



# Einführung in die Evaluierung von Unterricht und Projektzielen

Emmerich Boxhofer

## Definition

Evaluation oder Evaluierung<sup>1</sup> bedeutet sach- und fachgerechte Untersuchung und Bewertung.  
(Quelle: wikipedia)

Als pädagogischer bzw. andragogischer<sup>2</sup> Fachbegriff bedeutet Evaluation „das methodische Erfassen und das begründete Bewerten von Prozessen und Ergebnissen zum besseren Verstehen und Gestalten einer Praxismaßnahme im Bildungsbereich durch Wirkungskontrolle, Steuerung und Reflexion.“

## Gütekriterien

- **Objektivität:** Das Messergebnis einer Variablen ist unabhängig vom Beobachter.
- **Zuverlässigkeit:** Wiederholte Messungen führen zu identischen Ergebnissen.
- **Validität:** Die Variablen sind eine gute Operationalisierung der in der Theorie verwendeten Größe.

Andreas Guschka (deutscher Pädagoge):

- **Kommunikation**
- **Intervention**
- **Transparenz**
- **Relevanz**

---

<sup>1</sup> aus lateinisch valere: „stark, wert sein“

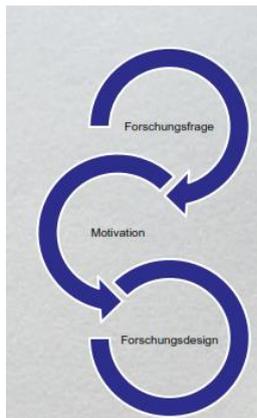
<sup>2</sup> Wissenschaft - Verstehen und Gestalten der lebenslangen Bildung

## Was ist Forschung?

- Methodische Erkenntnissuche
- Unter der Verwendung einer systematischen Vorgehensweise.
- Neue Erkenntnisse müssen **nachvollziehbar** und **nachprüfbar** sein.
  
- **Das Ziel:** Wissenszuwachs (auch wenn er klein ist)

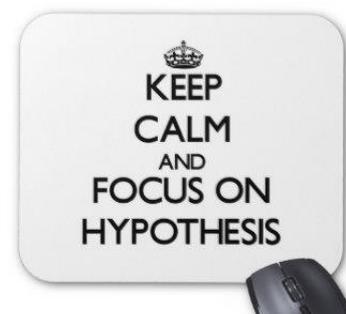
## Die 3 „W“

Was? Warum? Wie?



## Hypothese, was ist das?

Gute Forschung lebt von guten Forschungsfragen!  
Aus diesen Fragestellungen lassen sich Hypothesen bilden.  
Entscheidende Schritte eines Forschungsprozesses sind das Finden und Formulieren der Forschungshypothesen.  
Hypothesen sind Behauptungen, die durch wissenschaftliche Untersuchungen bewiesen oder widerlegt werden können.





Hypothesen sind so zu formulieren dass sie

- widerspruchsfrei sind (Eine Person kann nicht größer als 180 cm **und** kleiner als 120 cm sein
- an der Realität scheitern können (Es müssen Ereignisse vorstellbar sein, die die Hypothese falsifizieren<sup>3</sup> können.)
- operationalisierbar sind (Die Fragestellungen müssen überprüfbar sein, bzw. man muss beobachten können, was gefragt ist.)

### **Begriffe**

*Sachhypothesen* orientieren sich an Erfahrungen. („Je höher die Gewissenhaftigkeit, desto größer die Wahrscheinlichkeit, dass...“)

*Nullhypothese*  $H_0$  = „Es gibt keinen Zusammenhang zwischen ... und ...“

*Alternativhypothese*  $H_A$

zweiseitig: Es gibt einen Zusammenhang zwischen...

einseitig: je größer ..., desto ... (entspricht einer Sachhypothese)

*Zusammenhangshypothese*

Aussage über Zusammenhänge zwischen den Merkmalsausprägungen zweier (oder mehrerer) Variablen. Beispiel: Die Länge der Haare bei einem Burschen korreliert mit der Wahrscheinlichkeit, mit einem Mädchen verwechselt zu werden.

*Unterschiedshypothese*

Aussage über einen Unterschied zwischen zwei (oder mehreren) Gruppen hinsichtlich einer (oder mehrerer) abhängiger Variablen.

Beispiel: *Alkoholisiertere unterscheiden sich von Nichtalkoholisierten hinsichtlich der Kompetenz, ein Auto unfallfrei steuern zu können.*

*Veränderungshypothese*

Aussage über einen Unterschied zwischen einer Messung und einer zweiten, zeitlich späteren Messung. Beispiel: Ein forschendes Lehrkonzept erhöht die Methodenkompetenz von Lehrer/-innen.

### **Procedere**

Datenoperationalisierung - Datengewinnung - Datenmanagement (Datenverarbeitung)

---

<sup>3</sup> **Falsifikation**, auch **Falsifizierung** (von lat. *falsificare* „als falsch erkennen“) oder **Widerlegung**, ist der Nachweis der Ungültigkeit einer Aussage, Methode, These, Hypothese oder Theorie. Aussagen oder experimentelle Ergebnisse, die Ungültigkeit nachweisen können, heißen „Falsifikatoren“. (Quelle wikipedia)

## Fragebogenkonstruktion und Items

Ziel eines Fragebogens ist es, zu einer oder mehreren Merkmalsdimensionen (Kategorien, Dimension, ...) empirische Daten zu erheben.

Das Item ist eine einzelne, nicht weiter zerlegbare, statistische Variable.

Items sind konkrete Fragestellungen oder Behauptungen, die helfen sollen, eine Hypothese zu untersuchen.

Ratingskala: Der Abstand zwischen den Bereichen ist gleich groß

bipolare Fragestellung: es gibt zwei gegensätzliche Ausprägungen.

unipolare Fragestellungen: es gibt einen Nullpunkt. (vgl. Gehaltshöhe, „nie“, ...)

3 – 7 Kategorien, darüber Überforderung.

Problem der Vieldeutigkeit der „Mittelkategorie“ (Weiß ich es nicht oder bin ich neutral).

Gerade Anzahl daher sinnvoll, allerdings gibt es Nachteil, wenn Aussage wirklich nicht entscheidbar ist. Hier hilft eine Antwortkategorie: „Ich weiß nicht“.

Günstig sind 5 Kategorien.

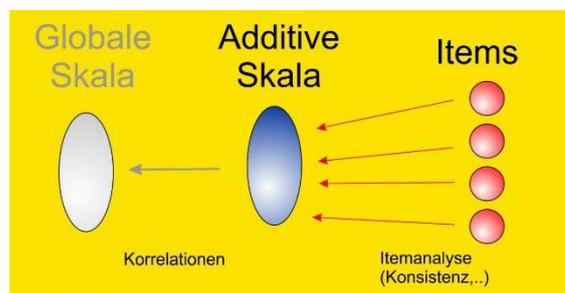
Kreuzen viele Probanden häufig zwischen den Kategorien an, deutet das auf schlechte Itemformulierung hin.

- Problematik bei „verneinten“ Fragestellungen – allerdings steigert sich die Konzentration (Ich bin mir nicht sicher, ob ich einen Parkschein gelöst habe?)
- keine Begriffe mit mehreren Bedeutungen verwenden
- Nur bekannte Formulierungen verwenden
- Je Item nur eine Frage verwenden
- keine doppelten Verneinungen (Ich bin mir nicht sicher, ob ich keinen Parkschein gelöst habe?)
- keine Kürzel verwenden
- eindeutige Zeitpunkte angeben, wenn zeitliche Relevanz vorhanden ist. (In Ihrer Kindheit waren Sie viel in der frischen Luft – besser: Im Alter von ...)
- Keine Meinungsbilder bzw. Werturteile („Sollen Krankenschwestern...“) abfragen, sondern Wirklichkeitsurteile (Ich glaube, dass ...)

Ein Fragebogen etwas soll herausfinden, z.B. ob ein Mensch sozial ist oder verträumt.

Dazu entwickelt man verschieden Items. Items kann man zu Clustern zusammenfassen.

Items können einen bestimmten Zahlenwert haben, den man summieren kann.





Beispiel: Big 5

Inwieweit treffen die folgenden Aussagen auf Sie zu?

	trifft über- haupt nicht zu	trifft eher nicht zu	weder noch	eher zutreffend	trifft voll und ganz zu
(1) Ich bin eher zurückhaltend, reserviert.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
(2) Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
(3) Ich bin bequem, neige zur Faulheit.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
(4) Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
(5) Ich habe nur wenig künstlerisches Interesse.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
(6) Ich gehe aus mir heraus, bin gesellig.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
(7) Ich neige dazu, andere zu kritisieren.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
(8) Ich erledige Aufgaben gründlich.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
(9) Ich werde leicht nervös und unsicher.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
(10) Ich habe eine aktive Vorstellungskraft, bin fantasievoll.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

## Anonymität

### Namenscode

„ Phantasiename, den man auch noch nach Jahren erkennt“ zB „Kriachal“.  
Keine Begriffe mit Farben oder häufigen Begriffen wie „Sonne“,...

### Zifferncode

eigenes Geburtsdatum und Geburtsdatum der Mutter addieren.

Eigenes Geburtsdatum: 12. 4. 1987 → 120487

Geburtsdatum Ihrer Mutter: 3. 5. 1959 → 030559

---

**Summe = Kennzahl 151046**



## Was noch zu beachten wäre?

### INUS

Insufficient, but nonredundant part of an unnecessary but sufficient condition.

Unzureichender, aber nicht überflüssiger Teil einer nicht notwendigen aber ausreichenden Bedingung.

Jede Wirkung kann verschiedene Ursachen haben.

Jede Ursache besteht aus vielen Faktoren.

Wir kennen selten alle Faktoren.

Aussagen zu Ursache-Wirkungs-Beziehungen sind immer Wahrscheinlichkeitsaussagen.

Ursache-Wirkungsbeziehungen sind immer situativ.

*Beispiel:*

*Die Ursache „Umgang mit Smartphone“ ist ein nicht hinreichender Teil (insufficient) der komplexen Bedingung „Kompetenz im Umgang mit Digitalen Medien“.*

*Aber der Umgang mit einem Smartphone ist ein notwendiger (nicht überflüssiger - nonredundant) Teil der komplexen Bedingung „Kompetenz im Umgang mit Digitalen Medien“, weil ohne sie die Bedingung nicht erfüllt ist.*

*„Kompetenz im Umgang mit Digitalen Medien“ ist eine nicht notwendige Bedingung für eine allgemeine medienpädagogische Kompetenz, weil noch viele andere, komplexe Bedingungen denkbar sind (Videospiele, WhatsApp,...) aber sie ist ausreichend (sufficient), weil sie zur Förderung der medienpädagogischen Kompetenz beiträgt.*

### SMART-Kriterien

**spezifisch, messbar, attraktiv, realistisch, terminierbar**

SMART-Kriterien werden bei der Bestimmung der Ziele von Maßnahmen und Interventionen herangezogen. Die Ziele werden so formuliert, dass sie SMART sind.

### MAX-KON-MIN

**Maximiere** die Wirkung einer unabhängigen Variablen (Unterrichtsmethode) auf die abhängige Variable (Lernerfolg).

**Kontrolliere** die Einflüsse von unerwünschten systematischen Störvariablen.

Sie hängt mit der unabhängigen Variablen zusammen. Eine systematische Störvariable ist zum Beispiel ein kognitives Leistungsniveau, weil dieses den Lernerfolg beeinflusst.

**Minimiere** die Einflüsse von unsystematischen Störvariablen, die zwar die abhängige Variable beeinflusst, aber nicht mit der unabhängigen Variable zusammen hängt.

Baustellenlärm beeinflusst den Lernerfolg, aber beide Klassen sind gleichermaßen betroffen.



### **Hawthorne-Effekt**

Phänomen, dass die Teilnehmer einer Studie ihr natürliches Verhalten ändern, weil sie wissen, dass sie an einer Studie teilnehmen und unter Beobachtung stehen. Dies kann zu einer falschen Einschätzung der Wirksamkeit führen.

Beispiel:

Einfluss der Lichtverhältnisse auf Arbeitsleistung (Roethlisberger, Dickson, 1924- 1932)

Gruppe A: Lichtverhältnisse wurden verbessert

Gruppe B: Lichtverhältnisse blieben gleich

Auch bei Gruppe B stieg die Arbeitsleistung – die Beobachtung durch die Forscher beeinflusste die Leistungssteigerung – Reaktivität.

### **Qualitativ oder Quantitativ ?**

Quantitative Methoden	Qualitative Methoden
<ul style="list-style-type: none"><li>• Repräsentativität</li><li>• Zahlen</li><li>• Was passiert?</li></ul>	<ul style="list-style-type: none"><li>• Tiefe</li><li>• Menschen</li><li>• Warum passiert etwas?</li></ul>

<b>strukturiert</b> <b>standardisiert</b> <b>kontrolliert</b> <b>systematisch</b>	<b>unstrukturiert</b> <b>nicht standardisiert</b> <b>unkontrolliert</b> <b>anekdotisch</b>
--	---



## **Weitere Fragen**

### **Kontextualisierung**

Wie gliedert sich das Ergebnis in die Literatur ein?

Gibt es eine Folgeforschung – oder weitere nicht im Projekt beantwortete Fragen?

Was sollte auf Grund Ihrer Arbeit verändert werden?

### **Was haben Sie herausgefunden?**

Zusammenfassung der Ergebnisse

Interpretation der Ergebnisse

Was haben Sie NICHT herausgefunden?

## Ein paar Grundlagen der beschreibenden Statistik



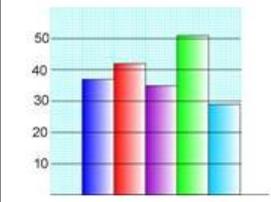
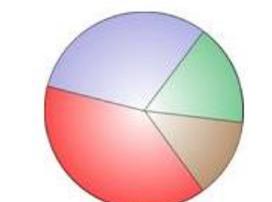
### Häufigkeiten

absolute Häufigkeit = Anzahl, wie oft ein Wert oder Eigenschaft vorkommt.

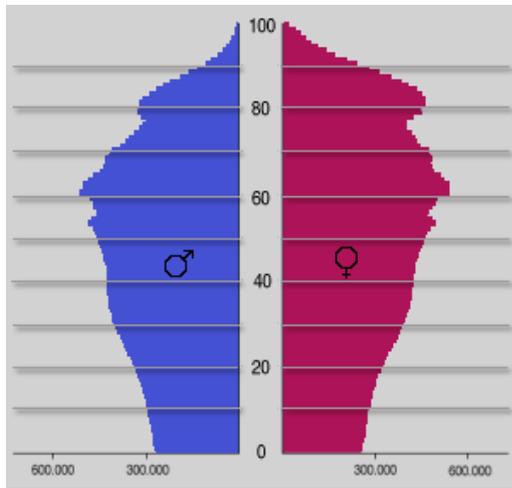
relative Häufigkeit = Vergleich von absoluter Häufigkeit zur Gesamtanzahl

prozentuelle Häufigkeit = Angabe der relativen Häufigkeit in Prozent

Lieblingswintersportart	Strichliste	Absolute Häufigkeit	Relative Häufigkeit bei 25 Personen	Prozentuelle Häufigkeit
Schi alpin		8	$8 : 25 = 0,32$	$0,32 \cdot 100 = 32\%$
Snowboard	-	12	$12 : 25 = 0,48$	$0,48 \cdot 100 = 48\%$
Eislaufen		5	$5 : 25 = 0,2$	$0,2 \cdot 100 = 20\%$

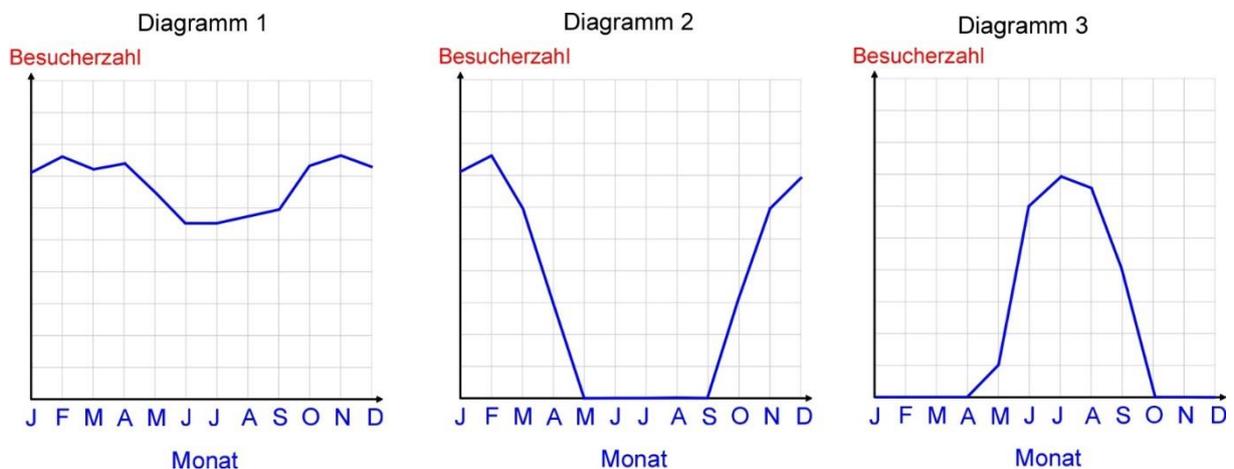
Balken- oder Säulendiagramm	Liniendiagramm	Kreisdiagramm	Piktogramme
Hier kann man die absolute Häufigkeit gut ablesen.	Hier kann man die Veränderung eines Wertes gut ablesen.	Hier kann man die relative Häufigkeit gut ablesen.	Auch hier kann man absolute Häufigkeiten gut ablesen.
			

Beispiel einer absoluten Häufigkeitsverteilung: Prognose der Altersverteilung für Deutschland im Jahr 2050:



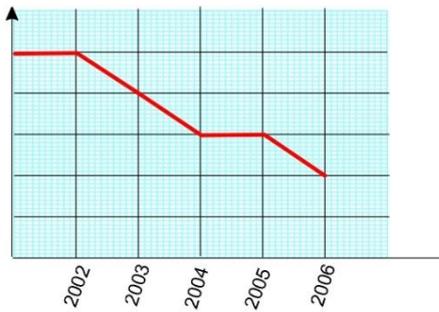
### Liniendiagramme

Die folgenden 3 Diagramme geben die Besucherzahlen eines Kinos, eines öffentlichen Freibades und eines Schiliftes an. Welcher Graf gehört zu welcher Einrichtung?

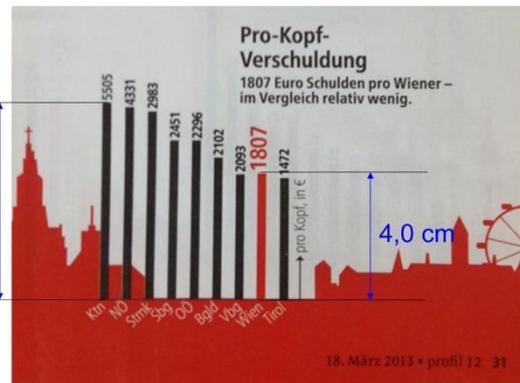


## Manipulationen

Preise



Preise



$$5505 \dots 6,2 \text{ cm}$$

$$1807 \dots x = 6,2 / 5505 \cdot 1807 = 2,03 \text{ cm}$$



## Mittelwerte

### *Arithmetisches Mittel*

Der Mittelwert ist also ein Parameter, der die zentrale Tendenz einer Verteilung in der Form eines Schwergewichts erfasst.

Schreibweise mit Summenzeichen:

$$x_{\text{mittel}} := \frac{\sum_{i=1}^N x_i}{N}$$

In der Mathematik wird das große Sigma (großes S des griechischen Alphabets) als Symbol für eine Summationsanweisung verwendet.

Werden z.B. 80 Werte addiert, so schreibt man anstelle von  $x_1 + x_2 + x_3 + \dots + x_{80}$ :

$$\sum_{i=1}^{80} x_i$$

Messwerte werden allgemein mit  $x_i$  bezeichnet.

Der Querstrich über dem Messwertsymbol zeigt eine Mittelwertbildung an:

Der Strichprobenumfang wird mit N bezeichnet.

### *Median*

Der arithmetische Mittelwert kennzeichnet das Verteilungsschwergewicht.

(Würde man alle Messwerte auf eine Waage legen, dann ist der Mittelwert der Waagemittelpunkt)

Der Median kennzeichnet den Punkt der Messwertachse, bei dem die Verteilung halbiert wird.

(Der Median berücksichtigt nur die Anzahl der Messwerte)

Taschengeld pro Woche in Euro:

5, 6, 6, **7**, 7, 8, 9 - Die Zahl 7 liegt genau in der Mitte - **7** ist daher der Median.

Bei einer geraden Anzahl von Werten, berechnet man den Median: Er ist das arithmetische Mittel der beiden mittleren Werte.

5, 6, 6, 7, 8, 8, 9, 9 - 7 und 8 liegen in der Mitte - **7,5** ist daher der Median.  $(7 + 8) : 2 = 7,5$

### **STREUUNG (Kennzahlen der Streuung)**

Mittelwerte allein können eine Verteilung nicht charakterisieren.  
Hier haben beide Gruppen bezüglich der Körpergröße den gleichen Mittelwert.



### *Standardabweichung*

In der Statistik wird meist die so genannte „Standardabweichung“ verwendet, um Streuung zu beschreiben.

Um keine negativen Werte zu erhalten, quadriert man die Abstände, addiert sie, dividiert durch die Anzahl der Messwerte und zieht zum Schluss die Wurzel. (Excel: STABWNA)

(In der empirischen Statistik dividiert man vor dem Wurzelziehen nicht durch „n“ sondern durch „n – 1“ ( Excel: STABWA); Schätzung der Standardabweichung der Grundgesamtheit aus einer Stichprobe).

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

(Den Ausdruck unter der Wurzel nennt man auch VARIANZ)

Mit Mittelwerten kann man sehr gut manipulieren. Zum Beispiel bei Durchschnittseinkommen sagt die Standardabweichung, ob die Gehälter dicht oder weit gestreut vom Mittelwert liegen.



### Variationskoeffizenz

Quotient aus Standardabweichung und Mittelwert.

$$V = \frac{s}{\bar{X}}$$

Die Standardabweichung gibt zwar die Streuung eines Merkmals an, zum Vergleich der Streuungen verschiedener Merkmale, deren Mittelwerte sich stark unterscheiden, ist sie aber nicht geeignet. Sind zum Beispiel die Varianzen von 50 kg-Zementsäcken und 0,5 kg Zuckerpackungen gleich groß, dann ist die Streuung beim Abfüllen von Zementsäcken vielleicht akzeptabel, bei den Zuckerpackungen aber inakzeptabel groß.

### Quartile und Kastenschaubilder

Mit so genannten Quartilen und Kastenschaubildern (Boxplots) können weitere Aussagen über Datenmengen getroffen werden.

Die Daten (zB das Alter von Kinobesuchern – 140 Besuchern) werden aufsteigend nach ihrer Größe sortiert. Dadurch ist der größte Wert (Maximum = 70) und der kleinste Wert (Minimum = 15) sofort ersichtlich.

Die Spannweite ist die Differenz von Maximum und Minimum (70 – 15 = 55)

15, 15, 16, 16, 16, 17, 17, 17, 17, 18, 18, 18, 18, 18, 19, 19, 19, 19, 19, 19,  
20, 20, 20, 20, 20, 20, 21, 21, 21, 21, 21, 22, 22, 22, 22, 23, 23, 23, 24, 24,  
25, 25, 25, 26, 26, 26, 27, 27, 28, 30, 32, 33, 35, 37, 38, 38, 39, 39, 39, 39,  
40, 40, 40, 40, 41, 41, 41, 41, 41, 41, 42, 42, 42, 42, 42, 42, 42, 42, 43,  
43, 43, 43, 43, 44, 44, 44, 44, 44, 44, 44, 45, 45, 45, 45, 46, 46, 46, 46, 46,  
46, 47, 47, 47, 48, 48, 48, 48, 49, 49, 49, 49, 50, 50, 50, 52, 52, 52, 52, 53,  
53, 54, 54, 54, 56, 56, 57, 58, 58, 59, 60, 60, 61, 64, 65, 67, 67, 68, 68, 70

Sie sind sehr nützlich, um sich über einen Datensatz schnell einen Überblick zu verschaffen oder verschiedene Datensätze miteinander zu vergleichen.

