Statistik Zusammenfassung

Gerrit Konrad

2024-08-17

Inhaltsverzeichnis

0.1 Vorwort

Statistik ist ein Teilgebiet der Mathematik, das Methoden bereitstellt, um Daten zu beschreiben und aus Stichproben Schlussfolgerungen über eine Grundgesamtheit zu ziehen. Die Statistik gliedert sich in drei Hauptbereiche:

- **Deskriptive Statistik:** Darstellung und Charakterisierung umfangreicher Datensätze durch einfache Maßzahlen.
- **Analytische Statistik:** Verallgemeinerung von Stichprobendaten auf die Grundgesamtheit unter Berücksichtigung zufälliger Schwankungen.
- **Wahrscheinlichkeitstheorie:** Grundlage für Schätz- und Testverfahren der Analytischen Statistik, berechnet Wahrscheinlichkeiten zukünftiger Beobachtungen.

0.2 Was ist Statistik?

Statistik umfasst Methoden zur Sammlung, Analyse, Interpretation und Präsentation von Daten. Hierbei unterscheidet man zwischen:

- **Deskriptive Statistik:** Ziel ist es, Daten verständlich und übersichtlich darzustellen.
- **Analytische Statistik:** Erlaubt es, aus Stichproben Rückschlüsse auf die Grundgesamtheit zu ziehen, indem zufällige Schwankungen quantifiziert werden.
- **Wahrscheinlichkeitstheorie:** Dient als Basis für die Analytische Statistik, indem sie Wahrscheinlichkeitsmodelle und Berechnungen für zukünftige Ereignisse liefert.

Die Anwendung statistischer Methoden erfolgt häufig mit Computerprogrammen, wobei in diesem Buch Excel für die Durchführung statistischer Analysen genutzt wird.

1 Grundlegende Begriffe

1.1 Statistische Einheit

Die statistische Einheit ist das Objekt, an dem Messungen oder Beobachtungen durchgeführt werden. Beispiele sind Personen, Unternehmen oder Produkte.

Beispiel: Ein einzelner Schüler in einer Schule.

1.2 Grundgesamtheit, Teilgesamtheiten, Stichproben

- **Grundgesamtheit:** Gesamtheit aller möglichen Untersuchungseinheiten. *Beispiel:* Alle Schüler einer Schule.
- **Teilgesamtheiten:** Untergruppen der Grundgesamtheit. *Beispiel:* Alle Schüler einer bestimmten Klasse.
- **Stichprobe:** Eine Auswahl von Einheiten aus der Grundgesamtheit, die zur Analyse herangezogen wird. *Beispiel:* 30 zufällig ausgewählte Schüler aus der gesamten Schule.

1.3 Merkmale und Merkmalsausprägungen

- **Merkmale:** Eigenschaften, die an den statistischen Einheiten gemessen werden. *Beispiel:* Körpergröße der Schüler.
- **Merkmalsausprägungen:** Konkrete Werte, die ein Merkmal annehmen kann. *Beispiel:* 160 cm, 170 cm, 180 cm, etc.

1.4 Merkmalstypen und Messniveaus

- **Merkmalstypen:** Während nominal- und ordinalskalierte Merkmale nur diskret aufgefasst werden können, lassen sich metrisch skalierte Merkmale sowohl diskret als auch stetig dar- stellen.
  - **Diskrete Merkmale:** Können nur bestimmte, abzählbare Werte annehmen. *Beispiel:* Anzahl der Kinder in einer Familie (0, 1, 2, . . .).
  - **Stetige Merkmale:** Können jeden beliebigen Wert innerhalb eines Intervalls annehmen, je nachdem wie genau die Messung erfolgt. *Beispiel:* Körpergröße in Zentimetern.
- **Messniveaus:**
  - **Nominal:** Kategorische Daten ohne natürliche Reihenfolge. *Beispiel:* Geschlecht (männlich, weiblich).
  - **Ordinal:** Kategorische Daten mit natürlicher Reihenfolge, aber ohne festen Abstand zwischen den Kategorien. *Beispiel:* Schulnoten (sehr gut, gut, befriedigend).
  - **Metrisch:** Umfasst sowohl intervallskalierte als auch verhältnisskalierte Daten, da beide kontinuierliche und messbare Größen darstellen.
    - **Intervallskaliert:** Numerische Daten mit gleichen Abständen zwischen den Werten, aber ohne natürlichen Nullpunkt. *Beispiel:* Datumsangaben.
    - **Verhältnisskaliert:** Numerische Daten mit gleichen Abständen und einem natürlichen Nullpunkt. Man kann sagen "Ein Wert ist doppelt so groß wie ein anderer". *Beispiel:* Gewicht in Kilogramm.

1.5 Methoden der deskriptiven Statistik

Methoden zur Darstellung und Analyse von Daten umfassen Tabellen, Grafiken und statistische Kennzahlen wie Mittelwert, Median und Standardabweichung.

Beispiele:

- *Tabellen:* Frequenzverteilung einer Umfrage.
- *Grafiken:* Säulendiagramm der Häufigkeit von Schulnoten.
- *Statistische Kennzahlen:* Mittelwert der Körpergröße, Median des Alters, Standardabweichung des Gewichts.

2 Begriffe der Häufigkeit

2.1 Absolute und relative Häufigkeit

Die **absolute Häufigkeit** H_j gibt an, wie oft ein bestimmtes Ereignis j in einer Datenreihe vorkommt.

H_j = Anzahl der Beobachtungen von j

Wenn n statistische Einheiten beobachtet werden, gilt:

$$\sum_{j=1}^{k} H_j = n$$

Die **relative Häufigkeit** h_j ist das Verhältnis der absoluten Häufigkeit zur Gesamtzahl der Beobachtungen n. Wird oft in Prozent ausgedrückt:

$h_j = \frac{H_j}{n}$

Die relative Häufigkeit nimmt Werte zwischen 0 und 1 an und es gilt:

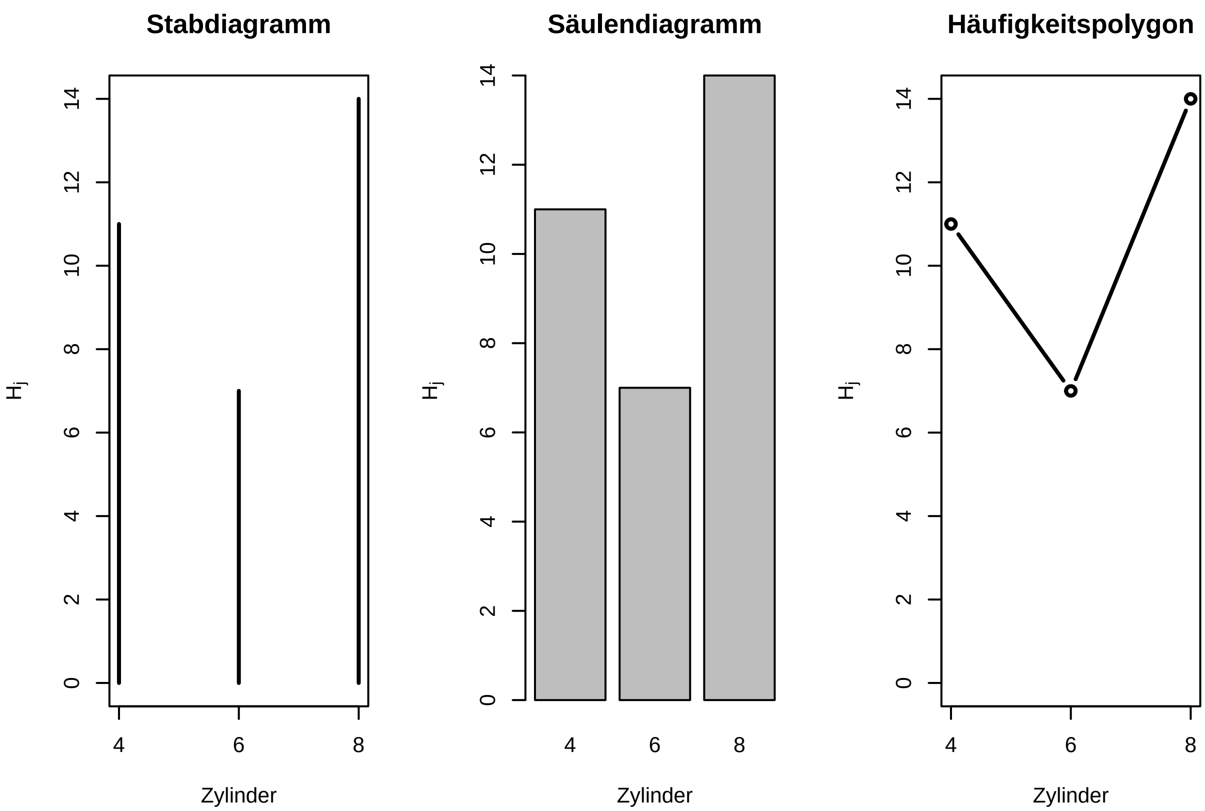
$$\sum_{j=1}^{k} h_j = 1$$

Beispiel: Die Zylinderzahlen der untersuchten Autos können folgendermaßen zusammengefasst werden:

Zylinder	Absolute_Häufigkeit	Relative_Häufigkeit
4	11	0.344
6	7	0.219
8	14	0.438
Summe	32	1.000

Wenn Merkmalsausprägungen kontinuierlich oder zu zahlreich sind, werden sie in Klassen zusammengefasst. Jede Klasse wird durch eine untere und obere Grenze, eine Klassenmitte und eine Klassenbreite definiert. Die Häufigkeiten werden für die Klassen berechnet. Formel bleibt also gleich.

2.2 Graphische Darstellungsmöglichkeiten von Häufigkeiten



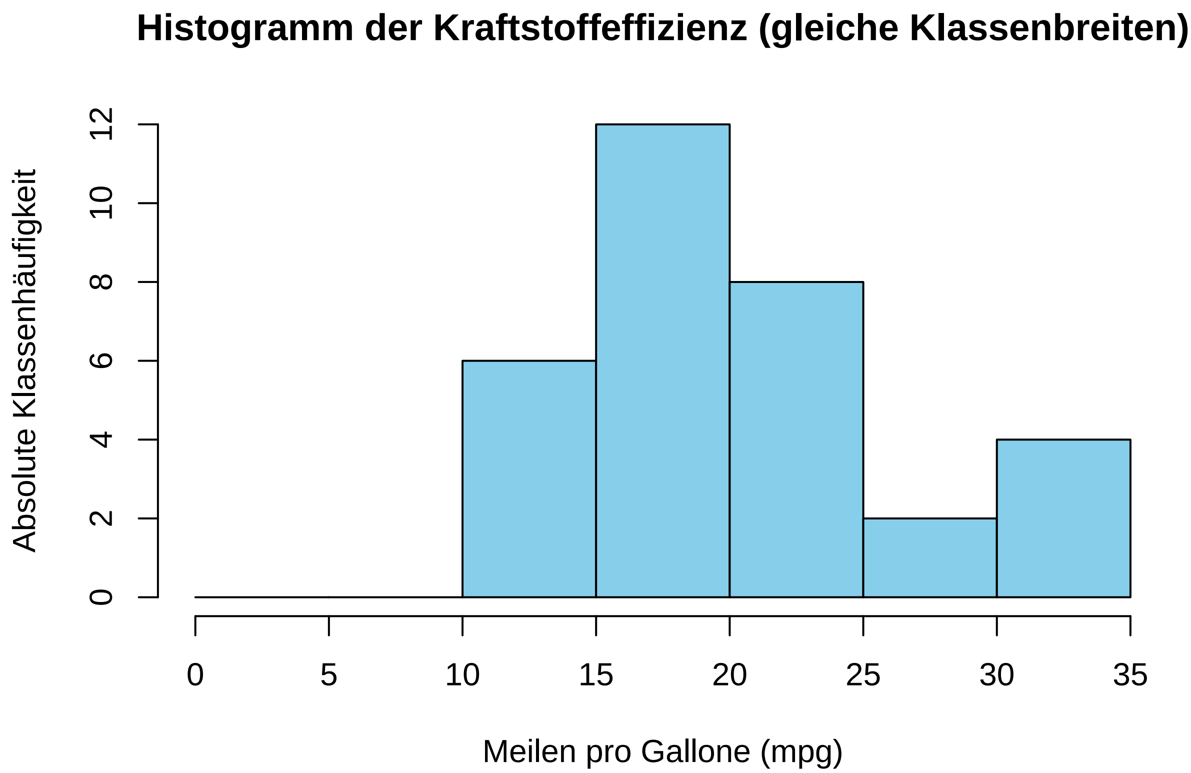
Histogramm

Ein **Histogramm** ist eine Möglichkeit, die Häufigkeitsverteilung klassierter Daten graphisch darzustellen.

Im folgenden Beispiel analysieren wir die Kraftstoffeffizienz von Fahrzeugen, gemessen in Meilen pro Gallone (mpg).

Histogramme mit äquidistanten Klassen: Bei Histogrammen mit äquidistanten Klassen sind die Intervalle (Klassenbreiten) gleich groß. Dies bedeutet, dass jeder Balken die gleiche Breite hat. Der Vorteil dieser Methode ist die einfache Vergleichbarkeit der Häufigkeiten in den einzelnen Klassen.

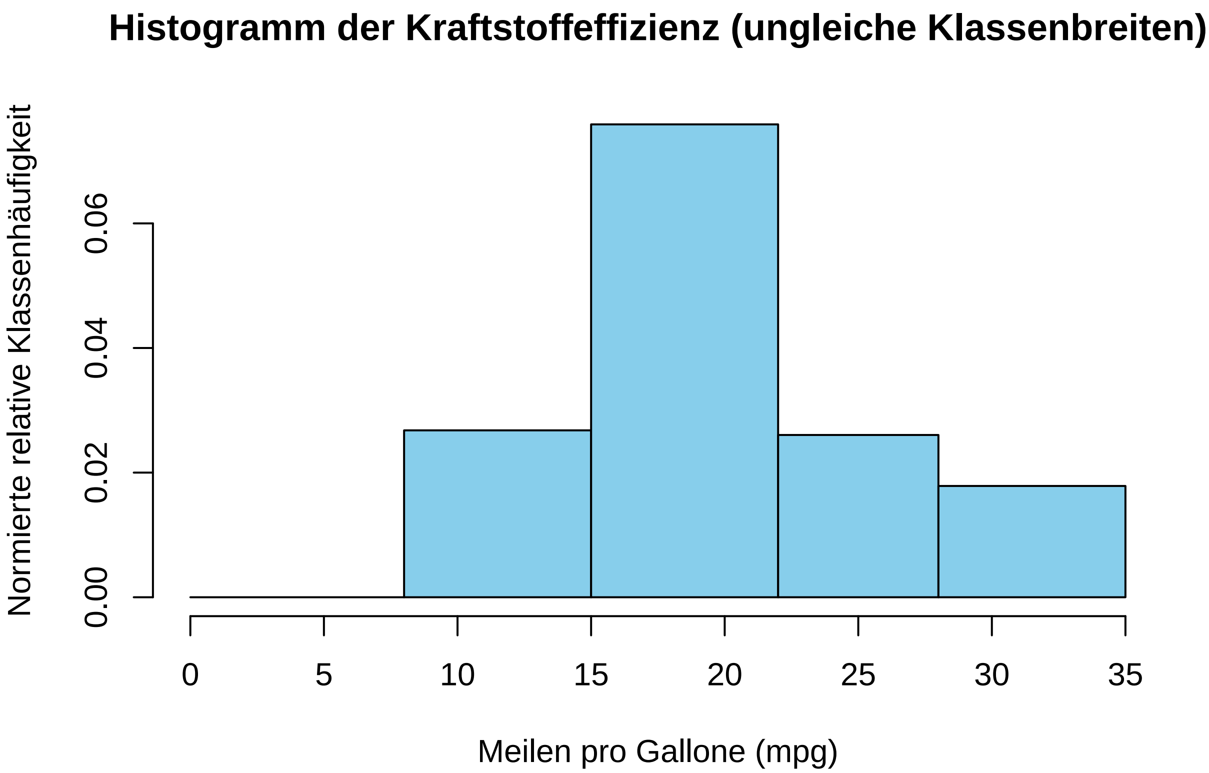
j	Klasse	Hj	hj	Klassenbreite
1	[0;5)	0	0.0000	5
2	[5;10)	0	0.0000	5
3	[10;15)	6	0.1875	5
4	[15;20)	12	0.3750	5
5	[20;25)	8	0.2500	5
6	[25;30)	2	0.0625	5
7	[30;35)	4	0.1250	5



Histogramme mit ungleichen Klassenbreiten: Bei Histogrammen mit ungleichen Klassenbreiten variieren die Intervalle in ihrer Breite. Um die Vergleichbarkeit der Häufigkeiten zu gewährleisten, wird in diesen Histogrammen oft die Höhe der Balken nicht direkt durch die Häufigkeit, sondern durch die sogenannte "normierte Häufigkeit", oder auch genannt "Häufigkeitsdichte" bestimmt. Die normierte Häufigkeit wird berechnet als:

$$\text{Normierte Häufigkeit} = \frac{Häufigkeit}{Klassenbreite}$$

j	Klasse	Hj	hj	Klassenbreite	norm_Hj	norm_hj
1	[0;8)	0	0.00000	8	0.00000	0.00000
2	[8;15)	6	0.18750	7	0.85714	0.02679
3	[15;22)	17	0.53125	7	2.42857	0.07589
4	[22;28)	5	0.15625	6	0.83333	0.02604
5	[28;35)	4	0.12500	7	0.57143	0.01786



Kreisdiagramm (Pie Chart)

Darstellung der relativen Häufigkeiten als Segmente eines Kreises. Der Kreis repräsentiert die Gesamtheit der Daten, und die Segmente repräsentieren die Anteile der verschiedenen Ausprägungen.

Piktogramm

Darstellung der Häufigkeiten durch Bilder oder Symbole, die proportional zur Häufigkeit sind.

2.3 Empirische Verteilungsfunktionen S_j

Eine empirische Verteilungsfunktion beschreibt die Verteilung von Merkmalsausprägungen in einer Beobachtungsreihe. Sie gibt die kumulierte Häufigkeit bis zu einem bestimmten Wert an und ist besonders nützlich für die Analyse von ordinalen oder metrischen Merkmalen.

Die **absolute Summenhäufigkeit** H_j ist die kumulierte Anzahl der Beobachtungen bis zu einer bestimmten Merkmalsausprägung.

$$H_j = \sum_{i=1}^{j} H_i$$

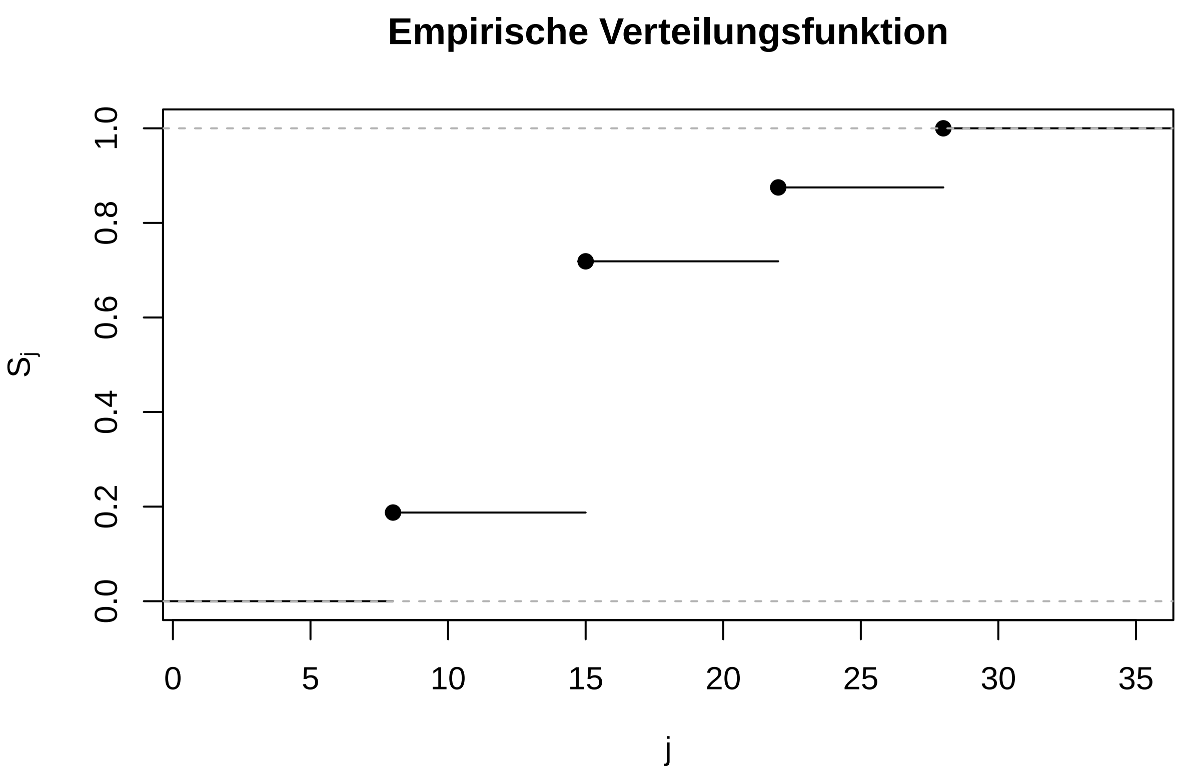
Beispiel: Wenn $H_1, H_2, H_3, \ldots, H_j$ die absoluten Häufigkeiten sind, dann ist H_j die Summe der absoluten Häufigkeiten bis zur j-ten Ausprägung.

Die **relative Summenhäufigkeit** ist das Verhältnis der absoluten Summenhäufigkeit zur Gesamtzahl der Beobachtungen und gibt den Anteil der Beobachtungen an, die höchstens den Wert x haben.

$$S_j = \sum_{i=1}^{j} h_i$$

Als Beispiel nehmen wir wieder die Kraftstoffeffizienz von Fahrzeugen:

j	Klasse	Hj	hj	emp.Verteilungsfkt.
1	[0;8)	0	0.00000	0.00000
2	[8;15)	6	0.18750	0.18750
3	[15;22)	17	0.53125	0.71875
4	[22;28)	5	0.15625	0.87500
5	[28;35)	4	0.12500	1.00000



3 Maßzahlen zur Beschreibung von Häufigkeitsverteilungen

Häufigkeitsverteilungen können durch Maßzahlen oder Parameter charakterisiert werden, die es ermöglichen, zentrale Eigenschaften einer Verteilung zusammenzufassen. Dazu gehören Lagemaße, Streuungsmaße und Formmaße.

3.1 Lagemaße

Beschreiben das Zentrum / die Mitte einer Beobachtungsreihe

3.1.1 Arithmetisches Mittel

Es ist das am häufigsten verwendete Lagemaß für metrische Merkmale. Es wird berechnet, indem die Summe aller beobachteten Werte durch die Anzahl der Beobachtungen geteilt wird.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

wobei x_i die einzelnen Beobachtungswerte und n die Anzahl der Beobachtungen sind.

3.1.2 Gewogenes arithmetisches Mittel

Wenn die Beobachtungen unterschiedliche Gewichte w_i haben, ergibt sich das gewogene arithmetische Mittel

$$\bar{x}_w = \frac{1}{n} \sum_{i=1}^{n} w_i \cdot x_i$$

wobei w_i die absolute Häufigkeit ist.

3.1.3 Geometrische Mittel

Wird verwendet um das durchschnittliche Wachstum über Zeiträume zu messen

$$\bar{x}_{\text{geo}} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

3.1.4 Harmonische Mittel

Wenn ein funktionaler Zusammenhang besteht, z.B. Durchschnittsgeschwindigkeit

$$\frac{n}{\sum_{i=1}^{n} 1/x_i}$$

3.1.5 Median

Wird als Zentralwert bezeichnet und teilt die Daten in zwei gleich große Hälften. Kann für metrisch und ordinal skalierte Merkmale verwendet werden und ist robust gegenüber Ausreißern.

$$\bar{x}_{Median} = \begin{cases} x_{\frac{n+1}{2}} & \text{falls } n \text{ ungerade} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{falls } n \text{ gerade} \end{cases}$$

3.1.6 Quantile

Index berechnung: $I = n \cdot p$

Berechnung des Quantils p:

$$x_p = \begin{cases} \frac{1}{2}(x_I + x_{I+1}) & \text{falls } I \text{ ganzzahlig} \\ x_{[I]+1} & \text{falls } I \text{ nicht ganzzahlig} \end{cases}$$

$[] =$ abrunden.

3.1.7 Modus

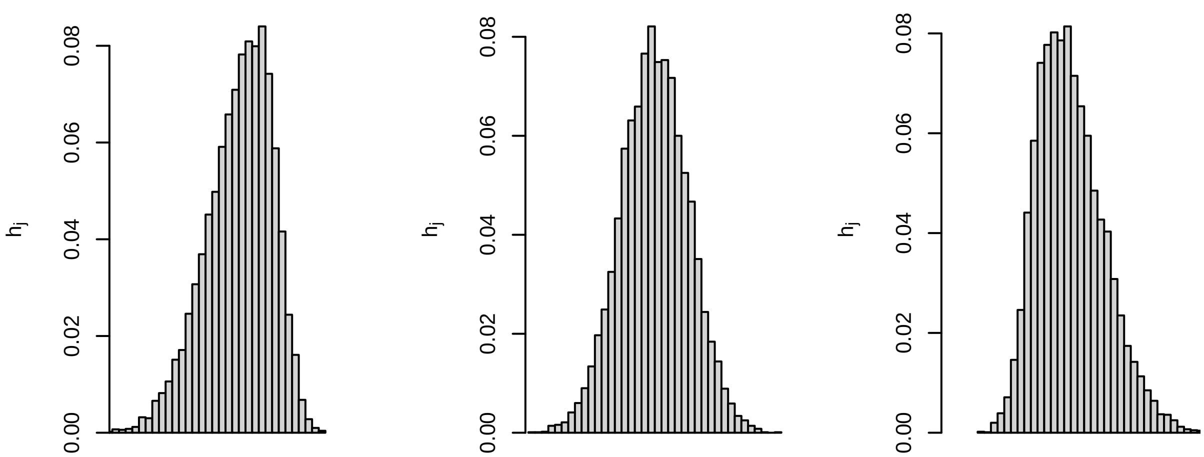
Der **Modus** oder **Modalwert** ist die häufigste Ausprägung einer Verteilung.

$$h_{Modus} \geq h_j$$

3.1.8 Zusammenfassung Formeln

Lagemaß	Symbol	Berechnung
Modus	$\bar{x}_{Modus}$	$h_{Modus} \geq h_j$
Median	$\bar{x}_{Median}$	$x_{\frac{n+1}{2}}$ oder $\frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$
Quantil	Q_α	Wert der Verteilungsfunktion
Arithmetisches Mittel	$\bar{x}$	$\frac{1}{n}\sum_{i=1}^{n} x_i$
Geometrisches Mittel	$\bar{x}_{geo}$	$\sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$
Harmonisches Mittel	$\bar{x}_{harm}$	$\frac{n}{\sum_{i=1}^{n} 1/x_i}$

3.2 Zusammenhang zwischen Lagemaßen und Form



- Linksschiefe Häufigkeitsverteilung: $\bar{x} < \bar{x}_{Median} < \bar{x}_{Modus}$
- Symmetrische Häufigkeitsverteilung: $\bar{x} = \bar{x}_{Median} = \bar{x}_{Modus}$
- Rechtsschiefe Häufigkeitsverteilung: $\bar{x} > \bar{x}_{Median} > \bar{x}_{Modus}$

3.3 Streuungsmaße

3.3.1 Spannweite

Ist die differenz zwischen dem Maximum und Minimum einer Beobachtungsreihe.

$$R = x_{max} - x_{min}$$

3.3.2 Quartilsabstand

Gibt die Größe des Bereiches zwischen dem oberen und unteren Quartil einer Beobachtungsreihe an.

$$Q_o - Q_u = Q_{0.75} - Q_{0.25}$$

- Zwischen dem oberen und dem unteren Quartil liegen 50% der Beobachtungen. - Kann auch für ordinalskalierte Merkmale bestimmt werden. - Ist robust (unempfindlich gegenüber Ausreißern)

3.3.3 Varianz

Mittlere quadrierte Abweichung vom arithmetischen Mittel

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{oder} \quad s^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

- Es gilt immer $s^2 \geq 0$
- Wird unterschiedlich für die Stichprobe und die Grundgesamtheit (Population) berechnet.
- Grundidee: Einbezug aller Abweichungen vom Mittelwert
- Beobachtungen, die weit von $\bar{x}$ entfernt liegen, werden überproportional stark gewichtet.

3.3.4 Standardabweichung

Ist die Wurzel aus der Varianz

$$s = \sqrt{s^2}$$

- Weist die gleiche Maßeinheit wie die Daten auf

3.3.5 Variationskoeffizient

Ist der Quotient aus Standardabweichung und arithmetische Mittel

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

- Ist vergleichbar

3.3.6 Zusammenfassung Formeln

Lagemaß	Symbol	Berechnung
Spannweite	R	$x_{max} - x_{min}$
Interquartilsabstand	IQR	$Q_{0.75} - Q_{0.25}$
(empirische) Varianz	s^2	$\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$
Standardabweichung	s	$\sqrt{s^2}$
Variationskoeffizient	V	$s/\bar{x}$

3.4 Konzentrationsmaße

Man spricht von Konzentration oder Ungleichheit, falls zu einem bestimmten Zeitpunkt ein relativ kleiner Anteil der Merkmalsträger einen hohen Anteil an der Summe der Merkmalswerte besitzt.

3.4.1 Lorenzkurve

$$k_j = \sum_{i=1}^{j} \frac{H_j}{n} = \sum_{i=1}^{j} h_j \quad und \quad l_j = \frac{\sum_{i=1}^{j} a_i H_i}{\sum_{i=1}^{q} a_i H_i} = \frac{\sum_{i=1}^{j} a_i H_i}{\sum_{i=1}^{n} x_i}$$

$\sum_{i=1}^{n} x_i$ (Gesamtsumme aller Beobachtungswerte)

- Die Lorenzkurzve verläuft durch die Punkte $(0,0)$ und $(1,1)$
- Die Lorenzkurve verläuft immer **unterhalb** der Winkelhalbierenden.
- Die Lorenzkurve ist winkelhalbierend, wenn alle Mermalsausprägungen gleich häufig vorkommen. Dann liegt keine Konzentration vor. Je weiter die Lorenzkurve sich von der Winkelhalbierenden entfernt, desto größer ist die Ungleichheit.

Beispiel Einkommensverteilung der Länder A, B, C

```
A <- c(1000, 3000, 4000, 4000, 8000)
B <- c(2000, 2000, 4000, 8000)
C <- c(1000, 2000, 5000, 8000)
```

Lorenzkurve für Land A

j	Werte	H_j	h_j	k	l
0	0	0	0.0	0.0	0.00
1	1000	1	0.2	0.2	0.05
2	3000	1	0.2	0.4	0.20
3	4000	2	0.4	0.8	0.60
4	8000	1	0.2	1.0	1.00

3.5 Gini Koeffizient

Das Doppelte der Fläche zwischen der Lorenzkurve und der Winkelhalbierenden heißt **Gini-Koeffizient** G und wird als Konzentrationsmaß einer Häufigkeitsverteilung verwendet.

$$G = \sum_{i=1}^{n} (k_i + k_{i-1})(l_i - l_{i-1}) - 1$$

mit $(k_i,\, l_i)$, $i = 0, ..., n$ Stützpunkte der Lorenzkurve

- Um den Gini-Koeffizienten zu berechnen, sind alle Stützpunkte der Lorenzkurve erforderlich. Es gilt $0 \leq G \leq \frac{n-1}{n} < 1$.
- Wenn die Lorenzkurve winkelhalbierend ist, gilt $G = 0$. In diesem Fall gibt es keine Einkommensunterschiede.
- Werden *alle* Ausgangswerte x_i mit einem Faktor a multipliziert, sodass $y_i = a \cdot x_i$, dann gilt $G_y = G_x$.

```
## [1] 0.3
```

4 Zweidimensionale Verteilung

4.1 Streudiagramm

Ein **Streudiagramm** (Scatter-Plot) ist eine grafische Darstellung von zweidimensionalen Merkmalen. Dabei werden die Wertepaare (x_i, y_i) als Punkte in einem Koordinatensystem eingezeichnet:

- **X-Achse**: Merkmalsausprägung für das Merkmal X.
- **Y-Achse**: Merkmalsausprägung für das Merkmal Y.

Beispiel: Die monatliche Sonnenscheindauer am Vormittag (X) und Nachmittag (Y) wird über ein Jahr hinweg beobachtet und als Punkte im Diagramm dargestellt.

4.2 Kontingenztafel

Eine **Kontingenztafel** ist eine tabellarische Darstellung der gemeinsamen Häufigkeitsverteilung zweier Merkmale:

- **Absolute Häufigkeit**: H_{ij} ist die Anzahl der Beobachtungen mit der Merkmalskombination (a_i, b_j).
- **Relative Häufigkeit**: $h_{ij} = \frac{H_{ij}}{n}$ ist der Anteil der Beobachtungen mit (a_i, b_j).

Randhäufigkeiten: - Absolute Randhäufigkeit von a_i: $H_{i\cdot} = \sum_j H_{ij}$. - Relative Randhäufigkeit von a_i: $h_{i\cdot} = \sum_j h_{ij}$.

4.3 Korrelationskoeffizient

Der **Korrelationskoeffizient** nach Bravais-Pearson misst den linearen Zusammenhang zwischen zwei Merkmalen X und Y:

- **Formel**:

  $$r_{XY} = \frac{s_{XY}}{s_X \cdot s_Y}$$

  wobei s_{XY} die Kovarianz ist, definiert als:

  $$s_{XY} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

- **Standardabweichungen**:

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

  $$s_Y = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2}$$

- **Interpretation**:
  - $r_{XY} > 0$: Positive Korrelation.
  - $r_{XY} < 0$: Negative Korrelation.
  - $|r_{XY}| \leq 0.5$: Vermutlich kein signifikanter Zusammenhang.

5 Kombinatorik

Die Kombinatorik beschäftigt sich mit dem Zählen von Möglichkeiten, Elemente aus einer Menge zusammenzustellen. Wichtige Konzepte sind:

5.1 Fakultäten und Binomialkoeffizienten

- **Fakultät (n!)**: Das Produkt der ersten n natürlichen Zahlen. Wichtig für Permutationen.
  - Beispiel: $5! = 1 \times 2 \times 3 \times 4 \times 5 = 120$
  - $0! = 1$ per Definition.
- **Binomialkoeffizient** $\binom{n}{k}$: Gibt die Anzahl der Möglichkeiten an, k Elemente aus n auszuwählen.
  - Formel: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

5.2 Permutationen

- **Permutation**: Anordnung aller Elemente einer Menge.
  - Beispiel: Für 3 verschiedenfarbige Kugeln gibt es $3! = 6$ Anordnungen.
- **Permutationen bei gleichen Elementen**: Wenn es gleiche Elemente gibt, reduziert sich die Anzahl der Permutationen.
  - Formel: $\frac{n!}{n_1!n_2!\ldots n_k!}$ (für k Gruppen von gleichen Elementen)
  - Beispiel: Das Wort "ANANAS" hat 6 Buchstaben, wobei A und N jeweils 2-mal vorkommen. Die Anzahl der Permutationen ist $\frac{6!}{3!2!} = 60$

5.3 Kombinationen

- **Kombination ohne Wiederholung**: Auswahl von k Elementen aus n, ohne Berücksichtigung der Reihenfolge.
  - **Formel**: $\binom{n}{k}$
  - **Beispiel**: Wähle 3 Früchte aus einer Schale mit 5 verschiedenen Früchten (Apfel, Banane, Orange, Traube, Kiwi).
    - $\binom{5}{3} = 10$
    - Erklärung: Die Auswahl könnte z. B. {Apfel, Banane, Orange} sein.
- **Kombination mit Wiederholung**: Auswahl von k Elementen aus n, ohne Berücksichtigung der Reihenfolge.
  - **Formel**: $\binom{n+k-1}{k}$
  - **Beispiel**: Wähle 3 Kugeln aus einer Schale mit 2 verschiedenen Kugeln (rot und blau), wobei Kugeln zurückgelegt werden.
    - $\binom{2+3-1}{3} = \binom{4}{3} = 4$
    - Erklärung: Mögliche Kombinationen sind {rot, rot, rot}, {rot, rot, blau}, {rot, blau, blau}, {blau, blau, blau}.
- **Variation ohne Wiederholung**: Kombinationen mit Berücksichtigung der Reihenfolge.
  - **Formel**: $\frac{n!}{(n-k)!}$
  - **Beispiel**: Ordne 3 von 5 unterschiedlich nummerierten Büchern.
    - $\frac{5!}{(5-3)!} = 60$
    - Erklärung: Eine mögliche Reihenfolge könnte {Buch 3, Buch 1, Buch 5} sein.
- **Variation mit Wiederholung**: Kombinationen mit Berücksichtigung der Reihenfolge.
  - **Formel**: n^k
  - **Beispiel**: Erstelle eine 3-stellige Zahl aus den Ziffern 1 bis 3, wobei Ziffern wiederholt werden können.
    - $3^3 = 27$
    - Erklärung: Eine mögliche Zahl könnte 312 sein.

5.4 Zusammenfassung

Modell	**Ohne Wiederholung**	**Mit Wiederholung**
Permutation	$P(n) = n!$	$P(n_1, n_2, \ldots, n_k) = \frac{n!}{n_1! n_2! \ldots n_k!}$
Kombinationen mit Reihenfolge	$V(n, k) = \frac{n!}{(n-k)!}$	$V_w(n, k) = n^k$
Kombinationen ohne Reihenfolge	$C(n, k) = \binom{n}{k}$	$C_w(n, k) = \binom{n+k-1}{k}$

6 Grundbegriffe der Wahrscheinlichkeiten

6.1 Zufallsexperimente und Ereignisse

Definition

Ein **Zufallsexperiment** ist ein Experiment, das folgende Bedingungen erfüllt:

1. Es wird nach einer genau festgelegten Vorschrift durchgeführt.
2. Es kann beliebig oft unter den gleichen Bedingungen wiederholt werden.
3. Sein Ergebnis ist nicht mit Sicherheit vorhersehbar.

Der **Ereignisraum** (Symbol: Ω) ist die Menge aller möglichen Ergebnisse.

Sicheres und Unmögliches Ereignis

- **Sicheres Ereignis**: Tritt immer ein: $E = \Omega$.
- **Unmögliches Ereignis**: Tritt nie ein: $\varnothing = \{\}$.

Beispiele

- **Münzwurf**: $\Omega = \{Kopf, Zahl\}$
- **Würfelwurf**: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- **Lebensdauer einer Glühlampe**: $\Omega = \{x | x \geq 0\}$

6.2 Rechnen mit Ereignissen

Definition

- **Komplementärereignis**: Das Ereignis $\bar{A}$, das eintritt, wenn A nicht eintritt: $\bar{A} \subset \Omega$.
- **Disjunkte Ereignisse**: Zwei Ereignisse A und B, die nicht gleichzeitig eintreten können: $A \cap B = \varnothing$.

Mengenarten

- **Vereinigungsmenge**: $A \cup B$ Tritt ein, wenn A oder B oder beide eintreten.
- **Schnittmenge**: $A \cap B$ Tritt ein, wenn A und B beide eintreten.
- **Differenzmenge**: $A - B$ Tritt ein, wenn A eintritt, aber nicht B.

Beispiele

- **Komplementär**: „ungerade Zahl“ und „gerade Zahl“ beim Würfelwurf
- **Disjunkt**: „Kugel fällt auf eine gerade Zahl“ und „Kugel fällt auf Zahl 15“ beim Roulette

6.3 Wahrscheinlichkeiten

Wahrscheinlichkeit (von Mises)

Die **statistische Wahrscheinlichkeit** eines Ereignisses A ist die relative Häufigkeit dieses Ereignisses bei einer sehr großen Anzahl n der Wiederholungen des Zufallsexperiments:

$$P(A) = \lim_{n \to \infty} \frac{H_n(A)}{n} = \lim_{n \to \infty} h_n(A)$$

Beispiel Beim Würfeln eines fairen Würfels interessiert das Ereignis „Werfen einer geraden Zahl“ {2, 4, 6}. Mit zunehmender Anzahl von Würfen nähert sich die relative Häufigkeit dem Wert 0,5.

Wahrscheinlichkeit (Laplace)

Bei endlich vielen gleichwahrscheinlichen Ergebnissen.

$$P(A) = \frac{|A|}{|\Omega|} \frac{\text{Anzahl der für A günstigen Ergebnisse}}{\text{Anzahl der möglichen Ergebnisse}}$$

Beispiel Bei zwei fairen Würfeln ist die Wahrscheinlichkeit für einen Pasch:

$$P(\text{Pasch}) = \frac{6}{36} = \frac{1}{6}$$

6.4 Kolmogorovsche Axiome

1. **Positivität**: $P(A) \geq 0$
2. **Normiertheit**: $P(\Omega) = 1$
3. **Additivität**: Für disjunkte Ereignisse A und B, also $A \cap B = \varnothing$:

$$P(A \cup B) = P(A) + P(B)$$

Rechenregeln für Wahrscheinlichkeiten

- **Gesamtheit aller möglichen Ergebnisse**:

$$\sum_{i=1}^{n} P(A_i) = 1$$

- **Unmögliches Ereignis**:

$$P(\varnothing) = 0$$

- **Komplementäre Ereignisse**:

$$P(\bar{A}) = 1 - P(A)$$

- **Vereinigung beliebiger Ereignisse**:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- **Für disjunkte Ereignisse** A und B:

$$P(A \cap B) = 0$$

- **Ereignisse mit der Beziehung** $A \subseteq B$:

$$P(A) \leq P(B)$$

6.5 Bedingte Wahrscheinlichkeiten

Seien A und B zwei Ereignisse. Die Wahrscheinlichkeit des Ereignisses A unter der Bedingung, dass das Ereignis B bereits eingetreten ist, heißt **bedingte Wahrscheinlichkeit** von A unter B oder Wahrscheinlichkeit von A gegeben B und ist definiert als

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{für} \quad P(B) \neq 0$$

$P(A \cap B)$ kann als die Wahrscheinlichkeit, dass beide Ereignisse gleichzeitig eintreten, interpretiert werde

Durch Umformung der Definition der bedingten Wahrscheinlichkeit erhält man den **Multiplikationssatz** der Wahrscheinlichkeitsrechnung:

$$P(A \cap B) = P(A) \cdot P(B|A) \quad \text{sowie} \quad P(B \cap A) = P(B) \cdot P(A|B)$$

Es gilt zudem:

$$P(B \cap A) + P(B \cap \bar{A}) = P(B)$$
$$P(\bar{A}|B) = 1 - P(A|B)$$

6.6 Unabhängigkeit von Ereignissen

Zwei Ereignisse A und B heißen **unabhängig**, wenn das Eintreten des einen Ereignisses die Wahrscheinlichkeit des Eintretens des anderen Ereignisses nicht beeinflusst. Dann gilt

$$P(A \cap B) = P(A) \cdot P(B)$$

6.7 Satz der totalen Wahrscheinlichkeit

Sei $A_1, A_2, \ldots, A_k$ eine disjunkte Zerlegung des Ereignisraumes Ω. Wenn $B \subset \Omega$, dann gilt

$$P(B) = \sum_{i=1}^{k} P(A_i) \cdot P(B|A_i)$$

6.8 Satz von Bayes

Ist die Verbindung zwischen den Wahrscheinlichkeiten $P(A|B)$ und $P(B|A)$. Diese bedingten Wahrscheinlichkeiten sind nicht gleich!

Wenn der Ereignisraum Ω in A und $\bar{A}$ partitioniert wird und B ein beliebiges Ereignis aus Ω ist, gilt

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$$

Der Satz von Bayes ergibt sich aus Kombination von $P(A|B) = \frac{P(A \cap B)}{P(B)}$ und $P(B|A) = \frac{P(B \cap A)}{P(A)}$.

7 Test

[2] [1]

Literatur

[1] Benjamin Buchwitz. *Statistics*. 2024. URL: https://bchwtz.github.io/bchwtz-stat/ (besucht am 25. 07. 2024).

[2] Dipl. Stat. Adriane Sommer Prof. Dr. Monika Reimpell. *Studienbuch Statistik*. 2st. Frauenstuhlweg 31, 58644 Iserlohn: Wisschenschaftliche Genossenschaft Südwestfahlen eG, 2014.