Etwas Admin
Häufigkeitsverteilung (Frequency distributions)
Lagemaße
Modelle der Mitte und Verteilung
Viele, viele Graphen und ein Datenbeispiel

VO Statistik

Sitzung 3: Gut verteilt: Deskriptive Statistik II

Dominik Duell

Universität Innsbruck

Etwas Admin

Material online

- Stellt sicher, dass ihr die Folien, das Vorlesungsskript und die Lehrbücher findet
- In Sitzung 4, nächste Woche erstes Pop Quiz

Plan für heute

1. Beschreibung unserer Daten in Form von

a. Häufigkeitsverteilungen
b. Lagemaße

2. Nachdenken über Statistisches Modelle

3. Ein Datenbeispiel: Emissionen

Häufigkeitsverteilung (Frequency distributions)

Häufigkeitsverteilung (Frequency distributions)

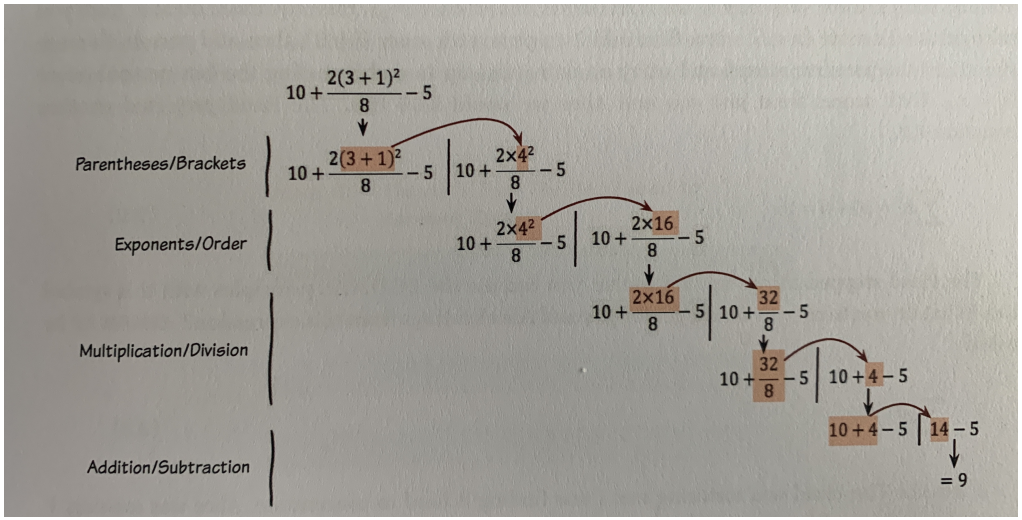
High salary	Finished education	Kind	Humour	Ambitious
4	9	10	8	7
5	8	10	10	7
9	5	8	8	8
4	4	9	8	5
5	7	10	10	8
3	3	10	8	7
10	10	10	10	10
10	7	9	9	6
6	6	8	7	7
1	4	7	6	7
3	4	8	6	4
8	8	9	8	6
8	9	9	7	5
2	1	10	8	9
6	7	9	7	7
4	3	9	8	6
8	7	9	5	7
4	6	10	9	8
7	10	10	10	5
7	9	10	9	6

Quelle: Fields, S.73, Table 3.1

Rating (X)	Frequency (f)
10	2
9	1
8	3
7	2
6	2
5	2
4	4
3	2
2	1
1	1
	$N = \sum f = 20$

Quelle: Fields, S.74, Table 3.2

Exkurs: Notation



Quelle: Fields, S.45, Fig 2.1

$\rightarrow$, dazu $\sum_{i=1}^{N}$ und $\prod_{i=1}^{N}$

Absolute Häufigkeit

Definition: Bei einer Stichprobe vom Umfang N wird ausgezählt, wie häufig jeder der k Ausprägungen auftritt. Diese Anzahl bezeichnet man als **absolute Häufigkeit** einer Ausprägung n_j (wobei j für eine bestimmte der k Ausprägungen steht).

$$\sum_{j=1}^{k} n_j = N$$

Die Summe der absoluten Häufigkeit (frequency) von allen k Ausprägungen ergibt die Stichprobengröße N. Quelle: Sibbertsen/Lehne, S.13/14

Sex (X)	Frequency (f)
Female	20
Male	0
Non-binary	0

Quelle: Fields, S.75, Table 3.3

Rating (*X*)	Frequency (*f*)	Cumulative frequency (up)	Relative frequency (*f*/*N*)	Percentage	Cumulative percentage (up)	Cumulative percentage (down)
10	2	20	0.10	10%	100%	10%
9	1	18	0.05	5%	90%	15%
8	3	17	0.15	15%	85%	30%
7	2	14	0.10	10%	70%	40%
6	2	12	0.10	10%	60%	50%
5	2	10	0.10	10%	50%	60%
4	4	8	0.20	20%	40%	80%
3	2	4	0.10	10%	20%	90%
2	1	2	0.05	5%	10%	95%
1	1	1	0.05	5%	5%	100%

Quelle: Fields, S.78, Table 3.4

Relative Häufigkeit

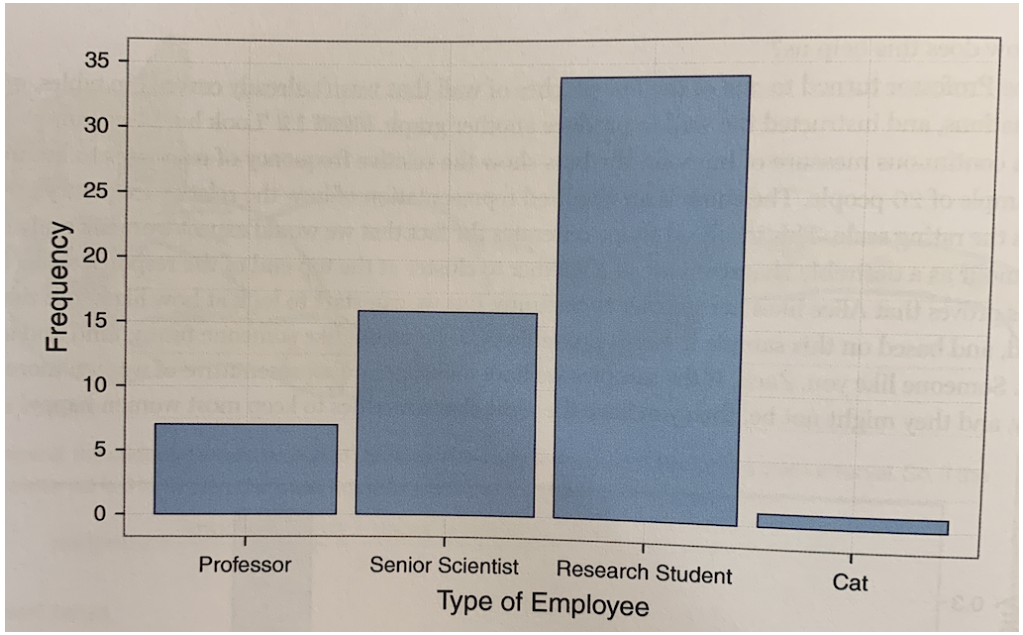
Definition: Die absolute Häufigkeit bezogen auf den Stichprobenumfang

$$\frac{n_j}{N}$$

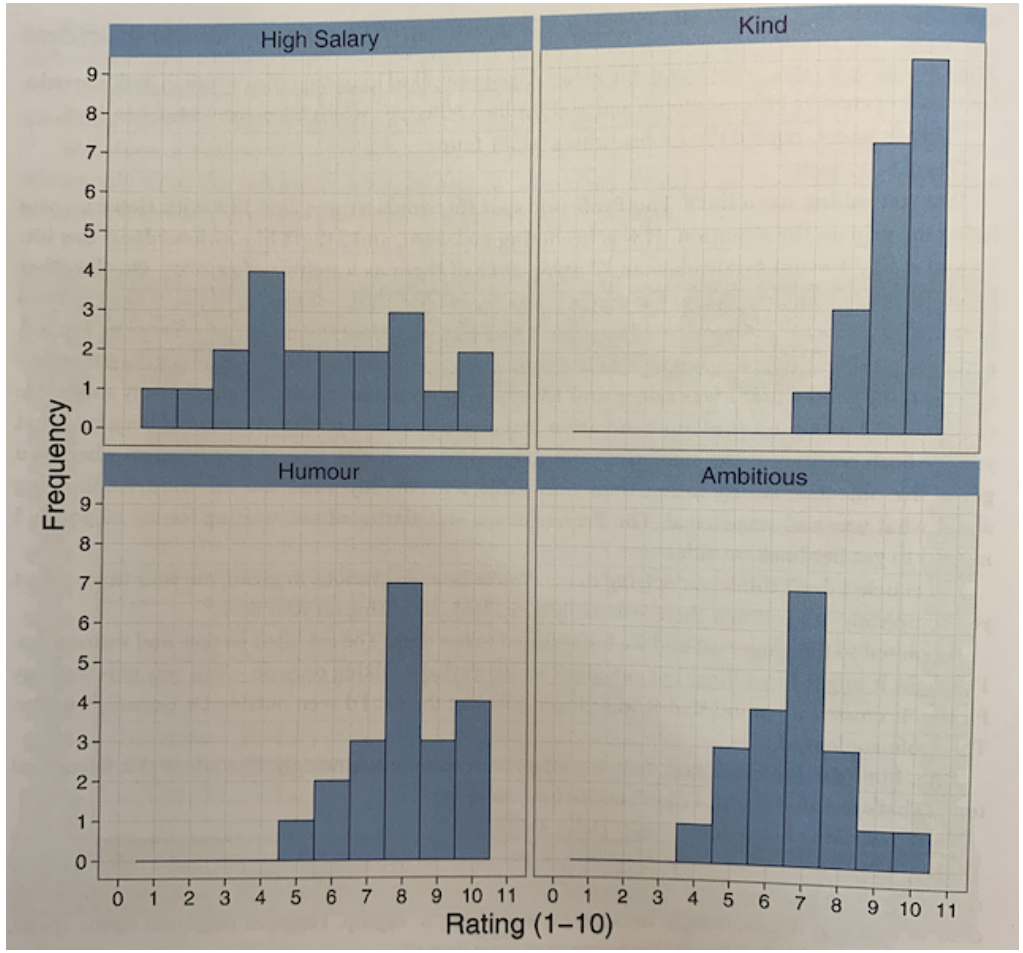
ist die **relative Häufigkeit** (relative frequency) einer Ausprägung.
Die Summe der relativen Häufigkeiten von allen k Ausprägungen ergibt 1.

$$\sum_{j=1}^{k} \frac{n_j}{N} = 1$$

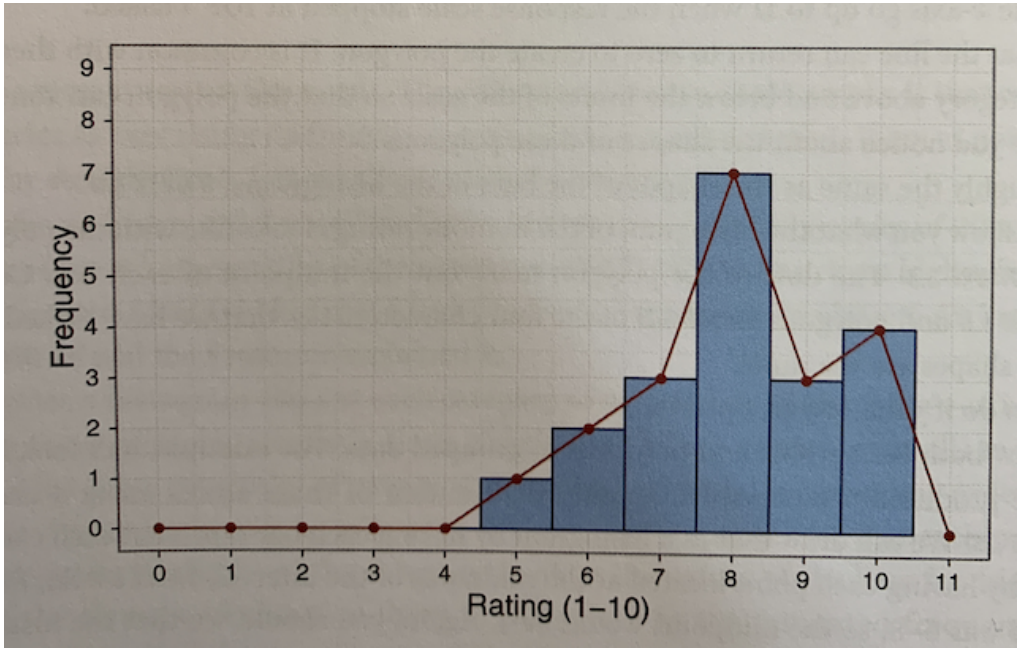
Quelle: Sibbertsen/Lehne, S.13/14



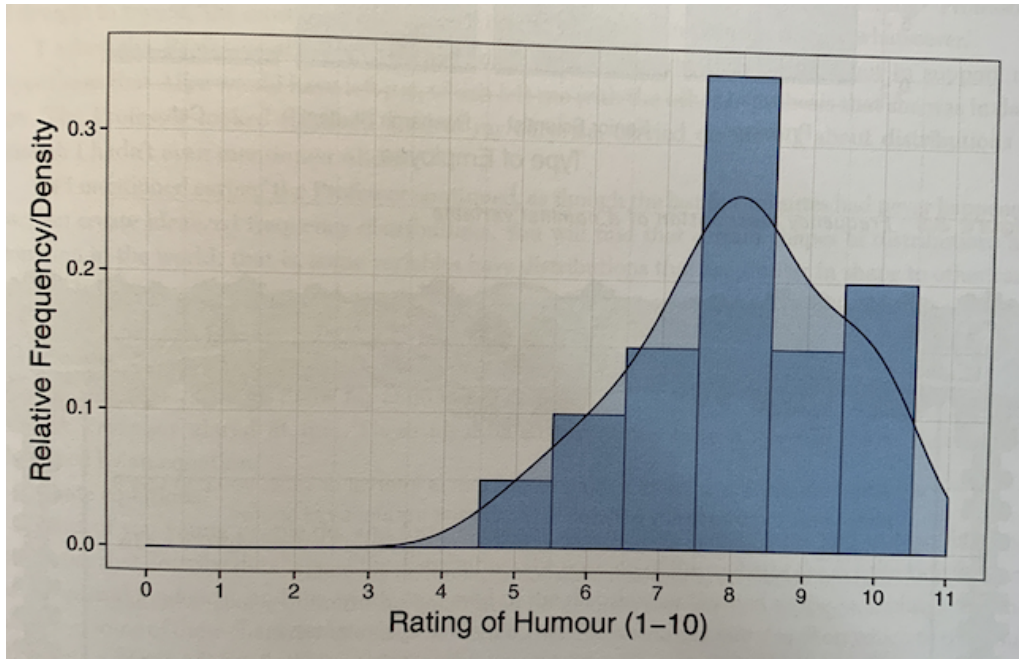
Quelle: Fields, S.81, Table 3.6



Quelle: Fields, S.84, Figure 3.1

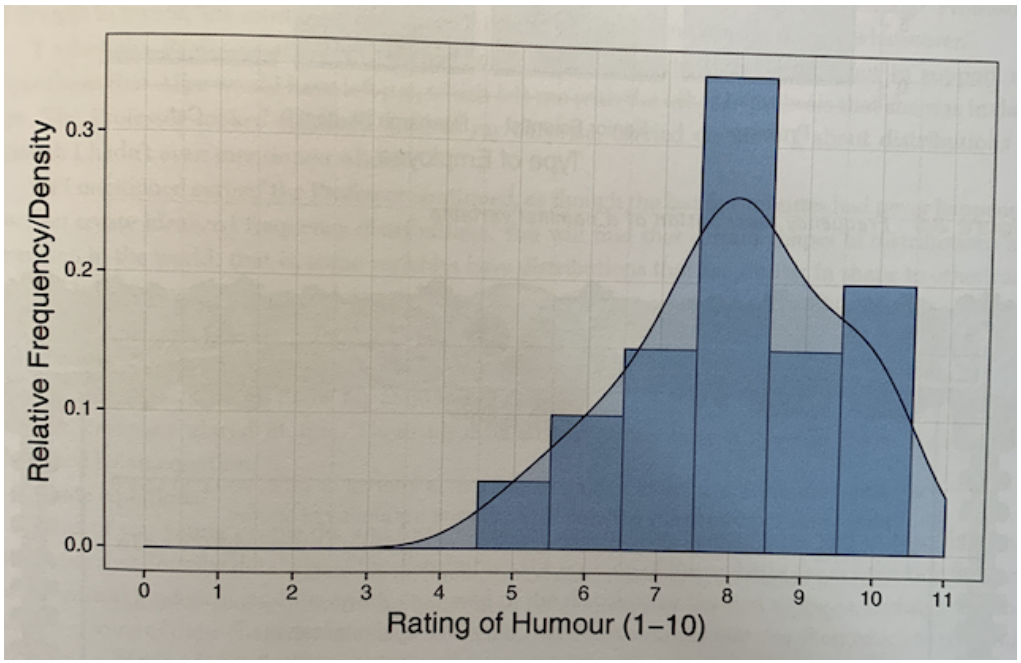


Quelle: Fields, S.86, Figure 3.3



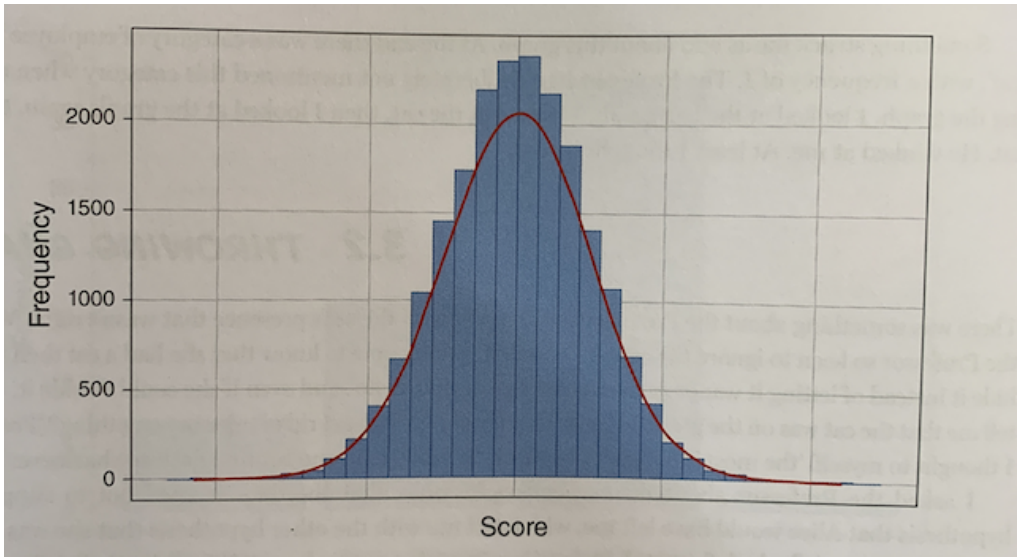
Quelle: Fields, S.89, Figure 3.5

Graphische Darstellung von Häufigkeit



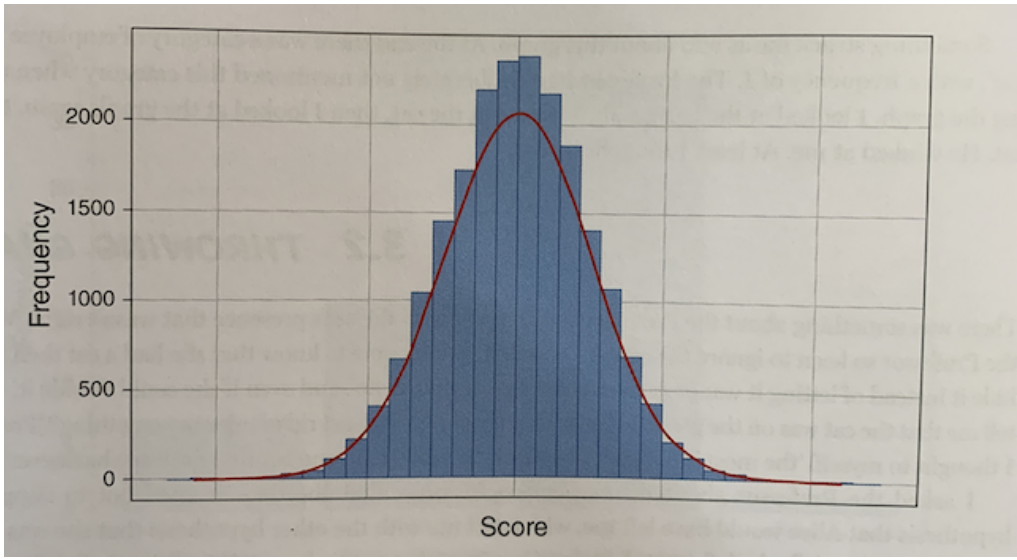
Quelle: Fields, S.90, Figure 3.6

Wichtige Verteilungen



Quelle: Fields, S.92, Figure 3.7

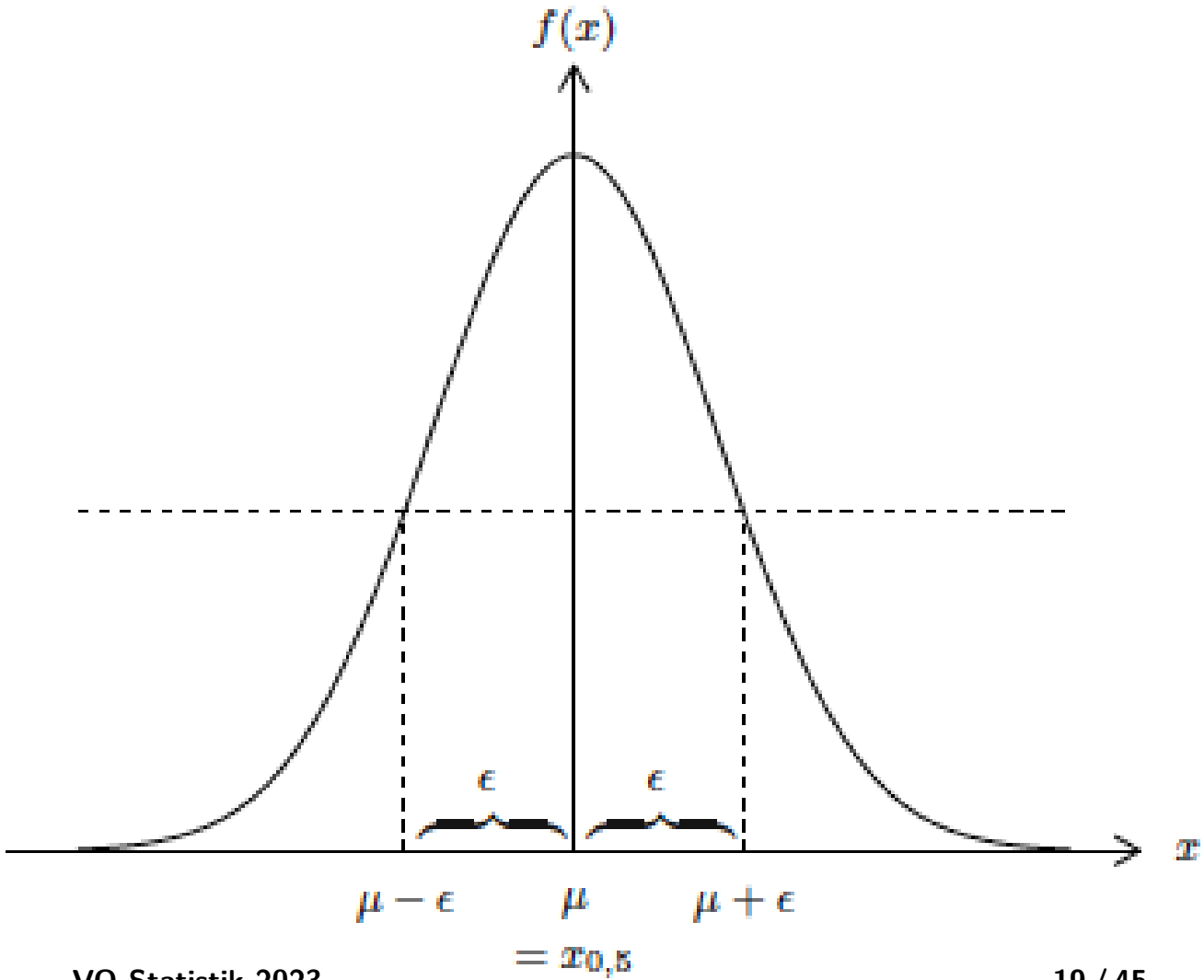
Wichtige Verteilungen



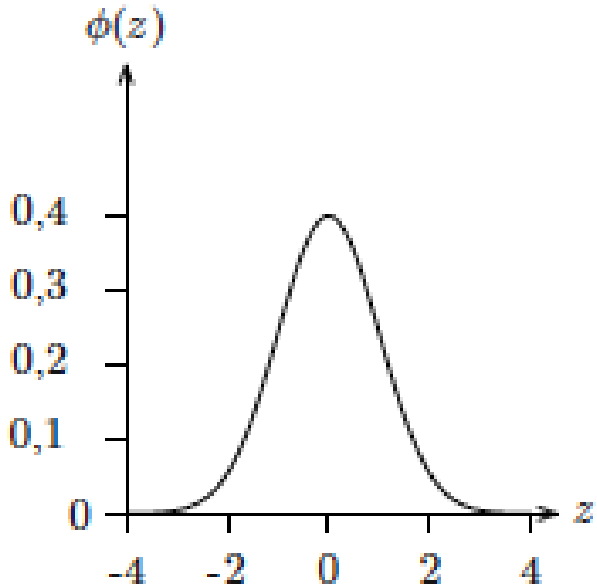
Quelle: Fields, S.92, Figure 3.7

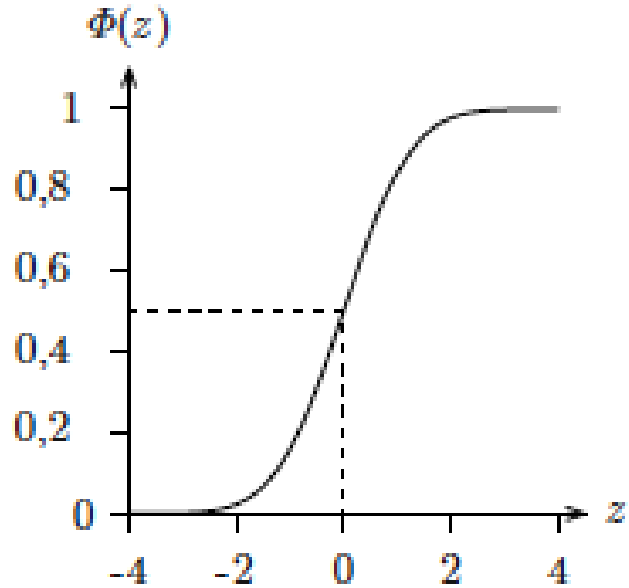
$\rightarrow$ Annäherung an die **Normalverteilung**

Normalverteilung



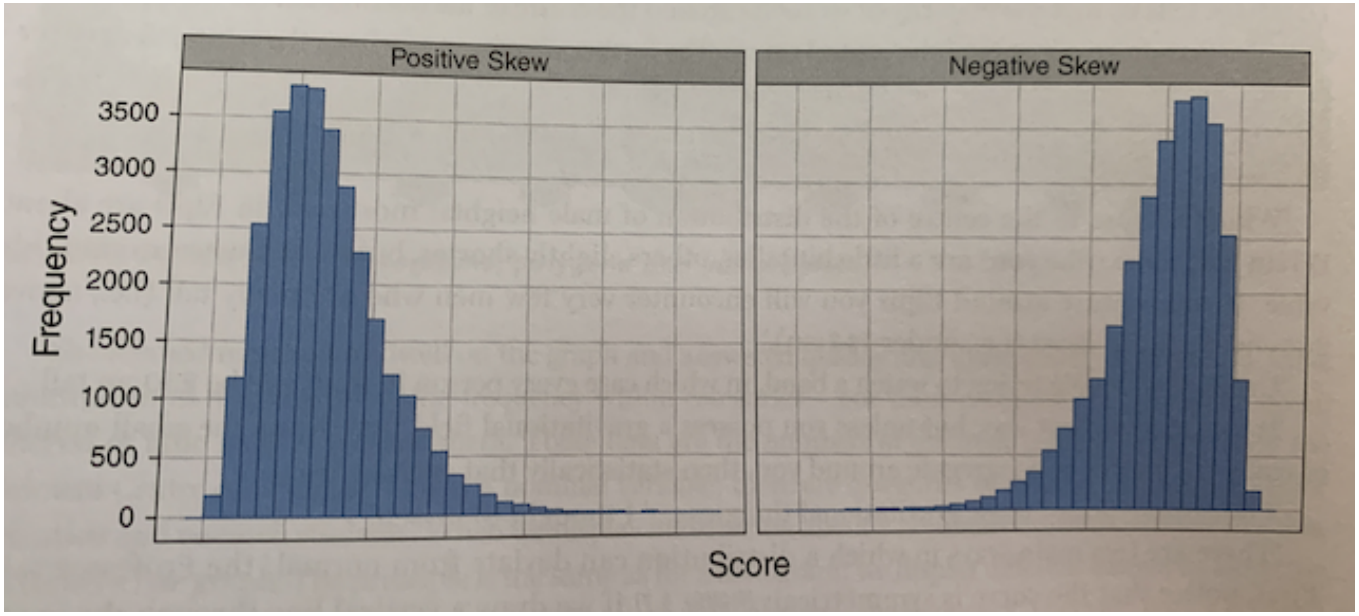
Standardnormalverteilung





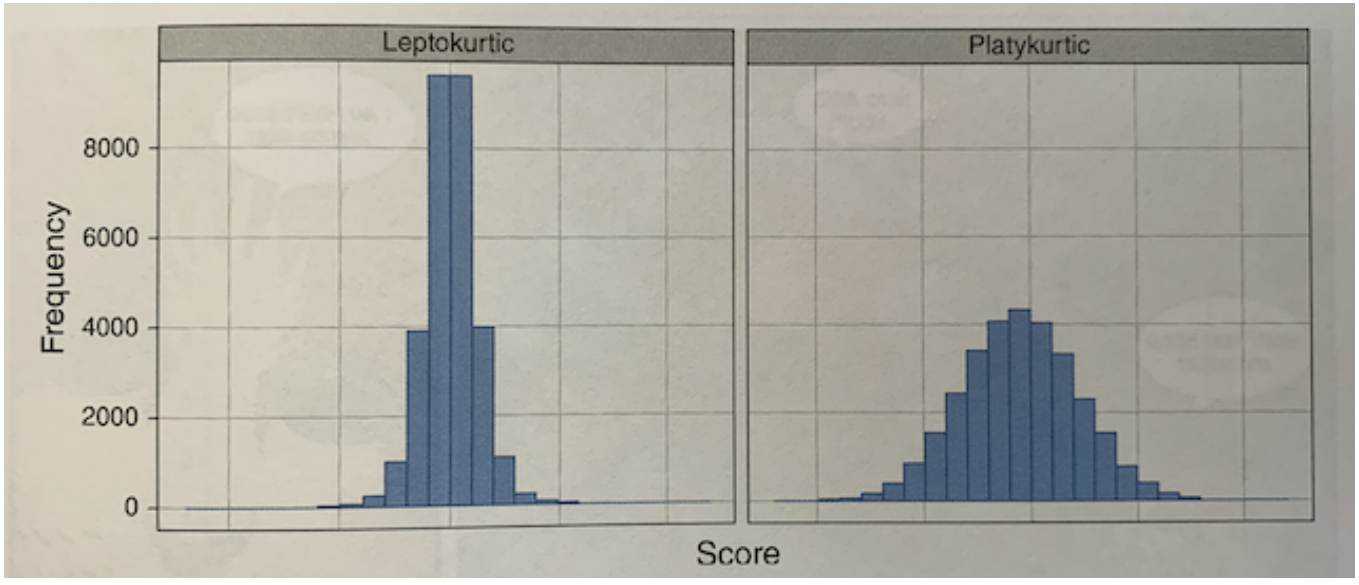
Quelle: Sibbertsen/Lehne, S.259

Schiefmaße: links- und rechtssteile Verteilung (skew)



Quelle: Fields, S.92, Figure 3.8

Wölbungsmaße (kurtosis)



Quelle: Fields, S.93, Figure 3.9

→ Sitzungsnotizen

Lagemaße

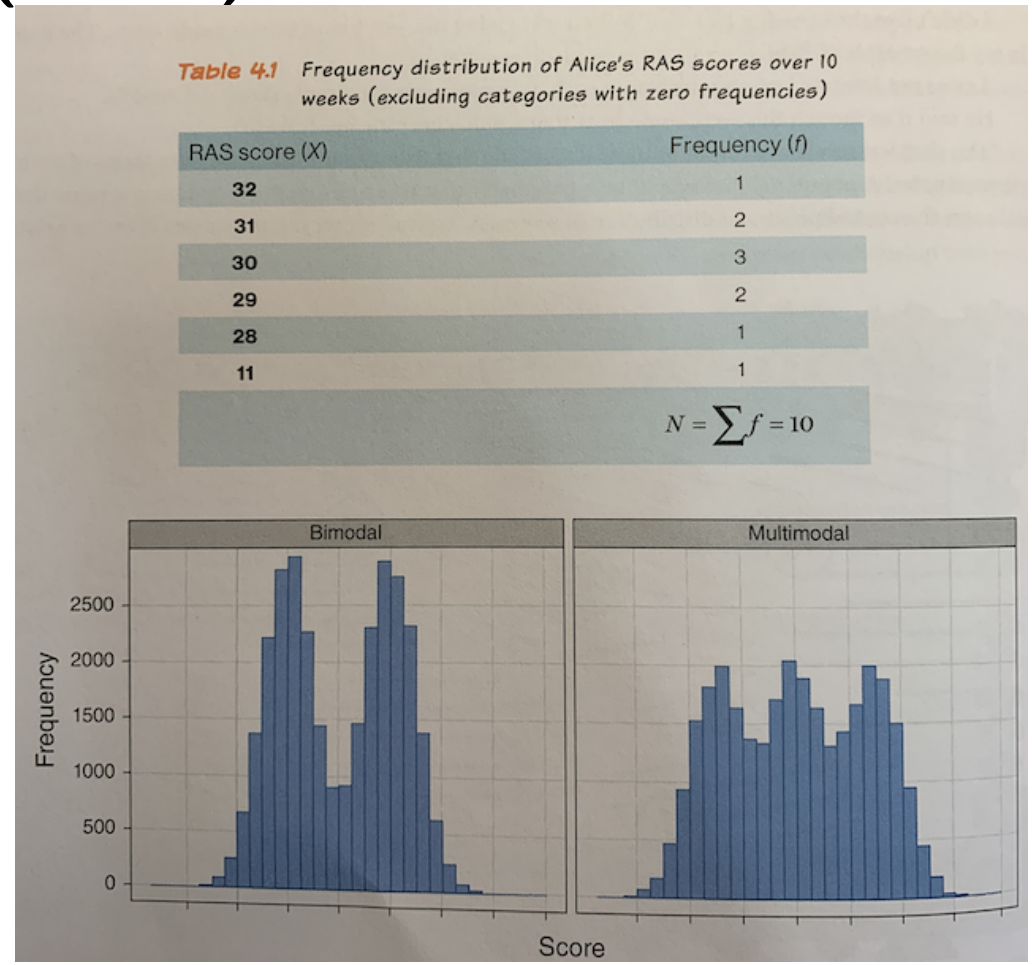
Modalwert (mode)

Definition: Die Ausprägung, die am häufigsten im Datensatz enthalten ist, bezeichnet man als **Modus**.
Quelle: Sibbertsen/Lehne, S.54

Modalwert (mode)

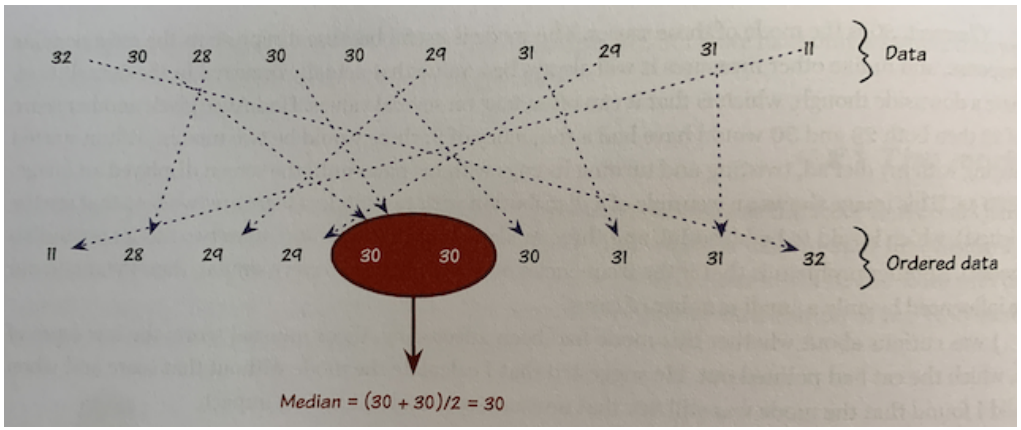
Table 4.1 Frequency distribution of Alice's RAS scores over 10 weeks (excluding categories with zero frequencies)

RAS score (X)	Frequency (f)
32	1
31	2
30	3
29	2
28	1
11	1
	$N = \sum f = 10$

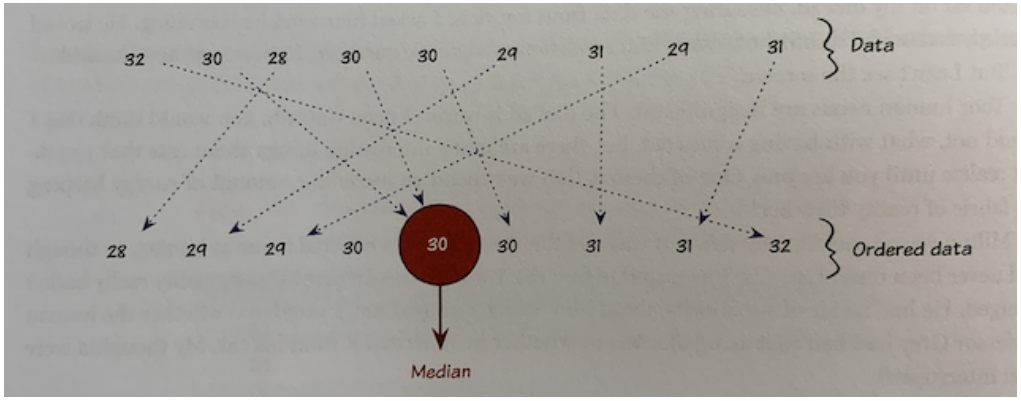


Quelle: Fields, S.110, Figure 4.4

Median



Quelle: Fields, S.112, Figure 4.5



Quelle: Fields, S.112, Figure 4.6

Arithmetische Mittel (mean)

Definition:

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

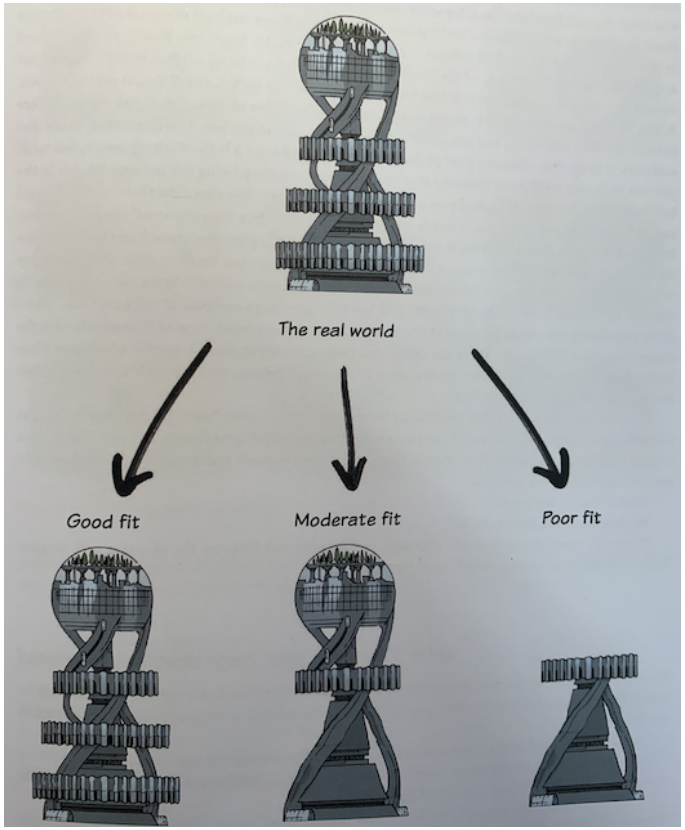
Quelle: Sibbertsen/Lehne, S.54

Modelle der Mitte und Verteilung

Statistische Modelle

→ Sitzungsnotizen

Statistische Modelle



Quelle: Fields, S.103, Figure 4.1

Grundstruktur eines statistischen Modells

Ausprägung in der Grundgesamtheit = Modell + Fehler

$$\text{outcome}_i = \textit{model} + \textit{error}_i$$

Grundstruktur eines statistischen Modells

Ausprägung in der Grundgesamtheit = Modell + Fehler

$$\text{outcome}_i = \textit{model} + \textit{error}_i$$

Was ist das richtige Modell, welches unsere Daten am besten beschreibt und sich der Ausprägung in der Grundgesamtheit am besten annähert?

Grundstruktur eines statistischen Modells

Ausprägung in der Grundgesamtheit = Modell + Fehler

$$\text{outcome}_i = \textit{model} + \textit{error}_i$$

Was ist das richtige Modell, welches unsere Daten am besten beschreibt und sich der Ausprägung in der Grundgesamtheit am besten annähert?

Fangen wir mal einem wirklich guten Modell der mittleren Lage an!

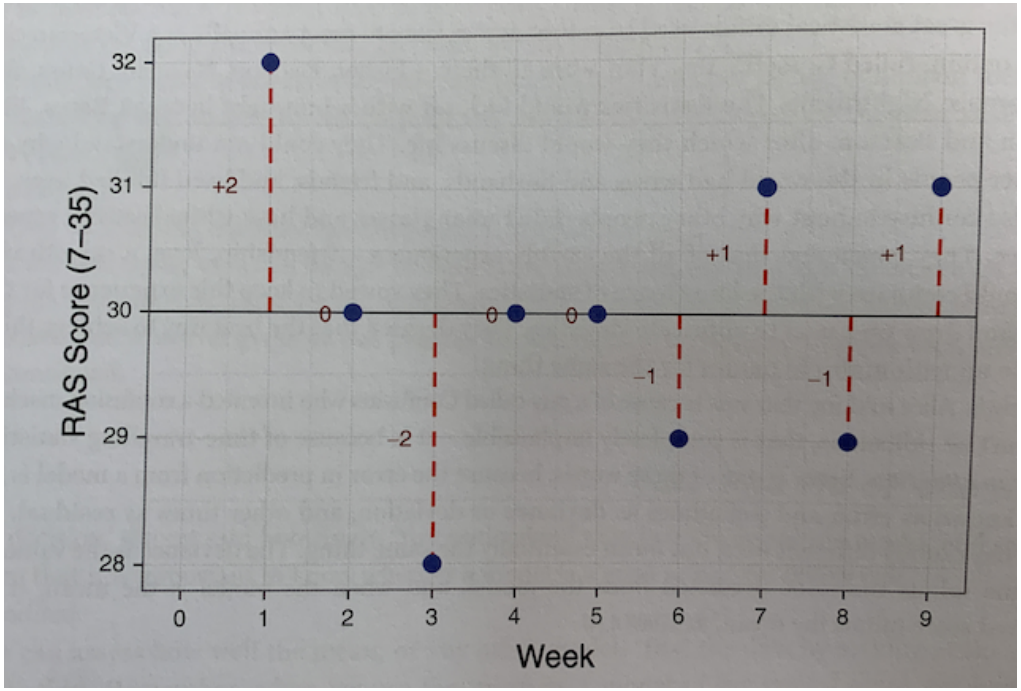
Grundstruktur eines statistischen Modells

Ausprägung in der Grundgesamtheit = Modell + Fehler

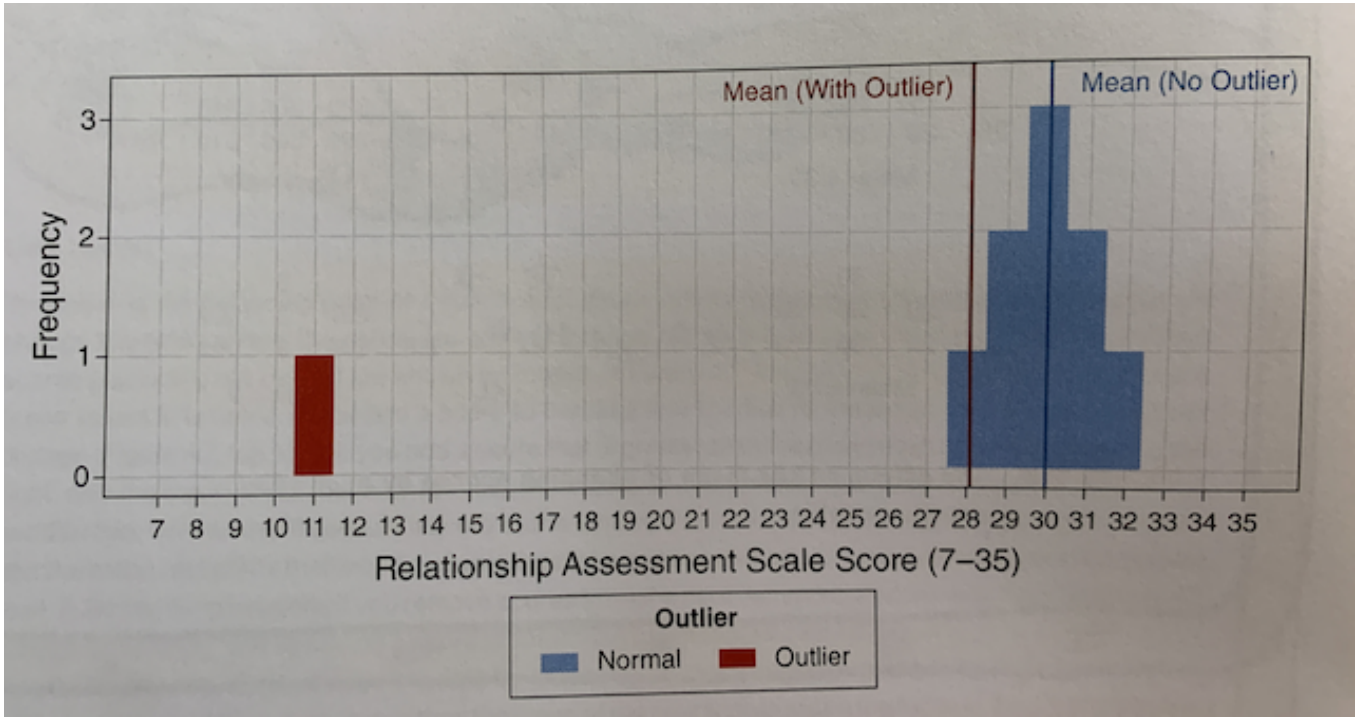
$$\text{outcome}_i = \textit{model} + \textit{error}_i$$

Was ist das richtige Modell, welches unsere Daten am besten beschreibt und sich der Ausprägung in der Grundgesamtheit am besten annähert?

Fangen wir mal einem wirklich guten Modell der mittleren Lage an! Kennt ihr alle.

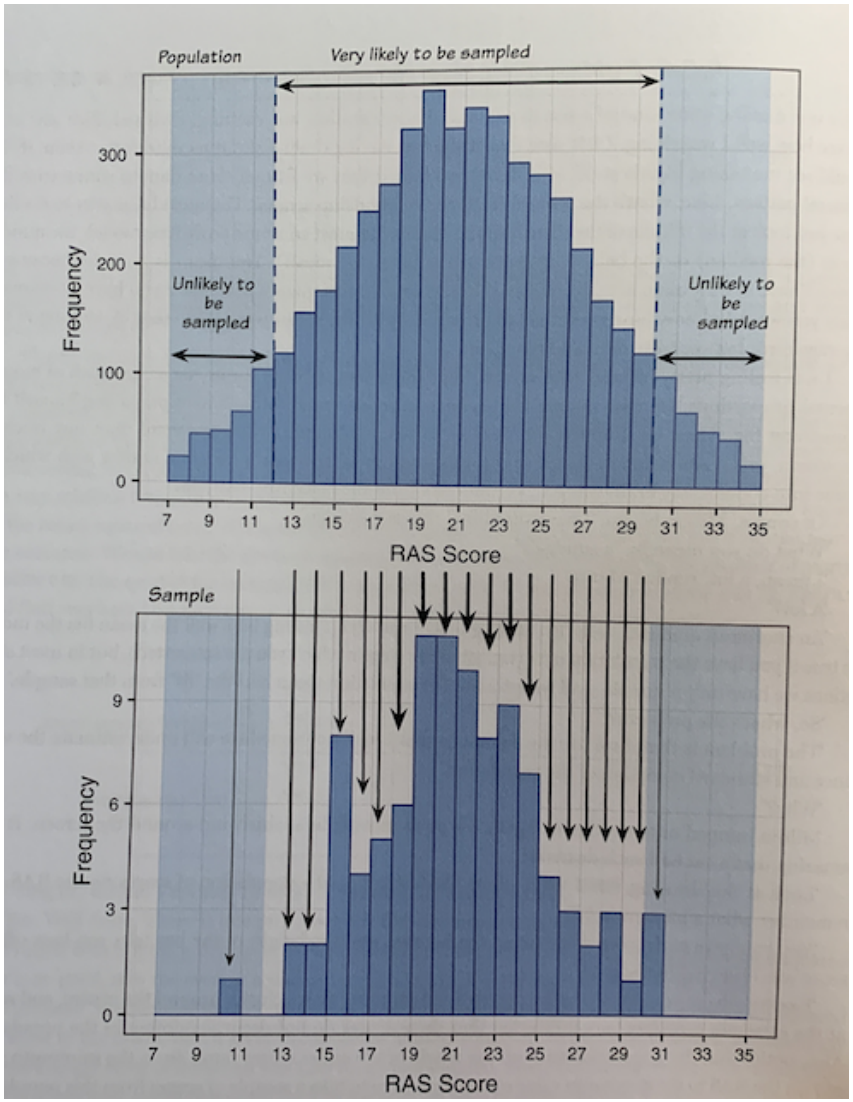


Quelle: Fields, S.120, Figure 4.10

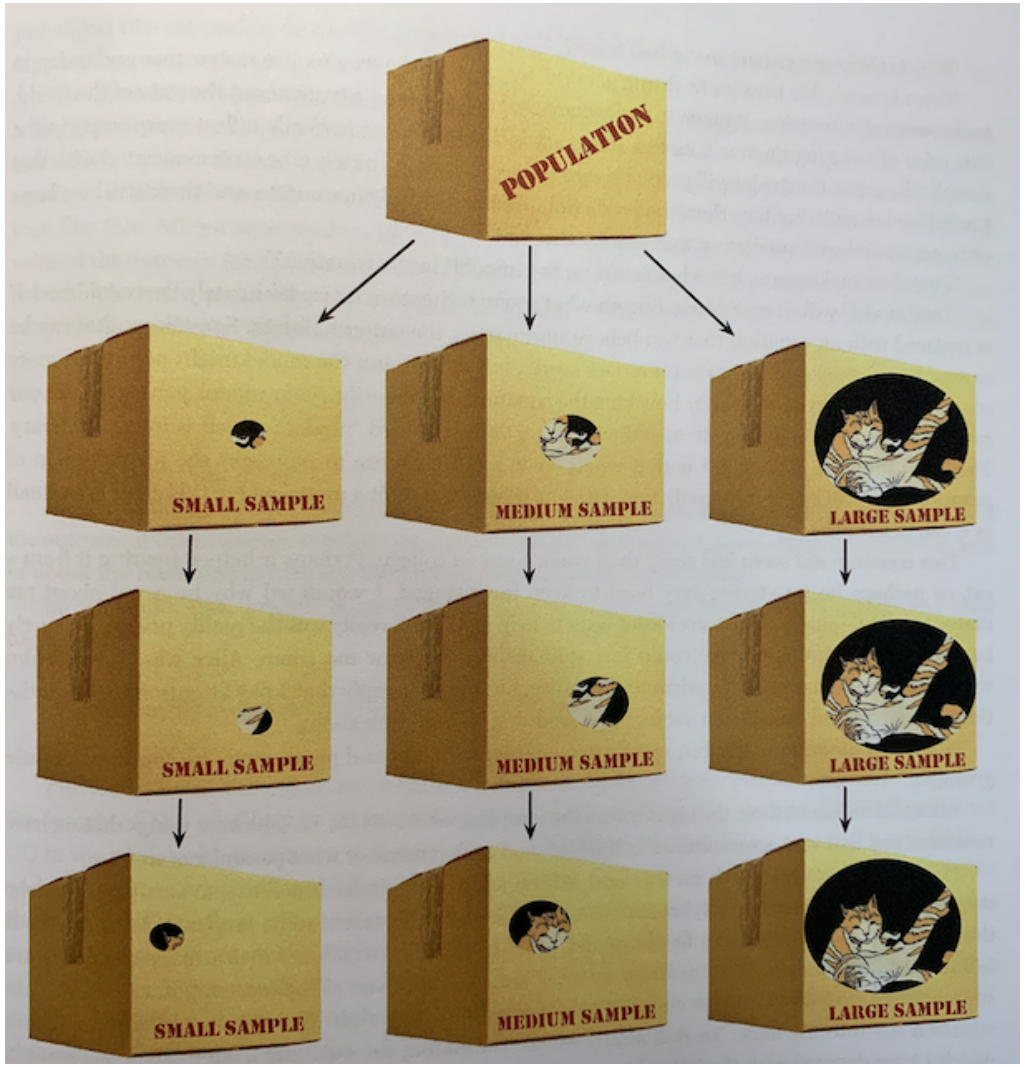


Quelle: Fields, S.116, Figure 4.9

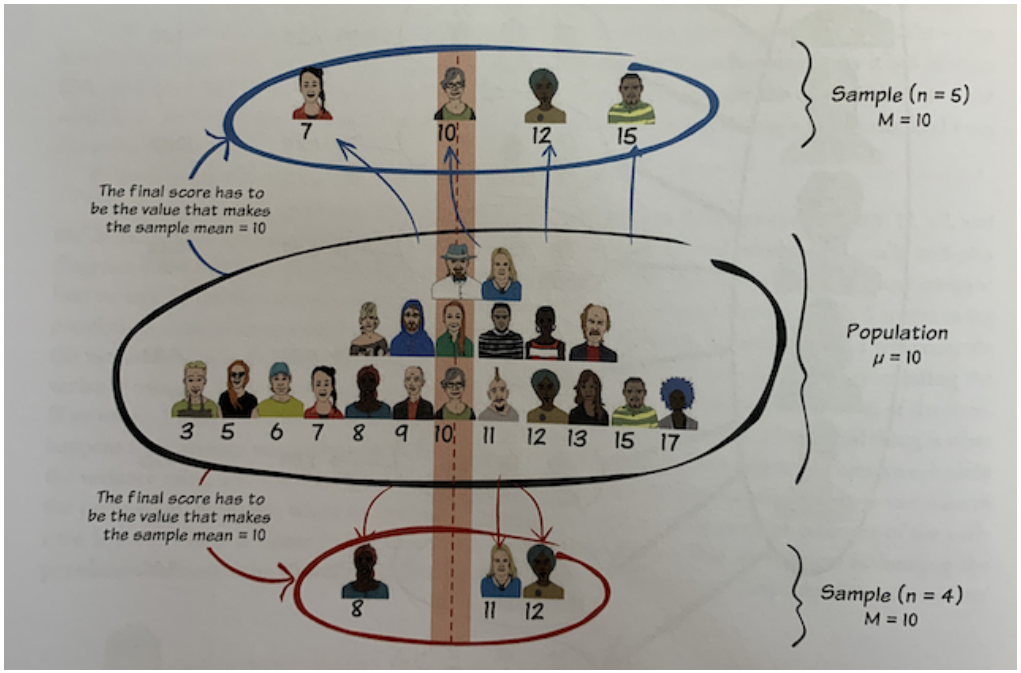
$$\text{outcome}_i = \text{Sample mean} + \text{error}_i$$



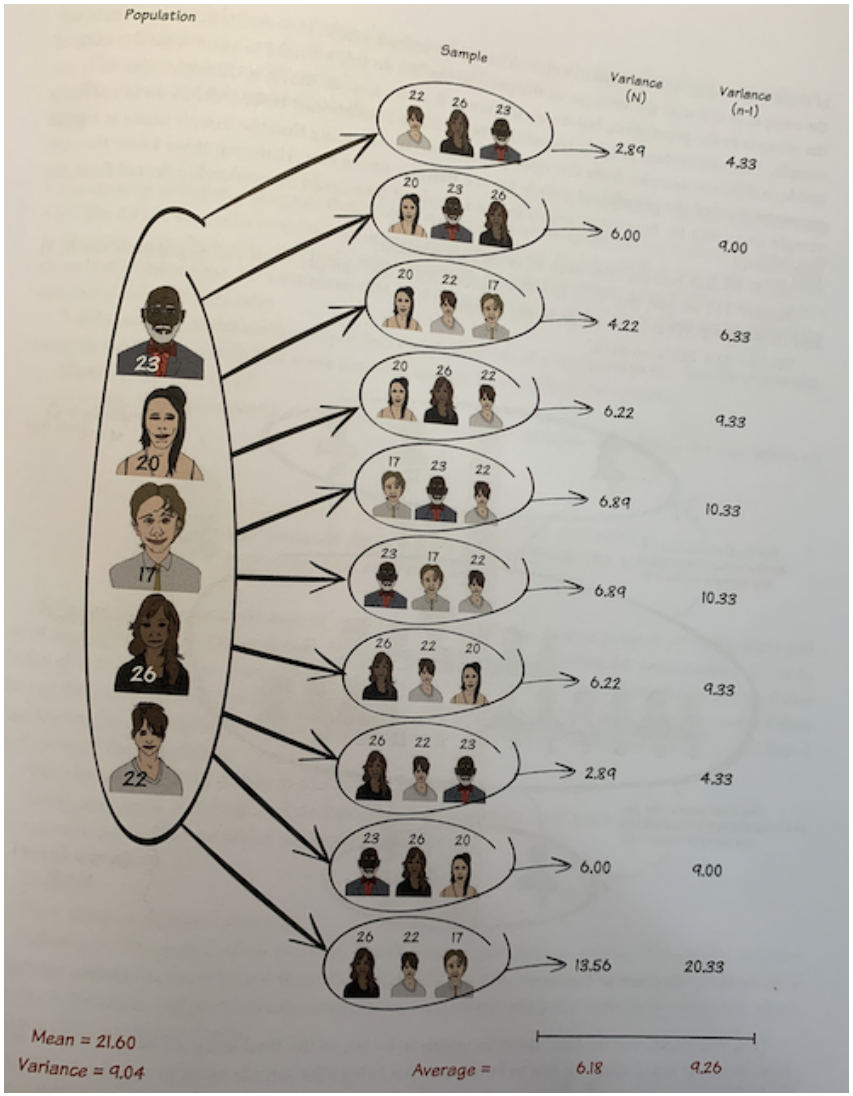
Quelle: Fields, S.124, Figure 4.11



Quelle: Fields, S.105, Figure 4.2



Quelle: Fields, S.128, Figure 4.13



Quelle: Fields, S.127, Figure 4.12

Viele, viele Graphen und ein Datenbeispiel

Wie visualisiert man welche Statistiken?

$\rightarrow$ Viele, viele Graphen

Datenbeispiel: Wer ist der Saubär?

→ Sitzungsnotizen

Was solltet ihr von heute mitnehmen?

1. Verstehen, wie die in einer Stichprobe erhobenen Ausprägungen einer Variable eine Häufigkeitsverteilung ergeben.
2. Wissen, welche wichtigen Lagemaße existieren und für welche Arten von Merkmale diese aussagekräftig sind.
3. Erklären können, wie das ziehen eine Stichprobe zur empirischen Häufigkeitsverteilung führt und wann wir aus einer solchen Verteilung lernen können.
4. Wissen warum wir von einem statistischen Modell sprechen, wenn wir Stichprobenstatistiken berechnen (schätzen).
5. Ideen haben, wie man Daten am besten visuell beschreibt.

Was gibt's nächste Woche

1. Streuungsmaße (Varianz, Dispersion, Quantilabstand etc.)
2. Modelle statistischer Zusammenhänge