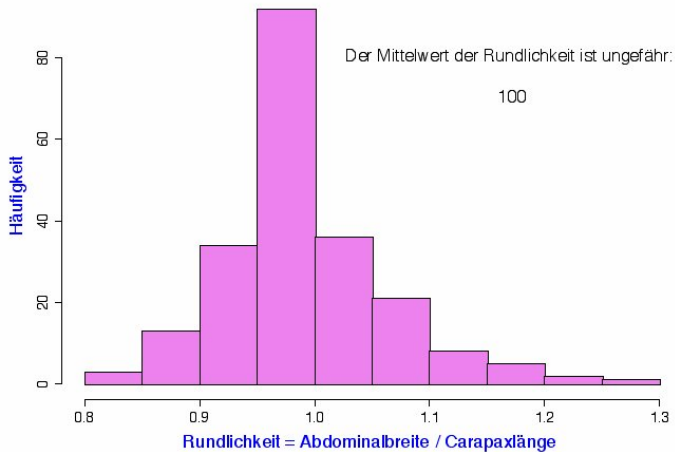
*Nichteiertragende Weibchen 6.9.88*

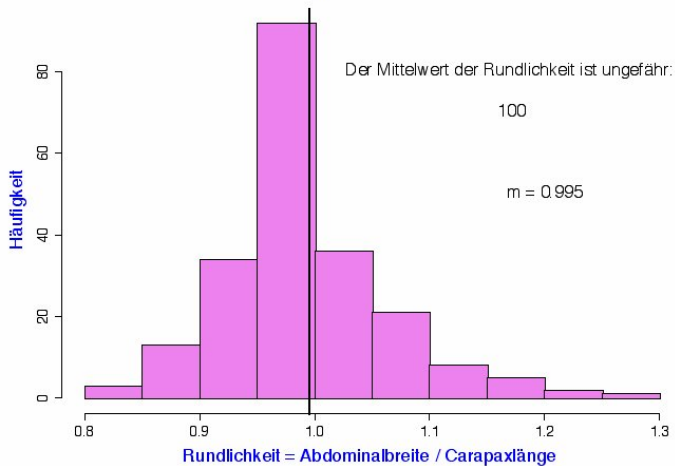
*Nichteiertragende Weibchen 6.9.88*



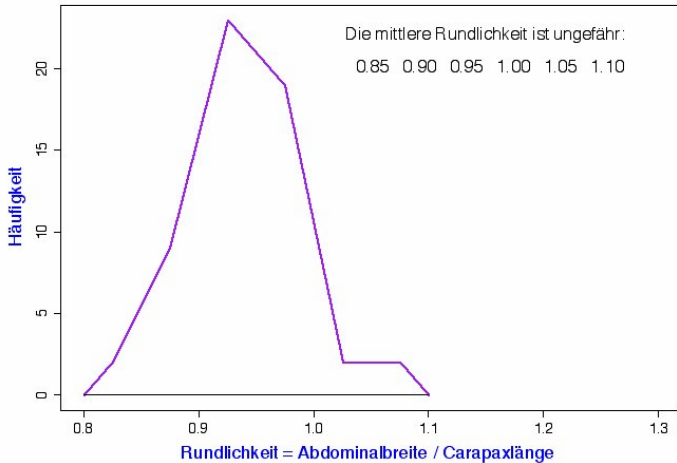
*Nichteiertragende Weibchen 6.9.88*

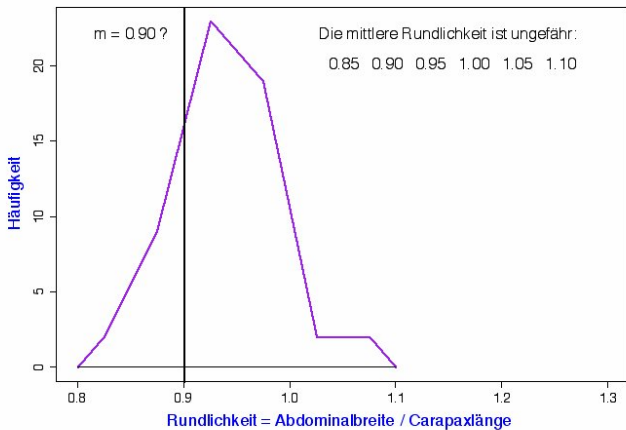
*Nichteiertragende Weibchen 6.9.88*

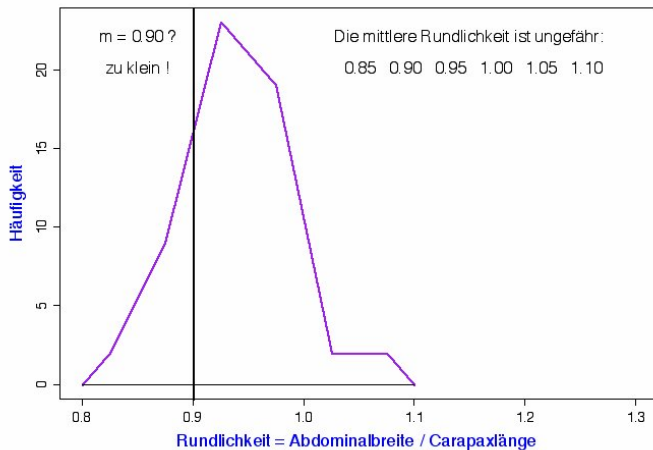
*Nichteiertragende Weibchen 6.9.88*

*Nichteiertragende Weibchen 6.9.88*

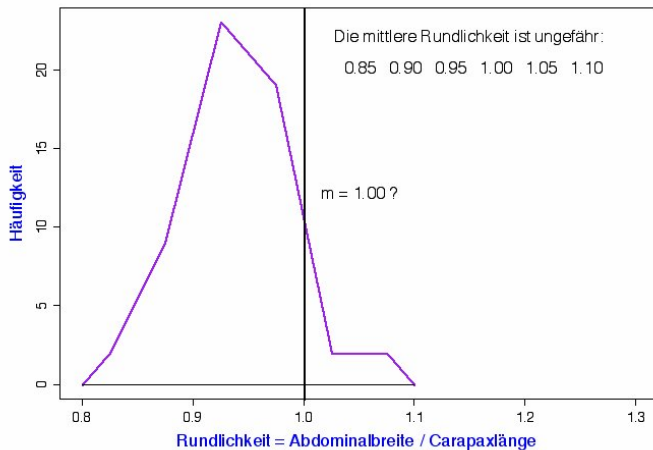
Beispiel:  
3.11.88

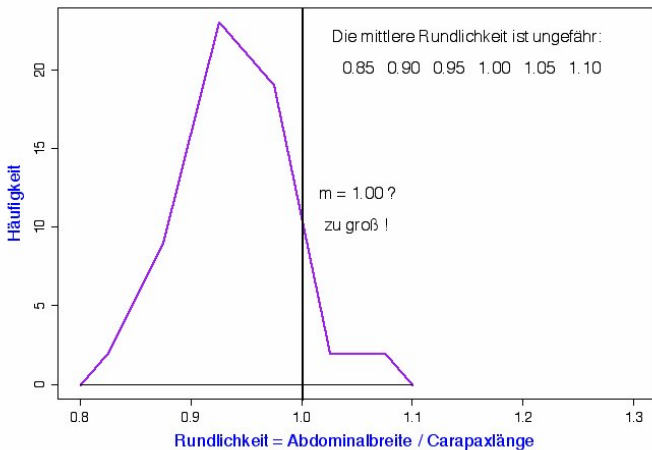
*Nichteiertragende Weibchen 3.11.88*

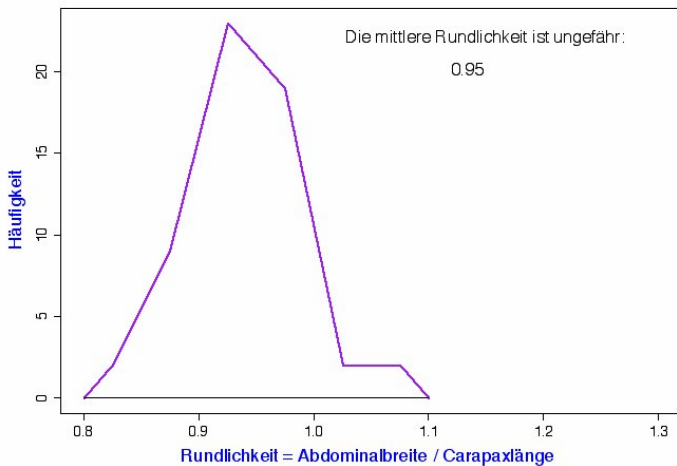
*Nichteiertragende Weibchen 3.11.88*

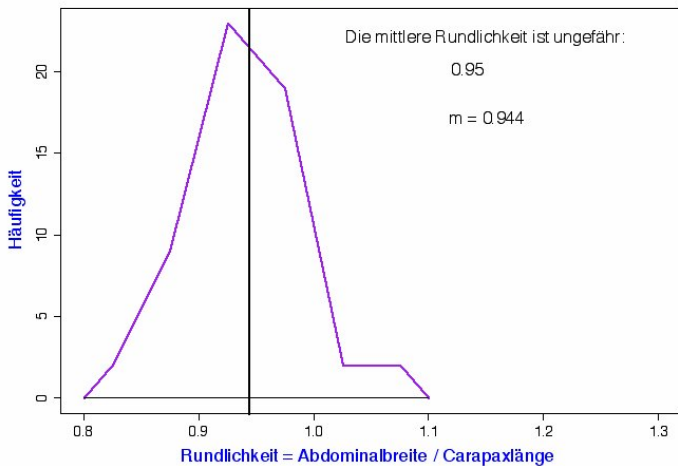
*Nichteiertragende Weibchen 3.11.88*



*Nichteiertragende Weibchen 3.11.88*

*Nichteiertragende Weibchen 3.11.88*

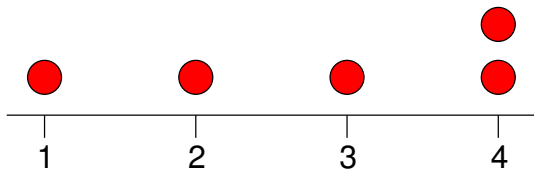
*Nichteiertragende Weibchen 3.11.88*

*Nichteiertragende Weibchen 3.11.88*

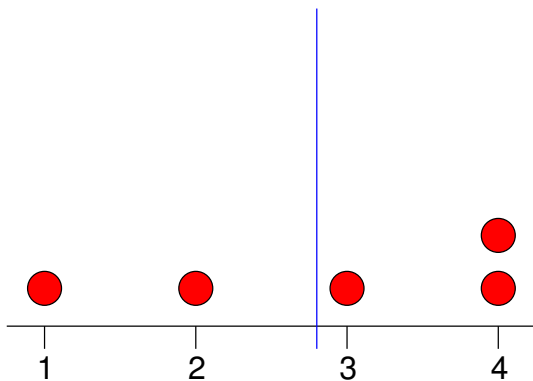
# Die Standardabweichung

# Die Standardabweichung

Wie weit weicht  
eine typische Beobachtung  
vom  
Mittelwert  
ab ?

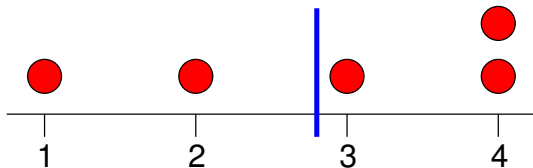


Mittelwert=2,8

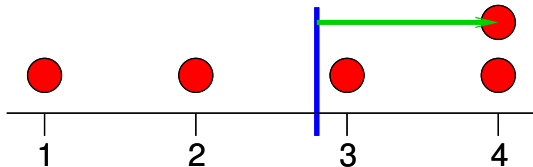




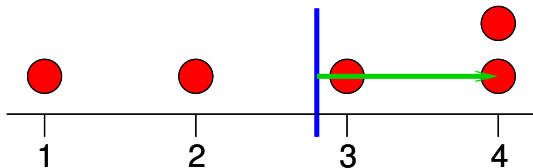
typische  
Abweichung =?



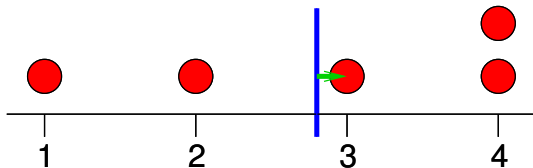
$$\text{Abweichung} = 4 - 2,8 = 1,2$$



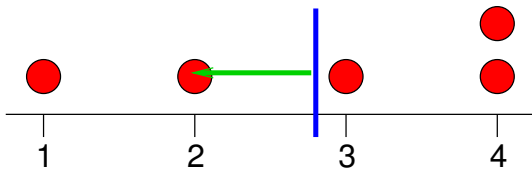
$$\text{Abweichung} = 4 - 2,8 = 1,2$$



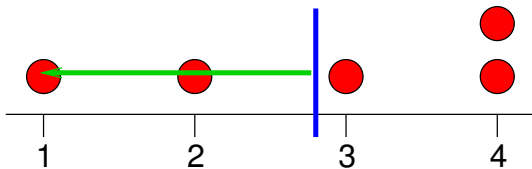
$$\text{Abweichung} = 3 - 2,8 = 0,2$$



$$\text{Abweichung} = 2 - 2,8 = -0,8$$



$$\text{Abweichung} = 1 - 2,8 = -1,8$$



Die **Standardabweichung**  $\sigma$  (“sigma”)  
[auch *SD* von engl. *standard deviation*]  
ist ein  
etwas komisches  
gewichtetes Mittel  
der Abweichungsbeträge

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:

Die **Standardabweichung** von  $x_1, x_2, \dots, x_n$   
als Formel:

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

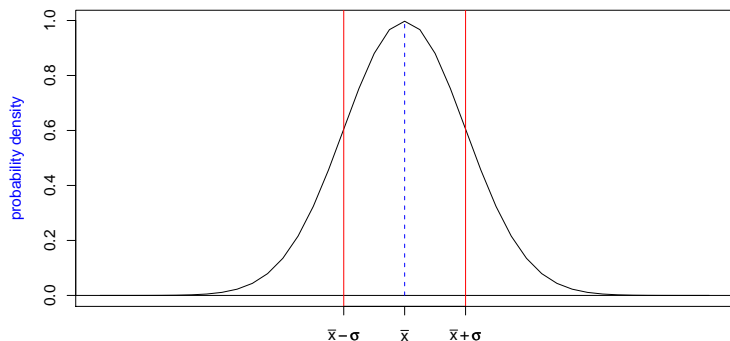
Die **Standardabweichung** von  $x_1, x_2, \dots, x_n$   
als Formel:

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

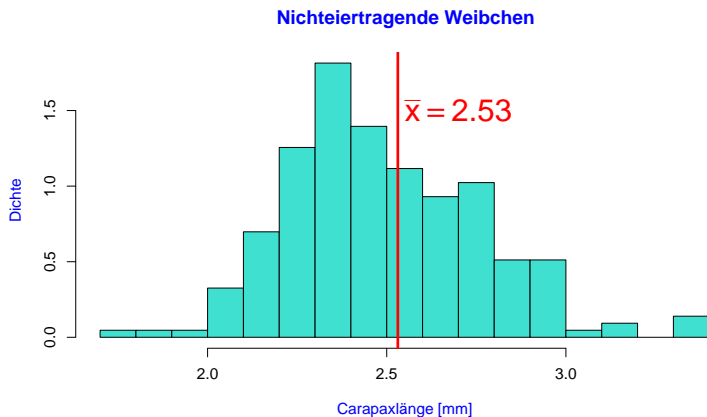
$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  heißt **Varianz**.

# Faustregel für die Standardabweichung

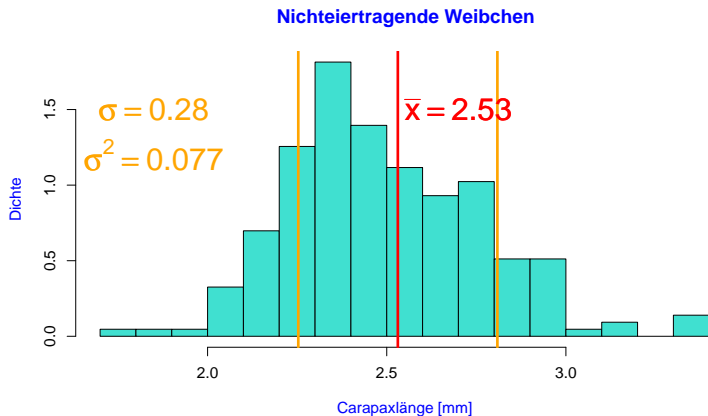
Bei ungefähr glockenförmigen (also eingipfligen und symmetrischen) Verteilungen liegen ca. 2/3 der Verteilung zwischen  $\bar{x} - \sigma$  und  $\bar{x} + \sigma$ .



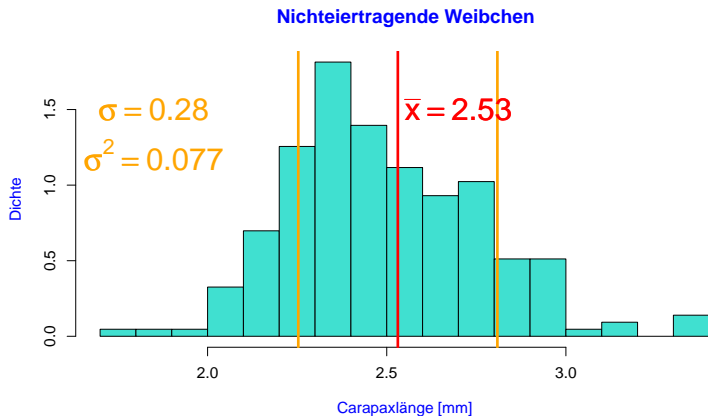
# Standardabweichung der Carapaxlängen nichteiertragender Weibchen vom 6.9.88



# Standardabweichung der Carapaxlängen nichteierttragender Weibchen vom 6.9.88



# Standardabweichung der Carapaxlängen nichteierttragender Weibchen vom 6.9.88



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

# Varianz der Carapaxlängen nichteiерtragender Weibchen vom 6.9.88

Alle Carapaxlängen im Meer:  $\mathcal{X} = (X_1, X_2, \dots, X_N)$ .



# 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:  $\mathcal{S} = (S_1, S_2, \dots, S_{n=215})$

# 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:

$$\sigma_S^2 = \frac{1}{n} \sum_{i=1}^{215} (S_i - \bar{S})^2 \approx 0,0768$$

# Varianz der Carapaxlängen nichtteiertragender 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  $\sigma_X^2$  in der ganzen Population verwenden?

# 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  $\sigma_X^2$  in der ganzen Population verwenden?

Ja, können wir machen.

# 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  $\sigma_{\mathcal{X}}^2$

# Varianzbegriffe

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

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

# Varianzbegriffe

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

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

korrigierte Stichprobenvarianz:

$$s^2 = \frac{n}{n-1} \sigma_S^2$$

# Varianzbegriffe

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

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

korrigierte Stichprobenvarianz:

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



# Varianzbegriffe

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

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

korrigierte Stichprobenvarianz:

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

Mit “Standardabweichung von  $S$ ” ist meistens das korrigierte  $s$  gemeint.

# Beispiel

# Die Daten

$x$             1    3    0    5    1

$$\bar{x} = ?$$

Summe

$x$

1

3

0

5

1

10

$$\bar{x} = 10/5 = 2$$

Summe

|     |   |   |   |   |   |    |
|-----|---|---|---|---|---|----|
| $x$ | 1 | 3 | 0 | 5 | 1 | 10 |
|-----|---|---|---|---|---|----|

$$x - \bar{x}$$

$$\bar{x} = 10/5 = 2$$

Summe

|     |   |   |   |   |   |    |
|-----|---|---|---|---|---|----|
| $x$ | 1 | 3 | 0 | 5 | 1 | 10 |
|-----|---|---|---|---|---|----|

|               |    |   |    |   |    |   |
|---------------|----|---|----|---|----|---|
| $x - \bar{x}$ | -1 | 1 | -2 | 3 | -1 | 0 |
|---------------|----|---|----|---|----|---|

$$\bar{x} = 10/5 = 2$$

Summe

|     |   |   |   |   |   |    |
|-----|---|---|---|---|---|----|
| $x$ | 1 | 3 | 0 | 5 | 1 | 10 |
|-----|---|---|---|---|---|----|

|               |    |   |    |   |    |   |
|---------------|----|---|----|---|----|---|
| $x - \bar{x}$ | -1 | 1 | -2 | 3 | -1 | 0 |
|---------------|----|---|----|---|----|---|

$$(x - \bar{x})^2$$

$$\bar{x} = 10/5 = 2$$

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 |



$$\bar{x} = 10/5 = 2$$

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 |

$$s^2 = \text{Summe}((x - \bar{x})^2) / (n - 1)$$

$$\bar{x} = 10/5 = 2$$

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 |

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

$$\bar{x} = 10/5 = 2$$

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 |

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

$$\bar{x} = 10/5 = 2$$

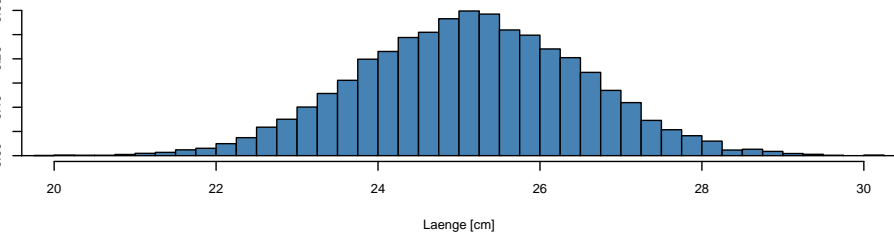
Summe

|                   |    |   |    |   |    |    |
|-------------------|----|---|----|---|----|----|
| $x$               | 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 |

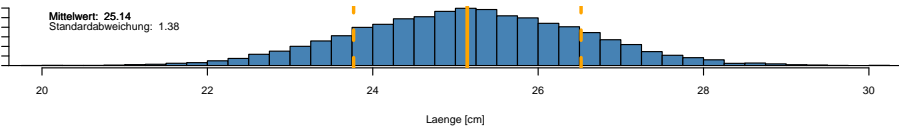
$$\begin{aligned}s^2 &= \text{Summe}((x - \bar{x})^2) / (n - 1) \\ &= 16 / (5 - 1) = 4\end{aligned}$$

$$s = 2$$

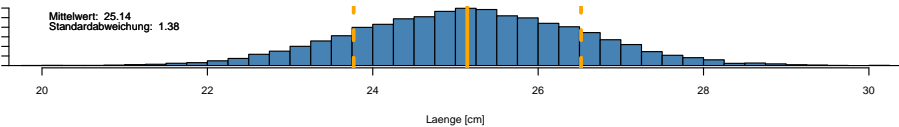
## Eine simulierte Fischpopulation (N=10000 adulte)



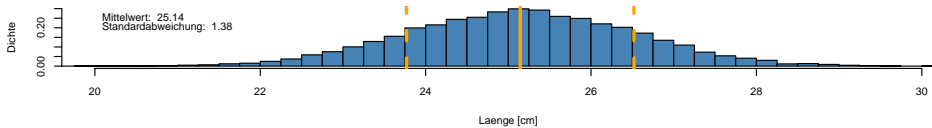
## Eine simulierte Fischpopulation (N=10000 adulte)



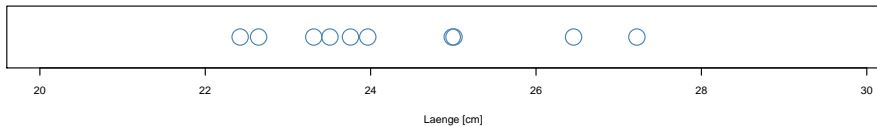
## Eine simulierte Fischpopulation (N=10000 adulte)



## Eine simulierte Fischpopulation (N=10000 adulte)

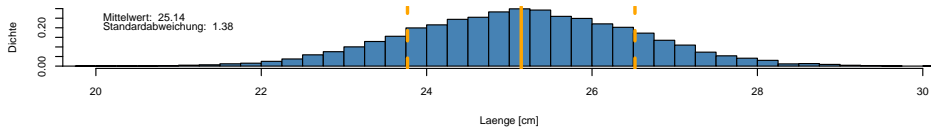


## Eine Stichprobe aus der Population (n=10)





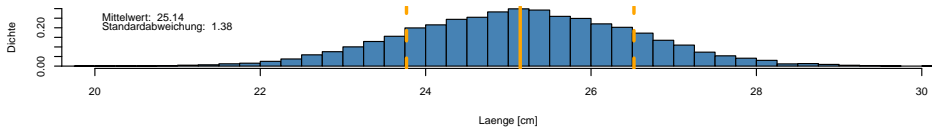
## Eine simulierte Fischpopulation (N=10000 adulte)



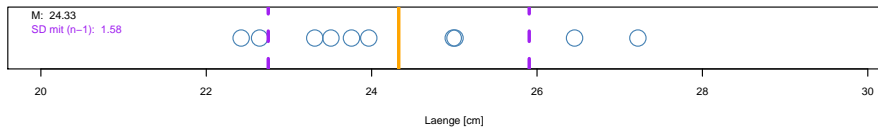
## Eine Stichprobe aus der Population (n=10)



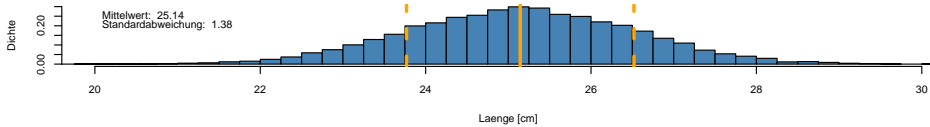
## Eine simulierte Fischpopulation (N=10000 adulte)



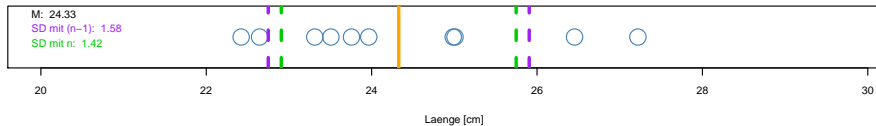
## Eine Stichprobe aus der Population (n=10)



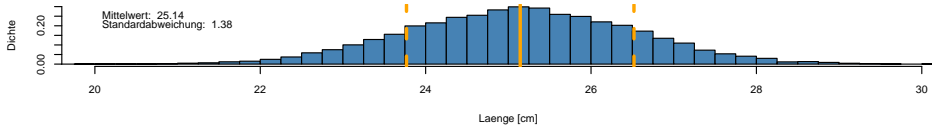
## Eine simulierte Fischpopulation (N=10000 adulte)



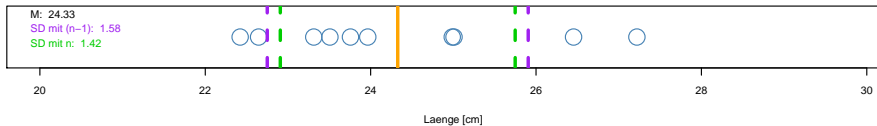
## Eine Stichprobe aus der Population (n=10)



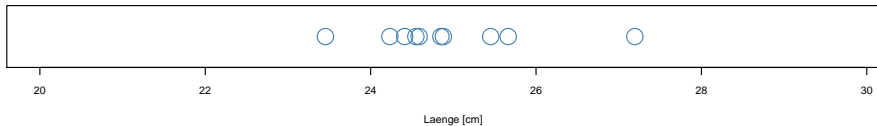
## Eine simulierte Fischpopulation (N=10000 adulte)



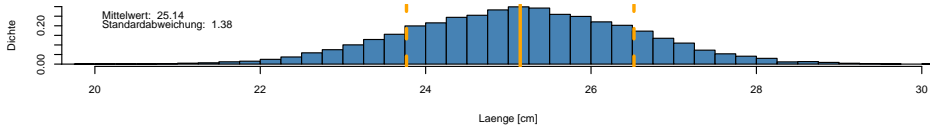
## Eine Stichprobe aus der Population (n=10)



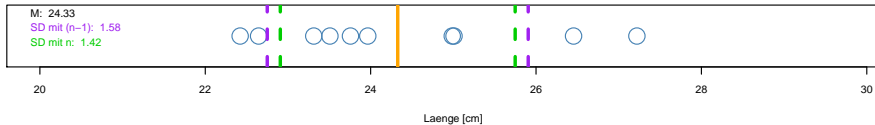
## Noch eine Stichprobe aus der Population (n=10)



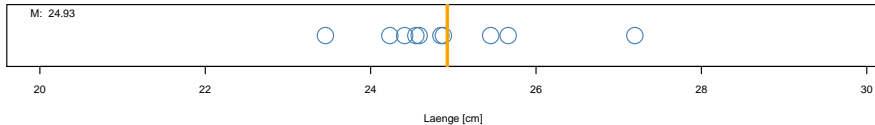
## Eine simulierte Fischpopulation (N=10000 adulte)



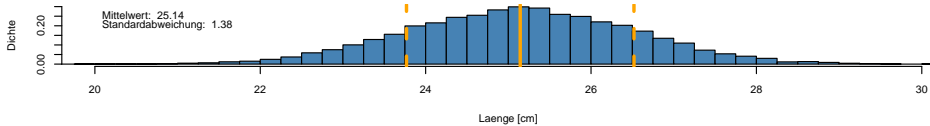
## Eine Stichprobe aus der Population (n=10)



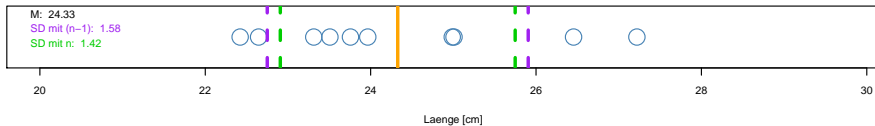
## Noch eine Stichprobe aus der Population (n=10)



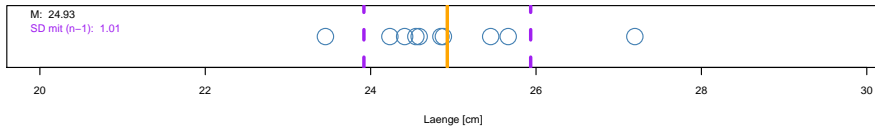
## Eine simulierte Fischpopulation (N=10000 adulte)



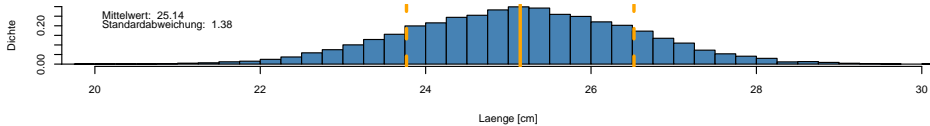
## Eine Stichprobe aus der Population (n=10)



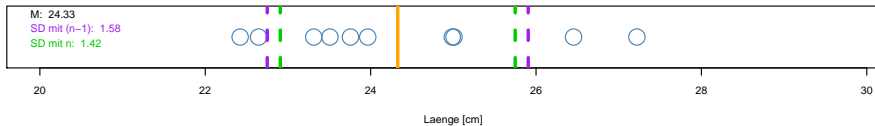
## Noch eine Stichprobe aus der Population (n=10)



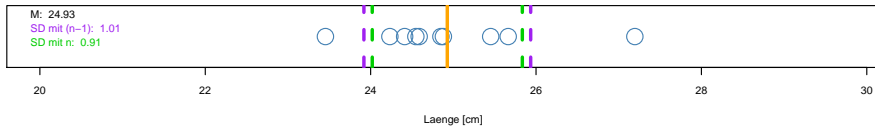
## Eine simulierte Fischpopulation (N=10000 adulte)

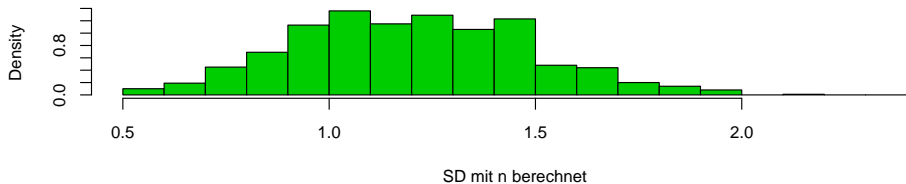
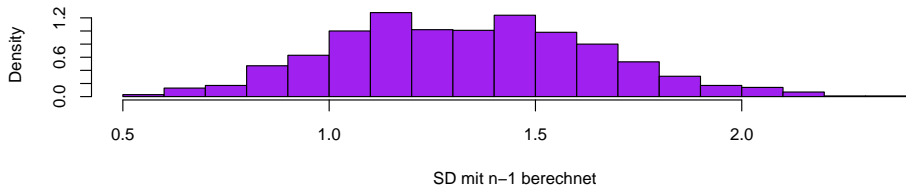


## Eine Stichprobe aus der Population (n=10)

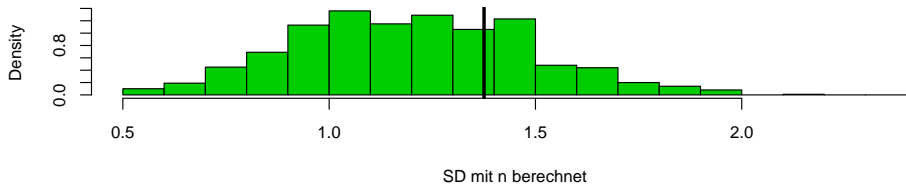
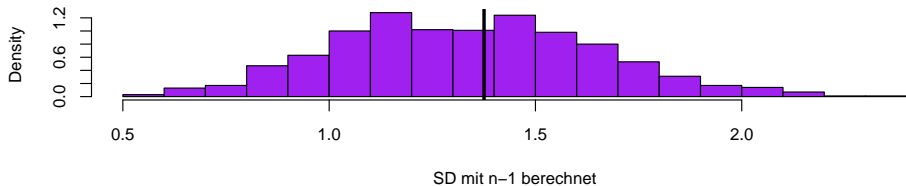


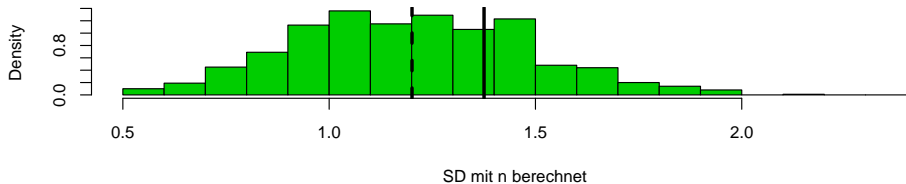
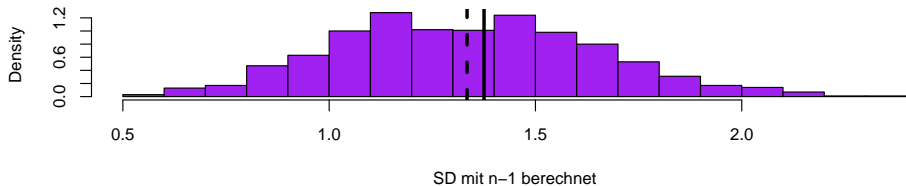
## Noch eine Stichprobe aus der Population (n=10)



1000 Stichproben, jeweils vom Umfang  $n=10$ 



1000 Stichproben, jeweils vom Umfang  $n=10$ 

1000 Stichproben, jeweils vom Umfang  $n=10$ 

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

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

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

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

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

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

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

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

verwenden.

- 1 Einführung
  - Konzept und Quellen
  - Plan
- 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik
- 3 Graphische Darstellungen
  - Histogramme und Dichtepolygone
  - Stripcharts
  - Boxplots
  - Beispiel: Ringeltaube
  - Beispiel: Darwin-Finken
- 4 Statistische Kenngrößen
  - Median und andere Quartile
  - Mittelwert und Standardabweichung
- 5 **Vom Sinn und Unsinn von Mittelwerten**
  - **Beispiel: Wählerische Bachstelzen**
  - **Beispiel: Spiderman & Spiderwoman**
  - **Beispiel: Kupfertoleranz beim Roten Straußgras**

## Mittelwert und Standardabweichung. . .

- charakterisieren die Daten gut, falls deren Verteilung glockenförmig ist


## Mittelwert und Standardabweichung. . .

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

## 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.


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



## 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.

 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.

- 1 Einführung
  - Konzept und Quellen
  - Plan
- 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik
- 3 Graphische Darstellungen
  - Histogramme und Dichtepolygone
  - Stripcharts
  - Boxplots
  - Beispiel: Ringeltaube
  - Beispiel: Darwin-Finken
- 4 Statistische Kenngrößen
  - Median und andere Quartile
  - Mittelwert und Standardabweichung
- 5 **Vom Sinn und Unsinn von Mittelwerten**
  - **Beispiel: Wählerische Bachstelzen**
  - Beispiel: Spiderman & Spiderwoman
  - Beispiel: Kupfertoleranz beim Roten Straußgras

# Bachstelzen fressen Dungfliegen

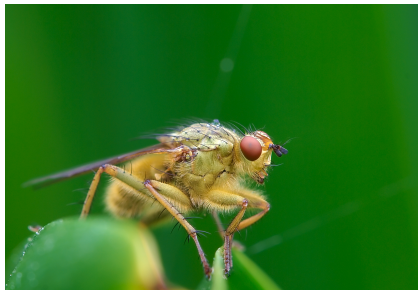
Räuber



Bachstelze (White Wagtail)  
*Motacilla alba alba*

image (c) by Artur Mikołajewski

Beute



Gelbe Dungfliege  
*Scatophaga stercoraria*

image (c) by Viatour Luc

# 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.

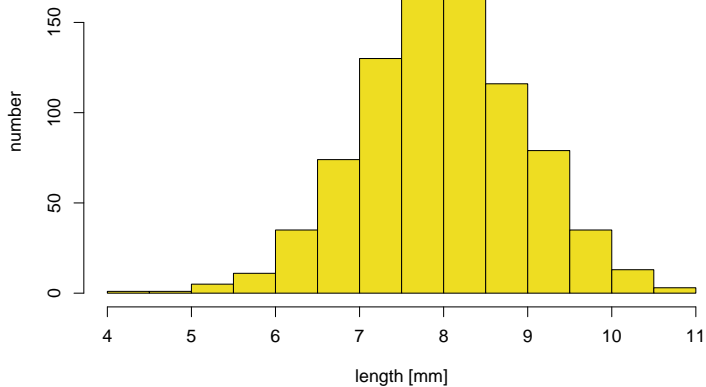


N.B. Davies.

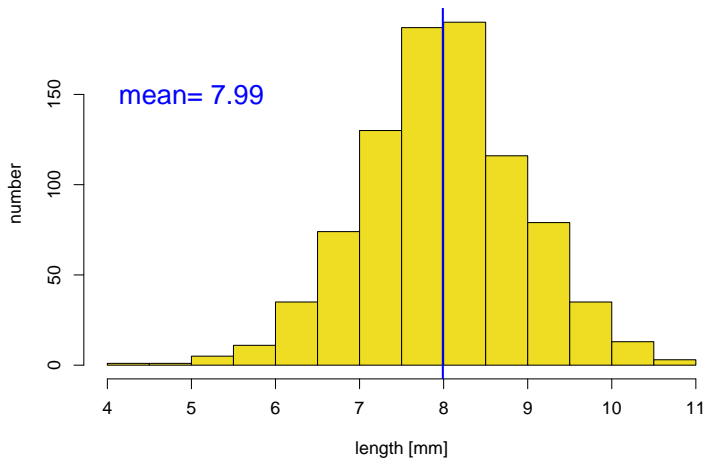
Prey selection and social behaviour in wagtails (Aves: Motacillidae).

*J. Anim. Ecol.*, 46:37–57, 1977.

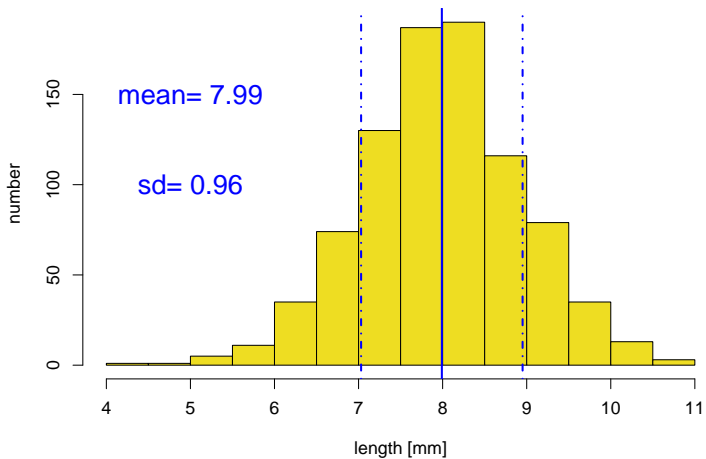
## available dung flies



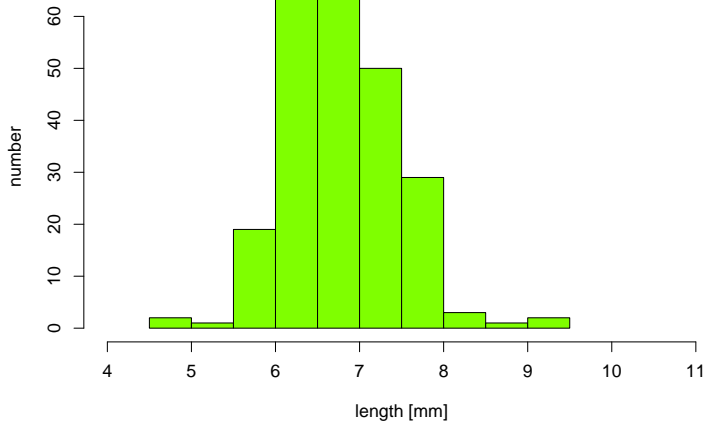
## available dung flies



## available dung flies

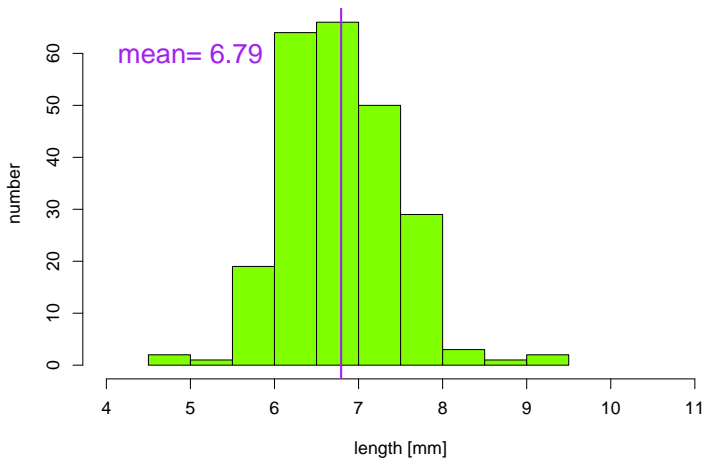


## captured dung flies

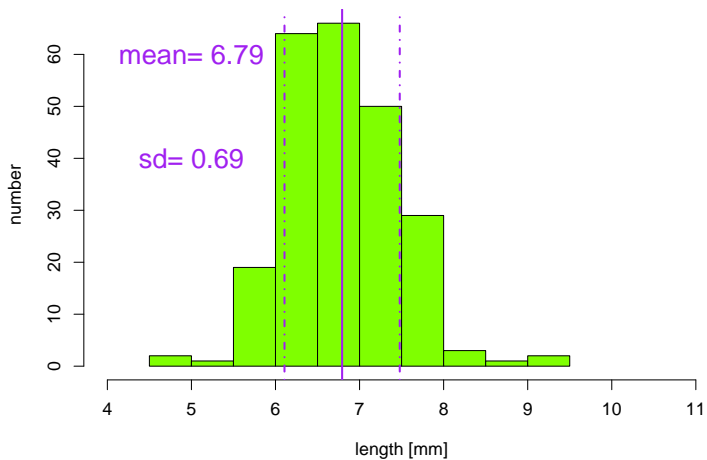




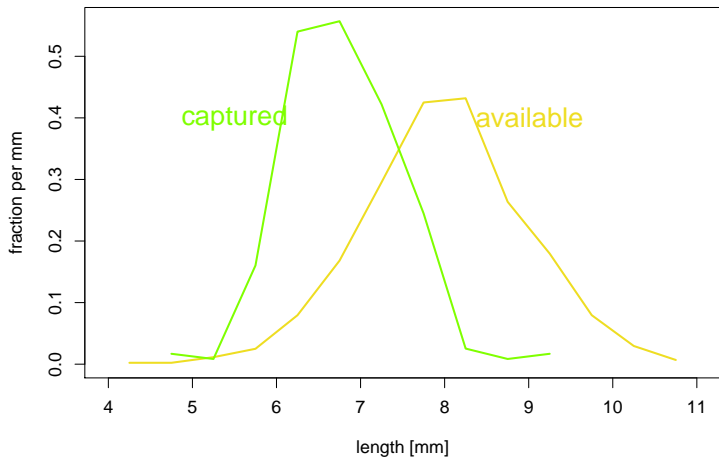
## captured dung flies



## captured dung flies



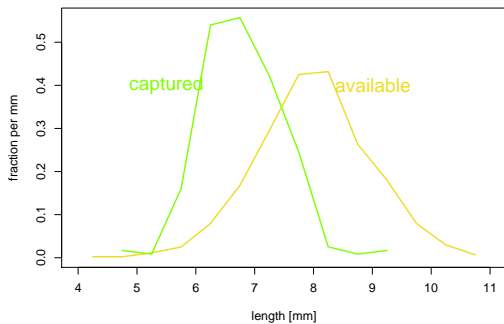
## dung flies: available, captured



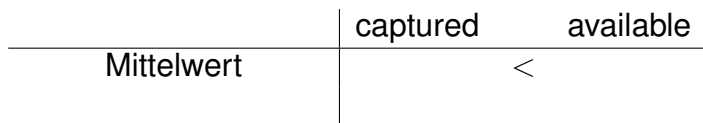
# Vergleich der Größenverteilungen

|            | captured | available |
|------------|----------|-----------|
| Mittelwert |          |           |

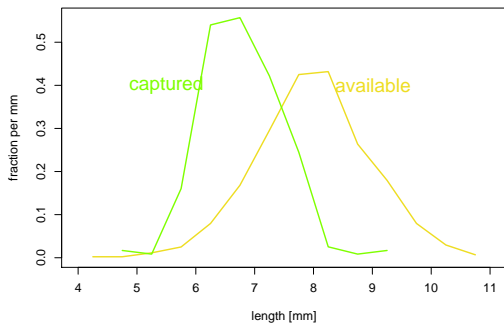
dung flies: available, captured



# Vergleich der Größenverteilungen



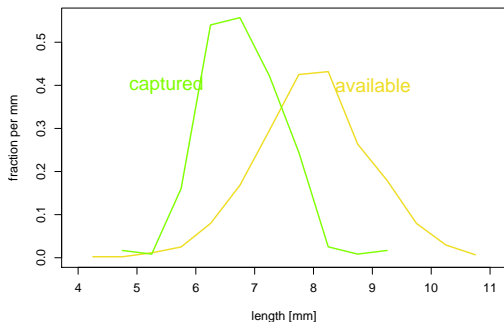
dung flies: available, captured



# Vergleich der Größenverteilungen

|            | captured |   | available |
|------------|----------|---|-----------|
| Mittelwert | 6.29     | < | 7.99      |

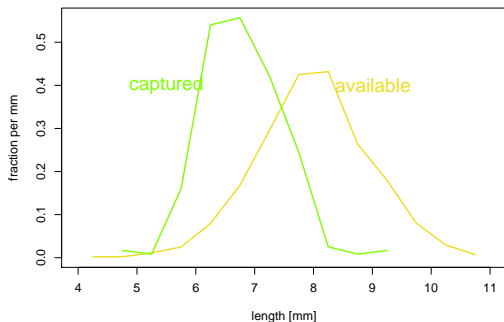
dung flies: available, captured



# Vergleich der Größenverteilungen

|                    | captured | available |
|--------------------|----------|-----------|
| Mittelwert         | 6.29     | 7.99      |
| Standardabweichung |          |           |

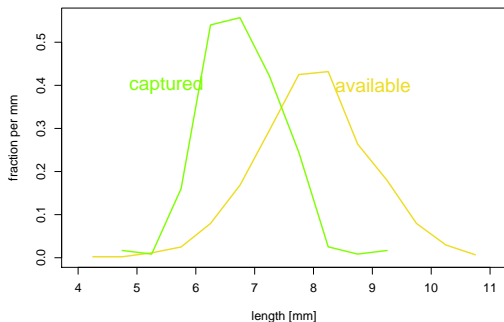
dung flies: available, captured



# Vergleich der Größenverteilungen

|                    | captured |   | available |
|--------------------|----------|---|-----------|
| Mittelwert         | 6.29     | < | 7.99      |
| Standardabweichung |          | < |           |

dung flies: available, captured

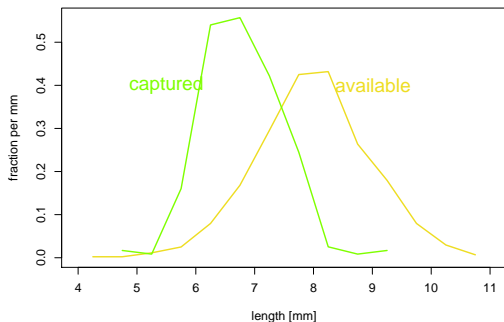




# Vergleich der Größenverteilungen

|                    | captured |   | available |
|--------------------|----------|---|-----------|
| Mittelwert         | 6.29     | < | 7.99      |
| Standardabweichung | 0.69     | < | 0.96      |

dung flies: available, captured



# Interpretation

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

# 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.

- 1 Einführung
  - Konzept und Quellen
  - Plan

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

- 3 Graphische Darstellungen

- Histogramme und Dichtepolygone
- Stripcharts
- Boxplots
- Beispiel: Ringeltaube
- Beispiel: Darwin-Finken

- 4 Statistische Kenngrößen

- Median und andere Quartile
- Mittelwert und Standardabweichung

- 5 Vom Sinn und Unsinn von Mittelwerten

- Beispiel: Wählerische Bachstelzen
- **Beispiel: Spiderman & Spiderwoman**
- Beispiel: Kupfertoleranz beim Roten Straußgras



## *Nephila madagascariensis*

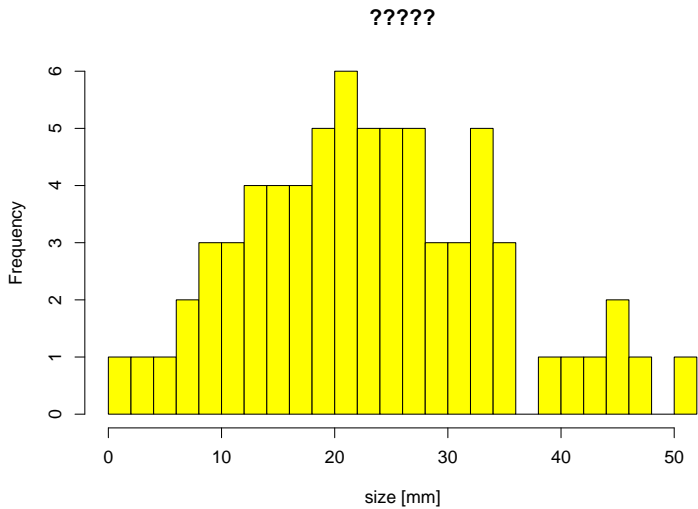
image (c) by Bernard Gagnon

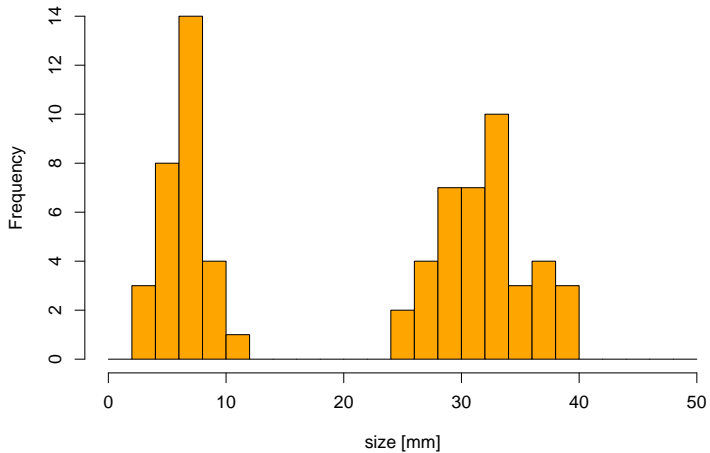
## Simulated Data:

Eine Stichprobe von 70 Spinnen

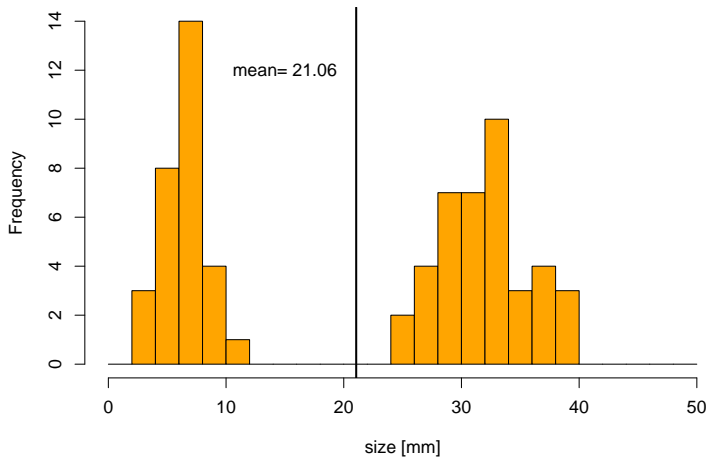
Mittlere Größe: 21,06 mm

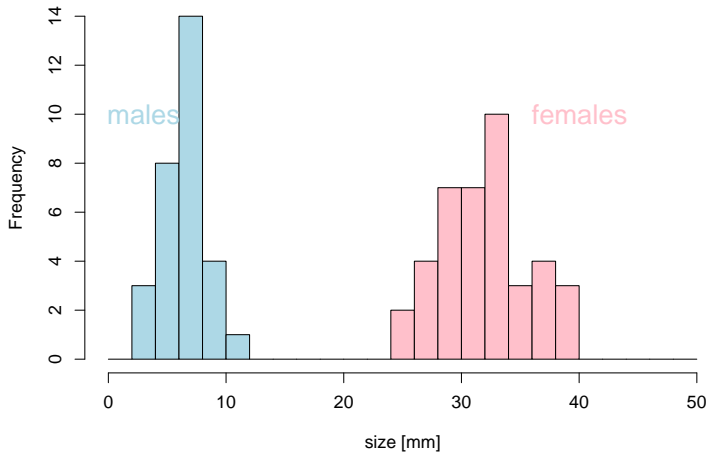
Standardabweichung der Größe: 12,94 mm

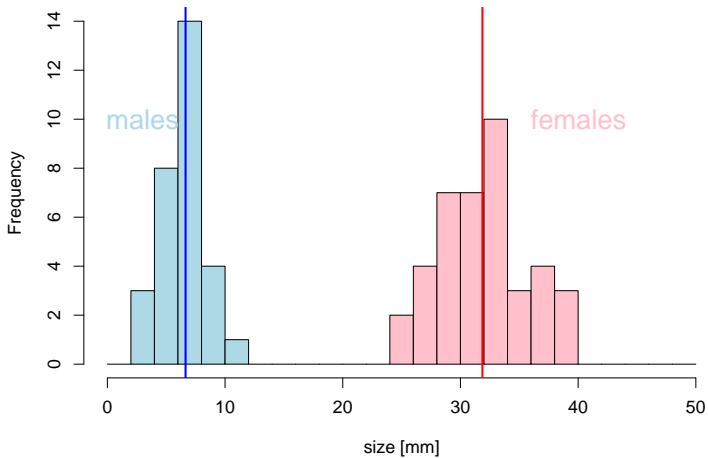


***Nephila madagascariensis* (n=70)**



***Nephila madagascariensis* (n=70)**

***Nephila madagascariensis* (n=70)**

**Nephila madagascariensis (n=70)**



*Nephila madagascariensis*

image (c) by Arthur Chapman

# Fazit des Spinnenbeispiels

Wenn die Daten aus verschiedenen Gruppen zusammengesetzt sind, die sich bezüglich des Merkmals deutlich unterscheiden, kann es sinnvoll sein, Kenngrößen wie den Mittelwert für jede Gruppe einzeln zu berechnen.

- 1 Einführung
  - Konzept und Quellen
  - Plan
- 2 Ziele der deskriptiven (d.h. beschreibenden) Statistik
- 3 Graphische Darstellungen
  - Histogramme und Dichtepolygone
  - Stripcharts
  - Boxplots
  - Beispiel: Ringeltaube
  - Beispiel: Darwin-Finken
- 4 Statistische Kenngrößen
  - Median und andere Quartile
  - Mittelwert und Standardabweichung
- 5 **Vom Sinn und Unsinn von Mittelwerten**
  - Beispiel: Wählerische Bachstelzen
  - Beispiel: Spiderman & Spiderwoman
  - **Beispiel: Kupfertoleranz beim Roten Straußgras**

# Kupfertolerantes Rotes Straußgras



Rotes Straugras  
*Agrostis tenuis*

image (c) Kristian Peters



Kupfer  
*Cuprum*

Hendrick met de Bles



A.D. Bradshaw.

Population Differentiation in *agrostis tenuis* Sibth. III.  
populations in varied environments.

*New Phytologist*, 59(1):92 – 103, 1960.



T. McNeilly and A.D Bradshaw.

Evolutionary Processes in Populations of Copper Tolerant  
*Agrostis tenuis* Sibth.

*Evolution*, 22:108–118, 1968.





A.D. Bradshaw.

Population Differentiation in *agrostis tenuis* Sibth. III.  
populations in varied environments.

*New Phytologist*, 59(1):92 – 103, 1960.



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.

# Anpassung an Kupfer?

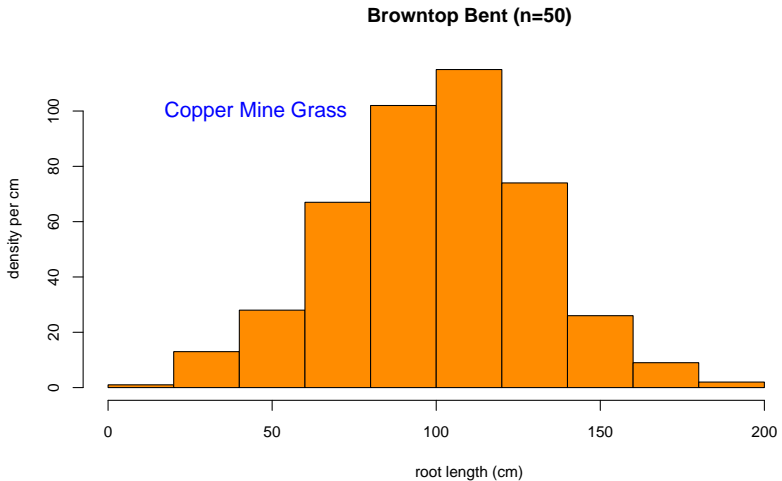
- Pflanzen, denen das Kupfer schadet, haben kürzere Wurzeln.
- Die Wurzellängen von Pflanzen aus der Umgebung von Kupferminen wird gemessen.

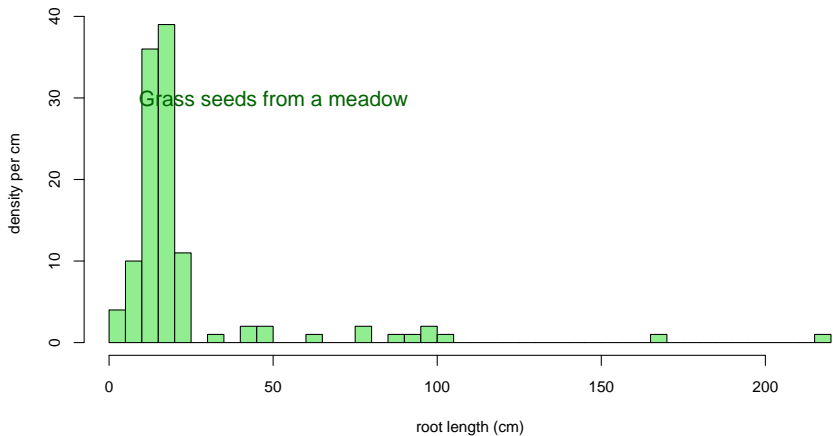
# Anpassung an Kupfer?

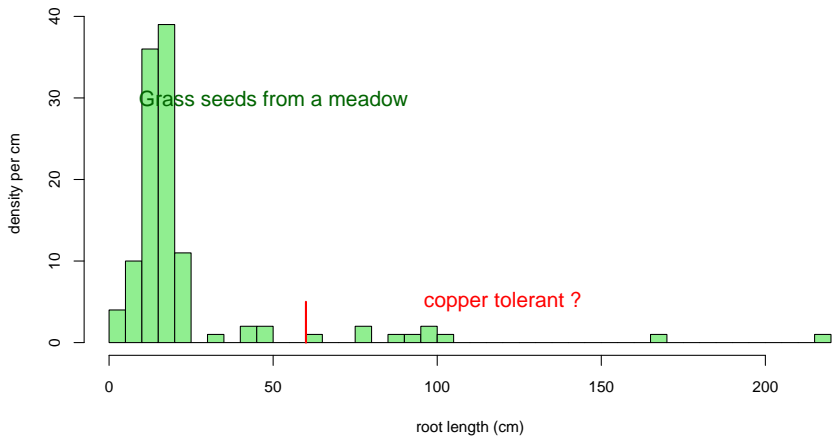
- Pflanzen, denen das Kupfer schadet, haben kürzere Wurzeln.
- Die Wurzellängen von Pflanzen aus der Umgebung von Kupferminen wird gemessen.
- Samen von unbelasteten Wiesen werden bei Kupferminen eingesäht.

# Anpassung an Kupfer?

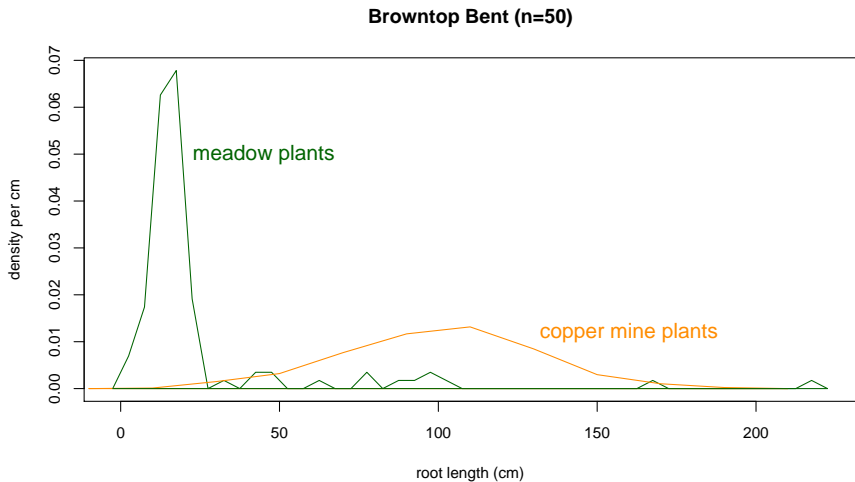
- Pflanzen, denen das Kupfer schadet, haben kürzere Wurzeln.
- Die Wurzellängen von Pflanzen aus der Umgebung von Kupferminen wird gemessen.
- Samen von unbelasteten Wiesen werden bei Kupferminen eingesät.
- Die Wurzellängen dieser “Wiesenpflanzen” werden gemessen.

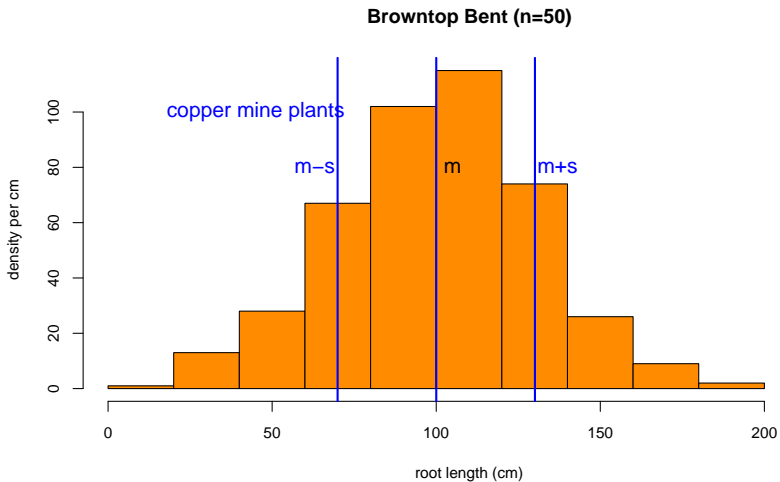


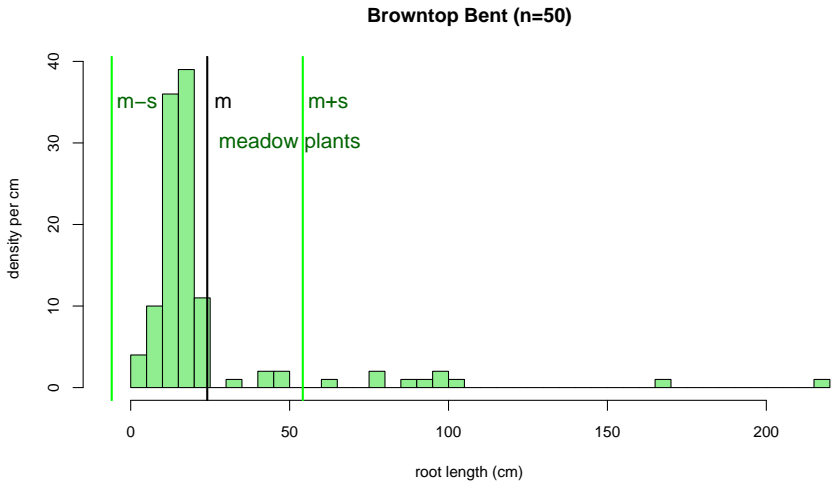
**Browntop Bent (n=50)**

**Browntop Bent (n=50)**









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.

# 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.

# 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!



# 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!