Grundlagen der Statistik

Prof. Dr. Jörg große Schlarmann

Einführung in die deskriptive und inferentielle Statistik

Version vom 19.06.2025

Lizenz

Dies ist ein grundlegender Leitfaden der Statistik, der die deskriptive Statistik, Regression, Wahrscheinlichkeit und die Grundlagen der Inferenz abdeckt (Konfidenzintervalle und parametrische Hypothesentests). Grundlage bildet dabei das Werk von Sánchez Alberca et al. (2023)[1], dessen Kapitel in der vorliegenden Version umfangreich ergänzt wurden.

Der Quelltext des Buches steht unter https://github.com/produnis/statistics-manual zur Verfügung.

Zitationsvorschlag

große Schlarmann, J (2025): “Grundlagen der Statistik. Einführung in die deskriptive und inferentielle Statistik”, Hochschule Niederrhein, https://www.produnis.de/manual/

```
@book{grSchl_exeRcise,
    author = {{große Schlarmann}, Jörg},
    title = {Grundlagen der Statistik. Einführung in die deskriptive und inferentielle Statistik},
    year = {2025},
    publisher = {Hochschule Niederrhein},
    address = {Krefeld},
    copyright = {CC BY-NC-SA 4.0},
    url = {https://www.produnis.de/manual},
    language = {de},
}
```

[1] siehe https://github.com/asalber/statistics-manual
[2] siehe https://creativecommons.org/licenses/by-nc-sa/4.0/

Inhaltsverzeichnis

1 Einführung

1.1 Was ist Statistik?

Definition "Statistik"

Die Statistik ist ein mathematisches Gebiet, das sich mit der Sammlung, Zusammenfassung, Analyse und Interpretation von Daten befasst. Ihr Zweck besteht darin, Informationen aus Daten zu extrahieren, um Entscheidungen zu treffen.

Die Statistik spielt eine wichtige Rolle in jeder wissenschaftlichen oder technischen Disziplin, die mit Datenarbeit zu tun hat, besonders bei großen Datenvolumen, wie z.B. Physik, Chemie, Medizin, Psychologie, Wirtschaft oder Sozialwissenschaften.

Statistik…

- verschafft Überblick.
- bringt Ordnung ins Chaos.
- gibt Sicherheit.
- quantifiziert mögliche Fehler.
- ermöglicht Entscheidungen.

1.1.1 Warum ist Statistik notwendig?

- **Eine sich ändernde Welt:** Wissenschaftler versuchen die Welt zu studieren. Die Welt ist hoch variabil, was es erschwert, das Verhalten von Dingen vorherzusagen.
- **Variabilität:** Diese Variabilität ist der Grund, warum Statistik existiert. Sie dient als Brücke zwischen der realen Welt und den mathematischen Modellen, die diese Welt erklären möchten. Die Statistik bietet eine Methode zur Beurteilung der Unterschiede zwischen Realität und theoretischen Modellen.
- **Evidenzbasierung:** wissenschaftliche *Beweise* (Evidenz) sind mathematische Beweise. Wenn beispielsweise die Überlegenheit von Anwendung A gegenüber Anwendung B gezeigt werden soll, muss dies mathematisch erfolgen. Statistik ist ein Dialekt der Mathematik, der hierfür verwendet werden kann.

1.2 Population und Stichprobe

Definition "Population"

Eine Population ist eine Gruppe von Elementen (z.B. Menschen), die durch eine oder mehrere Eigenschaften definiert sind. Diese Eigenschaften treffen auf jedes Element zu und sind nur bei ihnen vorhanden.

Definition "Grundgesamtheit"

Die Anzahl der Individuen in einer Population wird als Grundgesamtheit bezeichnet und durch N dargestellt.

Manchmal sind nicht alle Individuen für die Studie erreichbar. Dann unterscheiden wir zwischen:

- Theoretischer Grundgesamtheit: Die Individuen, auf die wir die Studienergebnisse übertragen möchten.
- Beforschbare Grundgesamtheit: Die Individuen, die wirklich in der Studie erreichbar sind.

1.2.1 Vollerhebung

Wissenschaftlerinnen studieren ein Phänomen in einer Bevölkerungsgruppe, um es zu verstehen, um Wissen über es zu erlangen und es dadurch möglichst zu kontrollieren.

Allerdings ist ein vollständiges Verständnis der Grundgesamtheit nur dann möglich, wenn alle ihre Individuen studiert werden (so genannte "Vollerhebung"). Doch dies ist aus mehreren Gründen nicht immer möglich:

- Die Population ist zu groß, um sämtliche Individuen untersuchen zu können.
- Die Interventionen, denen die Individuen unterzogen werden, sind gefährlich oder schädlich.
- Die Kosten – sowohl in Bezug auf Geld als auch Zeit –, die notwendig wären, um alle Individuen in der Bevölkerung zu studieren, sind unverhältnismäßig hoch.

1.2.2 Stichproben

Wenn es nicht möglich oder zu aufwendig ist, alle Individuen einer Population zu studieren, untersuchen wir nur eine Teilmenge von ihnen.

Definition "Stichprobe"

Eine Stichprobe ist eine Teilmenge der Population.

Definition "Stichprobenumfang"

Die Anzahl der Individuen in einer Stichprobe wird als Stichprobenumfang bezeichnet und mit n dargestellt.

Definition "Proband"

Probanden sind Menschen, die sich freiwillig dazu bereit erklärt haben, an einer Studie teilzunehmen.

Im Allgemeinen werden Studien anhand von Stichproben, die aus der Grundgesamtheit entnommen wurden, durchgeführt.

Zwar bieten die Ergebnisse einer Stichprobenanalyse nur Annäherungen über die Gegebenheiten in der Grundgesamtheit, doch in den meisten Fällen ist dies ausreichend.

1.2.3 Bestimmung des Stichprobenumfangs

Eine der interessantesten Fragen, die sich stellen, lautet:

> Wie viele Probanden werden benötigt, um eine Stichprobe zu ziehen, die annehmbare Aussagen über die Grundgesamtheit zulässt?

Die Antwort hängt von mehreren Faktoren ab, wie z.B. der Variabilität innerhalb der Population oder der gewünschten Zuverlässigkeit für Schlussfolgerungen über die Grundgesamtheit. Im Allgemeinen gilt jedoch:

> Je größer die Stichprobe ist, desto zuverlässiger werden die Schlussfolgerungen über die Grundgesamtheit sein, jedoch wird die Studie dadurch länger und teurer werden.

Beispiel

Um zu verstehen, was “ausreichende” Stichprobe bedeutet, können wir ein Beispiel mit einem Bild verwenden. Eine digitale Fotografie besteht aus vielen kleinen Punkten (Pixel), die in einer großen Matrix aus Zeilen und Spalten angeordnet sind. Je mehr Zeilen und Spalten vorhanden sind, desto höher die Auflösung des Bildes. In dieser Analogie entspricht das Bild der Grundgesamtheit und jeder Pixel stellt ein Individuum dar. Im Bild hat jeder Pixel eine Farbe, und es ist die Variabilität dieser Farben, die das Motiv des Bildes bildet.
Wie viele Pixel müssen wir in einer Stichprobe aufnehmen, um das Bild erkennen zu können?
Die Antwort hängt von der Variabilität der Farben im Bild ab. Wenn alle Pixelex im Bild dieselbe Farb haben, reicht nur ein Pixel aus, um das Motiv zu kennen. Wenn es eine hohe Variabilität in den Farben gibt, wird ein größerer Stichprobenumfang benötigt.

Die unten stehende Abbildung enthält eine kleine Stichprobe von Pixeln eines Fotos. Können Sie das Motiv des Bildes erkennen?

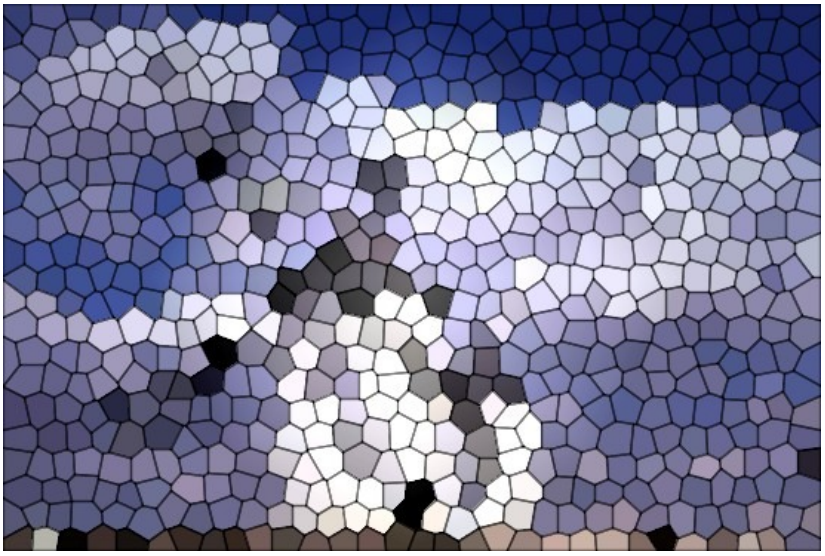


Abbildung 1.1: Mit einem kleinen Stichprobenumfang ist es schwierig, das Bildmotiv zu ermitteln!

Wahrscheinlich ist es Ihnen schwer gefallen das Motiv zu erkennen, weil die Anzahl der ausgewählten Pixel in der Stichprobe zu gering ist, um die Farbenvielfalt im Bild zu verstehen.

Die nächste Abbildung unten enthält eine größere Stichprobe. Könnten Sie das Motiv des Bildes jetzt erkennen?

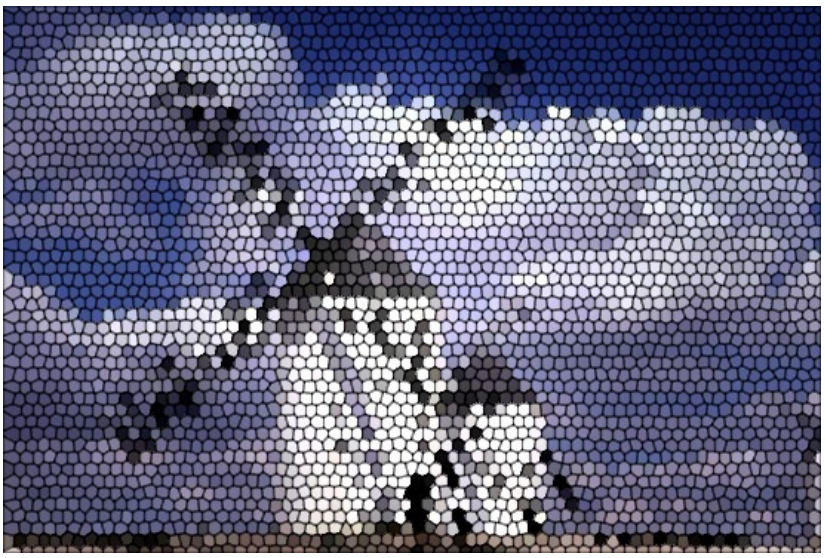


Abbildung 1.2: Mit einer großen Stichprobe ist es leichter, das Motiv des Bildes zu erkennen!

Abbildung 1.3: Originalauflösung

Der Vollständigkeit halber enthält nachfolgende Abbildung alle Pixeln der Grundgesamtheit.

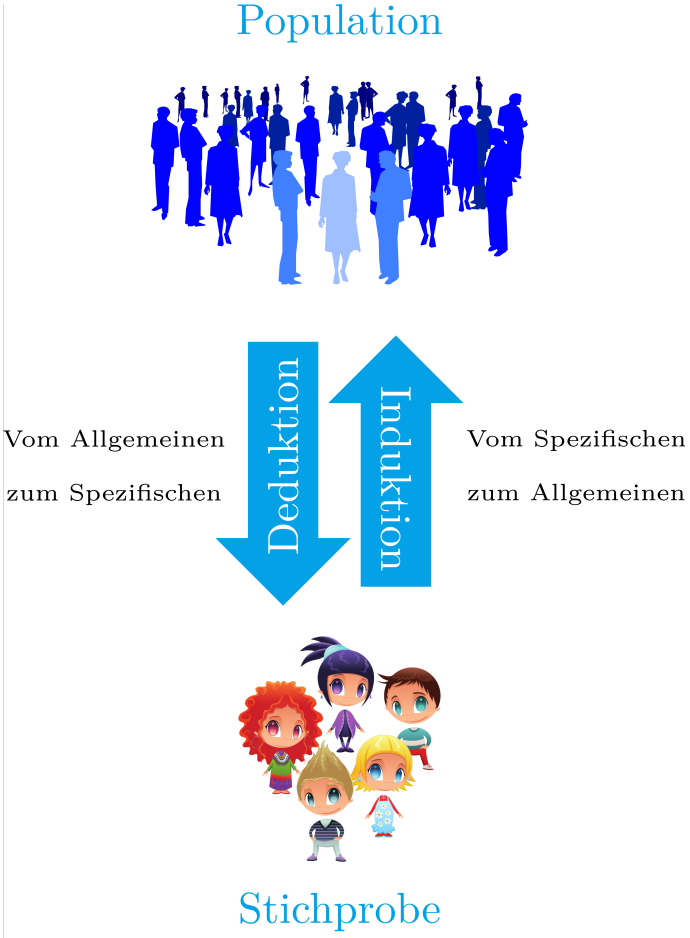
Es ist häufig nicht erforderlich, alle Pixel eines Fotos zu verwenden, um das Motiv zu erkennen!

1.2.4 Typen des logischen Denkens

Deduktive Eigenschaften: Wir schließen vom Allgemeinen auf das Spezifische. Wenn etwas in der Population *vorhanden* ist, dann ist es auch in der Stichprobe *vorhanden*. Diese Art des Schlussfolgerns bringt keine wirklich *neuen* Kenntnisse, sondern verifiziert lediglich die Vorannahmen.

Induktive Eigenschaften: Wir schließen vom Spezifischen auf das Allgemeine. Wenn etwas in der Stichprobe *vorhanden* ist, muss es zwar nicht *zwangsläufig* in der Grundgesamtheit *vorhanden* sein (also vorsichtig mit Extrapolationen). Allerdings ist dies der einzige Weg, um neue Kenntnisse zu generieren.

Die Statistik basiert grundlegend auf der induktiven Schlußweise, weil sie die Informationen, die von Stichproben gewonnen wurden, verwendet, um Rückschlüsse über die Grundgesamtheit zu ziehen.



1.3 Sampling

Definition "Sampling"

Sampling beschreibt den Prozess, wie Elemente der Grundgesamtheit (Probanden) in die Stichprobe aufgenommen werden.

Um zuverlässige Informationen über die Grundgesamtheit zu erhalten, muss die Stichprobe *repräsentativ* sein (siehe hierzu von der Lippe (2002) (2011)). Das bedeutet, dass die Stichprobe die Variabilität der Grundgesamtheit auf kleinerem Maßstab nachbildet. Das Ziel des Sampling ist es, eine solche Stichprobe zu erhalten (was streng genommen nur bei Zufallsstichproben möglich ist, siehe von der Lippe (2002), (2010))!

1.3.1 Stichprobentypen

Es existieren viele Samplingmethoden, doch diese erzeugen immer eine der beiden nachfolgend genannten Stichprobentypen:

- **Zufallsstichprobe:** Die Individuen der Stichprobe werden zufällig ausgewählt. Jedes Mitglied der Grundgesamtheit hat dabei die selbe Chance, ausgewählt zu werden.
- **keine Zufallsstichprobe:** Die Individuen der Stichprobe werden nicht *wirklich* zufällig ausgewählt. Einige Mitglieder der Grundgesamtheit haben eine höhere Chance ausgewählt zu werden als andere.

Beispiel “Gelegenheitsstichprobe”

Sie wollen das Durchschnittseinkommen Ihrer Stadt ermitteln. Hierzu begeben Sie sich zum Hauptbahnhof, und befragen dort *zufällig* 100 Personen. In diesem Fall handelt es sich **nicht** um eine Zufallsstichprobe, da nicht alle Einwohner Ihrer Stadt (Grundgesamtheit) die selbe Chance hatten, von Ihnen befragt zu werden. Jedes Individuum, das nicht zu Ihrem Befragungszeitraum am Hauptbahnhof war, hatte eine Samplingchance von 0%. Man spricht in diesem Fall auch von einer *Gelegenheitsstichprobe*, weil Sie “nur” diejenigen befragt haben, die “gerade da waren”.
Die meisten klinischen Studien sind Gelegenheitsstichproben. Man nimmt die Patienten, die gerade auf Station liegen.

Nur Zufallsstichproben können einer gewissen Repräsentativität nahe kommen.

Nicht-zufällige Stichproben eignen sich nicht zur Generalisierung, da sie wahrscheinlich von den Gegebenheiten in der Grundgesamtheit abweichen. Dennoch gibt es gerade in der klinischen Forschung häufig keine andere Möglichkeit zur Probandenrekrutierung.

1.3.2 Einfache Zufällige Stichprobenentnahme

Die beliebteste zufällige Samplingmethode ist die einfache zufällige Stichprobenentnahme, die folgende Eigenschaften aufweist:

1. **Chancengleichheit:** Jedes Individuum in der Grundgesamtheit hat die selbe Chance zur Auswahl.
2. **Ziehen mit Zurücklegen:** Die ausgewählten Individuen werden zurück in die Grundgesamtheit gegeben, bevor das nächste Individuum ausgewählt wird. Auf diese Weise bleibt die Population während der gesamten Entnahme unverändert.
3. **Unabhängigkeit:** Die Auswahl eines Individuals ist unabhängig von der Auswahl anderer Individuen.

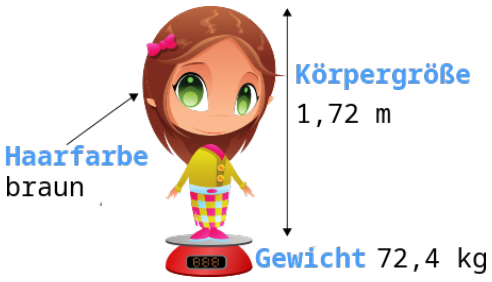
Die einzige Möglichkeit, eine Zufallsstichprobe zu ziehen, besteht darin, jedem Mitglied der Population eine eindeutige Identifikationsnummer zuzuordnen (was einem Zensus entspricht) und danach eine zufällige Auswahl vorzunehmen.

1.4 Statistische Variablen

In jeder statistischen Untersuchung interessieren wir uns für bestimmte Eigenschaften oder Merkmale von Individuen.

Definition “Statistische Variable”

Eine statistische Variable ist eine Eigenschaft oder ein Merkmal, das bei den Individuen einer Population gemessen wird.



⚠ Definition “Daten”

Die Daten sind die tatsächlichen Werte oder Ergebnisse, die für eine bestimmte Variable erhoben wurden.

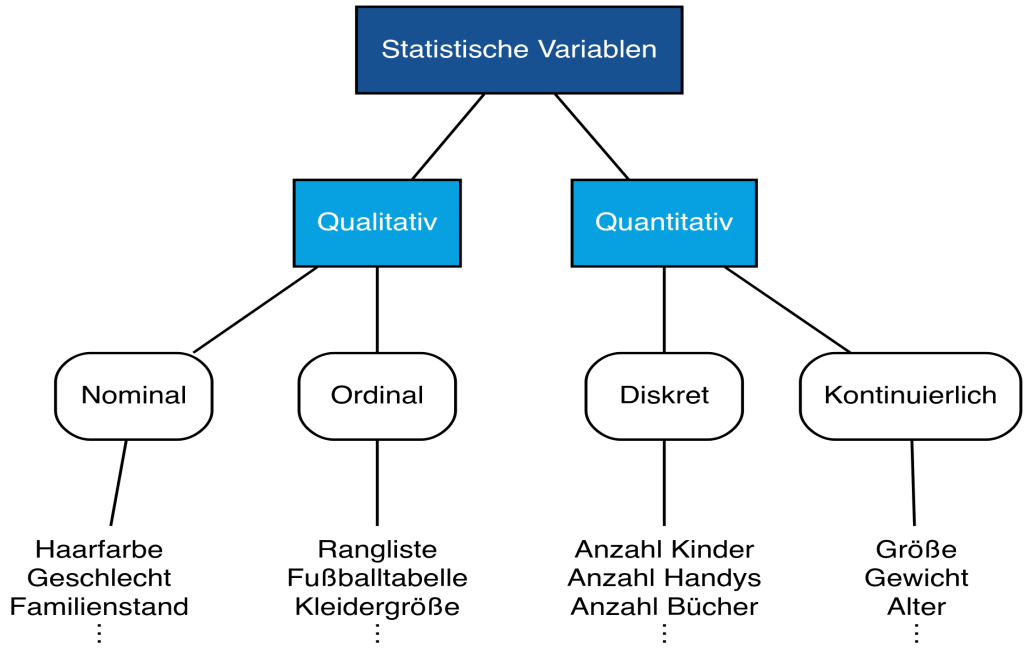
1.4.1 Skalenniveau

Eine Skala dient dazu, irgend etwas einzuteilen, zu ordnen oder zu sortieren. In der Statistik unterscheiden sich die 3 Skalen in ihrem Informationsgehalt und in der Möglichkeit der statistischen Auswertung.

Laut der Art ihrer Werte und ihrem Skalenniveau lassen sich Variablen wie folgt unterscheiden:

- **Qualitative Variablen:** enthalten keine numerischen Werte, sondern *Text*.
  - **Nominal:** Es gibt keine natürliche Reihenfolge zwischen den Werten. Beispiel: Haarfarbe oder Geschlecht.
  - **Ordinal:** Es gibt eine natürliche Reihenfolge zwischen den Werten. Beispiel: Die Wochentage oder Schulnoten, oder alphabetisch sortierte Nachnamen.
- **Quantitative Variablen:** messen numerische Werte.
  - **Metrisch:** alles, was man messen oder zählen kann.
    - **Diskret:** Die Werte sind isolierte Zahlen (häufig ganze Zahlen). Beispiel: Die Anzahl der Kinder oder Autos in einer Familie.
    - **Kontinuierlich:** Die Werte können jeden Wert in einem reellen Intervall annehmen. Beispiel: Körpergröße, Gewicht oder Alter eines Menschen.

Qualitative und diskrete Variablen werden auch als kategoriale Variablen bezeichnet, und ihre Werte entsprechen Kategorien.



1.4.2 Nominalskala

Bei einer Nominalskala (lat. Nomen = Name, Bedeutung) entsprechen die erhobenen Werte nur einer Art von Etikett, "man gibt dem Kind einen Namen", ohne dass irgendeine Wertigkeit oder Reihenfolge zugrunde liegt. Man kann auch sagen, dass Objekte oder Ereignisse in Kategorien eingeteilt werden, die sich gegenseitig ausschließen. Entweder hat ein Objekt ein bestimmtes Merkmal oder nicht.

Beispiele für nominalskalierte Daten

- Geschlecht (männlich, weiblich, …)
- Augenfarbe (blau, grün, braun, …)
- Körperbau (leptosom, pyknisch, athletisch, …)
- bestanden (ja, nein)

1.4.3 Ordinalskala

Zusätzlich zu den Eigenschaften der Nominalskala unterliegen Daten einer Ordinalskala (lat. Ordinatus = Ordnung, ordentlich) einer vorgegebenen Reihenfolge oder Rangfolge.

Beispiele für ordinalskalierte Daten

- Steigerung (gut, besser, am besten)
- Schulnote (sehr gut, gut, befriedigend, ausreichend, mangelhaft, ungenügend)
- Siegerehrung (erster, zweiter, dritter)

1.4.4 metrische Skala

Das höchste Niveau besitzt die metrische Skala. Die Werte unterliegen nicht nur einer Reihenfolge, benachbarte Werte weisen auch gleiche Abstände auf.

Beispiele für metrische Daten

- Lebensalter in Jahren
- Gewicht in kg
- Länge in cm
- Lohn in Euro

Unterschied zwischen metrisch und ordinal

Die Abstände bei z.B. Schulnoten sind **nicht** gleich bzw. konstant! Darf man sich beispielsweise für die Note 1 nur drei Fehler erlauben, sind es bei der Note 2 schon sieben, und für die Note 3 schon zwanzig Fehler!
Die Leistungsdifferenz zwischen den Noten ist also **nicht konstant**!

1.4.5 Wählen der passenden Variablen

Manchmal kann eine Eigenschaft auf verschiedene Arten gemessen werden.

Beispiel Rauchen

Ob ein Mensch raucht oder nicht könnte auf folgende Weise gemessen werden:

- **Raucht:** Ja/Nein. (Nominal)
- **Rauchstufe:** Nichtrauchen/unüblich/mäßig/quasi/heftig. (Ordinal)
- **Anzahl der Zigaretten pro Tag:** 0,1,2,… (Diskret)

In solchen Fällen sind quantitative Variablen gegenüber qualitativen vorzuziehen, kontinuierliche Variablen werden den diskreten vorgezogen und ordinalen Variablen werden nominalen vorgezogen, da sie mehr Information liefern.

Laut ihrer Rolle in der Untersuchung unterscheiden wir:

- **Unabhängige Variablen:** Variablen, die nicht von anderen Variablen abhängig sind. Sie werden im Experiment oft manipuliert, um ihren Einfluss auf eine abhängige Variable zu beobachten. Sie werden auch als Prädiktionsvariablen bezeichnet.
- **Abhängige Variablen:** Variablen, die von anderen Variablen abhängig sind. Sie werden im Experiment nicht manipuliert und werden auch als Ergebnisvariablen bezeichnet.

Beispiel

In einer Studie über das Leistungsniveau von Studierenden in einem Kurs sind die Intelligenz der Schüler und die tägliche Lernzeit unabhängige Variablen, während die Note im Kurs eine abhängige Variable ist.

1.4.6 Arten statistischer Studien

- **Experimentelle Studien:** unabhängige Variablen werden manipuliert, um den Einfluss dieser Änderung auf die abhängigen Variablen zu beobachten.

Beispiel

In einer Studie über das Leistungsniveau von Studierenden bei einer Prüfung manipuliert die Dozentin bestimmte Rahmenbedingungen im Prüfungsraum. Sie bildet zwei oder mehr Gruppen mit unterschiedlichen Rahmenbedingungen (z.B. Kekse und Getränke zur Verfügung stellen) und vergleicht die Ergebnisse beider Gruppen.

- **Nicht-experimentelle Studien:** unabhängige Variablen werden **nicht** manipuliert. Das bedeutet nicht, dass es unmöglich wäre, sondern dass es entweder praktisch nicht möglich oder ethisch nicht vertretbar wäre.

Beispiel

In einer Studie interessieren sich Forschende für den Einfluss von Rauchen auf Lungenkrebs. Während es möglich ist, Menschen dazu zu bitten, zu rauchen, um den Einfluss dieser Tätigkeit auf ihre Lungen zu studieren, wäre es jedoch ethisch höchst fragwürdig. In diesem Fall können die Forschenden zwei Gruppen von Menschen untersuchen: Raucher und Nichtraucher, und deren Lungenfunktionen und Krebsquoten miteinander vergleichen.

Merke

Experimentelle Studien ermöglichen es, einen **kausalen** Zusammenhang zwischen Variablen zu identifizieren, während nicht-experimentelle Studien nur Beziehungen oder Verhältnisse zwischen Variablen identifizieren können.

1.4.7 Datentabelle

Die Variablen einer Studie werden bei jedem Individuum der Stichprobe gemessen. Das ergibt eine Datensammlung, die üblicherweise in tabellarischer Form angeordnet wird und als **Datenstrom** bezeichnet wird.

In dieser Tabelle enthält jede Spalte Informationen zu einer Variablen, während jede Zeile Informationen zu einem Individuum enthält.

Beispiel

Name	Alter	Geschlecht	Gewicht (kg)	Größe (cm)
Anna Tomie	42	W	85	179
Bud Zillus	31	M	115	173
Dieter Mietenplage	27	M	79	181
Hella Scheinwerfer	29	W	60	170
Inge Danken	36	W	57	158
Jason Zufall	54	M	96	174

1.5 Phasen einer statistischen Untersuchung

Eine statistische Untersuchung durchläuft meist die folgenden Phasen:

1. Die Studie beginnt mit einer vollständigen Planung, in der die Studienziele, die Population, die zu messenden Variablen und der benötigte Stichprobenumfang festgelegt werden.

2. Anschließend wird eine Stichprobe aus der Population gezogen (Sampling), und die Variablen werden bei den Individuen der Stichprobe gemessen (was die Datentabelle ergibt).
3. Der nächste Schritt besteht darin, die Informationen der Stichprobe zu beschreiben und zu zusammenzufassen. Das ist die Aufgabe der *deskriptiven Statistik*.

4. Anschließend werden die erlangten Informationen auf ein mathematisches Modell projiziert, das beabsichtigt, die Vorgänge innerhalb der Grundgesamtheit zu erklären. Das Modell wird auf seine Gültigkeit überprüft. Dies ist Aufgabe der *inferentiellen Statistik*.

5. Schließlich wird das validierte Modell verwendet, um Vorhersagen zu treffen und Schlussfolgerungen für die Grundgesamtheit zu ziehen.

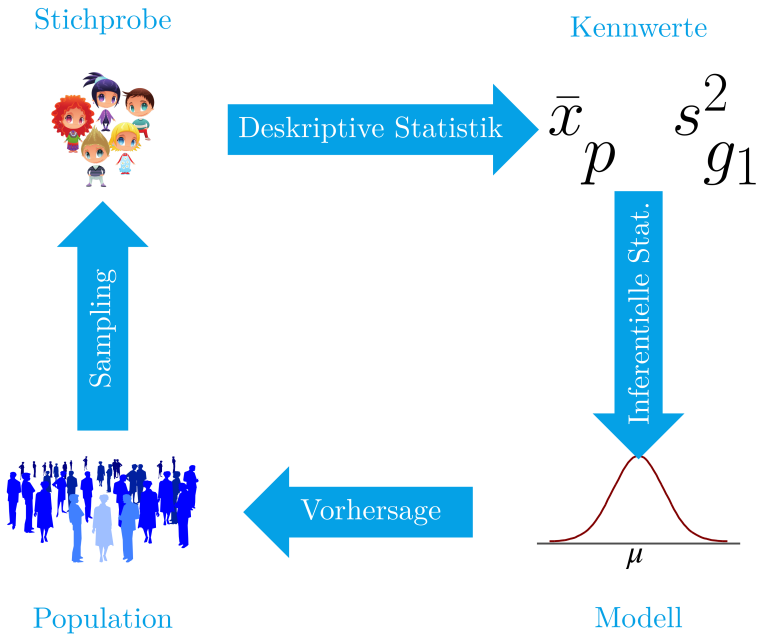


Abbildung 1.4: Der statistische Kreislauf

2 Häufigkeitsverteilungen

2.1 Deskriptive Statistik

Die **deskriptive Statistik** ist der Bereich innerhalb der Statistik, der sich mit dem Darstellen, Analysieren und Zusammenfassen von Informationen in einer Stichprobe beschäftigt. Nach dem Samplingprozess ist dies der nächste Schritt in jeder Studie und besteht meist aus folgenden Teilen:

1. Ordnen und Klassifizieren der Daten sowie Einteilung in Gruppen.
2. Tabellarische und grafische Darstellung der Daten nach ihrer Häufigkeit.

3. Berechnung numerischer Maßzahlen, die die Informationen in der Stichprobe zusammenfassen (Stichprobenstatistiken).

> ! Merke
>
> die Stichprobenstatistik hat kein Inferenzpotential, weshalb keine Verallgemeinerungen auf die Population angewendet werden sollten!

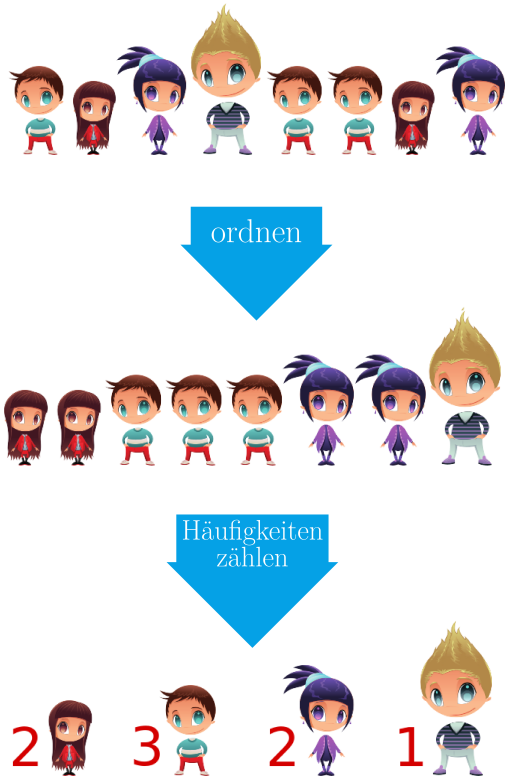
2.2 Häufigkeitsverteilung

2.2.1 Stichprobenklassifizierung

Das Studium einer statistischen Variablen beginnt mit der Messung der Variable bei den Individuen der Stichprobe und der Klassifizierung der Werte. Es gibt zwei Weisen, Daten zu klassifizieren:

- **Klassifizierung ohne Klassierung**: Sortieren der Werte vom kleinsten zum größten (wenn eine Reihenfolge existiert). Diese Methode wird bei qualitativen Variablen und diskreten Variablen mit wenigen eindeutigen Werten verwendet.

- **Klassifizierung mit Klassierung**: Gruppieren der Werte in Intervalle (Klassen) und Sortieren dieser Intervalle vom kleinsten zum größten. Diese Methode wird bei kontinuierlichen Variablen und diskreten Variablen mit vielen eindeutigen Werten verwendet.



2.2.2 Häufigkeiten

Definition "Häufigkeiten"

Bei einer Stichprobe mit n Werten der Variable X können wir für jedes x_i bestimmen:

1. **Absolute Häufigkeit (n_i):** ist die Anzahl, mit der ein bestimmter Wert x_i in der Stichprobe auftritt.
2. **Relative Häufigkeit (f_i):** ist das Verhältnis der Anzahl eines bestimmten Wertes x_i zur Gesamtanzahl der Beobachtungen in der Stichprobe.

$$f_i = \frac{n_i}{n}$$

3. **Kumulative absolute Häufigkeit (N_i):** ist die Gesamtanzahl der Werte in der Stichprobe, die kleiner oder gleich x_i sind

$$N_i = n_1 + \cdots + n_i$$

4. **Kumulative relativeHäufigkeit (F_i):** ist das Verhältnis der kumulativen absoluten Häufigkeit zur Gesamtanzahl der Beobachtungen in der Stichprobe.

$$F_i = \frac{N_i}{n}$$

2.2.3 Häufigkeitstabelle

Die Menge der Werte einer Variablen mit ihren jeweiligen Häufigkeiten nennt man *Häufigkeitsverteilung*, und sie wird üblicherweise als *Häufigkeitstabelle* dargestellt.

Tabelle 2.1: Häufigkeitstabelle

Variable X	**absolute Häufigkeit**	**relative Häufigkeit**	**kumulierte absolute Häufigkeit**	**kumulierte relative Häufigkeit**
x_1	n_1	f_1	N_1	F_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_i	f_i	N_i	F_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	f_k	N_k	F_k

Beispiel: Kinder in Familien

Die Anzahl an Kindern wurde bei 25 Familien wie folgt erhoben:

1, 2, 4, 2, 2, 2, 3, 2, 1, 1, 0, 2, 2, 0, 2, 2, 1, 2, 2, 3, 1, 2, 2, 1, 2

Die Häufigkeitstabelle für die Anzahl an Kindern in der Stichprobe ist

Tabelle 2.2: Kinder in Familien

x_i	n_i	f_i	N_i	F_i
0	2	0.08	2	0.08
1	6	0.24	8	0.32
2	14	0.56	22	0.88
3	2	0.08	24	0.96
4	1	0.04	25	1.00
$\sum$	**25**	**1.00**		

2.2.4 Klassierung

Beispiel Körpergröße

Von 30 Studierenden wurde die Körpergröße wie folgt gemessen:

179, 173, 181, 170, 158, 174, 172, 166, 194, 185, 162, 187, 198, 177, 178, 165, 154, 188, 166, 171, 175, 182, 167, 169, 172, 186, 172, 176, 168, 187.

Die dazugehörige Häufigkeitstabelle wäre sehr lang, da sich nur sehr wenige Werte wiederholen. Es ist daher üblich, Werte aus bestimmten Bereichen in so genannten “Klassen” zusammenzufassen.

Definition “Klasse”

Eine Klasse ist die Menge sämtlicher Messwerte, die innerhalb festgelegter Grenzen liegt. Diese festgelegten Klassen*grenzen* werden durch den kleinsten und den größten Messwert einer Klasse gebildet. Die Klassen*mitte* ist das arithmetische Mittel aus beiden Klassengrenzen. Die Klassen*breite* ist die Differenz der Klassengrenzen.

Bei der Klassierung sollten grundsätzlich folgende Regeln beachtet werden:

- die Anzahl der Klassen sollte nicht zu groß und nicht zu klein sein. Als Fausregel kann die Klassenanzahl auf $\sqrt{n}$ bzw. $log_2(n)$ gelegt werden. Für das Beispiel der Körpergrößen von 30 Studierenden gergeben sich $\sqrt{30} = 5,48 \approx$ **5** Klassen.
- die Klassen dürfen sich nicht überschneiden und müssen den gesamten Wertebereich abdecken. Ob die Intervalle links offen und rechts geschlossen sind oder umgekehrt, spielt keine Rolle.
- Der kleinste Wert muss in der ersten Klasse liegen und der größte Wert in der letzten.
- Die Klassengrenzen werden mir runde und eckigen Klammer angegeben. Eine eckige Klammer bedeutet, dass der nachstehen Wert in die Klasse eingeschlossen wird, eine runde Klammer schließt die nebenstehende Zahl aus.

- (oder) bedeutet größer oder kleiner **als**.
- [oder] bedeutet größer-**gleich** oder kleiner-**gleich**.
- bei (und) gehört der Wert **nicht** mehr zu der Klasse.
- bei [und] **gehört** der Wert zu der Klasse.

Eine Häufigkeitstabelle der klassierten Körpergrößen mit 5 Klassen sieht so aus:

x_i	n_i	f_i	N_i	F_i
(150,160]	2	0.07	2	0.07
(160,170]	8	0.27	10	0.34
(170,180]	11	0.36	21	0.70
(180,190]	7	0.23	28	0.93
(190,200]	2	0.07	30	1.00
$\sum$	**30**	**1.00**		

2.2.5 Häufigkeitstabellen qualitativer Daten

Beispiel Blutgruppen

Bei 30 Studierenden wurden die Blutgruppen wie folgt bestimmt:

A, B, B, A, AB, 0, 0, A, B, B, A, A, A, A, AB, A, A, A, B, 0, B, B, B, A, A, A, 0, A, AB, 0

Die Häufigkeitstabelle der Blutgruppen ist:

x_i	n_i	f_i
0	5	0.16
A	14	0.47
B	8	0.27
AB	3	0.10
$\sum$	30	1

Warum gibt es keine kumulativen Werte in der Tabelle?

Da es sich um nominale Werte handelt, gibt es keine natürliche Reihenfolge. Theoretisch gäbe es 24 Kombinationsmöglichkeiten, um eine Reihenfolge zu erzwingen. Diese sind aber alle aussagelos.

2.3 Diagramme

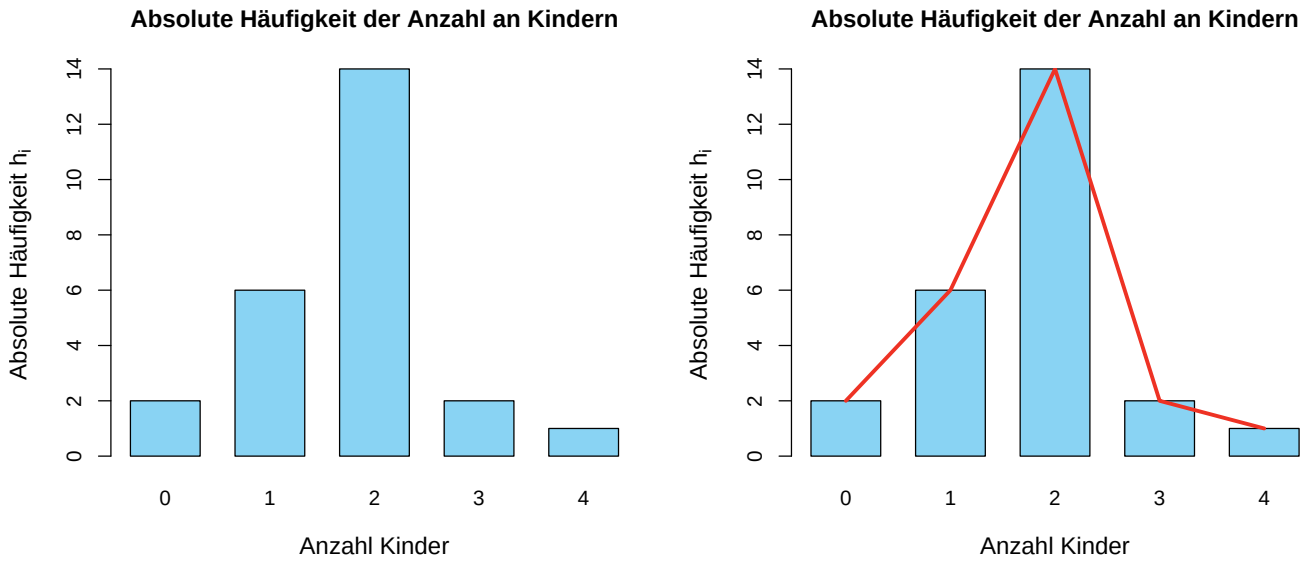
Häufigkeitsverteilungen werden üblicherweise auch grafisch dargestellt. Abhängig von dem Typ der Variablen und ob die Daten klassiert wurden, gibt es verschiedene Arten von Diagrammen:

- Säulendiagramm
- Histogramm
- Liniendiagramm
- Kreisdiagramm

2.3.1 Säulendiagramm

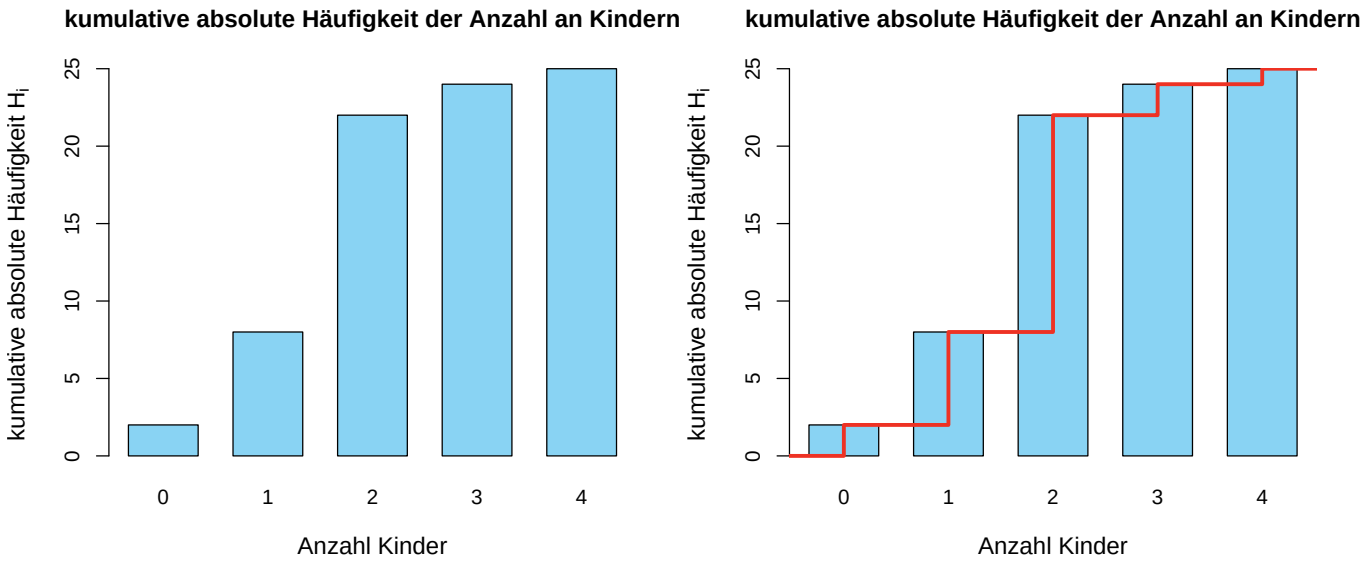
Ein Säulendiagramm besteht aus einer Reihe von Balken, einem für jeden Wert oder jede Kategorie der Variablen, die auf einem Koordinatensystem dargestellt sind. Üblicherweise werden die Werte oder Kategorien der Variable auf der x-Achse und die Häufigkeiten auf der y-Achse dargestellt. Für jeden Wert oder jede Kategorie der Variablen wird ein Balken in der Höhe seiner Frequenz gezeichnet. Die Breite des Balkens ist nicht wichtig, aber die Balken sollten klar voneinander getrennt sein.

Abhängig von dem Typ der auf der y-Achse dargestellten Häufigkeit erhalten wir verschiedene Arten von Säulendiagrammen. Manchmal wird ein Polygon, bekannt als Frequenzpolygon, gezeichnet, indem die Spitzen jeder Bar durch gerade Linien verbunden werden.

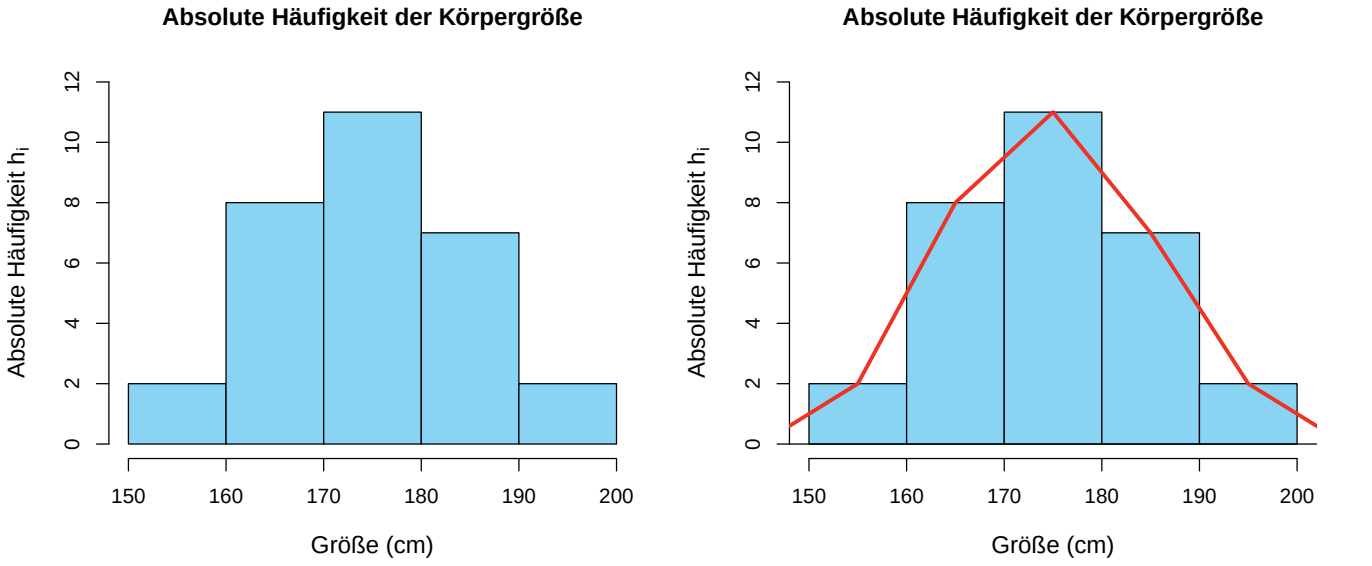


2.3.2 Histogramm

Ein Histogramm ist ähnlich wie ein Säulendiagramm, aber für klassierte Daten.
Die Klassen oder Gruppierungsintervalle werden üblicherweise auf der x-Achse dargestellt, und die Häufigkeiten auf der y-Achse. Für jede Klasse wird ein Balken in der Höhe seiner Frequenz gezeichnet.



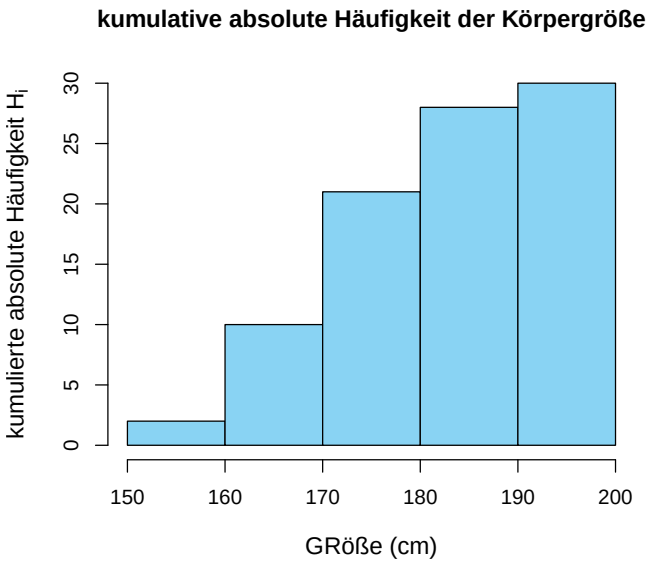
Im Gegensatz zu Säulendiagrammen entspricht die Breite eines Balkens der Breite einer Klasse, und es gibt keinen Platz zwischen zwei aufeinanderfolgenden Balken. Abhängig von dem Typ der auf der y-Achse dargestellten Häufigkeit erhalten wir verschiedene Arten von Histogrammen. Manchmal wird ein Polygon, bekannt als Frequenzpolygon, gezeichnet, indem die Spitzen jeder Bar verbunden werden.

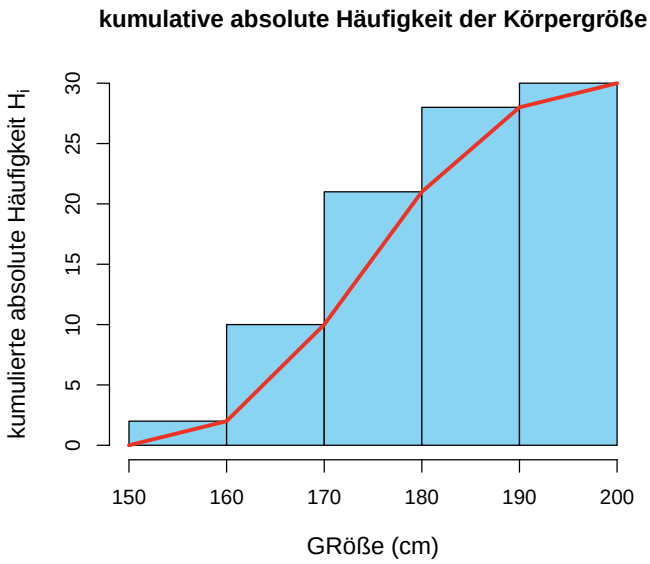


Das kumulative Häufigkeitspolynom (für absolute oder relative Häufigkeiten) wird als **Ogive** bezeichnet. Beachten Sie, dass in der Ogive die rechte obere Ecke jeder Bar mit der Linien verbunden wird, anstatt des mittleren Punkts, weil wir die akkumulierte Häufigkeit einer Klasse nur am Ende des Intervalls erreichen.

2.3.3 Kreisdiagramm

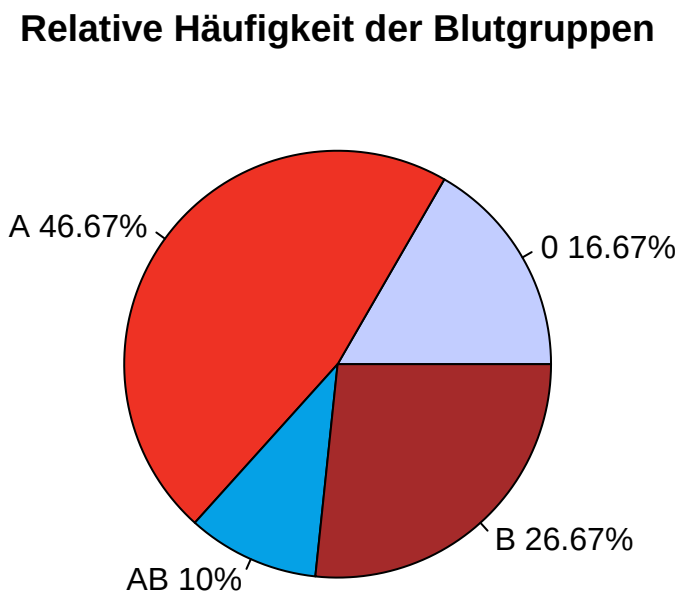
Ein Kreisdiagramm (auch *Tortendiagramm*) besteht aus einem Kreis, der in Scheiben geteilt ist, einer für jeden Wert oder jede Kategorie der Variablen. Jede Scheibe wird als Sektion bezeichnet, und ihr Winkel oder Gebiet ist proportional zur Häufigkeit des entsprechenden Wertes.





Tortendiagramme können absolute oder relative Häufigkeiten darstellen, aber nicht kumulative Häufigkeiten, und sie werden bei nominalen Variablen verwendet.

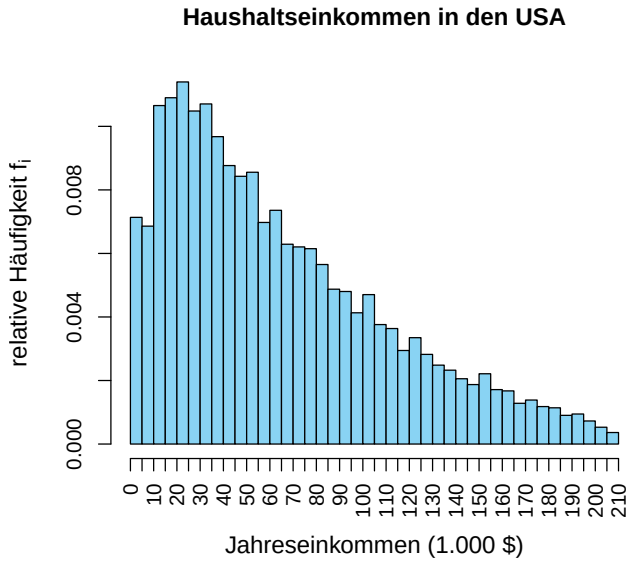
Für ordinal oder quantitative Variablen ist es besser, Säulendiagramme oder Histogramme zu verwenden, weil es einfacher ist, Unterschiede in einer Dimension (Länge der Balken) wahrzunehmen als in zwei Dimensionen (Gebiete der Sektionen).



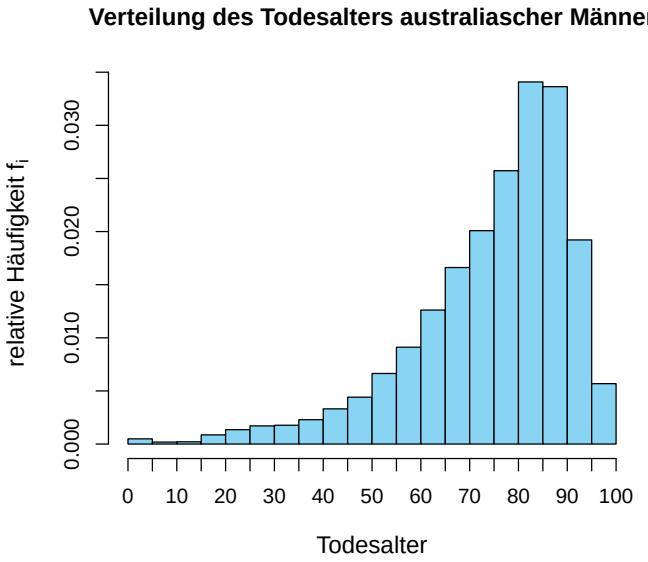
2.3.4 Verteilungsformen

Wenn wir metrische Daten erheben, wie z.B. die Anzahl der Krankenhausaufenthalte, dann stellen wir diese Daten als Häufigkeitstabelle oder Säulendiagramm dar.

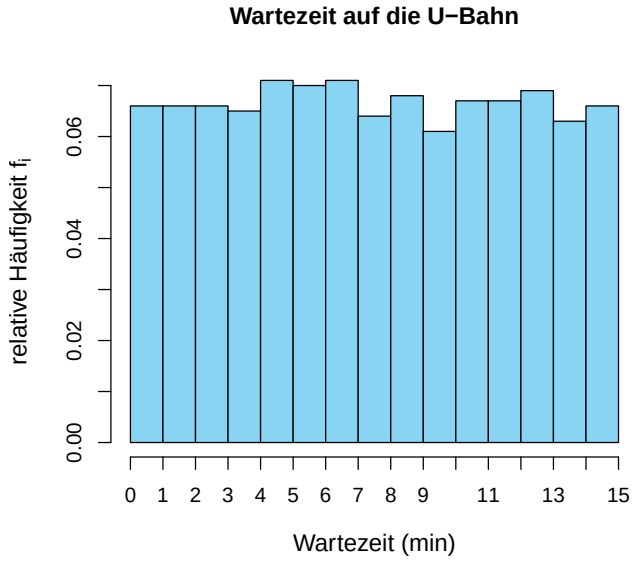
Dabei nehmen die Diagramme häufig *typische* Formen an.



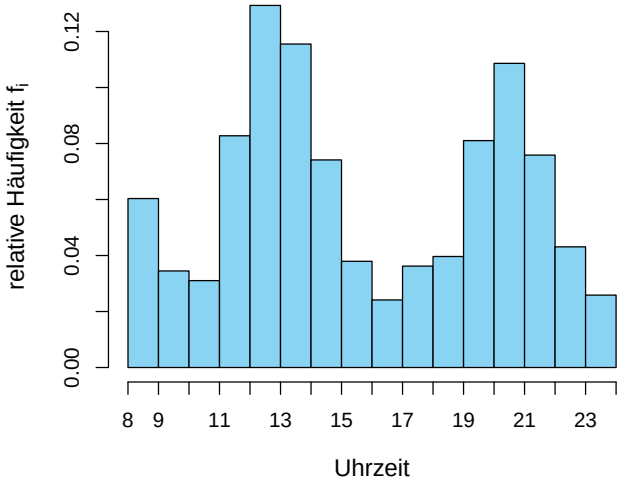
(a) linksgipflig (rechts-schief)



(a) rechtsgipflig (links-schief)

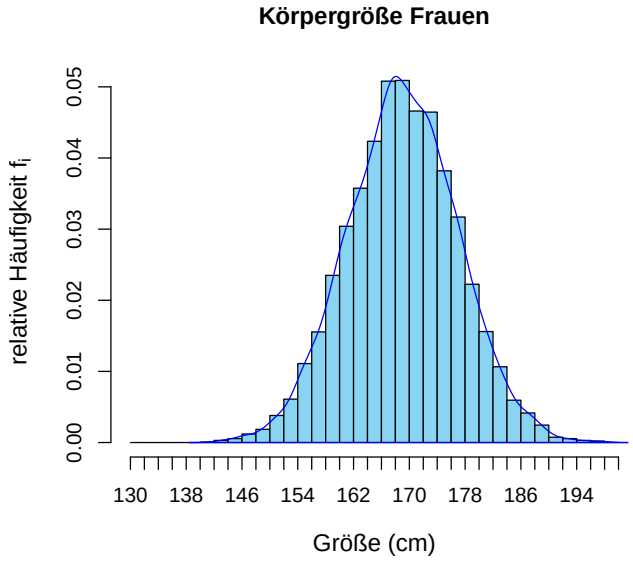


(a) gleichverteilt

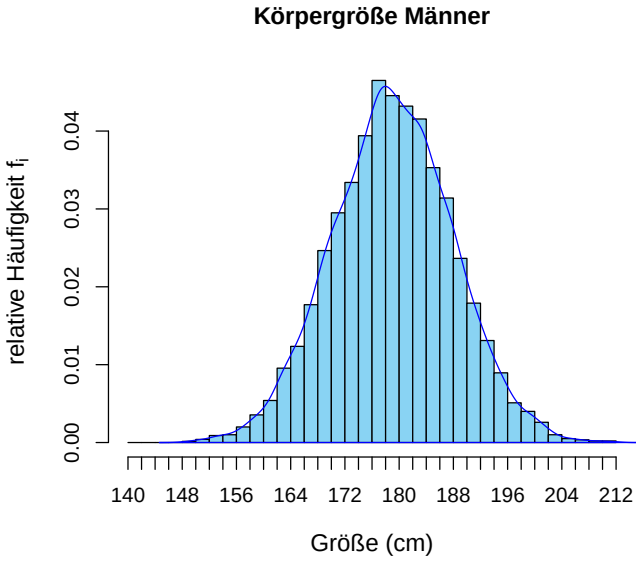


(a) mehrgipflig

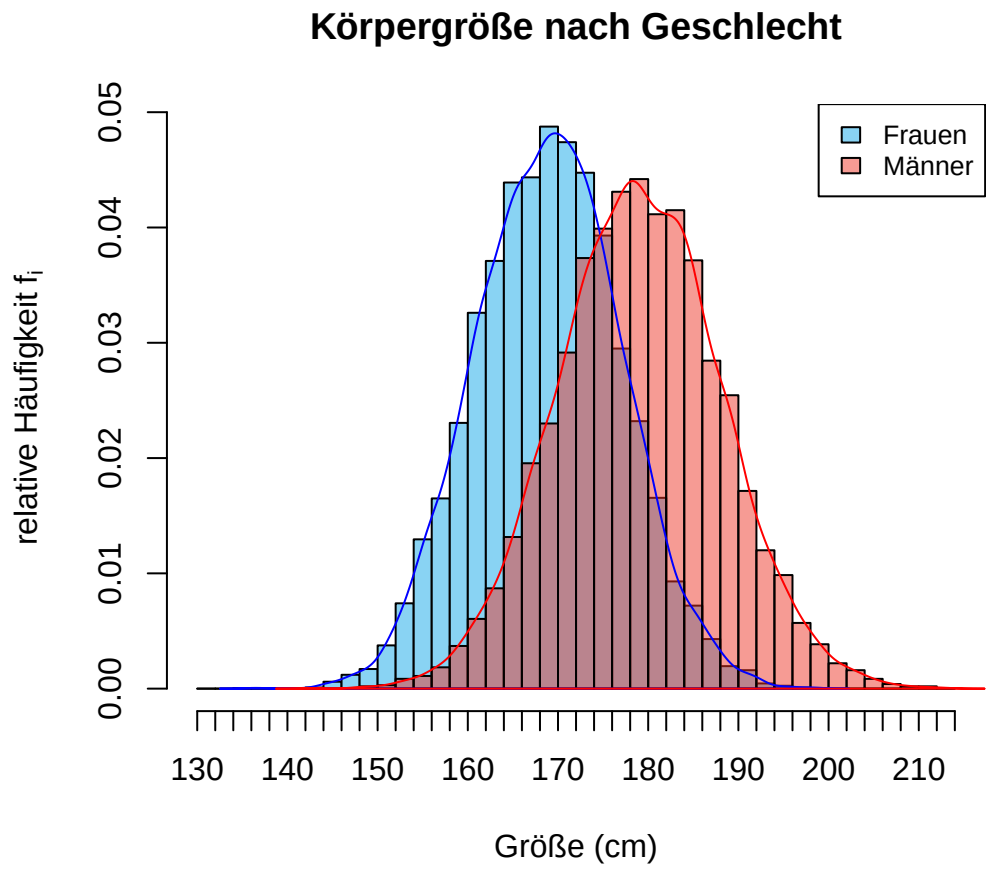
2.3.5 Glockenkurven



(a) normalverteilt



(a) normalverteilt

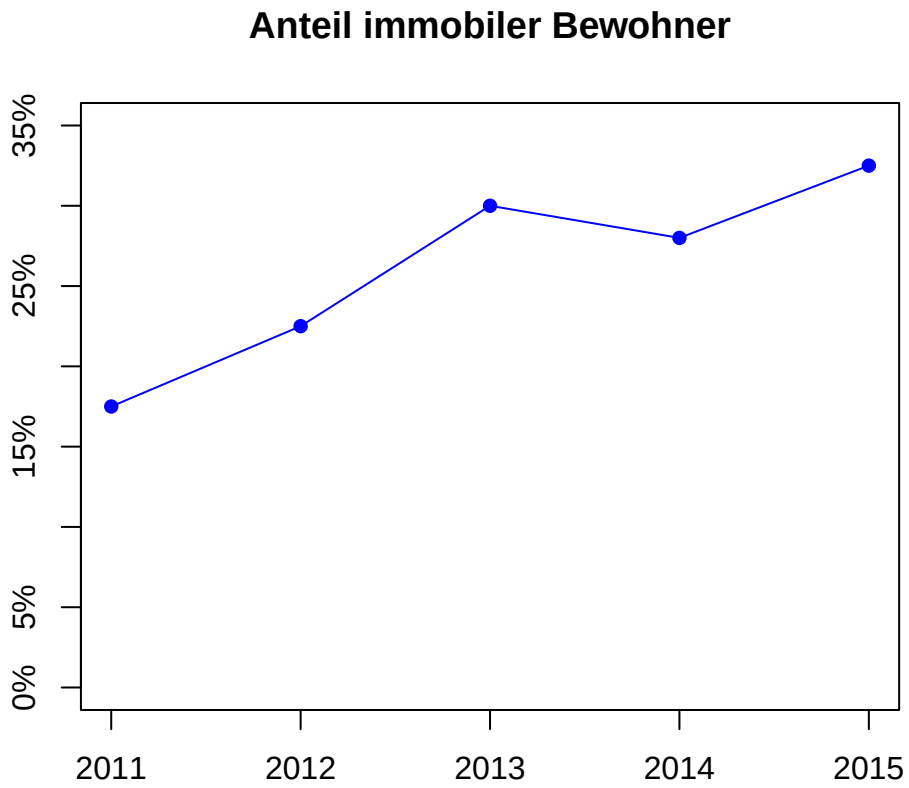


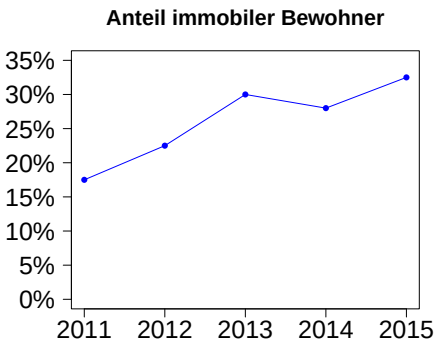
Die Glockenkurvenverteilung tritt so häufig in der Natur auf, dass sie als **Normalverteilung** oder **Gaußsche Verteilung** bekannt ist.

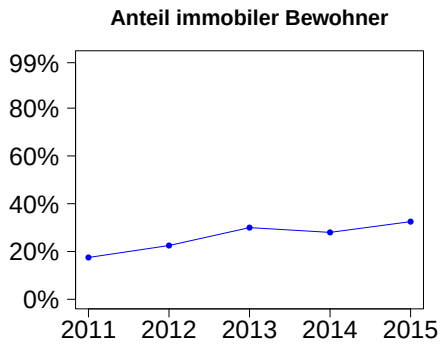
Gauß'sche Glockenkurve

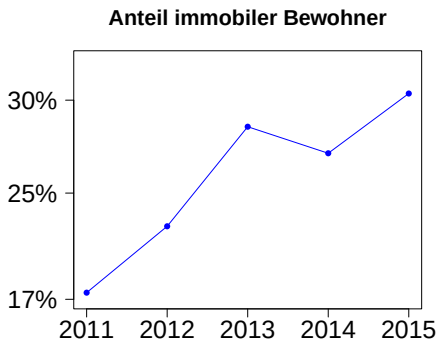
2.3.6 Liniendiagramme

Liniendiagramme (auch *Polygonzüge*) eignen sich, um zeitliche Verläufe darzustellen – und sollten auch nur dafür verwendet werden.









In allen drei Diagrammen sind identische Werte abgetragen. Durch die unterschiedliche Skalierung der y-Achse entsteht aber ein sehr unterschiedlicher Eindruck.

2.4 Ausreißer

Eines der größten Probleme in Stichproben sind Ausreißer, also Werte, die sehr von den restlichen Werten der Stichprobe abweichen. Es ist wichtig, Ausreißer zu identifizieren, bevor man irgendwelche Analysen durchführt, denn Ausreißer verzerren meistens die Ergebnisse.

Ausreißer tauchen immer am Ende der Verteilung auf und können mit einem Box-und-Whisker-Diagramm leicht identifiziert werden (wie später gezeigt wird).

2.4.1 Ausreißerhandhabung

Bei großen Stichproben haben Ausreißer weniger Bedeutung und können unbeachtet gelassen werden. Bei kleinen Stichproben gibt es mehrere Optionen:

- **Entferne die Ausreißer**, da es sich um *fehlerhafte* Werte handelt.

- **Ersetze die Ausreißer** durch den kleinsten oder größten Wert in der Verteilung, der selbst kein Ausreißer ist.
- **Lasse die Ausreißer**, und ändere das theoretische Modell, so dass die Ausreißer erklärbar werden.

3 Stichprobenstatistik

Die deskriptive Statistik hat zum Ziel, empirische Daten durch Tabellen und Grafiken übersichtlich darzustellen und zu ordnen, sowie durch geeignete grundlegende Kenngrößen zahlenmäßig zu beschreiben.

Eine Tabelle oder Grafik informiert über die gesamte Verteilung eines Merkmals mit seinen Ausprägungen. Demgegenüber sagen statistische Kenngrößen etwas über spezielle Eigenschaften der Verteilung aus.

- **Lagekenngrößen:** beschreiben Stellen, an denen sich die Daten konzentrieren oder die die Verteilung in gleiche Teile unterteilen.
- **Streuungskenngrößen:** messen die Ausbreitung und Variabilität der Daten.
- **Formmaße:** messen die Symmetrie und den "Schwanz" der Verteilung.

Diese Werte werden auch genannt **Stichprobenstatistiken** genannt.

Die zahlenmäßige Darstellung durch statistische Kenngrößen beruht auf einer Reduktion der Daten. Die Gesamtheit des Datenmaterials wird hierbei durch typische Vertreter repräsentiert.

3.1 Lagekenngrößen

Es werden zwei Gruppen von Lagekenngrößen unterschieden:

- Zentrale Lagewerte messen die Werte, an denen sich die Daten konzentrieren, meist im Zentrum der Verteilung. Diese Werte sind diejenigen, die die Stichprobe am besten darstellen. Die wichtigsten sind:
  - Arithmetisches Mittel (Mittelwert)
  - Median
  - Modal (Modus)
- Nicht-zentrale Lagewerte teilen die Stichprobe in gleichgroße Teile. Die wichtigsten sind:
  - Quartile.
  - Dezile.
  - Perzentile.

3.1.1 Arithmetischer Mittelwert

⚠ Definition "Arithmetisches Mittel $\overline{x}$"

Das arithmetische Mittel (umgangssprachlich *Mittelwert*) einer Variablen X ist die Summe der beobachteten Werte in der Stichprobe geteilt durch die Stichprobengröße n und wird durch das Symbol $\overline{x}$ (sprich "x quer") dargestellt.

$$\overline{x} = \frac{\sum x_i}{n}$$

Er kann wie folgt auch aus der Häufigkeitstabelle berechnet werden:

$$\overline{x} = \frac{\sum x_i \cdot n_i}{n} = \sum x_i \cdot f_i$$

In den meisten Fällen ist der arithmetische Mittelwert derjenige Wert, der die beobachteten Werte in der Stichprobe am besten darstellt.

! Merke

Das arithmetische Mittel kann nur bei metrischen Daten berechnet werden!

Für das Beispiel der Erhebung von Kindern in Familien lautet das arithmetische Mittel:

$$\overline{x} = \frac{1+2+4+2+2+2+3+2+1+1+0+2+2}{25} + \\ + \frac{0+2+2+1+2+2+3+1+2+2+1+2}{25} = \frac{44}{25} = 1.76 \text{ Kinder.}$$

Die Berechnung von $\overline{x}$ über die Häufigkeitstabelle

x_i	n_i	f_i	$x_i \cdot n_i$	$x_i \cdot f_i$
0	2	0.08	0	0
1	6	0.24	6	0.24
2	14	0.56	28	1.12
3	2	0.08	6	0.24
4	1	0.04	4	0.16
$\sum$	25	1	44	1.76

erfolgt mit:

$$\overline{x} = \frac{\sum x_i \cdot n_i}{n} = \frac{44}{25} = 1.76 \text{ Kinder} \qquad \overline{x} = \sum x_i \cdot f_i = 1.76 \text{ Kinder.}$$

Dies ist die Anzahl an Kinder, durch welche die Familien in der Stichprobe am besten repräsentiert werden.

Verwenden wir die Daten der Körpergröße von Studierenden, berechnet sich $\overline{x}$ wie folgt:

$$\overline{x} = \frac{179 + 173 + \cdots + 187}{30} = 175.07 \text{ cm.}$$

Über die klassierte Häufigkeitstabelle kann $\overline{x}$ bestimmt werden, indem jeweils die Klassenmitte als x_i verwendet wird.

X	x_i	n_i	f_i	$x_i \cdot n_i$	$x_i \cdot f_i$
(150,160]	155	2	0.07	310	10.33
(160,170]	165	8	0.27	1320	44.00
(170,180]	175	11	0.36	1925	64.17
(180,190]	185	7	0.23	1295	43.17

X	x_i	n_i	f_i	$x_i \cdot n_i$	$x_i \cdot f_i$
(190,200]	195	2	0.07	390	13.00
$\sum$		30	1	5240	174.67

$$\overline{x} = \frac{\sum x_i \cdot n_i}{n} = \frac{5240}{30} = 174.67 \text{ cm} \qquad \overline{x} = \sum x_i \cdot f_i = 174.67 \text{ cm}.$$

Beachten Sie, dass der Mittelwert, der aus der Tabelle berechnet wird, etwas von dem realen Wert abweicht. Da in den Berechnungen jeweils die **Klassenmitten** anstelle der tatsächlichen Werte verwendet werden, ergeben sich dadurch leichte Rundungsfehler.

3.1.2 Gewichtetes Mittel

In einigen Fällen haben die Werte der Stichprobe unterschiedlich starke Bedeutungen. In diesem Fall muss das jeweilige *Gewicht* der Werte berücksichtigt werden, wenn das Mittel berechnet wird.

⚠ Definition "Gewichtetes Mittel $\overline{x}_w$"

Das gewichtete Mittel der Werte $x_1, \ldots, x_n$ - wobei jeder Wert x_i ein Gewicht w_i hat - berechnet sich durch die Summe der Produkte jedes Wertes mit seinem Gewicht, dividiert durch die Summe der Gewichtwerte:

$$\overline{x}_w = \frac{\sum x_i w_i}{\sum w_i}$$

Aus der Häufigkeitstabelle kann das gewichtete Mittel mit folgender Formel berechnet werden:

$$\overline{x}_w = \frac{\sum x_i w_i n_i}{\sum w_i}$$

Ein Student möchte den Punktedurchschnitt seiner Studienleistungen in den folgenden drei Kursen berechnen

Kurs	Credits	Punkte
Mathe	8	15
Wirtschaft	5	10
Chemie	6	12

Das arithmetische Mittel der Punkte beträgt

$$\overline{x} = \frac{\sum x_i}{n} = \frac{15 + 10 + 12}{3} = 12,\overline{3} \text{ Punkte}.$$

Fächer mit mehr Creditpunkten erfordern mehr Arbeit und sollten bei der Berechnung des Durchschnitts ein größeres Gewicht erhalten. In diesem Fall ist es ratsam, das gewichtete Mittel mit den Creditpunkten als Gewichten zu berechnen.

$$\overline{x}_w = \frac{\sum x_i w_i}{\sum w_i} = \frac{15 \cdot 8 + 10 \cdot 5 + 12 \cdot 6}{8 + 5 + 6} = \frac{242}{19} = 12,74 \text{ Punkte}.$$

3.1.3 Median

Definition Median

Der Median einer Variablen X ist der Wert, der in der Mitte der geordneten Stichprobe liegt.

Der Median teilt die Stichprobenwerte in zwei gleich große Teile, das heißt, es gibt dieselbe Anzahl von Werten über und unter dem Median.

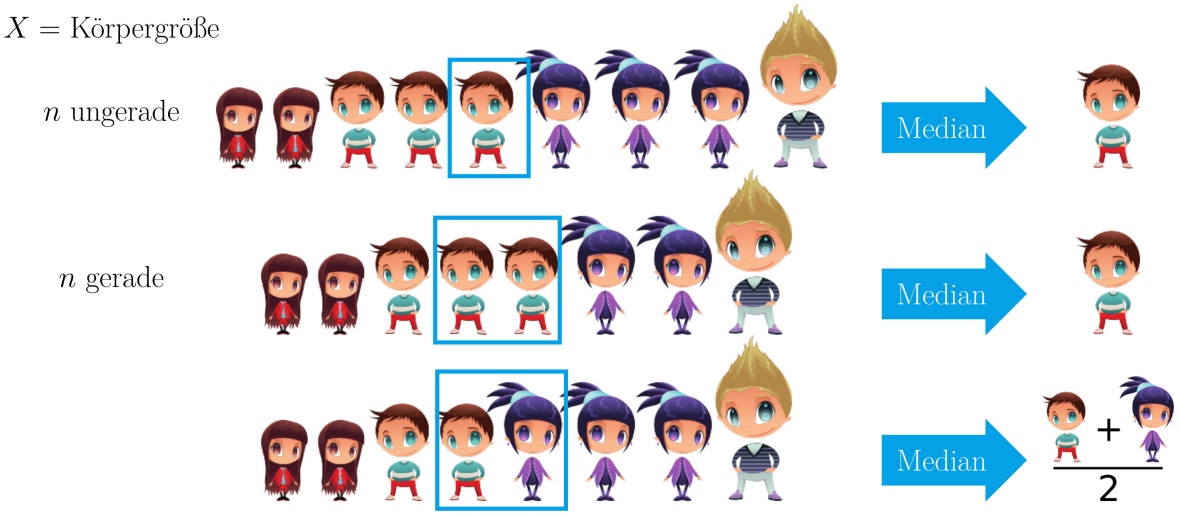
ACHTUNG!

Der Median kann **nicht** für nominale Variablen berechnet werden.

3.1.3.1 Nicht-klassierte Daten

Bei nicht-klassierten Daten gibt es zwei Möglichkeiten:

- **Gerade Stichprobengröße:** Der Median ist der Wert an Position $\frac{n}{2}$.
- **Ungerade Stichprobengröße:** Der Median ist das arithmetische Mittel der Werte an den Positionen $\frac{n}{2}$ und $\frac{n}{2}+1$.



Verwenden wir die Daten des Beispiels mit der Anzahl an Kindern in Familien. Die Stichprobengröße beträgt 25, was ungerade ist. Daher ist der Median der Wert an Position $\frac{25+1}{2} = 13$ der sortierten Stichprobe.

$$0, 0, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, \boxed{2}, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 4$$

Der Median liegt also bei 2 Kindern.

Mithilfe der relativen Häufigkeitstabelle ist der Median der kleinste Wert mit einer kumulativen absoluten Häufigkeit N_i größer oder gleich 13 bzw. einer kumulativen relativen Häufigkeit F_i größer oder gleich 0,5.

x_i	n_i	f_i	N_i	F_i
0	2	0.08	2	0.08
1	6	0.24	8	0.32
2	14	0.56	22	0.88
3	2	0.08	24	0.96
4	1	0.04	25	1
$\sum$	25	1		

3.1.3.2 klassierte Daten

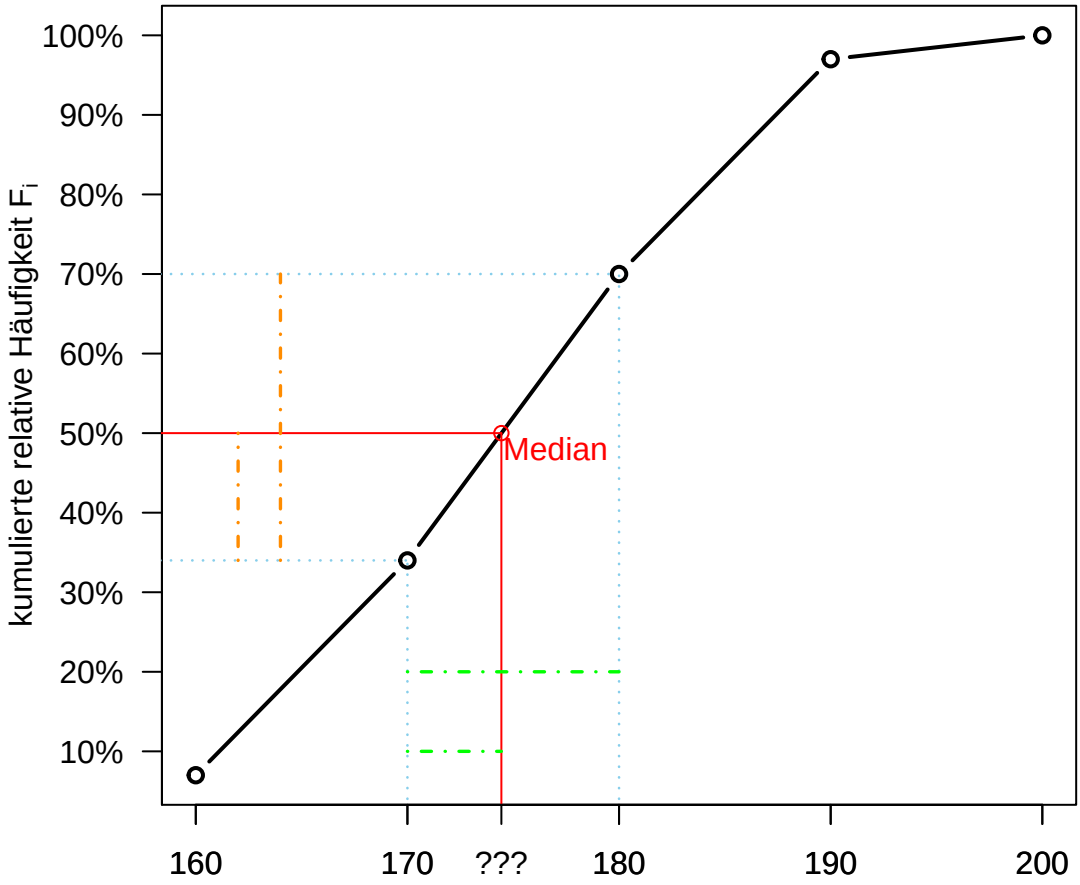
Sind die Werte in Klassen gruppiert, kann die Berechnung des Median geometrisch hergeleitet werden.

Die Werte unseres Beispiels der Körpergrößen liegen als klassierte Häufigkeitstabelle vor:

X	x_i	n_i	f_i	F_i
(150,160]	155	2	0.07	0.07
(160,170]	165	8	0.27	0.34
(170,180]	175	11	0.36	**0.70**
(180,190]	185	7	0.23	0.97
(190,200]	195	2	0.07	1
$\sum$		30	1	

Der Median liegt irgendwo zwischen 170 und 180.

Für die graphische Ermittlung des Medians benötigen wir die kumulierten Prozentwerte F_i, aus denen ein Summenpolygon angefertigt wird. Zusätzlich zeichnen wir ein paar Hilfslinien ein.



Der Median entspricht dem gesuchten Wert “???” auf der x-Achse.

Aus der Mathematik (Strahlensätze) wissen wir, dass das Verhältnis der beiden grünen Strecken auf der x-Achse dem Verhältnis der orangenen Strecken auf der y-Achse entspricht. Setzen wir also die Strecke ins Verhältnis, ergibt sich

$$\frac{x_Me - x_u}{x_o - x_u} = \frac{50 - y_u}{y_o - y - u}$$

Dabei entspricht

- x_Me dem Median
- x_u dem unteren Klassengrenze
- x_o der oberen Klassengrenz
- y_u der kumulierten relativen Häufigkeit der darunter liegenden Klasse
- y_o der kumulierten relativen Häufigkeit der Klasse

Die Gleichung kann nun nach x_Me umgestellt werden.

$$x_{Me} = x_u + (x_o - x_u) \cdot \frac{50 - y_u}{y_o - y_u}$$

Jetzt können wir unsere Werte einsetzen.

$$\begin{aligned}\text{Median} &= 170 + (180 - 170) \cdot \frac{50 - 34}{70 - 34} \\ &= 170 + 10 \cdot \frac{4}{9} \approx 174,44\end{aligned}$$

3.1.4 Modus

> **Definition "Modus"**
>
> Der Modus oder Modalwert ist der in einer Verteilung am häufigsten auftretende Wert.

Liegen klassierte Daten vor, entspricht der Modalwert der Klassenmitte der Klasse mit der größten Häufigkeit. Die Kategorie (Gruppe, Klasse, etc.) mit den häufigsten Werten nennt man die Modalklasse. Es ist durchaus denkbar, dass innerhalb einer Verteilung mehrere Modalwerte ermittelt werden können. Sind jedoch die Häufigkeiten der Merkmalsausprägungen annähernd gleich verteilt, macht es keinen Sinn, den Modalwert zu bestimmen. Der Modalwert, bzw. die Modalklasse lässt sich für Daten jeglichen Skalenniveaus angeben. Andererseits bietet er für nominale Merkmale die einzige Möglichkeit, etwas typisches über die Verteilung der Merkmalsausprägungen aufzuzeigen.

3.1.4.1 Modusberechnung

Unter Verwendung der Daten der Stichprobe zur Anzahl an Kindern in Familien ist der Wert mit der höchsten Häufigkeit = 2.

x_i	n_i
0	2
1	6
2	14
3	2
4	1

Der Modus beträgt also = 2 Kinder.

Unter Verwendung der Daten der Stichprobe zur Körpergröße von Studenten ist die Klasse mit der höchsten Häufigkeit (170, 180].

X	n_i
$(150, 160]$	2
$(160, 170]$	8
$(170, 180]$	11
$(180, 190]$	7
$(190, 200]$	2

Das bedeutet, die modale Klasse ist = $(170, 180]$.

3.1.5 Welchen Lagekennwert soll ich verwenden?

Welche Lagekenngröße hat die größte Aussagekraft bezüglich der Stichprobe? Als Faustregel kann folgende Reihenfolge verwendet werden:

1. **Mittelwert**: Der Mittelwert nutzt mehr Information aus der Stichprobe als die anderen Maße, da er die Größenordnung der Daten berücksichtigt.
2. **Median**: Der Median nutzt weniger Informationen als der Mittelwert, aber mehr als der Modus, da er die Ordnung der Daten berücksichtigt.
3. **Modus**: Der Modus ist das Maß, das am wenigsten Information aus der Stichprobe nimmt, da er nur die absolute Häufigkeit der Werte berücksichtigt.

> ! Achtung, Ausreisser!
>
> Der Mittelwert wird stärker von Ausreißern beeinflusst als der Median!

Stellen wir uns vor, die Befragung zur Anzahl an Kinder in 7 Familien ergäbe folgende Werte

0, 0, 1, 1, 2, 2, 15

Dann wären die Lagekenngrößen

- $\overline{x}$ = 3 Kinder
- Median = 1 Kind

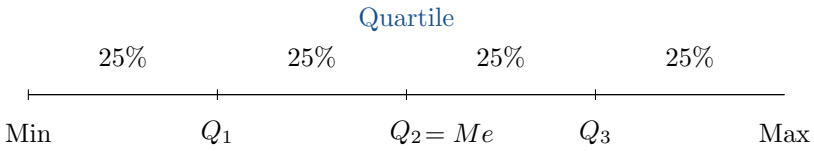
In diesem Fall hat der Median die größte Aussagekraft bezüglich der Stichprobe.

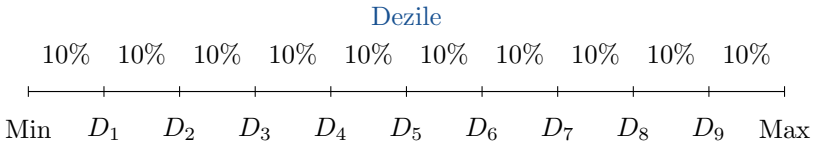
3.1.6 nicht-zentrale Lagewerte

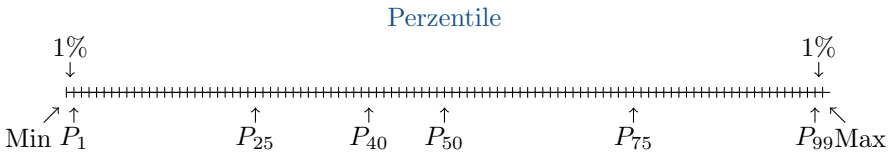
Die nichtzentralen Lagewerte (Quantile) teilen die Stichprobenverteilung in gleich große Teile. Die am häufigsten verwendeten sind:

- Quartile: Teilen die Verteilung in 4 gleich große Teile. Es gibt 3 Quartile:
  - Q_1 (25% kumuliert)
  - Q_2 (50% kumuliert),
  - Q_3 (75% kumuliert).
- Dezile: Teilen die Verteilung in 10 gleich große Teile. Es gibt 9 Dezile:
  - D_1 (10% kumuliert),
  - ⋮,
  - D_9 (90% kumuliert).
- Perzentile: Teilen die Verteilung in 100 gleich große Teile. Es gibt 99 Perzentile:

- P_1 (1% kumuliert),
- ⋮,
- P_{99} (99% kumuliert).







Beachten Sie, dass es Übereinstimmungen zwischen den Quartilen, Dezilen und Perzentilen gibt es. Zum Beispiel stimmt das erste Quartil mit dem 25. Perzentil überein, und das vierte Dezil stimmt mit dem 40. Perzentil überein.

3.1.6.1 Bestimmung der Quantile

Quantile werden ähnlich wie die Median berechnet. Der einzige Unterschied liegt in der kumulierenden relativen Häufigkeit F_i, die jedem Quantil zugeordnet ist.

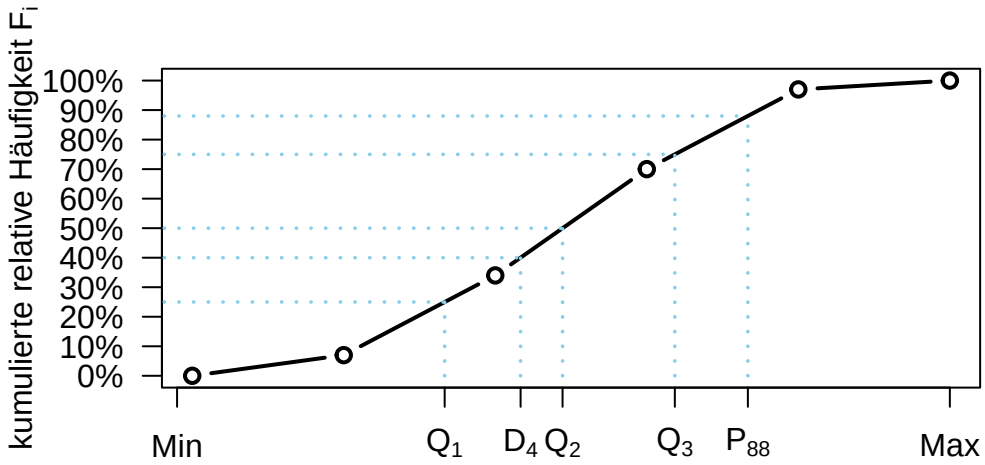
Die Häufigkeitstabelle der Anzahl an Kindern in Familien ist

x_i	F_i
0	0.08
1	0.32
2	0.88
3	0.96
4	1

Anhand der Spalte F_i kann das gewünschte Perzentil, Dezentil oder Quartil abgelesen werden.

$$
\begin{aligned}
F_{Q_1} = 0.25 &\Rightarrow Q_1 = 1 \text{ Kinder},\\
F_{Q_2} = 0.5 &\Rightarrow Q_2 = 2 \text{ Kinder},\\
F_{Q_3} = 0.75 &\Rightarrow Q_3 = 2 \text{ Kinder},\\
F_{D_4} = 0.4 &\Rightarrow D_4 = 2 \text{ Kinder},\\
F_{P_{92}} = 0.92 &\Rightarrow P_{92} = 3 \text{ Kinder}.
\end{aligned}
$$

Die Bestimmung kann ebenfalls (wie beim Median) geometrisch hergeleitet werden.



3.2 Streuungskenngrößen

Häufigkeitsverteilungen lassen sich nicht nur durch ihre Lage auf der Messwertskala, sonder auch durch ihre Streuung beschreiben.

Streuungskenngrößen sind Maßzahlen zur Bewertung oder Quantifizierung der Variablilität der Messwerte. Es gibt diverse Streuungskenngrößen. Die wichtigsten sind:

- Spannweite R
- Quartilsabstand Q
- Varianz s^2
- Standardabweichung s
- Variationskoeffizient vk

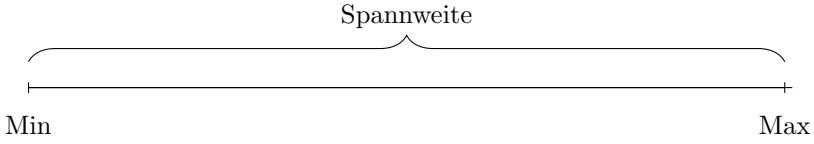
3.2.1 Spannweite

Ein besonders einfaches Streuungsmaß ist die Spannweite (auch: Variationsbreite), die mit "R" (englisch "range") abgekürzt wird.

Definition "Spannweite"

Die Spannweite R bestimmt sich aus der Differenz zwischen dem größten und dem kleinsten Messwert.

$$R = \max_{x_i} - \min_{x_i}$$



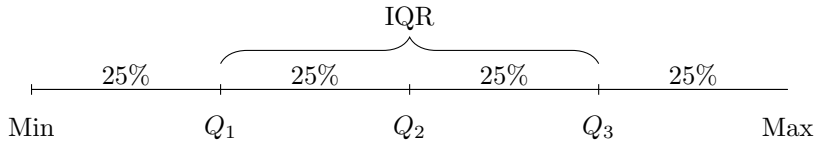
3.2.2 Interquartilsabstand

Definition "Interquartilsabstand"

Der Interquartilsabstand einer Variablen x ist der Unterschied zwischen dem dritten und dem ersten Stichproben-Quartil.

$$IQR = Q_3 - Q_1$$

Der IQR gibt die mittleren 50% der Verteilung an. Daher reagiert er - ähnlich wie der Median - robuster auf Ausreisser.



3.2.3 Boxplot

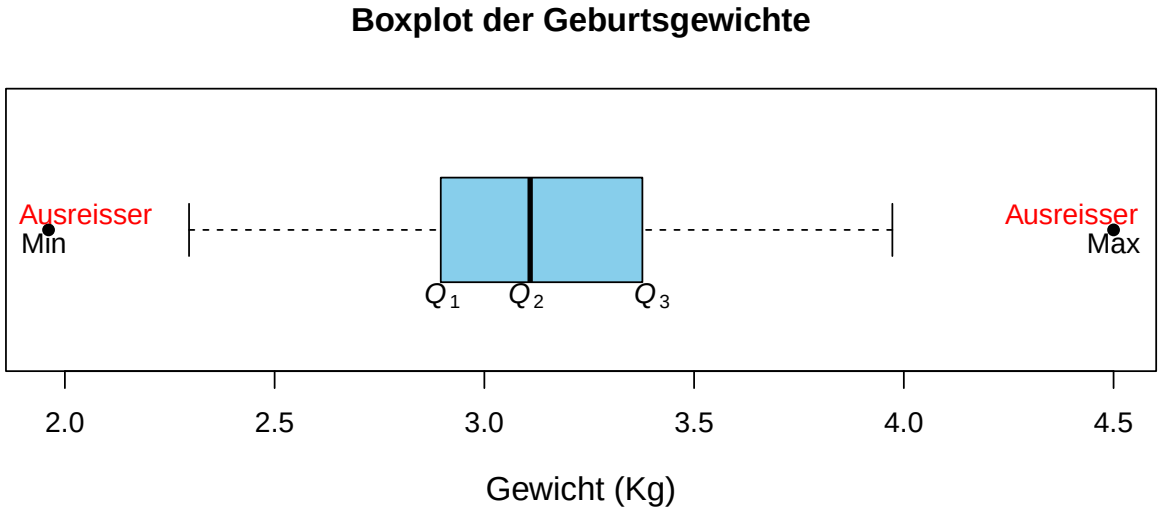
Die Streuung einer Variablen in einer Stichprobe kann graphisch mit einem **Boxplot** dargestellt werden. An ihm lassen sich die fünf wichtigsten Kennwerte (Minimum, Quartile und Maximum) ablesen.

Es besteht aus einem Block (Box), der vom ersten bis zum dritten Quartil gezeichnet wird. Dies entspricht dem Interquartilsabstand.

Zusätzlich werden an den Rändern zwei Segmenten, die als untere und obere Antennen (Whisker) bezeichnet werden. Normalerweise wird die Box durch den Median in zwei Teile gesplittet.

Das Boxplot erfreut sich großer Beliebtheit, da es mehrere Funktionen erfüllt:

- Es dient zur Einschätzung der Streuung, da es die Spannweite sowie den Interquartilabstand darstellt.
- Es dient zum Erkennen von Ausreißern, die außerhalb der Whiskers liegen.
- Es dient zur Bestimmung der Symmetrie der Verteilung, indem die Länge der Boxen und Whisker über und unter dem Median verglichen werden.



Um einen Boxplot zu erstellen, folgen Sie den nachfolgenden Schritten:

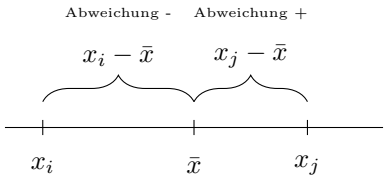
1. Berechnen Sie die Quartile.
2. Zeichnen Sie eine Box vom ersten zum dritten Quartil.
3. Teilen Sie die Box mit dem Median
4. Die Whiskers berechnen sich über die Hilfsgrößen f_1 und f_2 wie folgt:

$$f_1 = Q_1 - 1.5\,\text{IQR}$$
$$f_2 = Q_3 + 1.5\,\text{IQR}$$

Diese Grenzen definieren den Bereich, in dem die Daten als normal betrachtet werden. Jeder Wert außerhalb dieses Bereichs wird als Ausreisser betrachtet. Zeichnen Sie für den unteren Whisker einen Strich vom unteren Quartil zum kleinsten Wert in der Stichprobe, der größer oder gleich f_1 ist, und für den oberen Whisker einen Strich vom oberen Quartil zum größten Wert in der Stichprobe, der kleiner oder gleich f_2 ist. 5. Zeichnen Sie schließlich je einen Punkt für alle Werte kleiner f_1 oder größer f_2 (Ausreisser).

3.2.4 Abweichungen zur Mitte

Bei der *Abweichungen zur Mitte* wird die Entfernung von jedem Wert in der Stichprobe zum Mittelwert gemessen.



Wenn die Abweichungen groß sind, ist der Mittelwert weniger repräsentativ als bei kleinen Abweichungen.

Definition "Summe der quadrierten Abweichungen"

Üblicherweise wird an dieser Stelle die **Summe der quadrierten Abweichungen** (*SQA*) bestimmt, indem die Abweichungen quadriert und aufsummiert werden.

$$SQA = \sum (x_i - \overline{x})^2$$

3.2.5 Varianz

Definition "Varianz"

Die Varianz s^2 ist ein Maß für die Streuung oder Variabilität von Datenwerten. Sie gibt an, wie weit die einzelnen Werte eines Datensatzes im Durchschnitt vom Mittelwert (Durchschnitt) entfernt sind.

$$s^2 = \frac{SQA}{n-1} = \frac{\sum (x_i - \overline{x})^2}{n-1}$$

Die Varianz hat die quadrierte Einheiten der Variablen. Um ihre Deutung zu vereinfachen ist es üblich, ihre Wurzel zu berechnen.

3.2.6 Standardabweichung

Definition "Standardabweichung"

Die Standardabweichung s (auch: σ oder `sd` (englisch: standard deviation)) ist das wichtigste und bekannteste Maß für die Streuung der Messwerte um den Mittelwert.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum(x_i - \overline{x})^2}{n-1}}$$

Liegen metrische Messwerte vor, so sind das arithmetische Mittel $\overline{x}$ und die Standardabweichung s die wichtigsten Maßwerte, um die Eigenschaften der Messreihe ausreichend zu beschreiben.

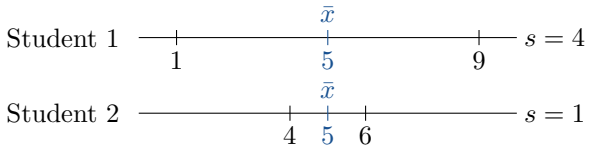
3.2.6.1 Interpretation

Varianz und Standardabweichung messen die Ausbreitung der Daten um den Mittelwert herum.

- Wenn die Varianz oder die Standardabweichung klein sind, sind die Werte um den Mittelwert herum konzentriert, und der Mittelwert ist ein gutes repräsentatives Maß.
- Wenn sie groß sind, sind die Stichprobenwerte weit von der Mitte entfernt, und der Mittelwert ist kein guter Repräsentant der Stichprobe.

Beispiel

Von 2 Studierenden wurden die Modulpunkte von 2 Kursen erhoben.



Welcher Mittelwert beschreibt die Leistung besser?

3.2.7 Variationskoeffizient

Um einschätzen zu können, ob die Standardabweichung eher groß oder klein ist, bietet sich der Variationskoeffizient an (englisch coefficient of variance).

Definition "Variationskoeffizient"

Der Variationskoeffizient setzt die Streuung (Standardabweichung) ins Verhältnis zum Mittelwert ($\overline{x}$).

$$vk = \frac{s}{|\overline{x}|}$$

Es entsteht ein dezimaler Prozentwert, der den Anteil der Standardabweichung am Mittelwert wiedergibt. Je größer vk ist, umso mehr streuen die Werte um den Mittelwert. Da er keine Einheiten hat, ist der Variationskoeffizient sehr hilfreich, um die Streuung in Verteilungen verschiedener Variablen zu vergleichen (je kleiner der vk, desto *besser*).

Bei der Anzahl an Kindern in Familien lag der Mittelwert bei $\overline{x} = 1,76$ Kindern und die Standardabweichung bei $s = 0,8616$ Kindern. Der Variationskoeffizient beträgt demnach:

$$vk = \frac{1,76 \text{ Kinder}}{0,8616 \text{ Kinder}} = 0,49$$

Bei der Stichprobe zur Körpergröße lag der Mittelwert bei $\overline{x} = 174,67$ cm und die Standardabweichung bei $s = 10,1$ cm. Der Variationskoeffizient beträgt demnach:

$$vk = \frac{174,67 \text{ cm}}{10,1 \text{ cm}} = 0,06$$

Das bedeutet, dass die Streuung in der Verteilung der Körpergrößen geringer ist als in der Verteilung der Anzahl an Kindern, und dass daher der Mittelwert der Körpergröße *repräsentativer* ist als der Mittelwert der Anzahl an Kindern.

3.3 Formmaße

Formmaße sind Maße, die die Form der Verteilung beschreiben. Besonders wichtige Aspekte sind:

- **Schiefe** (skewness): misst die Symmetrie der Verteilung hinsichtlich des Mittels.
- **Wölbung:** (kurtosis) misst die Länge der Verteilungsenden (Schwanz) oder die Spitzigkeit der Verteilung. Das am meisten verwendete

3.3.1 Schiefekoeffizient

⚠ Definition "Schiefekoeffizient" (coefficient of skewness)

Der Stichproben-Schiefekoeffizient g_1 einer Variablen X ist der Durchschnitt der hoch drei genommenen Abweichungen der Werte vom Stichprobenmittel, geteilt durch die hoch drei genommene Standardabweichung.

$$g_1 = \frac{\sum(x_i - \bar{x})^3 n_i/n}{s^3} = \frac{\sum(x_i - \bar{x})^3 f_i}{s^3}$$

Skewness misst die Symmetrie der Verteilung, das heißt, wie viele Werte in der Stichprobe über oder unter dem Mittel liegen und wie weit sie von diesem entfernt sind.

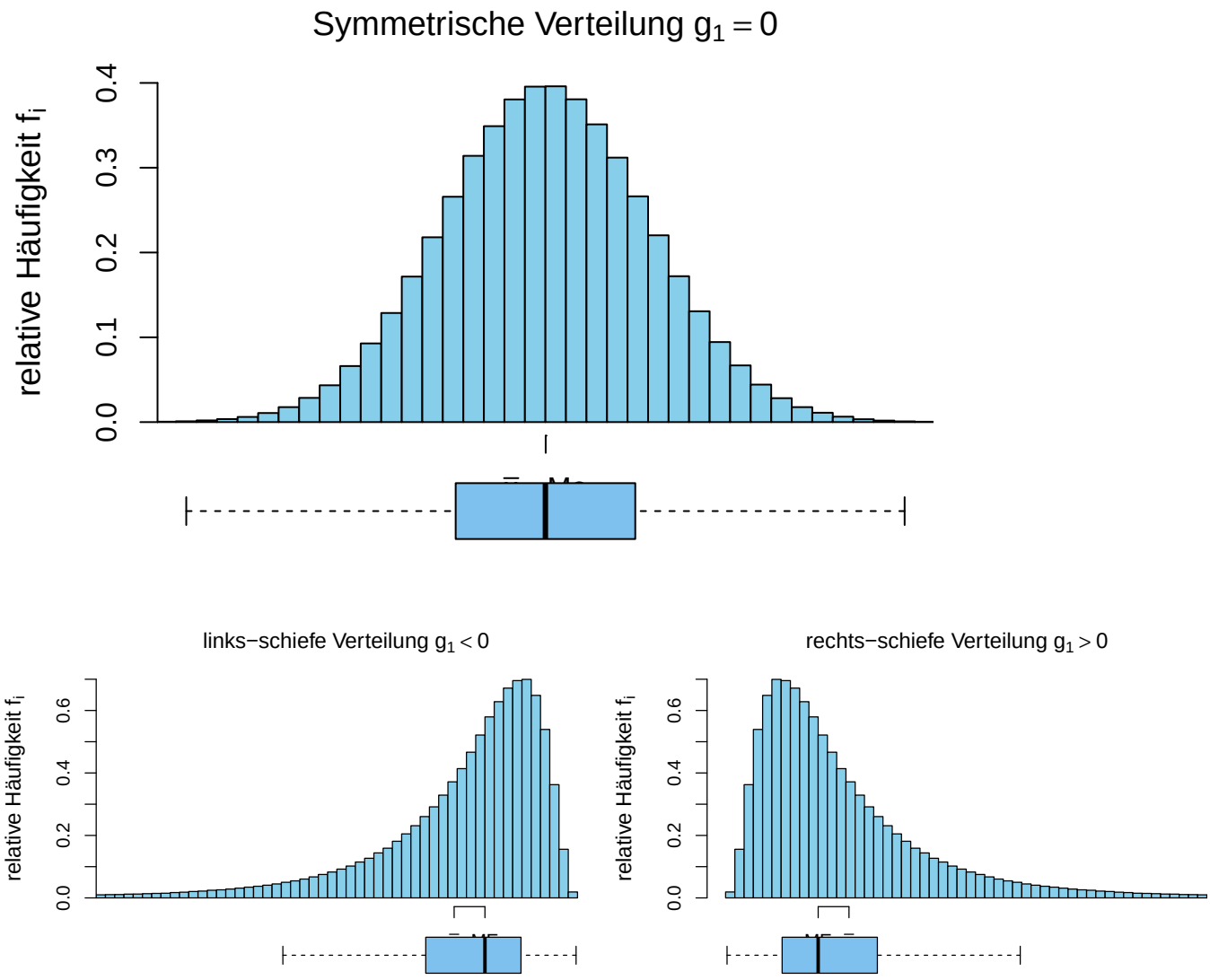
3.3.1.1 Interpretation des Schiefekoeffizienten

Je nach Vorzeichen von g_1 gilt:

- $g_1 = 0$ bedeutet, dass es dieselbe Anzahl an Werten im Stichprobenmaterial oberhalb und unterhalb des Mittelwerts gibt und diese gleich weit von ihm entfernt sind. Die Verteilung ist symmetrisch.

- $g_1 < 0$ bedeutet, dass es mehr Werte oberhalb des Mittelwerts als unterhalb davon gibt, aber die darunter liegenden Werte weiter entfernt sind. Die Verteilung ist links-schiefs (sie hat eine längere Schwanzspitze nach links).

- $g_1 > 0$ bedeutet, dass es mehr Werte unterhalb des Mittelwerts als oberhalb davon gibt, aber die darüber liegenden Werte weiter entfernt sind. Die Verteilung ist rechts-schiefs (sie hat eine längere Schwanzspitze nach rechts).



Verwenden wir die Daten der Körpergrößen von Studierenden, können wir der Häufigkeitstabelle eine weitere Spalte mit den kubierten Abweichungen vom Mittelwert $\overline{x} = 174,67$ cm hinzufügen:

X	x_i	n_i	$x_i - \bar{x}$	$(x_i - \bar{x})^3 \cdot n_i$
$(150, 160]$	155	2	-19.67	-15221.00
$(160, 170]$	165	8	-9.67	-7233.85
$(170, 180]$	175	11	0.33	0.40
$(180, 190]$	185	7	10.33	7716.12
$(190, 200]$	195	2	20.33	16805.14
$\sum$		30		2066.81

Die Skewness berechnet sich nun wie folgt:

$$g_1 = \frac{\sum (x_i - \bar{x})^3 n_i / n}{s^3} = \frac{2066,81/30}{10,1^3} = 0,07.$$

Der Wert liegt nahe bei 0, was bedeutet, dass die Verteilung nahezu symmetrisch (normalverteilt) ist.

3.3.2 Wölbungskoeffizient

Definition "Wölbungskoeffizient"

Der Wölbungskoeffizient g_2 (*coefficient of kurtosis*) einer Variablen X ist der Durchschnitt der zur vierten Potenz erhobenen Abweichungen der Werte vom Stichprobenmittelwert, geteilt durch die Standardabweichung hoch vier, und anschließend minus 3.

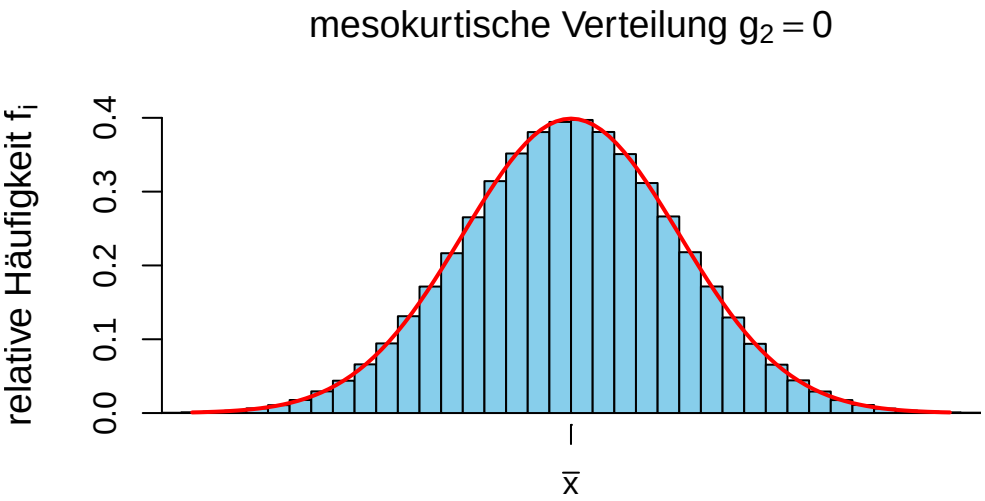
$$g_2 = \frac{\sum(x_i - \bar{x})^4 n_i/n}{s^4} - 3 = \frac{\sum(x_i - \bar{x})^4 f_i}{s^4} - 3$$

Der Koeffizient misst die Konzentration der Werte um die Mitte und die Länge der Schwanzenden der Verteilung. Die Normalverteilung dient als Bezugsgröße.

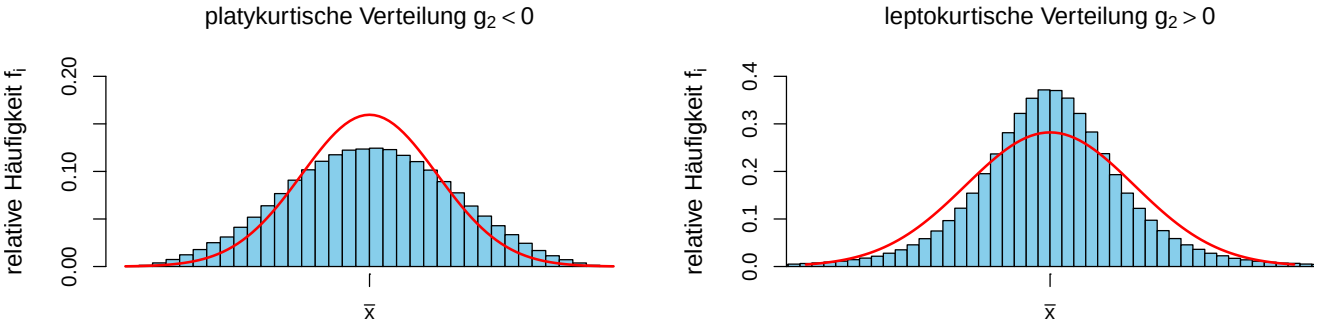
3.3.2.1 Interpretation des Wölbungskoeffizienten

Je nach Vorzeichen von g_1 gilt:

- $g_2 = 0$ zeigt an, dass die Wölbung *normal* ist, das heißt, die Konzentration der Werte um den Mittelwert wie bei einer gaußschen Glockenkurve (mesokurtisch).
- $g_2 < 0$ zeigt an, dass die Wölbung weniger als normal ist, das heißt, die Konzentration der Werte um den Mittelwert ist geringer als bei einer gaußschen Glockenkurve (plattikurtisch).
- $g_2 > 0$ zeigt an, dass die Kurtosis größer als normal ist, das heißt, die Konzentration der Werte um den Mittelwert ist größer als bei einer gaußschen Glockenkurve (leptokurtisch).



Verwenden wir die die Häufigkeitstabelle der Körpergrößen von Studierenden, können wir eine neue Spalte mit den Abweichungen vom Mittelwert $\overline{x} = 174,67$ cm zur vierten Potenz hinzufügen.



X	x_i	n_i	$x_i - \bar{x}$	$(x_i - \bar{x})^4 n_i$
$(150, 160]$	155	2	-19.67	299396.99
$(160, 170]$	165	8	-9.67	69951.31
$(170, 180]$	175	11	0.33	0.13
$(180, 190]$	185	7	10.33	79707.53
$(190, 200]$	195	2	20.33	341648.49
$\sum$		30		790704.45

Nun kann die Kurtosis wie folgt berechnet werden:

$$g_2 = \frac{\sum (x_i - \bar{x})^4 n_i / n}{s^4} - 3 = \frac{790704.45/30}{10.1^4} - 3 = -0.47.$$

Der Wert ist negativ, aber nahe bei 0. Das bedeutet, dass die Kurve leicht platykurtisch (niedriger als normalverteilt) ist.

Wie wir im Kapitel zur schließenden Statistik sehen werden, können viele statistische Tests nur auf normalverteilte (glockenförmigen) Verteilungen angewendet werden.

Normalverteilungen sind symmetrisch und mesokurtisch, weshalb ihre Symmetriekoeffizienten und Kurtosiskoeffizienten den Wert 0 haben. Daher ist eine Möglichkeit, eine Variable auf Normalverteilung zu testen, die Abweichungen der Skewness- und Kurtosis-Koeffizienten von 0 zu betrachten.

Im Allgemeinen wird die Normaverteilung abgelehnt, wenn g_1 oder g_2 außerhalb des Intervalls $[-2, 2]$ liegen. In diesem Fallist es üblich, eine Transformation anzuwenden, um diese Abweichung zu korrigieren.

3.4 Transformationen

In vielen Fällen werden die rohen Stichprobendaten *transformiert*, um eine passendere Skala zu erhalten, oder um Daten zu korrigieren, die nicht normalverteilt sind.

Nehmen wir z.B. folgende Werte zur Körpergröße

$$1.75m, 1.65m, 1.80m,$$

Wenn die Werte mit 100 multipliziert werden, ändert sich die Einheit von Meter in Zentimeter, und aus den Dezimalzahlen werden ganze Zahlen.

$$175cm, 165cm, 180cm,$$

Darüber hinaus ist es möglich, die Größen der Werte durch Subtraktion des kleinsten Wertes in der Stichprobe zu verringern. In unserem Beispiel würden wir von jedem Wert 165 cm abziehen:

$$10cm, 0cm, 15cm$$

Es ist offensichtlich, dass diese Daten leichter zu handhaben sind als die ursprünglichen. Im Grunde haben wir folgende Transformation auf die Daten angewendet:

$$Y = 100 \cdot X - 165$$

3.4.1 lineare Transformation

Eine der am häufigsten verwendeten Transformationen ist die lineare Transformation:

$$Y = a + b \cdot X$$

Bei der linearen Transformation veränderen sich das arithmetische Mittel und die Standardabweichung der transformierten Variablen wie folgt:

$$\bar{y} = a + b\bar{x},$$
$$s_y = |b|s_x$$

Zudem bleibt der Wölbungskoeffizient (Kurtosis) unverändert, während der Koeffizient der Schieflage (Skewness) nur das Vorzeichen ändert, falls b negativ ist.

3.4.1.1 Standardisierung und z-Transformation

⚠ Definition “Standardisierung”

Die standardisierte Variable Z einer Variablen X ist die Variable, die durch Abzug des Mittelwertes von X und anschließendes Dividieren durch die Standardabweichung entsteht:

$$Z = \frac{X - \bar{x}}{s_x}$$

Für jeden Wert x_i der Stichprobe ist der Standardiserungsscore (z-Wert) der Wert, der durch Anwendung der Standardisierungstransformation entsteht:

$$z_i = \frac{x_i - \bar{x}}{s_x}$$

Der Standardiserungsscore gibt die Anzahl an Standardabweichungen an, um welche ein Wert über oder unter dem Mittelwert liegt, und ist nützlich, um die Abhängigkeit der Variablen von ihren Maßeinheiten zu vermeiden. Die standardisierte Variable hat immer einen Mittelwert von 0 und eine Standardabweichung von 1.

$$\bar{z} = 0 \qquad s_z = 1$$

Beispiel

Die Leistungspunkte von 5 Studierenden in 2 Fächern sind

Student:	$Hans$	$Tina$	Kim	Flo	$Karl$		
$Mathe:$	2	5	4	**8**	6	$\bar{x} = 5$	$s_x = 2$
$Musik:$	1	9	**8**	5	2	$\bar{y} = 5$	$s_y = 3.16$

Hat Student Flo im Fach Mathe mit 8 Punkten eine *bessere* Leistung erbracht als Studentin Kim im Fach Musik (ebenfalls 8 Punkte)?

Es mag den Anschein haben, dass Kim und Flo die gleiche Leistung in den Fächern erbracht haben, da sie die gleichen Leistungspunkte erzielen. Aber um die Leistung der Studierenden zum Rest der Gruppe zu bestimmen, müssen wir die Abweichungen der Leitungspunkte in jedem Fach betrachten. Aus diesem Grund ist es besser, die z-Werte als Maßstab für die Leistungsfähigkeit zu verwenden.

$Mathe:$	-1.5	0	-0.5	**1.5**	0.5
$Musik:$	-1.26	1.26	**0.95**	0	-0.95

Nun ist erkennbar, dass Flo mit 8 Punkten im Fach Mathe um 1,5 Standardabweichungen über dem Leistungsdurchschnitt liegt, während Kim mit 8 Punkten im Fach Musik nur um 0,95 Standardabweichungen über dem Leistungsmittelwert liegt. In Relation zu allen Studierenden hat Flo eine *bessere* Leistung erbracht als Kim.

Beispiel

Fragen wir uns nun, wer der beste Studierende ist.

Student:	$Hans$	$Tina$	Kim	Flo	$Karl$
$Mathe:$	2	5	4	8	6
$Musik:$	1	9	8	5	2
$\sum$	3	**14**	12	13	8

Wenn wir nur den Leistungspunktsummen folgen, ist dies Studentin Tina mit insgesamt 14 Punkten.

Wenn wir allerdings die relative Leistung betrachten und die z-transformierten Werte aufsummieren…

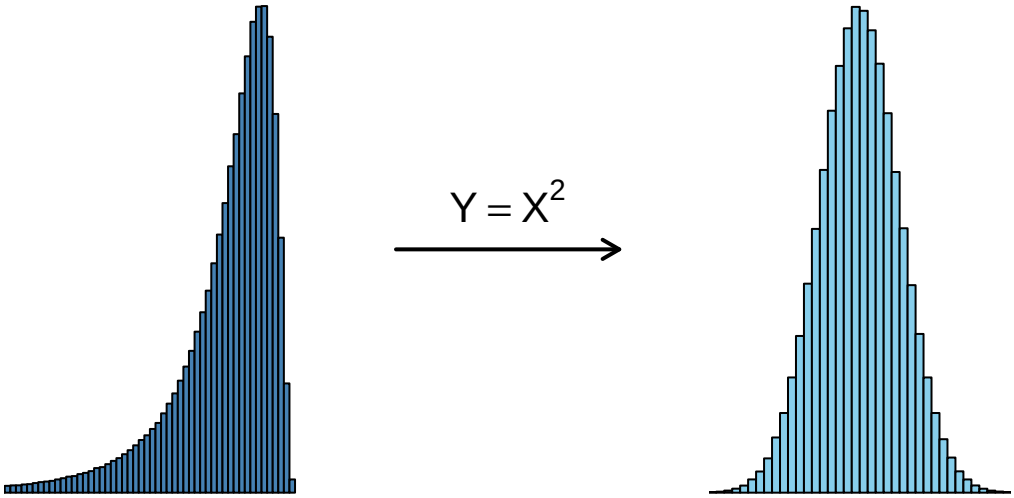
Student:	$Hans$	$Tina$	Kim	Flo	$Karl$
$Mathe:$	-1.5	0	-0.5	1.5	0.5
$Musik:$	-1.26	1.26	0.95	0	-0.95
$\sum$	-2.76	1.26	0.45	**1.5**	-0.45

…ist Student Flo der beste.

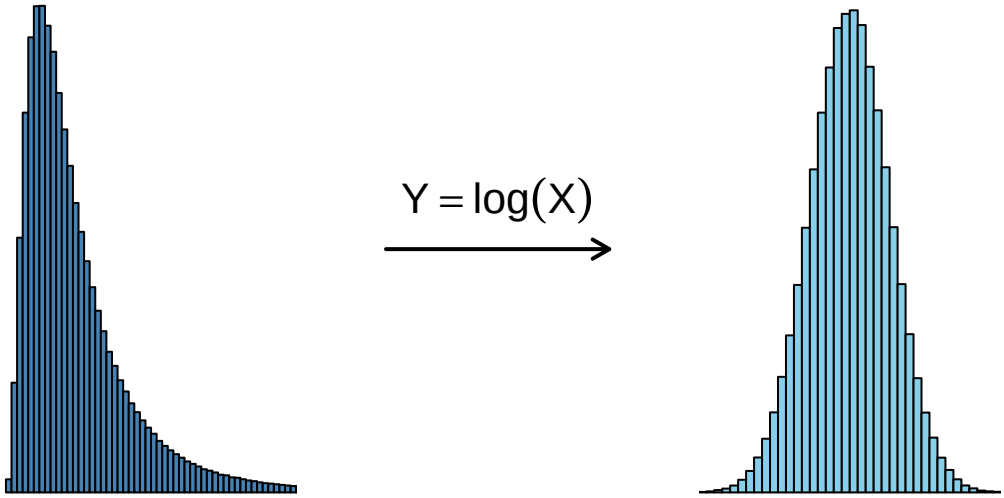
3.4.2 nicht-lineare Transformation

Nicht-lineare Transformationen werde verwendet, um die Nicht-Normalität der Verteilungen zu korrigieren.

Die Quadrat-Transformation $Y = X^2$ verringert kleine Werte und verstärkt große Werte. Daher wird sie verwendet, um links-schiefe Verteilungen zu korrigieren.



Die Wurzeltransformation $Y = \sqrt{X}$, die logarithmische Transformation $Y = \log(X)$ und die inverse Transformation $Y = \frac{1}{X}$ verringern große Werte und verstärken kleine Werte. Daher werden sie verwendet, um rechts-schiefe Verteilungen zu korrigieren.



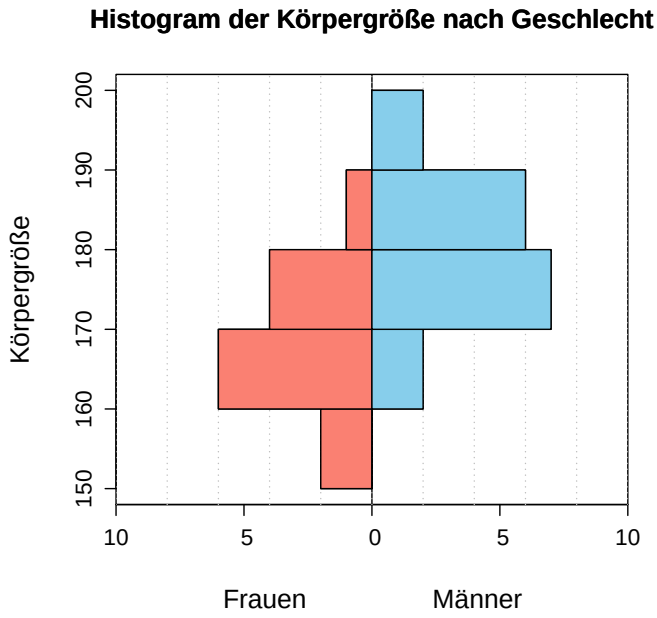
3.4.2.1 Faktoren

Manchmal ist es interessant, die Häufigkeitsverteilung der Stichprobe für verschiedene Teilproben zu beschreiben, die den Kategorien einer anderen Variablen entsprechen (z.B Geschlecht oder Beruf). Solche Klassifikationsvariablen werden als **Faktor** bezeichnet.

Beispiel

Das Aufteilen der Stichprobe "Körpergrößen" nach der Variable "Geschlecht" ergibt zwei Teilproben.

Frauen	173, 158, 174, 166, 162, 177, 165, 154, 166, 182, 169, 172, 170, 168.
Männer	179, 181, 172, 194, 185, 187, 198, 178, 188, 171, 175, 167, 186, 172, 176, 187.


Histogram der Körpergröße nach Geschlecht
Körpergröße
150
160
170
180
190
200
10
5
0
5
10
Frauen
Männer


Häufigkeitsverteilung Körpergröße nach Geschlech
Körpergröße
150
160
170
180
190
200
female
male
Geschlecht

4 Regression und Korrelation

Im vorherigen Kapitel haben wir gelernt, wie man die Verteilung einer einzelnen Variable in einer Stichprobe beschreibt. Allerdings müssen in den meisten Fällen mehrere Variablen beschrieben werden, die oft miteinander verbunden sind. Beispielsweise sollte eine Ernährungsstudie alle Variablen berücksichtigen, die mit dem Gewicht in Zusammenhang stehen könnten, wie z.B. Größe, Alter, Geschlecht, Rauchen, Ernährung und körperliche Betätigung.

Um ein Phänomen zu verstehen, das mehrere Variablen beinhaltet, reicht es nicht aus, jede Variable für sich allein zu studieren. Wir müssen alle Variablen gemeinsam untersuchen, um die Art ihrer Wechselbeziehungen und den Typ der Beziehung zwischen ihnen zu beschreiben.

In der Regel gibt es bei einer Abhängigkeitsstudie eine **abhängige** Variable Y, die von einem Satz an **unabhängigen** Variablen $X_1, \dots, X_n$ beeinflusst wird. Der einfachste Fall ist eine einfache Abhängigkeitsstudie mit nur einer unabhängigen Variable.

4.1 Gemeinsame Häufigkeiten

Um die Beziehung zwischen zwei Variablen X und Y zu untersuchen, müssen wir die gemeinsame Verteilung der zweidimensionalen Variable (X, Y) studieren, deren Werte Paare (x_i, y_j) sind, wobei das erste Element ein Wert von X und das zweite ein Wert von Y ist.

⚠ Definition “gemeinsame Häufigkeiten”

Bei einer Stichprobe mit n Werten und einer zweidimensionalen Variablen (X, Y), wird für jeden Wert der Variablen (x_i, y_j) folgendes definiert:

- Absolute Häufigkeit n_{ij}: Ist die Anzahl der Male, die das Paar (x_i, y_j) in der Stichprobe vorkommt.
- Relative Häufigkeit n_{ij}: Ist der Anteil der Male, die das Paar (x_i, y_j) in der Stichprobe vorkommt.

$$f_{ij} = \frac{n_{ij}}{n}$$

❗ Achtung!

Für zweidimensionale Variablen ergeben kumulative Häufigkeiten keinen Sinn.

4.1.1 gemeinsame Häufigkeitsverteilung

Die Werte der zweidimensionalen Variablen mit ihren Häufigkeiten werden als gemeinsame Häufigkeitsverteilung bezeichnet und in einer gemeinsamen Häufigkeitstabelle dargestellt.

$X\backslash Y$	y_1	$\cdots$	y_j	$\cdots$	y_q
x_1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1q}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_{i1}	$\cdots$	n_{ij}	$\cdots$	n_{iq}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_p	n_{p1}	$\cdots$	n_{pj}	$\cdots$	n_{pq}

Beispiel

Von 30 Studierenden wurden Körpergröße (in cm) und Gewicht (in kg) wie folgt gemessen:

(179,85), (173,65), (181,71), (170,65), (158,51), (174,66), (172,62), (166,60), (194,90), (185,75), (162,55), (187,78), (198,109), (177,61), (178,70), (165,58), (154,50), (183,93), (166,51), (171,65), (175,70), (182,60), (167,59), (169,62), (172,70), (186,71), (172,54), (176,68),(168,67), (187,80).

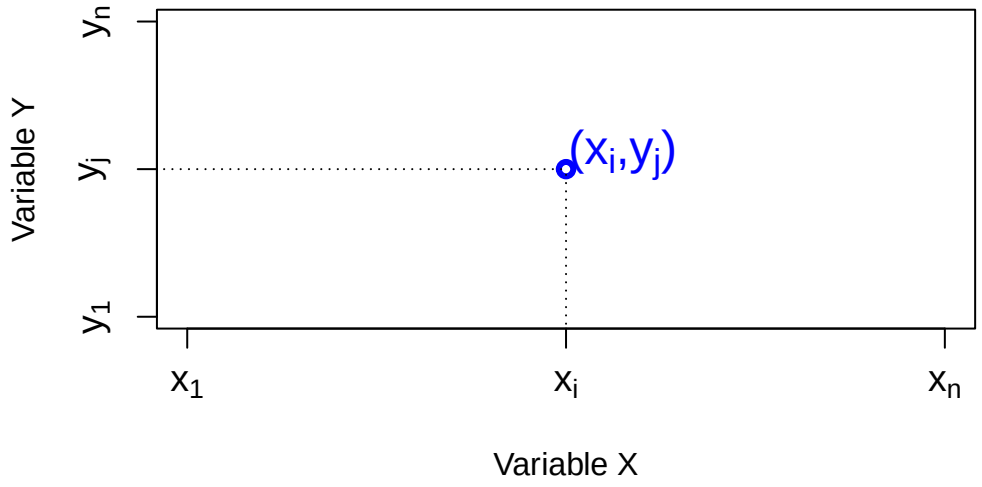
Die gemeinsame Häufigkeitstabelle ist entsprechend:

X/Y	$[50,60)$	$[60,70)$	$[70,80)$	$[80,90)$	$[90,100)$	$[100,110)$
$(150,160]$	2	0	0	0	0	0
$(160,170]$	4	4	0	0	0	0
$(170,180]$	1	6	3	1	0	0
$(180,190]$	0	1	4	1	1	0
$(190,200]$	0	0	0	0	1	1

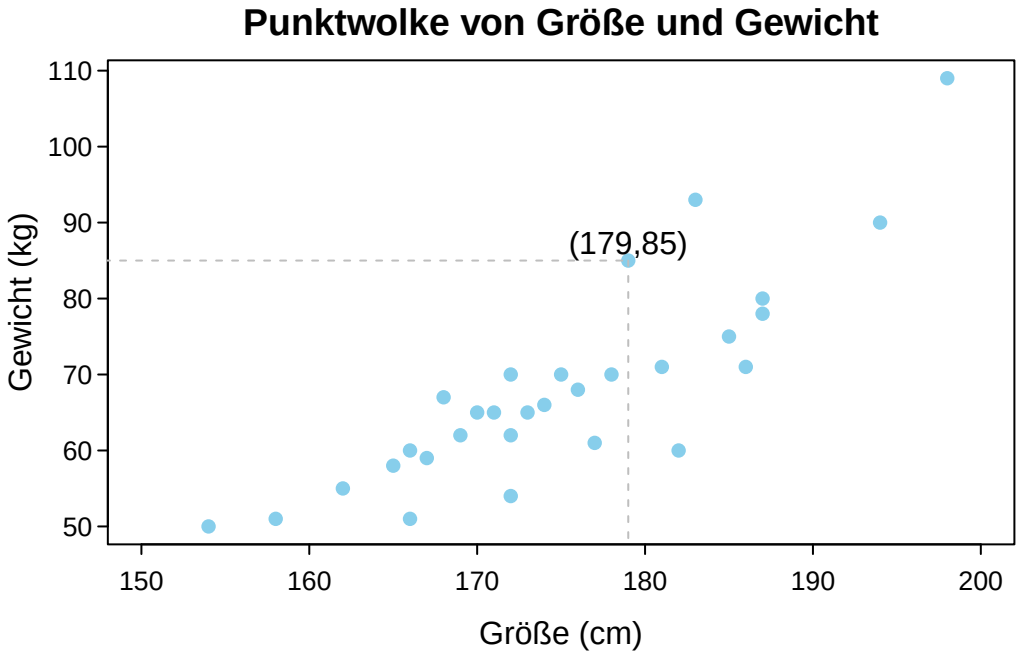
4.1.2 Streudiagramm

Die gemeinsame Häufigkeitsverteilung kann mit einem Streudiagramm (Punktwolke, Scatterplot) graphisch dargestellt werden, wobei die Daten als Sammlung von Punkten in einem XY-Koordinatensystem angezeigt werden.

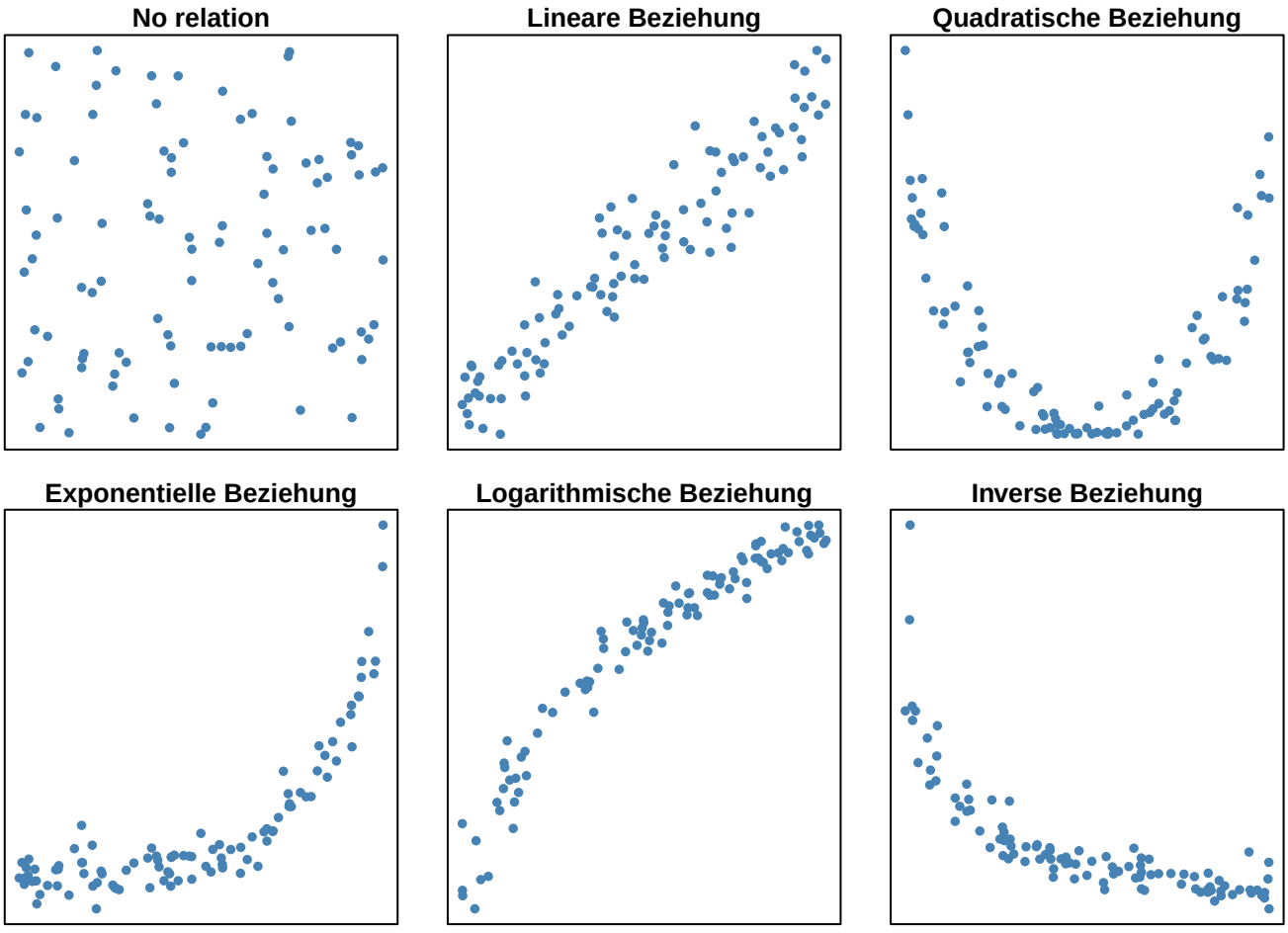
Üblicherweise wird die unabhängige Variable auf der X-Achse und die abhängige Variable auf der Y-Achse dargestellt. Für jedes Datenpaar (x_i, y_j) in der Stichprobe wird ein Punkt mit diesen Koordinaten auf der Ebene gezeichnet.



Das Ergebnis ist eine Ansammlung von Punkten, die auch *Punktwolke* genannt wird.



Die Form der Punktwolke gibt Aufschluss über die Art der Beziehung zwischen den Variablen X und Y.



4.1.2.1 Häufigkeitsverteilungen an den Rändern

Die Häufigkeitsverteilungen jeder der zweidimensionalen Variablen werden als Randhäufigkeitsverteilungen bezeichnet. Wir können die Randhäufigkeitsverteilungen aus der gemeinsamen Häufigkeitstabelle erhalten, indem wir die Häufigkeiten nach Zeilen und Spalten summieren.

$X\backslash Y$	y_1	$\cdots$	y_j	$\cdots$	y_q	n_x
x_1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1q}	n_{x_1}
$\vdots$	$\vdots$	$\vdots$	$+\downarrow$	$\vdots$	$\vdots$	$\vdots$
x_i	n_{i1}	$+\rightarrow$	n_{ij}	$+\rightarrow$	n_{iq}	n_{x_i}
$\vdots$	$\vdots$	$\vdots$	$+\downarrow$	$\vdots$	$\vdots$	$\vdots$
x_p	n_{p1}	$\cdots$	n_{pj}	$\cdots$	n_{pq}	n_{x_p}
n_y	n_{y_1}	$\cdots$	n_{y_j}	$\cdots$	n_{y_q}	n

Die Randhäufigkeitsverteilungen von Körpergröße und Gewicht sind

X/Y	$[50,60)$	$[60,70)$	$[70,80)$	$[80,90)$	$[90,100)$	$[100,110)$	n_x
$(150,160]$	2	0	0	0	0	0	2
$(160,170]$	4	4	0	0	0	0	8
$(170,180]$	1	6	3	1	0	0	11
$(180,190]$	0	1	4	1	1	0	7
$(190,200]$	0	0	0	0	1	1	2
n_y	7	11	7	2	2	1	30

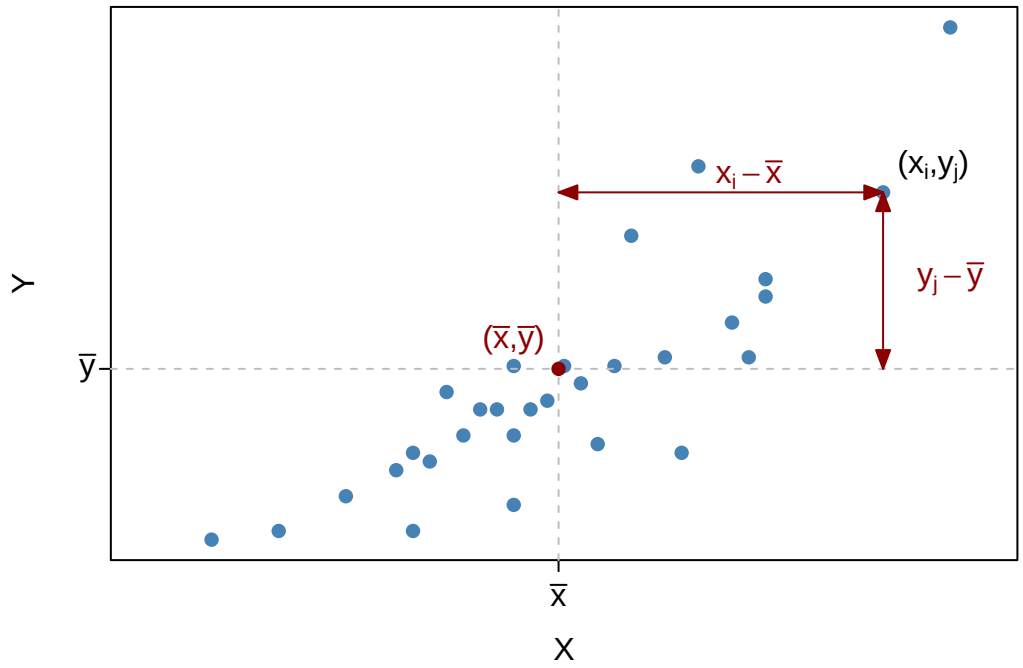
mit den den dazugehörigen Stichprobenstatistiken:

$$\bar{x} = 174.67 \text{ cm} \qquad s_x^2 = 102.06 \text{ cm}^2 \qquad s_x = 10.1 \text{ cm}$$
$$\bar{y} = 69.67 \text{ Kg} \qquad s_y^2 = 164.42 \text{ Kg}^2 \qquad s_y = 12.82 \text{ Kg}$$

4.2 Kovarianz

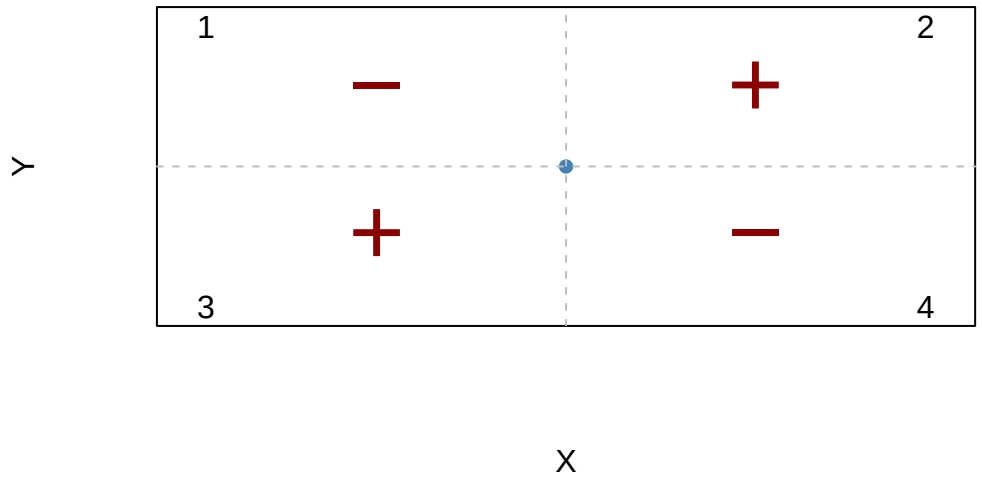
4.2.1 Abweichungen von den Mittelwerten

Um die Beziehung zwischen zwei Variablen zu untersuchen, müssen wir ihre gemeinsame Variation analysieren.



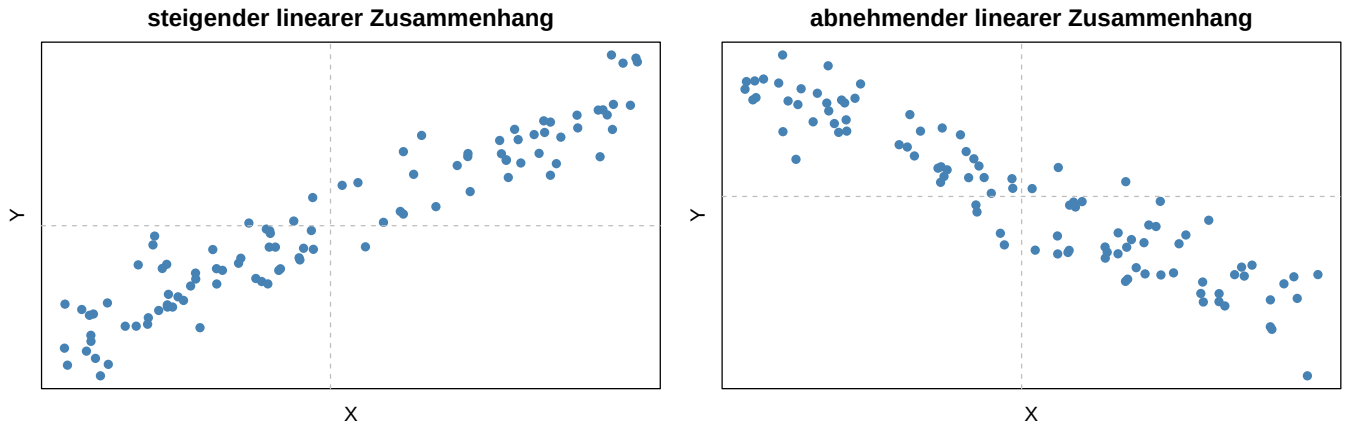
Wenn wir das Diagramm der Punktwolke in 4 Quadranten mit dem Mittelpunkt $(\bar{x}, \bar{y})$ unterteilen, so sind die Vorzeichen der Abweichungen vom Mittelwert:

Quadrant	$(x_i - \bar{x})$	$(y_j - \bar{y})$	$(x_i - \bar{x})(y_j - \bar{y})$
1	+	+	+
2	−	+	−
3	−	−	+
4	+	−	−



Wenn zwischen den Variablen eine *aufsteigende* lineare Beziehung besteht, werden die meisten Punkte in den Quadranten 1 und 3 liegen und die Summe der Produkte der Abweichungen vom Mittelwert wird *positiv* sein.

$$\sum (x_i - \bar{x})(y_j - \bar{y}) > 0$$



Wenn zwischen den Variablen eine *abnehmende* lineare Beziehung besteht, werden die meisten Punkte in den Quadranten 2 und 4 liegen und die Summe der Produkte der Abweichungen vom Mittelwert wird *negativ* sein.

$$\sum (x_i - \bar{x})(y_j - \bar{y}) < 0$$

4.2.2 Kovarianz

Definition "Kovarianz"

Die Kovarianz einer zweidimensionalen Variablen (X, Y) ist der Durchschnitt der Produkte der Abweichungen von den jeweiligen Mitteln.

$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_j - \bar{y}) n_{ij}}{n}$$

Sie kann auch unter Verwendung der folgenden Formel berechnet werden:

$$s_{xy} = \frac{\sum x_i y_j n_{ij}}{n} - \bar{x}\bar{y}$$

Die Kovarianz misst die lineare Beziehung zwischen zwei Variablen:

- Wenn $s_{xy} > 0$, besteht eine zunehmende lineare Beziehung.
- Wenn $s_{xy} < 0$, besteht eine abnehmende lineare Beziehung.
- Wenn $s_{xy} = 0$, gibt es keine lineare Beziehung.

Verwenden wir die Randhäufigkeitsverteilungen von Körpergröße und Gewicht

X/Y	$[50, 60)$	$[60, 70)$	$[70, 80)$	$[80, 90)$	$[90, 100)$	$[100, 110)$	n_x
$(150, 160]$	2	0	0	0	0	0	2
$(160, 170]$	4	4	0	0	0	0	8
$(170, 180]$	1	6	3	1	0	0	11
$(180, 190]$	0	1	4	1	1	0	7
$(190, 200]$	0	0	0	0	1	1	2
n_y	7	11	7	2	2	1	30

…mit den Mittelwerten.

$$\bar{x} = 174.67 \text{ cm} \qquad \bar{y} = 69.67 \text{ Kg}$$

So lässt sich die Kovarianz wie folgt berechnen:

$$\begin{aligned} s_{xy} &= \frac{\sum x_i y_j n_{ij}}{n} - \bar{x}\bar{y} \\ &= \frac{155 \cdot 55 \cdot 2 + 165 \cdot 55 \cdot 4 + \cdots + 195 \cdot 105 \cdot 1}{30} - 174.67 \cdot 69.67 = \\ &= \frac{368200}{30} - 12169.26 = 104.07 \text{cm} \cdot \text{Kg} \end{aligned}$$

Das bedeutet, dass eine positive lineare Beziehung zwischen den Variablen besteht.

4.3 Regression

In den meisten Fällen ist das Ziel einer Abhängigkeitsstudie nicht nur die Detektion einer Beziehung zwischen zwei Variablen, sondern auch die Expression dieser Beziehung mit einer mathematischen Funktion,

$$y = f(x)$$

um die abhängige Variable y für jeden Wert der unabhängigen x vorherzusagen.

Der Teil der Statistik, der sich mit dem Aufbau solch einer Funktion beschäftigt, wird Regression genannt, und die Funktion selbst heißt *Regressionsfunktion* oder *Regressionsmodell*.

4.3.1 Einfache Regressionsmodelle

Es gibt viele Arten von Regressionsmodellen. Die häufigsten Modelle sind in der folgenden Tabelle aufgeführt.

Model	**Gleichung**
$Linear$	$y = a + bx$
$Quadratisch$	$y = a + bx + cx^2$
$Kubisch$	$y = a + bx + cx^2 + dx^3$
$Potenz$	$y = a \cdot x^b$
$Exponentiell$	$y = e^{a+bx}$
$Logarithmisch$	$y = a + b \cdot \log(x)$
$Invers$	$y = a + \frac{b}{x}$
$Sigmoidal$	$y = e^{a+\frac{b}{x}}$

Die Wahl des Modells hängt von der Form der Punktwolke in der Streudiagramm ab.

4.3.2 Vorhersagefehler

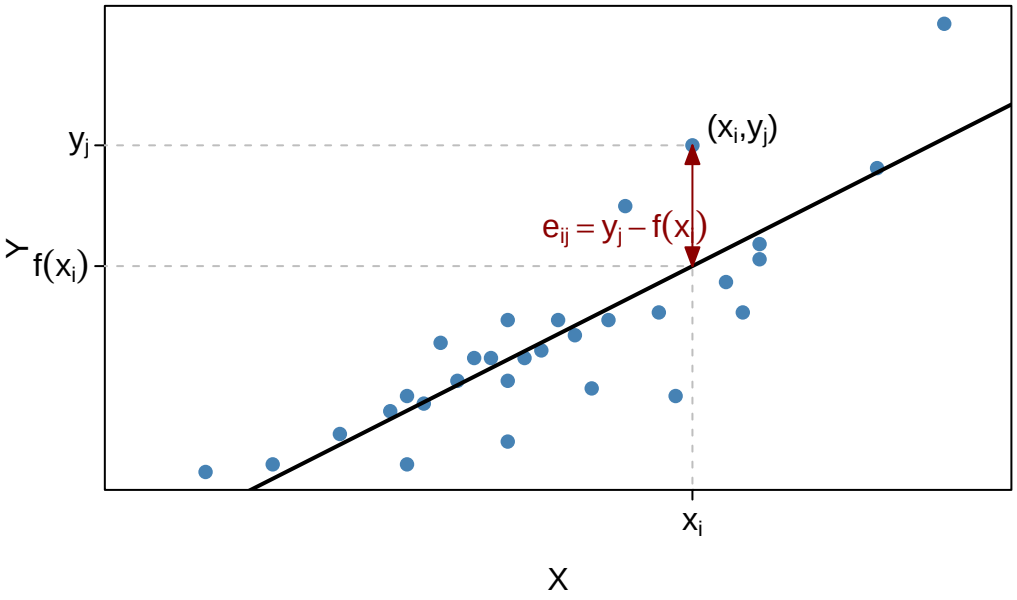
Sobald der Typ des Regressionsmodells gewählt wurde, müssen wir entscheiden, welche Funktion dieser Familie die Beziehung zwischen der abhängigen und unabhängigen Variablen am besten erklärt, d.h. welche Funktion die abhängige Variable am besten vorhersagt.

Diese Funktion ist diejenige, die die Abstände von den beobachteten Werten für Y in der Stichprobe zu den vorhergesagten Werten der Regressionsfunktion minimiert. Diese Abstände werden als *Restterme*, *Residuen* oder *Vorhersagefehler* bezeichnet.

Definition Vorhersagefehler:

Gegeben ein Regressionsmodell $y = f(x)$ für eine zweidimensionale Variable (X, Y), ist der Vorhersagefehler für jedes Paar (x_i, y_j) der Stichprobe die Differenz zwischen dem beobachteten Wert der abhängigen Variablen y_j und dem vorhergesagten Wert der Regressionsfunktion für x_i,

$$e_{i,j} = y_j - f(x_i).$$



4.3.2.1 Methode der kleinsten Quadrate

Um die Regressionsfunktion zu erhalten, kann die Methode der kleinsten Quadrate angewendet werde. Mit ihr wird die Funktion bestimmt, welche die quadrierten Residuen minimiert.

$$\sum e_{ij}^2.$$

Für ein lineares Modell $f(x) = a + bx$ hängt die Summe von zwei Parametern ab, dem Schnittpunkt durch die Y-Achse a und der Steigung der Geraden b,

$$\theta(a, b) = \sum e_{ij}^2 = \sum (y_j - f(x_i))^2 = \sum (y_j - a - bx_i)^2.$$

Dies reduziert das Problem darauf, geeignete Werte für a und b zu bestimmen.

4.4 Regressionsgerade

Um das Minimierungsproblem zu lösen, müssen wir die partiellen Ableitungen bezüglich a und b auf Null setzen.

$$\frac{\partial\theta(a,b)}{\partial a} = \frac{\partial\sum(y_j - a - bx_i)^2}{\partial a} = 0$$
$$\frac{\partial\theta(a,b)}{\partial b} = \frac{\partial\sum(y_j - a - bx_i)^2}{\partial b} = 0$$

Jetzt können wir das Gleichungssystem lösen zu:

$$a = \bar{y} - \frac{s_{xy}}{s_x^2}\bar{x} \qquad b = \frac{s_{xy}}{s_x^2}$$

Diese Formeln minimieren die Residuen von Y und ergeben das optimale lineare Modell.

Definition "Regressiongerade"

Für eine Stichprobe mit den zweidimensionalen Variablen (X, Y) ist die Regressionsgerade von Y auf X:

$$y = \bar{y} + \frac{s_{xy}}{s_x^2}(x - \bar{x})$$

Die Regressionsgerade von Y auf X ist die Gerade, die die Vorhersagefehler von Y minimiert und daher das lineare Regressionsmodell mit den besten Vorhersagen für Y darstellt.

4.4.1 Regressionlinienberechnung

Wir verwenden die Daten der Körpergrößen (X) und Körpergewichten (Y) mit den folgenden Kennwerten

$$\bar{x} = 174.67 \text{ cm} \qquad s_x^2 = 102.06 \text{ cm}^2 \qquad s_x = 10.1 \text{ cm}$$
$$\bar{y} = 69.67 \text{ Kg} \qquad s_y^2 = 164.42 \text{ Kg}^2 \qquad s_y = 12.82 \text{ Kg}$$
$$s_{xy} = 104.07 \text{ cm} \cdot \text{ Kg}$$

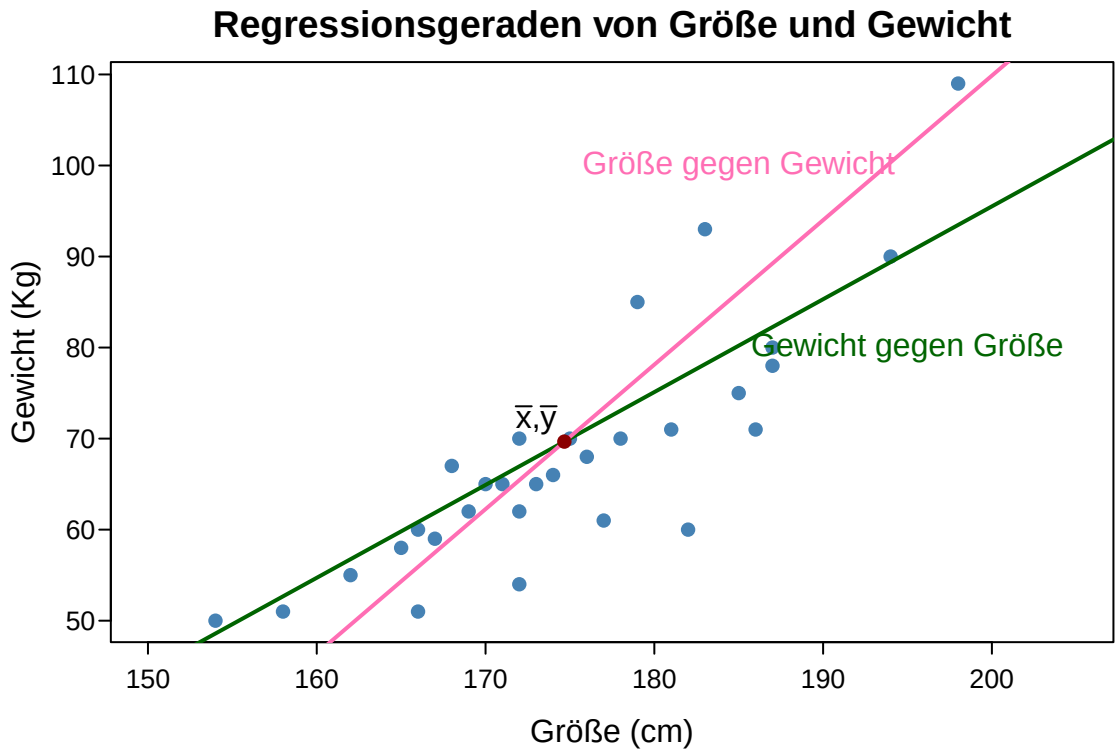
ergibt sich die Regressionsgerade für *Körpergröße erklärt durch Gewicht* (Größe gegen Gewicht, Größe $= f(\text{Gewicht})$) wie folgt:

$$y = \bar{y} + \frac{s_{xy}}{s_x^2}(x - \bar{x}) = 69.67 + \frac{104.07}{102.06}(x - 174.67) = -108.49 + 1.02x$$

Die Regressionsgerade für *Gewicht erklärt durch Körpergröße* (Gewicht gegen Größe, Gewicht $= f(\text{Größe})$) ist

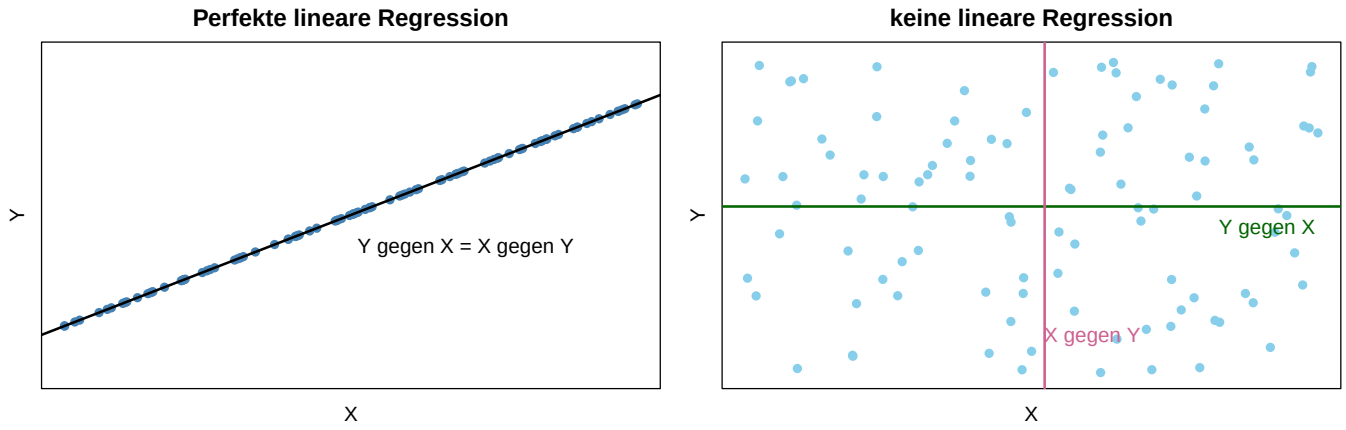
$$x = \bar{x} + \frac{s_{xy}}{s_y^2}(y - \bar{y}) = 174.67 + \frac{104.07}{164.42}(y - 69.67) = 130.78 + 0.63y$$

! Beachten Sie, dass die Regressionsgeraden unterschiedlich sind!



Üblicherweise sind die Regressionsgeraden von Y gegen X und X gegen Y **nicht** gleich, aber sie schneiden sich immer im Mittelpunkt $(\overline{x}, \overline{y})$.

Wenn es eine perfekte lineare Beziehung zwischen den Variablen gibt, dann sind beide Regressionsgeraden gleich, da diese Linie sowohl X-Residuen als auch Y-Residuen auf null setzt.



Wenn es keine lineare Beziehung zwischen den Variablen gibt, dann sind beide Regressionsgeraden konstant und gleich den jeweiligen Mittelwerten

$$y = \bar{y}, \qquad x = \bar{x}.$$

Daher schneiden sie sich senkrecht.

4.4.2 Regressionskoeffizient

Der wichtigste Parameter einer Regressionsgeraden ist die Steigung.

Definition Regressionskoeffizient b_{xy}

Für eine zweidimensionale Variable (X, Y) gibt der Regressionskoeffizient der Regressionsgeraden von Y gegen X deren Steigung an,

$$b_{yx} = \frac{s_{xy}}{s_x^2}$$

Der Regressionskoeffizient hat immer das gleiche Vorzeichen wie die Kovarianz. Er misst, wie sich die abhängige Variable Y in Bezug auf die unabhängige Variable X gemäß der Regressionsgerade verändert. Insbesondere gibt er an, um wie viele Einheiten die abhängige Variable pro Einheit, die die unabhängige Variable zunimmt, steigt oder fällt.

Beispiel: Körpergrößen und Gewichte

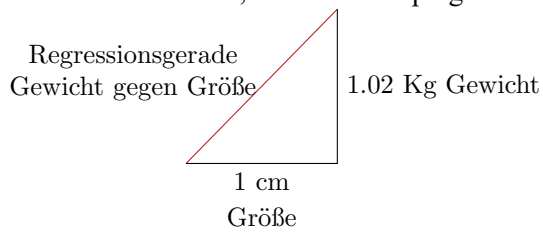
In unserer Stichprobe mit den Variablen Körpergröße und Gewicht war die Regressionsgerade für das Gewicht in Bezug auf die Körpergröße

$$y = -108.49 + 1.02x.$$

Daher ist der Regressionskoeffizient für das Gewicht in Bezug auf die Körpergröße

$$b_{yx} = 1.02 \text{ Kg/cm}.$$

Das bedeutet, dass das Gewicht gemäß der Regressionsgerade in Bezug auf die Körpergröße um 1.02 kg pro Zentimeter zunimmt, wenn die Körpergröße steigt.



4.4.3 Modellvorhersage

Üblicherweise werden Regressionsmodelle verwendet, um die abhängige Variable Y für bestimmte Werte der unabhängigen Variablen X vorherzusagen.

Denken Sie daran!

Um die besten Vorhersagen einer Variablen zu erhalten, müssen Sie die Regressionsgerade verwenden, bei der diese Variable die abhängige Rolle Y spielt.

Möchten wir das Gewicht einer Person mit einer Körpergröße von 180cm vorhersagen, müssen wir die Regressionsgerade für das Gewicht in Bezug auf die Körpergröße verwenden.

$$y = -108.49 + 1.02 \cdot 180 = 75.11 \text{ Kg}.$$

Um hingegen die Körpergröße einer Person mit einem Gewicht von 79kg vorherzusagen, müssen wir die Regressionsgerade für die Körpergröße in Bezug auf das Gewicht verwenden,

$$x = 130.78 + 0.63 \cdot 79 = 180.55 \text{ cm}.$$

Aber wie zuverlässig sind diese Vorhersagen?

4.5 Korrelation

Sobald wir ein Regressionsmodell haben, müssen wir seine Vorhersagefähigkeit bewerten, indem wir die Anpassungsgüte des Modells und die Stärke der Beziehung, die es aufbaut, untersuchen. Dieser Teil der Statistik wird als **Korrelation** bezeichnet.

Die Korrelationsanalyse untersucht die Residuen eines Regressionsmodells: Je kleiner die Residuen, desto besser passt das Modell und desto stärker ist die Beziehung, die es aufbaut. Um die Anpassungsgüte eines Regressionsmodells zu messen, wird häufig die *residuale Varianz* verwendet.

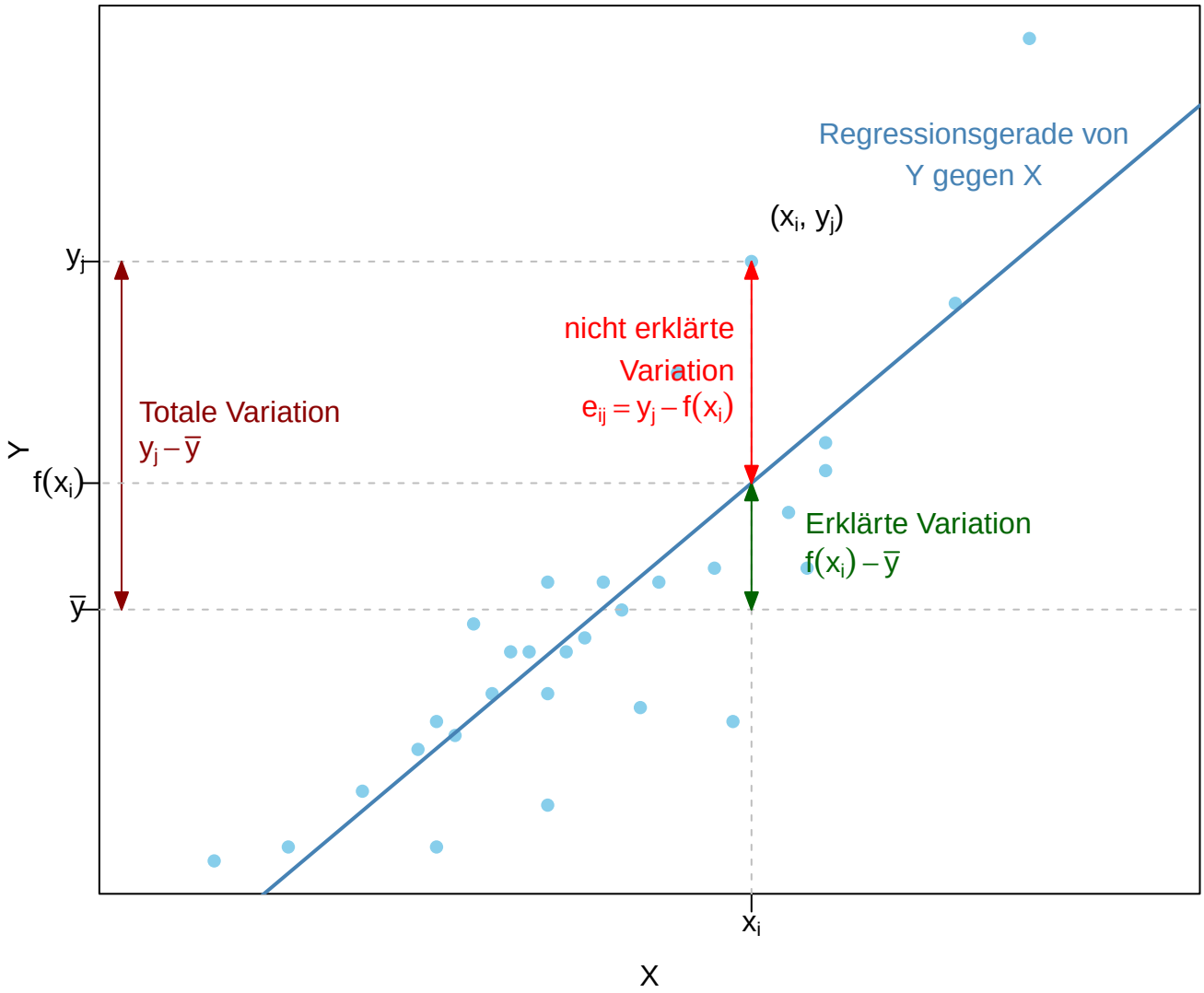
⚠ Definition "Residuale Varianz s^2_{ry}

Bei einem Regressionsmodell $y = f(x)$ für eine zweidimensionale Variable (X, Y) ist die residuale Varianz die durchschnittliche quadratische Abweichung der Residuen,

$$s^2_{ry} = \frac{\sum e^2_{ij} n_{ij}}{n} = \frac{\sum (y_j - f(x_i))^2 n_{ij}}{n}.$$

Je größer die Residuen, desto größer die residuale Varianz und desto kleiner die Anpassungsgüte. Wenn die lineare Beziehung perfekt ist, sind die Residuen Null und die residuale Varianz ist ebenfalls Null. Umgekehrt gilt, dass wenn es keine Beziehung gibt, die Residuen mit den Abweichungen vom Mittelwert übereinstimmen und die residuale Varianz gleich der Varianz der abhängigen Variable ist.

$$0 \leq s^2_{ry} \leq s^2_y$$



4.6 Korrelationskoeffizienten

4.6.1 Bestimmheitsmaß

Es ist möglich, aus der Reststreuung eine andere Korrelationsstatistik zu definieren, die leichter interpretierbar ist.

Definition Bestimmtheitsmaß R^2).

Für ein Regressionsmodell $y = f(x)$ einer zweidimensionalen Variablen (X, Y), ist sein Bestimmtheitsmaß

$$R^2 = 1 - \frac{s_{ry}^2}{s_y^2}$$

Da die Reststreuung von 0 bis s_y^2 reicht, gilt:

$$0 \leq R^2 \leq 1$$

Je größer R^2 ist, desto besser passt das Regressionsmodell und desto zuverlässiger werden seine Vorhersagen sein.

- Wenn $R^2 = 0$, gibt es keine Beziehung gemäß dem Regressionsmodell.
- Wenn $R^2 = 1$, ist die Beziehung gemäß dem Modell perfekt.

4.6.2 Lineares Bestimmtheitsmaß

Wenn das Regressionsmodell linear ist, beträgt die Reststreuung

$$\begin{aligned} s_{ry}^2 &= \sum e_{ij}^2 f_{ij} = \sum (y_j - f(x_i))^2 f_{ij} = \sum \left(y_j - \bar{y} - \frac{s_{xy}}{s_x^2}(x_i - \bar{x}) \right)^2 f_{ij} \\ &= \sum \left((y_j - \bar{y})^2 + \frac{s_{xy}^2}{s_x^4}(x_i - \bar{x})^2 - 2\frac{s_{xy}}{s_x^2}(x_i - \bar{x})(y_j - \bar{y}) \right) f_{ij} \\ &= \sum (y_j - \bar{y})^2 f_{ij} + \frac{s_{xy}^2}{s_x^4} \sum (x_i - \bar{x})^2 f_{ij} - 2\frac{s_{xy}}{s_x^2} \sum (x_i - \bar{x})(y_j - \bar{y}) f_{ij} \\ &= s_y^2 + \frac{s_{xy}^2}{s_x^4} s_x^2 - 2\frac{s_{xy}}{s_x^2} s_{xy} = s_y^2 - \frac{s_{xy}^2}{s_x^2}. \end{aligned}$$

… und das Bestimmtheitsmaß ist ….

$$R^2 = 1 - \frac{s_{ry}^2}{s_y^2} = 1 - \frac{s_y^2 - \frac{s_{xy}^2}{s_x^2}}{s_y^2} = 1 - 1 + \frac{s_{xy}^2}{s_x^2 s_y^2} = \frac{s_{xy}^2}{s_x^2 s_y^2}.$$

Beispiel

In der Stichprobe zu Körpergröße und Gewicht gelten folgende Kennwerte:

$$\begin{aligned} \bar{x} &= 174.67 \text{ cm} & s_x^2 &= 102.06 \text{ cm}^2 \\ \bar{y} &= 69.67 \text{ Kg} & s_y^2 &= 164.42 \text{ Kg}^2 \\ s_{xy} &= 104.07 \text{ cm} \cdot \text{ Kg} \end{aligned}$$

Das Bestimmtheitsmaß errechnet sich also per

$$R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} = \frac{(104.07 \text{ cm} \cdot \text{Kg})^2}{102.06 \text{ cm}^2 \cdot 164.42 \text{ Kg}^2} = 0.65.$$

Das bedeutet, dass das lineare Modell *Gewicht erklärt durch Größe* 65% der Variation des Gewichts erklärt und das lineare Modell *Körpergröße erklärt durch Gewicht* ebenfalls 65% der Variation der Körpergröße erklärt.

4.6.3 Korrelationskoeffizient

Definition Korrelationskoeffizient r^2

Gegeben eine Stichprobe zweidimensionaler Variablen (X, Y), ist der Korrelationskoeffizient der Stichprobe die Wurzel des linearen Bestimmtheitsmaßes mit dem Vorzeichen der Kovarianz:

$$r = \frac{s_{xy}}{s_x s_y}.$$

Das Bestimmheitsmaß R^2 reicht von 0 bis 1, und r reicht von -1 bis 1,

$$-1 \leq r \leq 1$$

Der Korrelationskoeffizient misst nicht nur die Stärke, sondern auch die Richtung (steigend oder fallend) der linearen Beziehung:

- Wenn $r = 0$, gibt es keine lineare Beziehung.
- Wenn $r = 1$, gibt es eine perfekte steigende lineare Beziehung.
- Wenn $r = -1$, gibt es eine perfekte fallende lineare Beziehung.

Beispiel

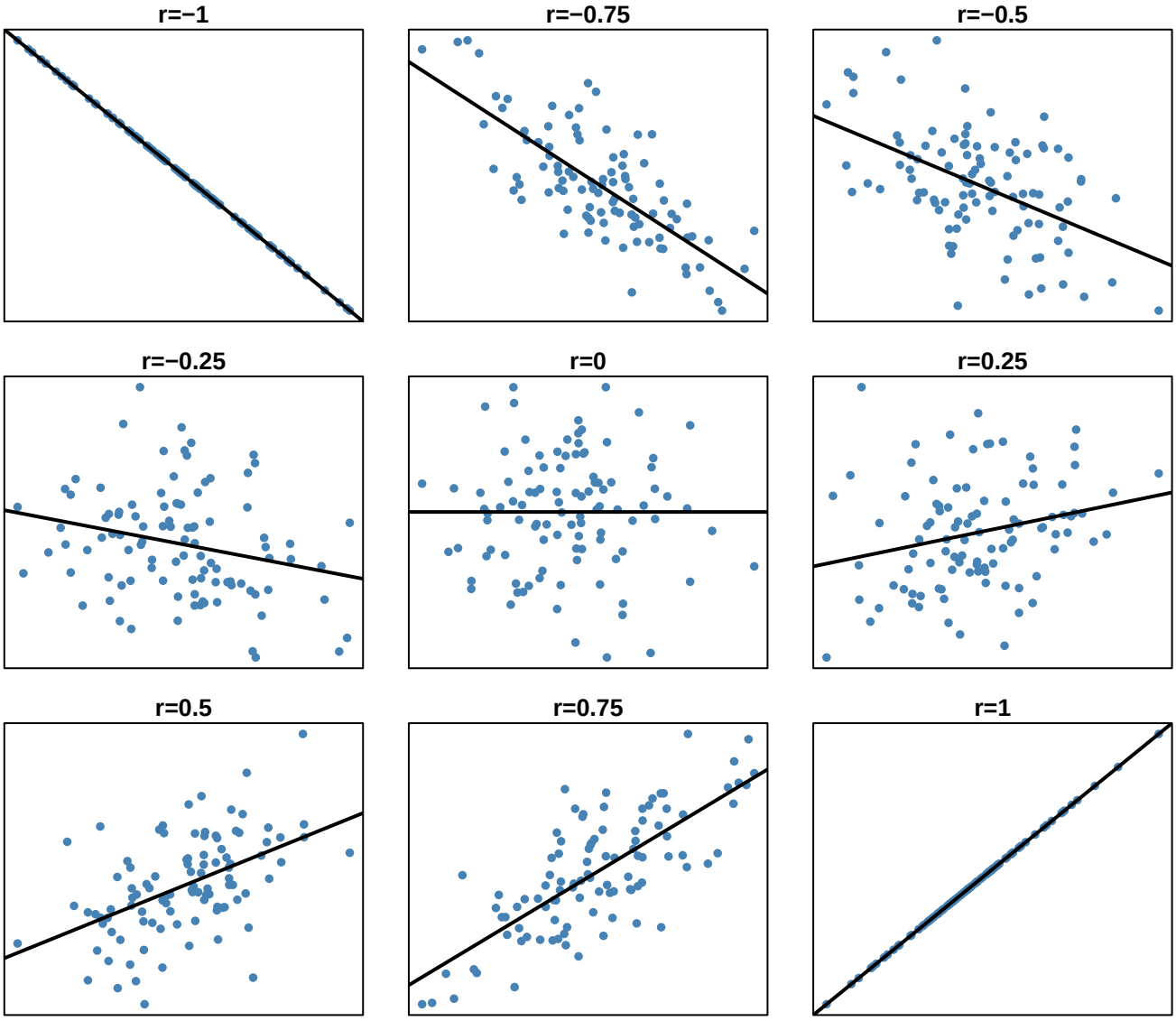
In der Stichprobe zu Körpergröße und Gewicht gelten folgende Kennwerte:

$$\begin{aligned} \bar{x} &= 174.67 \text{ cm} & s_x^2 &= 102.06 \text{ cm}^2 \\ \bar{y} &= 69.67 \text{ Kg} & s_y^2 &= 164.42 \text{ Kg}^2 \\ s_{xy} &= 104.07 \text{ cm} \cdot \text{ Kg} \end{aligned}$$

Der Korrelationskoeffizient kann nun berechnet werden mit:

$$r = \frac{s_{xy}}{s_x s_y} = \frac{104.07 \text{ cm} \cdot \text{Kg}}{10.1 \text{ cm} \cdot 12.82 \text{ Kg}} = +0.8.$$

Dies bedeutet, dass es eine relativ starke steigende lineare Beziehung zwischen Körpergröße und Gewicht gibt.



4.6.4 Zuverlässigkeit von Regressionsvorhersagen

Das Bestimmtheitsmaß erklärt, wie gut ein Regressionsmodell passt, aber andere Faktoren beeinflussen auch die Zuverlässigkeit von Regressionsvorhersagen:

- Ein höheres R^2 bedeutet eine bessere Passung und zuverlässigere Vorhersagen.
- Eine höhere Variabilität der Grundgesamtheit macht Vorhersagen schwieriger und weniger zuverlässig.
- Eine größere Stichprobengröße bietet mehr Informationen und ermöglicht zuverlässigere Vorhersagen.

Darüber hinaus muss man beachten, dass ein Regressionsmodell nur für den in der Stichprobe beobachteten Wertebereich gültig ist. Daher sollten keine Vorhersagen für Werte weit außerhalb dieses Bereichs gemacht werden, da wir für diese keine Informationen haben.

4.7 nicht-lineare Regression

Eine nichtlineare Regression kann auch mit der Methode der kleinsten Quadrate durchgeführt werden. Allerdings kann in einigen Fällen die Anpassung eines nichtlinearen Modells durch eine einfache Transformation seiner Variablen auf die Anpassung eines linearen Modells reduziert werden.

4.7.1 Transformationen von nichtlinearen Regressionsmodellen

- Logarithmisches Modell: Ein logarithmisches Modell $y = a + b \log x$ kann durch den Wechsel $t = \log x$ in ein lineares Modell überführt werden:

$$y = a + b \log x = a + bt$$

- Exponentielles Modell: Ein exponentielles Modell $y = e^{a+bx}$ kann durch den Wechsel $z = \log y$ in ein lineares Modell überführt werden:

$$z = \log y = \log(e^{a+bx}) = a + bx$$

- Potenzialmodell: Ein Potenzialmodell $y = ax^b$ kann durch die Wechsel $t = \log x$ und $z = \log y$ in ein lineares Modell überführt werden:

$$z = \log y = \log(ax^b) = \log a + b \log x = a' + bt.$$

- Inversmodell: Ein Inversmodell $y = a + b/x$ kann durch den Wechsel $t = 1/x$ in ein lineares Modell überführt werden:

$$y = a + b(1/x) = a + bt$$

- Sigmoidales Modell: Ein sigmoidales Modell $y = e^{a+b/x}$ kann durch die Wechsel $t = 1/x$ und $z = \log y$ in ein lineares Modell überführt werden:

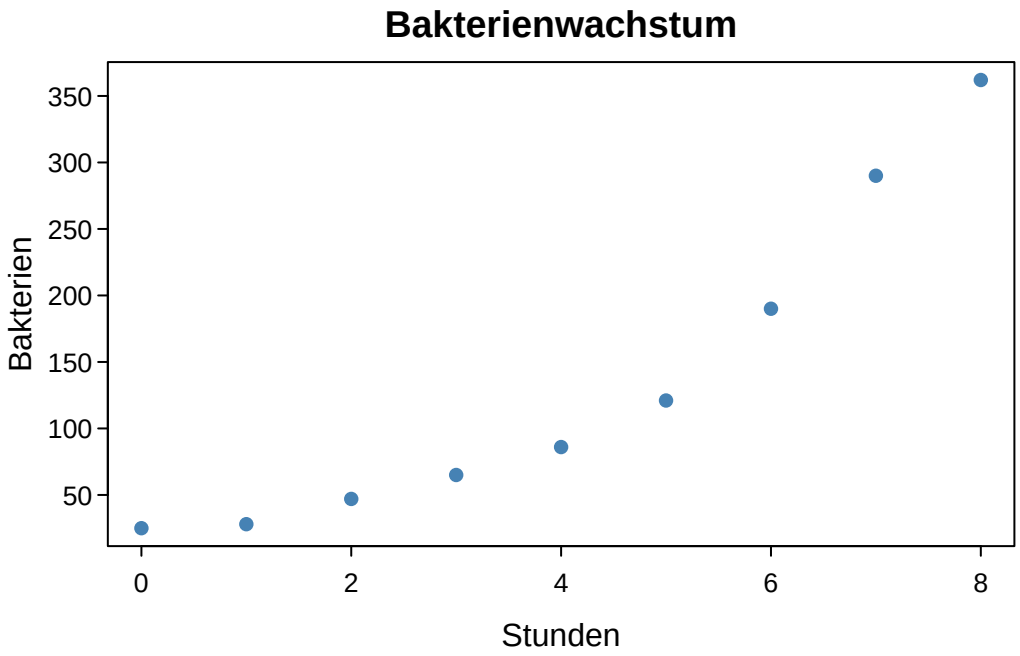
$$z = \log y = \log(e^{a+b/x}) = a + b(1/x) = a + bt$$

4.7.2 exponentieller Zusammenhang

Die nachfolgende Tabelle zeigt die Anzahl an Bakterien in einer Kultur über mehrere Stunden.

Stunden	Bakterien
0	25
1	28
2	47
3	65
4	86
5	121
6	190
7	290
8	362

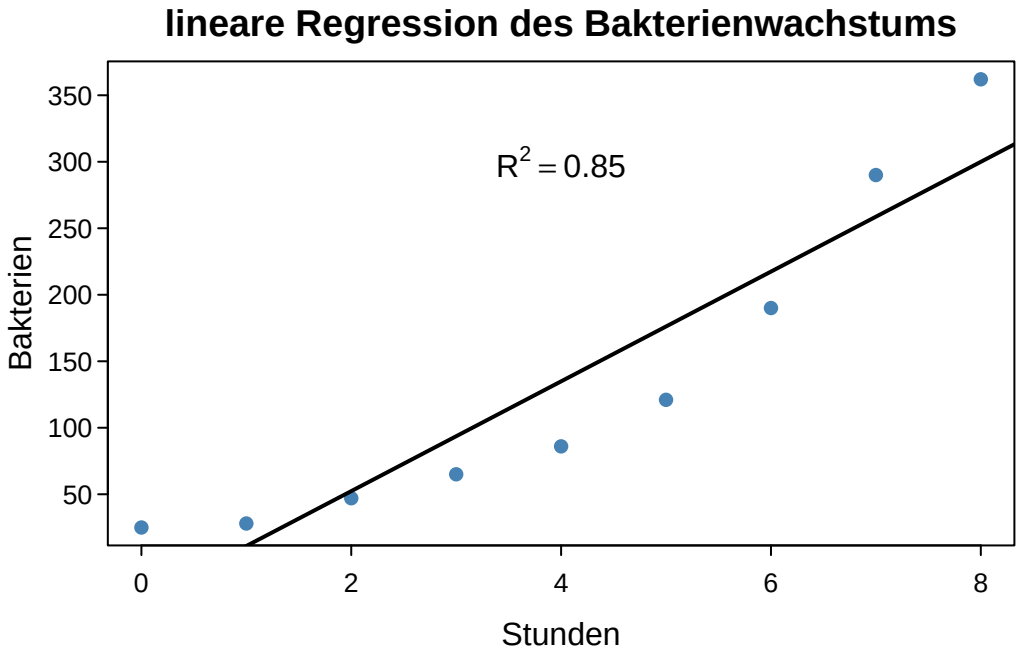
Das dazugehörige Streudiagramm sieht so aus:



Wenn wir ein lineares Modell erstellen (engl: *to fit*, auch *fitten* genannt), erhalten wir

$$\text{Bakterien} = -30,18 + 41,27 \text{ Stunden, mit } R^2 = 0,85.$$

Die Regressionsgerade sieht so aus:



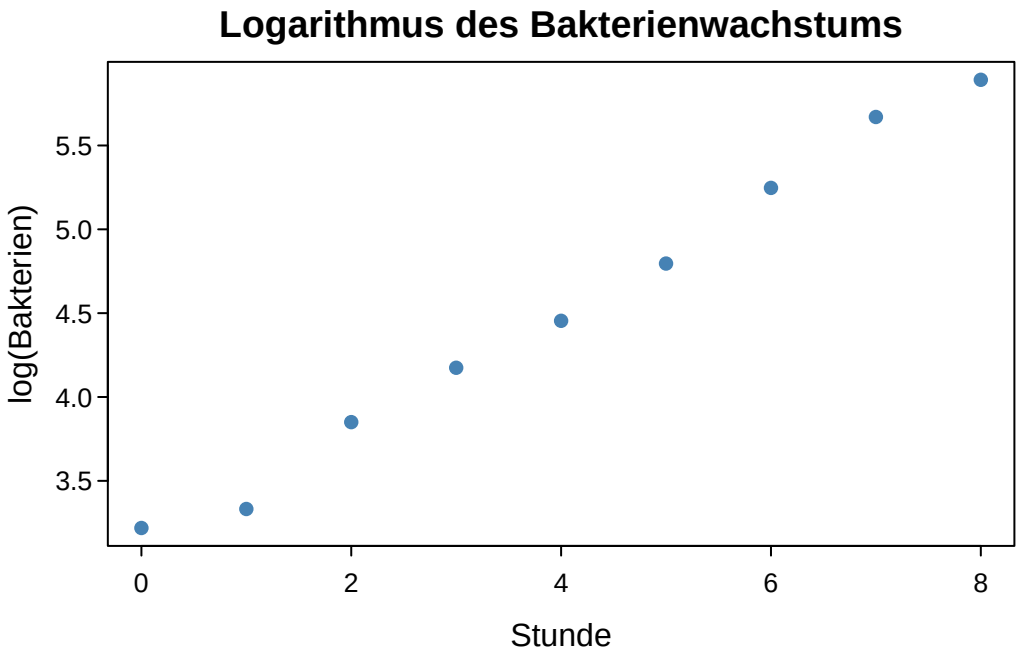
Ist dies ein *gutes* Modell?

4.7.2.1 exponentielles Modell

Obwohl das lineare Modell nicht schlecht ist, erscheint ein exponentielles Modell aufgrund der Form des Punktwolken-Clouds des Streuplots geeigneter zu sein.

Um ein exponentielles Modell $y = e^{a+bx}$ zu konstruieren, können wir die Transformation $z = \log y$ anwenden, indem wir also eine logarithmische Transformation der abhängigen Variablen durchführen.

Stunden	Bakterien	$\log(Bakterien)$
0	25	3.22
1	28	3.33
2	47	3.85
3	65	4.17
4	86	4.45
5	121	4.80
6	190	5.25
7	290	5.67
8	362	5.89



Jetzt bleibt nur noch, die Regressionslinie des Logarithmus der Bakterien über die Stunden zu berechnen:

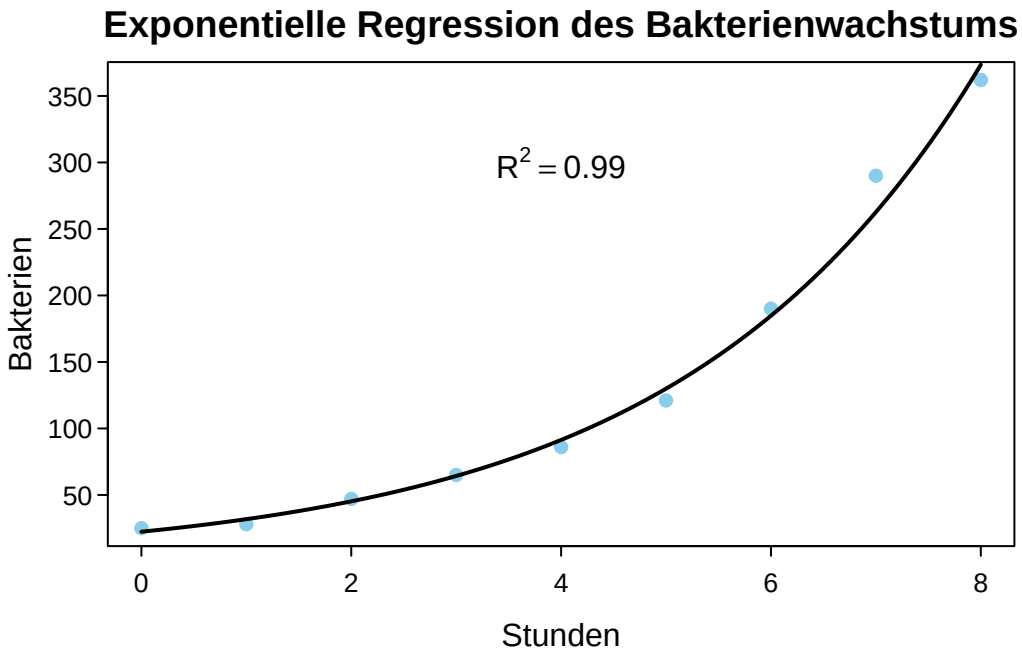
$$\log(Bakterien) = 3.107 + 0.352 \text{ Stunden},$$

und die Variable abschließend zurücktransformieren,

$$\text{Bakterien} = e^{3.107+0.352 \text{ Stunden}}$$

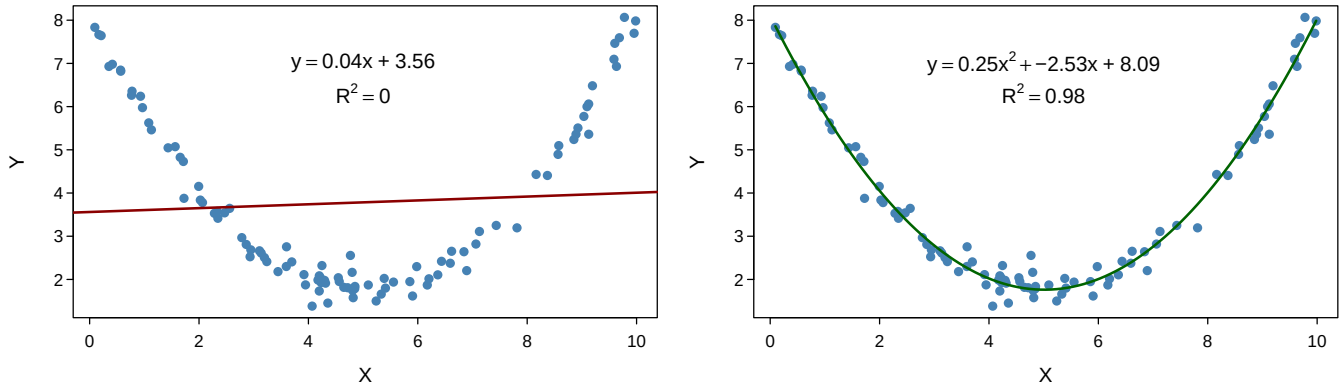
errechnet sich $R^2 = 0,99$.

Daher passt das exponentielle Modell deutlich besser als das lineare Modell.



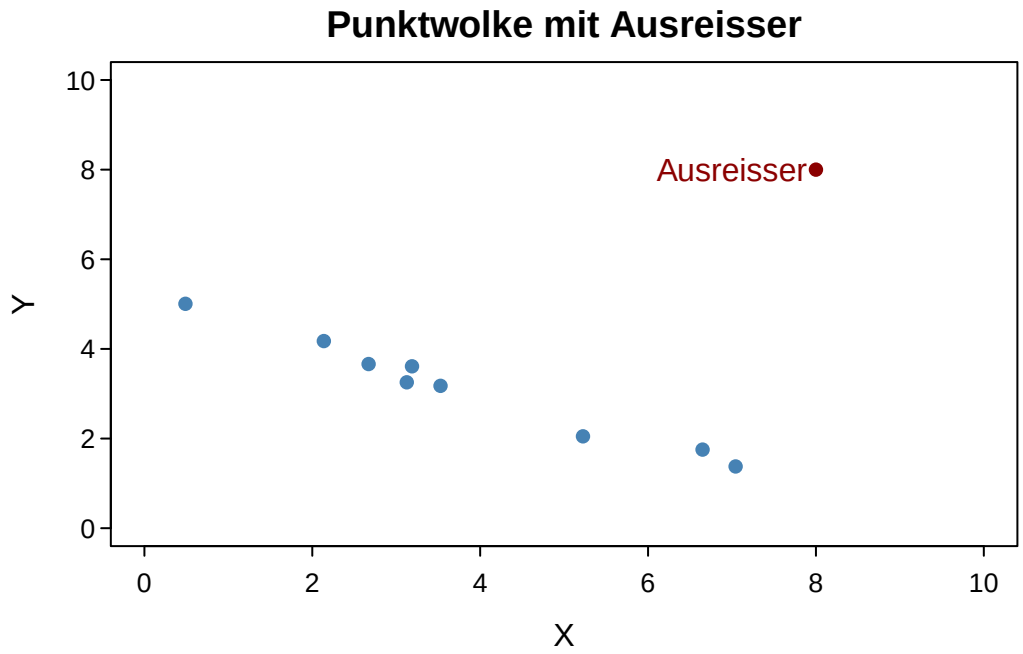
4.7.3 Regressionsrisiken

Es ist wichtig zu beachten, dass jedes Regressionsmodell seinen eigenen Bestimmtheitskoeffizienten hat. Daher bedeutet ein nah an Null liegender Bestimmtheitskoeffizient, dass es keine vom Modell festgelegte Beziehung gibt. Aber das bedeutet **nicht**, dass die Variablen *unabhängig* sind, da es eine andere Art von Beziehung geben könnte.

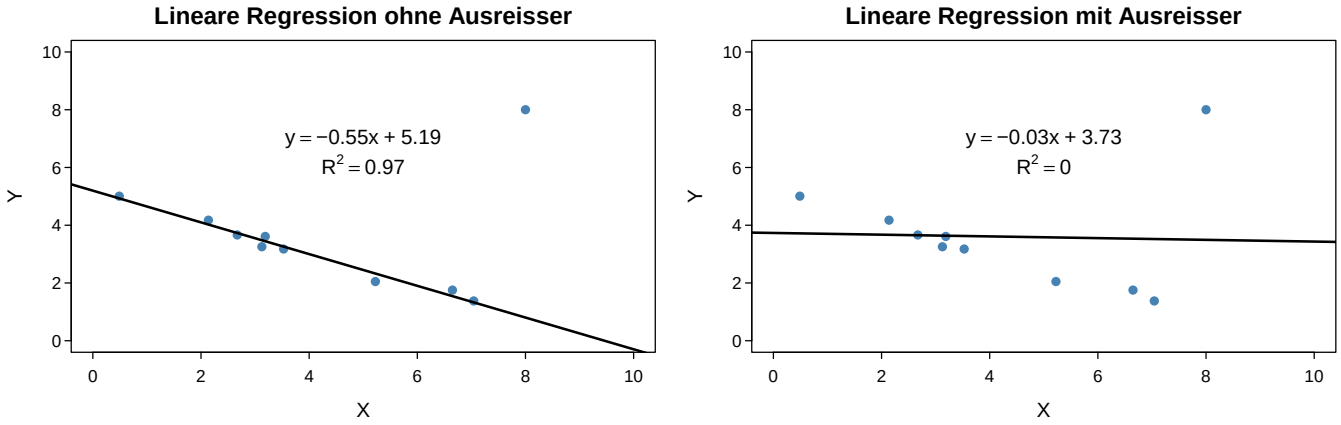


4.7.4 Ausreißer in Regressionsmodellen

In Regressionsstudien sind Ausreißer Punkte, die nicht der Tendenz der restlichen Punkte folgen, auch wenn die Werte des Paares für jede Variable separat keine Ausreißer sind.

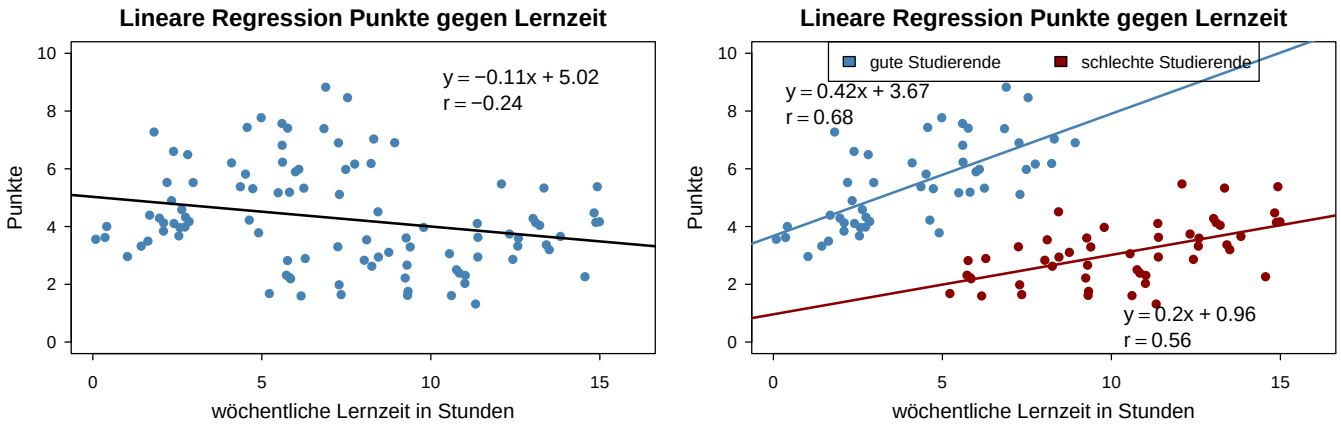


Ausreißer in Regressionsstudien können drastische Veränderungen in den Regressionsmodellen hervorrufen.



4.7.5 Simpsons Paradox

Manchmal verschwindet ein Trend oder kehrt sich sogar um, wenn man die Probe nach einer qualitativen Variable aufteilt, die mit der abhängigen Variablen zusammenhängt. Dies wird als Simpsons Paradox bezeichnet.



5 Wahrscheinlichkeit

Die deskriptive Statistik bietet Methoden zur Beschreibung der in der Probe gemessenen Variablen und ihrer Beziehungen, ermöglicht jedoch keine Schlüsse aus dieser Probe auf die Grundgesamtheit.

Jetzt ist es Zeit, von der Stichprobe zur Grundgesamtheit zu gelangen, und die Brücke dafür ist die **Wahrscheinlichkeitstheorie**.

Bitte beachten Sie, dass die Stichprobe nur begrenzte Informationen über die Grundgesamtheit hat, und um gültige Schlüsse für diese zu ziehen, muss die Probe *repräsentativ* sein. Dies ist strenggenommen nur bei Zufallsstichproben gegeben.

Die Wahrscheinlichkeitstheorie wird uns Werkzeuge bieten, um den Zufall in der Stichprobenziehung zu kontrollieren und das Vertrauensniveau der aus der Stichprobe gezogenen Schlüsse zu bestimmen.

5.1 Experiment mit zufälligem Ergebnis

Die Untersuchung einer Eigenschaft der Grundgesamtheit erfolgt durch Experimente mit zufälligem Ergebnis.

⚠ Definition “Zufallsexperiment”

Ein Zufallsexperiment ist ein Experiment, das zwei Bedingungen erfüllt:

1. Alle möglichen Ergebnisse sind bekannt.
2. Es ist unmöglich, das Ergebnis mit absoluter Sicherheit vorherzusagen.

💡 Beispiel: Glücksspiele

Glücksspiele sind typische Beispiele für Experimente mit zufälligem Ergebnis. Zum Beispiel ist das Werfen eines Würfels ein solches Experiment, weil:

1. Der Satz der möglichen Ergebnisse bekannt ist: $\{1, 2, 3, 4, 5, 6\}$.
2. Bevor der Würfel geworfen wird ist es unmöglich das Ergebnis mit absoluter Sicherheit vorherzusagen.

Ein weiteres Beispiel, das nichts mit Glücksspielen zu tun hat, ist die zufällige Auswahl einer Person aus einer menschlichen Grundgesamtheit und die Bestimmung ihrer Blutgruppe.

Grundsätzlich ist die Ziehung einer Stichprobe mittels einer Zufallsmethode ein Zufallsexperiment.

5.1.1 Wahrscheinlichkeitsraum

⚠ Definition “Wahrscheinlichkeitsraum Ω”

Der Satz Ω der möglichen Ergebnisse eines Zufallsexperiments wird als Wahrscheinlichkeitsraum (Probabilitätsraum) bezeichnet.

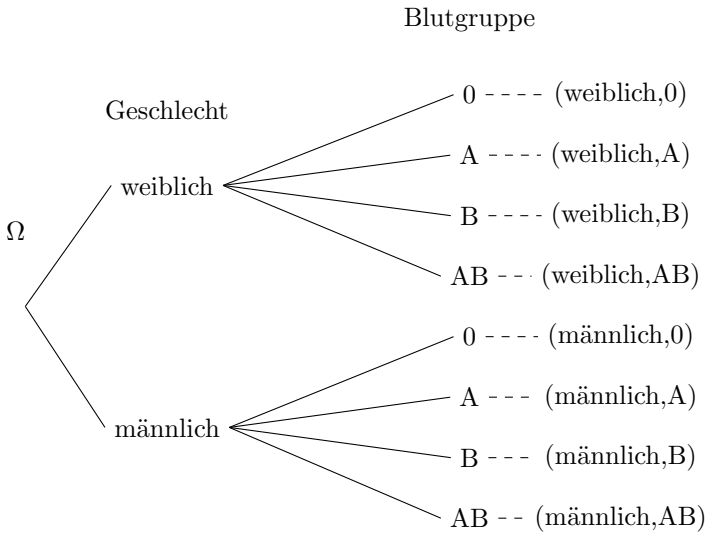
Beispiele für Probabilitätsräume:

- Beim Werfen einer Münze ist $\Omega = \{Kopf, Zahl\}$.
- Beim Werfen eines Würfels ist $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Beim Blutgruppentest einer zufällig ausgewählten Person ist $\Omega = \{0, A, B, AB\}$.
- Bei der Körpergröße einer zufällig ausgewählten Person ist $\Omega = \mathbb{R}^+$.

5.1.2 Baumdiagramme

In Experimenten, bei denen mehr als eine Variable gemessen wird, kann die Bestimmung des Wahrscheinlichkeitsraums schwierig sein. In solchen Fällen ist es ratsam, ein Baumdiagramm zur Konstruktion des Probabilitätsraums zu verwenden.

In einem Baumdiagramm wird jede Variable auf einer Ebene des Baums dargestellt und jeder mögliche Wert der Variablen als Zweig.



5.1.3 Zufallsereignis

Definition Zufallsereignis

Ein zufälliges Ereignis ist jede Teilmenge des Probabilitätsraums Ω eines Zufallsexperimentes.

Es gibt verschiedene Arten von Ereignissen:

- unmögliches Ereignis: Dies ist das Ereignis ohne Elemente $\emptyset$. Es hat keine Chance einzutreten.
- elementare Ereignisse: Dies sind Ereignisse mit nur einem Element.
- zusammengesetzte Ereignisse: Dies sind Ereignisse mit zwei oder mehr Elementen.
- sicheres Ereignis: Dies ist das Ereignis, das den gesamten Probabilitätsraum Ω enthält. Es tritt immer ein.

5.2 Mengentheorie

5.2.1 Ereignisraum

Definition 32 Ereignisraum

Gegeben einen Wahrscheinlichkeitsraum Ω eines Zufallsexperimentes, ist die Ereignismenge von Ω die Menge aller möglichen Ereignisse von Ω und wird mit $\mathcal{P}(\Omega)$ bezeichnet.

Beispiel

Für den Probabilitätsraum $\Omega = \{a, b, c\}$ ist dessen Ereignismenge:

$$\mathcal{P}(\Omega) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$$

5.2.2 Ereignisoperationen

Da Ereignisse Teilmengen des Wahrscheinlichkeitsraums sind, haben wir mittels der Mengentheorie folgende Operationen an Ereignissen:

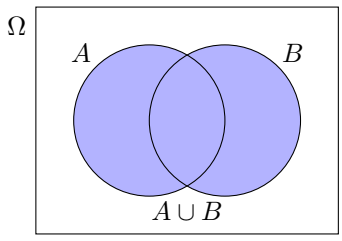
- Vereinigung
- Schnittmenge
- Komplementärmenge
- Differenzmenge

5.2.2.1 Vereinigung von Ereignissen

Definition "Vereinigungsereignis"

Gegeben zwei Ereignisse $A, B \subseteq \Omega$, ist die Vereinigung von A und B, bezeichnet mit $A \cup B$, das Ereignis aller Elemente, die Mitglieder von A oder B oder beiden sind.

$$A \cup B = \{x \mid x \in A \text{ oder } x \in B\}.$$



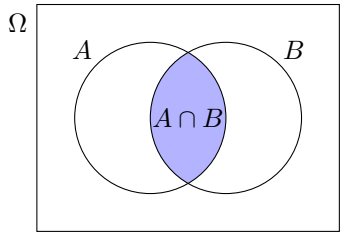
Das Vereinigungsereignis $A, B \subseteq \Omega$ tritt ein, wenn A oder B eintreten.

5.2.2.2 Schnittereignis

Definition "Schnittereignis"

Angenommen, wir haben zwei Ereignisse A und B, die Teilmengen von Ω sind. Das Schnittereignis (*Intersection*) von A und B, bezeichnet mit $A \cap B$, ist das Ereignis aller Elemente, die sowohl zu A als auch zu B gehören.

$$A \cap B = \{x \mid x \in A \text{ und } x \in B\}.$$



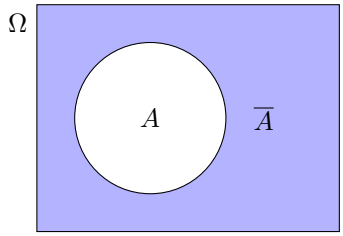
Das Ereignis $A \cap B$ tritt ein, wenn sowohl A als auch B eintreten. Zwei Ereignisse sind **unmöglich**, wenn ihr Schnitt leer ist.

5.2.2.3 Komplementärereignis

Definition "Komplementärereignis"

Angenommen, wir haben ein Ereignis A, das eine Teilmenge von Ω ist. Das komplementäre oder gegensätzliche Ereignis zu A, bezeichnet mit $\overline{A}$, ist das Ereignis aller Elemente in Ω außer denjenigen, die zu A gehören.

$$\overline{A} = \{x \mid x \notin A\}.$$



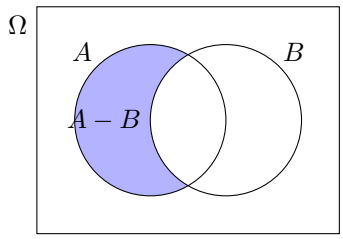
Das komplementäre Ereignis $\overline{A}$ tritt ein, wenn A nicht eintritt.

5.2.2.4 Differenzereignis

Definition "Differenzereignis"

Gegeben sind zwei Ereignisse A und B als Teilmengen von Ω. Die Differenz von A und B, bezeichnet mit $A - B$, ist das Ereignis aller Elemente, die zu A gehören, aber nicht zu B.

$$A - B = \{x \mid x \in A \text{ und } x \notin B\} = A \cap \overline{B}.$$



Das Differenzereignis $A - B$ tritt ein, wenn A eintritt, aber B nicht.

5.2.2.5 Beispiel

Beispiel

Der Ereignisraum beim Würfeln ist $\Omega = \{1, 2, 3, 4, 5, 6\}$ und die Ereignisse sind

- $A = \{2, 4, 6\}$ und
- $B = \{1, 2, 3, 4\}$,

dann gilt:

- Die Vereinigungsmenge von A und B ist $A \cup B = \{1, 2, 3, 4, 6\}$.
- Die Schnittmenge von A und B ist $A \cap B = \{2, 4\}$.
- Das Komplement von A ist $\overline{A} = \{1, 3, 5\}$
- Die Ereignisse A und $\overline{A}$ schließen sich gegenseitig aus.
- Die Differenz von A und B ist $A - B = \{6\}$ und die Differenz von B und A ist $B - A = \{1, 3\}$

5.2.3 Algebra

Sind die Ereignisse $A, B, C \subseteq \Omega$ gegeben, treffen die folgenden algebra'schen Eigenschaften zu:

- **Idempotenz:** $A \cup A = A, \quad A \cap A = A$
- **Kommutativität:** $A \cup B = B \cup A, \quad A \cap B = B \cap A$
- **Assoziativität:** $(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C)$
- **Distributivität:** $(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
- **Neutrales Element:** $A \cup \emptyset = A, \quad A \cap \Omega = A$
- **absorbierendes Element:** $A \cup \Omega = \Omega, \quad A \cap \emptyset = \emptyset$
- **komplementäres symmetrisches Element** $A \cup \overline{A} = \Omega, \quad A \cap \overline{A} = \emptyset$
- **doppelte Negation:** $\overline{\overline{A}} = A$
- **De Morgansche Gesetze:** $\overline{A \cup B} = \overline{A} \cap \overline{B}, \quad \overline{A \cap B} = \overline{A} \cup \overline{B}$

5.3 Wahrscheinlichkeit

Defintion "Wahrscheinlichkeit" (nach Laplace)

Für ein Zufallsexperiment mit einem Ereignisraum Ω, dessen Elemente alle gleich wahrscheinlich sind, ist die Wahrscheinlichkeit eines Ereignisses $A \subseteq \Omega$ der Quotient zwischen der Anzahl der Elemente von A und der Anzahl der Elemente von Ω

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl gewünschter Ergebnisse}}{\text{Anzahl aller möglichen Ergebnisse}}$$

Diese Definition ist gut bekannt, hat aber wichtige Einschränkungen:

- Es wird verlangt, dass alle Elemente des Ereignisraums gleich wahrscheinlich sind (Gleichwahrscheinlichkeit).
- Sie kann nicht bei unendlichen großen Ereignisräumen angewendet werden.

Vorsicht!

Diese Bedingungen werden in vielen echten Experimenten nicht erfüllt.

Beispiel

Angenommen, der Ereignisraum beim Werfen eines Würfels ist $\Omega = \{1, 2, 3, 4, 5, 6\}$ und das Ereignis $A = \{2, 4, 6\}$, dann beträgt die Wahrscheinlichkeit von A

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = 0.5.$$

Allerdings ist es beim Ereignisraum der Blutgruppe einer zufällig ausgewählten Person $\Omega = \{O, A, B, AB\}$ nicht möglich, die klassische Definition zu verwenden, um die Wahrscheinlichkeit für Gruppe A zu berechnen,

$$P(A) \neq \frac{|A|}{|\Omega|} = \frac{1}{4} = 0.25,$$

…weil Blutgruppen in der menschlichen Bevölkerung nicht *gleich* wahrscheinlich sind.

5.3.1 Häufigkeitswahrscheinlichkeit

Definition "Gesetz der großen Zahlen"

Wenn ein zufälliges Experiment eine große Anzahl von Malen wiederholt wird, nähert sich die relative Häufigkeit eines Ereignisses der Wahrscheinlichkeit des Ereignisses an.

Die folgende Definition der Wahrscheinlichkeit basiert auf diesem Satz.

Definition "Frequenzwahrscheinlichkeit"

Gegeben ein Ereignisraum Ω eines reproduzierbaren zufälligen Experiments, beträgt die Wahrscheinlichkeit eines Ereignisses $A \subseteq \Omega$ die relative Häufigkeit des Ereignisses A bei einer unbegrenzten Anzahl von Wiederholungen des Experiments.

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

Auch diese Definition hat einige Nachteile:

- Sie berechnet eine *Schätzung* der realen Wahrscheinlichkeit.
- Die Wiederholung des Experiments muss unter identischen Bedingungen stattfinden.

Beispiel Münzwurf

Gegeben ist der Ereignisraum beim Werfen einer Münze $\Omega = \{Kopf, Zahl\}$. Wenn nach 100 Würfen 54mal Kopf herauskam, beträgt die Wahrscheinlichkeit von $Kopf$

$$P(Kopf) = \frac{n_{Kopf}}{n} = \frac{54}{100} = 0.54.$$

Beispiel Blutgruppen

Gegeben ist der Ereignisraum der Blutgruppe einer zufällig ausgewählten Person $\Omega = \{O, A, B, AB\}$. Wenn nach einer Zufallsstichprobe von 1.000 Personen 412 mit der Blutgruppe A dabei waren, beträgt die Wahrscheinlichkeit von A

$$P(A) = \frac{n_A}{n} = \frac{412}{1000} = 0.412.$$

5.3.2 Axiomatische Wahrscheinlichkeit

Definition "Wahrscheinlichkeit" (nach Kolmogórov)

Gegeben ein Ereignisraum Ω eines zufälligen Experiments, ist eine Wahrscheinlichkeitsfunktion eine Funktion, die jeden Ereigniswert $A \subseteq \Omega$ auf eine reelle Zahl $P(A)$ (bekannt als Wahrscheinlichkeit von A), abbildet und folgende Axiome erfüllt:

1. Die Wahrscheinlichkeit jedes Ereignisses ist nicht negativ,
$$P(A) \geq 0.$$
2. Die Wahrscheinlichkeit des Ereignisraums beträgt 1,
$$P(\Omega) = 1$$
3. Die Wahrscheinlichkeit der Vereinigung zweier inkompatibler Ereignisse ($A \cap B = \emptyset$) ist die Summe ihrer Wahrscheinlichkeiten
$$P(A \cup B) = P(A) + P(B).$$

5.3.2.1 Eigenschaften der axiomatischen Wahrscheinlichkeit

Aus diesen Axiomen lassen sich einige wichtige Eigenschaften einer Wahrscheinlichkeitsfunktion ableiten. Gegeben ein Ereignisruam Ω eines zufälligen Experiments und die Ereignisse $A, B \subseteq \Omega$, gelten folgende Eigenschaften:

1. $P(\overline{A}) = 1 - P(A)$.
2. $P(\emptyset) = 0$.
3. Wenn $A \subseteq B$ dann $P(A) \leq P(B)$.
4. $P(A) \leq 1$. Das bedeutet $P(A) \in [0, 1]$.
5. $P(A - B) = P(A) - P(A \cap B)$.
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
7. Wenn $A = \{e_1, \ldots, e_n\}$, und $e_i\ i = 1, \ldots, n$ Elementarereignisse sind, dann gilt
$$P(A) = \sum_{i=1}^{n} P(e_i).$$

5.3.3 Wahrscheinlichkeitsinterpretation

Gemäß den vorherigen Axiomen ist die Wahrscheinlichkeit eines Ereignisses A eine reelle Zahl $P(A)$, die stets im Bereich von 0 bis 1 liegt.

Auf gewisse Weise drückt diese Zahl die Glaubwürdigkeit des Ereignisses aus, d.h. die Chancen, dass das Ereignis A in dem Experiment eintritt. Daher gibt sie auch ein Maß für die *Unsicherheit* bezüglich des Ereignisses an.

- Die maximale Unsicherheit entspricht der Wahrscheinlichkeit $P(A) = 0,5$ (A und $\overline{A}$ haben die gleiche Chance, einzutreten).
- Die minimale Unsicherheit entspricht der Wahrscheinlichkeit $P(A) = 1$ (A wird mit absoluter Sicherheit eintreten) und $P(A) = 0$ (A wird mit absoluter Sicherheit nicht eintreten).

Wenn $P(A)$ näher bei 0 als bei 1 liegt, sind die Chancen, dass A nicht eintritt, größer als die Chancen, dass A eintritt.

Im Gegensatz dazu, wenn $P(A)$ näher bei 1 als bei 0 liegt, sind die Chancen, dass A eintritt, größer als die Chancen, dass A nicht eintritt.

5.4 bedingte Wahrscheinlichkeit

Gelegentlich können wir bereits vor der Durchführung eines Experiments einige Informationen darüber erhalten. Üblicherweise wird diese Information als Ereignis B des gleichen Ereignisraums gegeben, von dem wir wissen, dass es wahr ist, bevor wir das Experiment durchführen.

In einem solchen Fall werden wir sagen, dass B ein *konditionierendes Ereignis* ist und die Wahrscheinlichkeit eines anderen Ereignisses A als *bedingte Wahrscheinlichkeit* bezeichnen. Diese wird als $P(A|B)$ ausgedrückt.

Dies muss als "Wahrscheinlichkeit von A unter der Bedingung B" oder "Wahrscheinlichkeit von A gegeben B" gelesen werden.

Beispiel

Konditionierende Ereignisse ändern normalerweise den Ereignisraum und damit die Wahrscheinlichkeiten der Ereignisse. Angenommen, wir haben eine Stichprobe von 100 Frauen und 100 Männern mit den folgenden Häufigkeiten:

	Nichtraucher	Raucher
Frauen	80	20
Männer	60	40

Dann beträgt die Wahrscheinlichkeit, ein Raucher zu sein, basierend auf der gesamten Probe:

$$P(\text{Raucher}) = \frac{60}{200} = 0.3.$$

Wenn wir jedoch wissen, dass die Person eine Frau ist, wird die Stichprobe auf die erste Zeile reduziert und die Wahrscheinlichkeit, eine Raucherin zu sein, beträgt:

$$P(\text{Raucherin}|\text{Frau}) = \frac{20}{100} = 0.2.$$

Definition "bedingte Wahrscheinlichkeit"

Gegeben ist ein Ereignisraum Ω eines zufälligen Experiments und zwei Ereignisse $A, B \subseteq \Omega$, so ist die Wahr-

scheinlichkeit von A unter der Bedingung, dass B eintritt:

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

solange $P(B) \neq 0$.

Diese Definition ermöglicht es, bedingte Wahrscheinlichkeiten zu berechnen, ohne den ursprünglichen Ereignisraum zu ändern.

Beispiel

In dem vorherigen Beispiel ist die bedingte Wahrscheinlichkeit, dass eine Person Raucher und weiblich ist:

$$P(\text{Raucher}|\text{weiblich}) = \frac{P(\text{Raucher} \cap \text{weiblich})}{P(\text{weiblich})} = \frac{20/200}{100/200} = \frac{20}{100} = 0.2.$$

5.4.1 Wahrscheinlichkeit des Schnittmengen-Ereignisses:

Aus der Definition der bedingten Wahrscheinlichkeit kann die Formel für die Wahrscheinlichkeit der Schnittmenge zweier Ereignisse abgeleitet werden:

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B).$$

Beispiel: Brustkrebs

In einer Bevölkerung gibt es 30 % Raucherinnen und wir wissen, dass 40 % dieser Raucher Krebs haben. Die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person raucht und Krebs hat, beträgt:

$$P(\text{Raucher} \cap \text{Krebs}) = P(\text{raucher})P(\text{Krebs}|\text{Raucher}) = 0.3 \times 0.4 = 0.12.$$

5.4.2 Unabhängigkeit der Ereignisse

Manchmal ändert das bedingende Ereignis die ursprüngliche Wahrscheinlichkeit des Hauptereignisses nicht.

Definition “Unabhängige Ereignisse”

Gegeben einen Ereignisraum Ω eines zufälligen Experiments, sind zwei Ereignisse $A, B \subseteq \Omega$ *unabhängig*, wenn die Wahrscheinlichkeit von A sich nicht ändert, wenn sie durch B bedingt ist, und umgekehrt, d.h.,

$$P(A|B) = P(A) \quad \text{und} \quad P(B|A) = P(B),$$

wenn $P(A) \neq 0$ und $P(B) \neq 0$.

Das bedeutet, dass das Eintreten eines Ereignisses keine relevante Information liefert, um die Unsicherheit des anderen Ereignisses zu ändern.

Wenn zwei Ereignisse unabhängig sind, ist die Wahrscheinlichkeit ihrer Schnittmenge gleich dem Produkt ihrer Wahrscheinlichkeiten,

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

Beispiel: Münzwurf

Der Ereignisraum für den doppelten Wurf einer Münze ist $\Omega = \{(K,K),(K,Z),(Z,K),(Z,Z)\}$ und alle Elemente sind gleich wahrscheinlich, wenn die Münze fair ist. Daher haben wir durch die klassische Definition der Wahrscheinlichkeit

$$P((K,K)) = \frac{1}{4} = 0.25.$$

Wenn wir $K_1 = \{(K,K),(K,Z)\}$, d.h., Kopf beim ersten Wurf, und $H_2 = \{(K,K),(Z,K)\}$ nennen, d.h., Kopf beim zweiten Wurf, können wir das gleiche Ergebnis erhalten, indem wir davon ausgehen, dass diese Ereignisse unabhängig sind,

$$P(K,K) = P(K_1 \cap K_2) = P(K_1) \cdot P(K_2) = \frac{2}{4} \cdot \frac{2}{4} = \frac{1}{4} = 0.25.$$

5.5 Wahrscheinlichkeitsraum

Definition "Wahrscheinlichkeitsraum"

Ein Wahrscheinlichkeitsraum eines zufälligen Experiments ist ein Tripel $(\Omega, \mathcal{F}, P)$ wobei

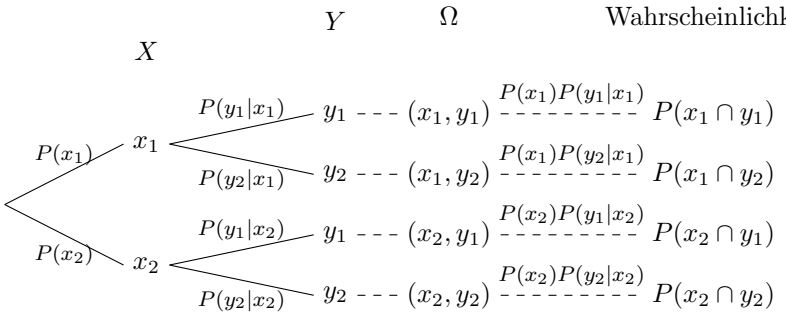
- Ω der Ereignisraum des Experiments ist.
- $\mathcal{F}$ eine Menge von Ereignissen des Experiments ist.
- P eine Wahrscheinlichkeitsfunktion ist.

Wenn wir die Wahrscheinlichkeiten aller Elemente von Ω kennen, können wir leicht den Wahrscheinlichkeitsraum konstruieren und die Wahrscheinlichkeit jedes Ereignisses in $\mathcal{F}$ berechnen.

5.5.1 Konstruktion des Wahrscheinlichkeitsraums

Um die Wahrscheinlichkeit jedes möglichen (elementaren) Ereignisses zu berechnen, können wir ein Baumdiagramm verwenden. Dabei gelten folgende Regeln:

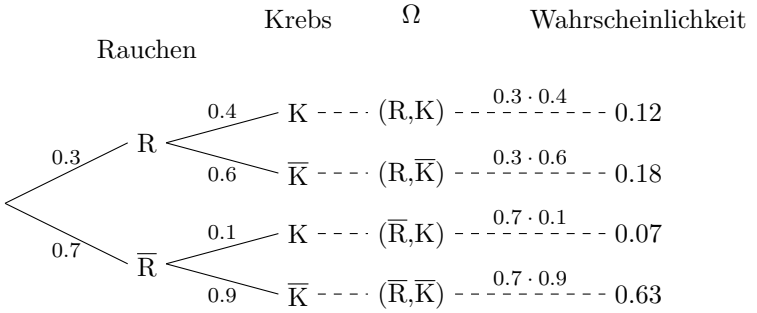
1. Jede Kante im Baum erhält eine Wahrscheinlichkeit. Diese gibt an, wie wahrscheinlich der jeweilige Wert unter der Bedingung der vorherigen Knoten ist.
2. Die Wahrscheinlichkeit eines vollständigen Pfads (also eines Ereignisses ganz am Ende des Baums) ergibt sich, indem man die Wahrscheinlichkeiten aller Kanten entlang dieses Pfads miteinander multipliziert – vom Start (Wurzel) bis zum Ende (Blatt).



5.5.2 Wahrscheinlichkeitsbaum mit abhängigen Variablen

Beispiel: Rauchen und Krebs:

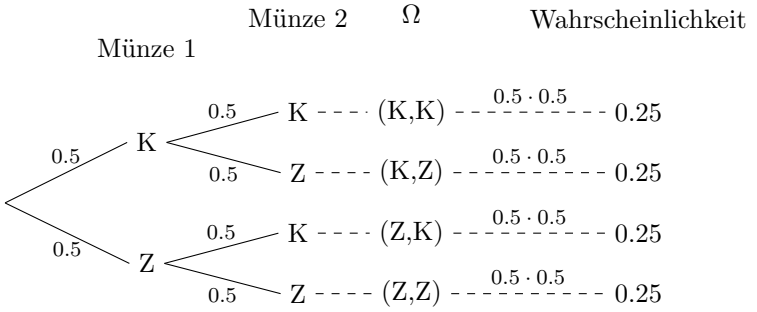
In einer Population gibt es 30 % Raucher und wir wissen, dass 40 % der Raucher Krebs haben, während nur 10 % der Nichtraucher an Krebs erkrankt sind. Der Wahrscheinlichkeitsbaum des Wahrscheinlichkeitsraums des zufälligen Experiments, das darin besteht, eine zufällig ausgewählte Person auszuwählen und die Variablen Rauchen und Brustkrebs zu messen, ist unten dargestellt.



5.5.3 Wahrscheinlichkeitsbaum mit unabhängigen Variablen

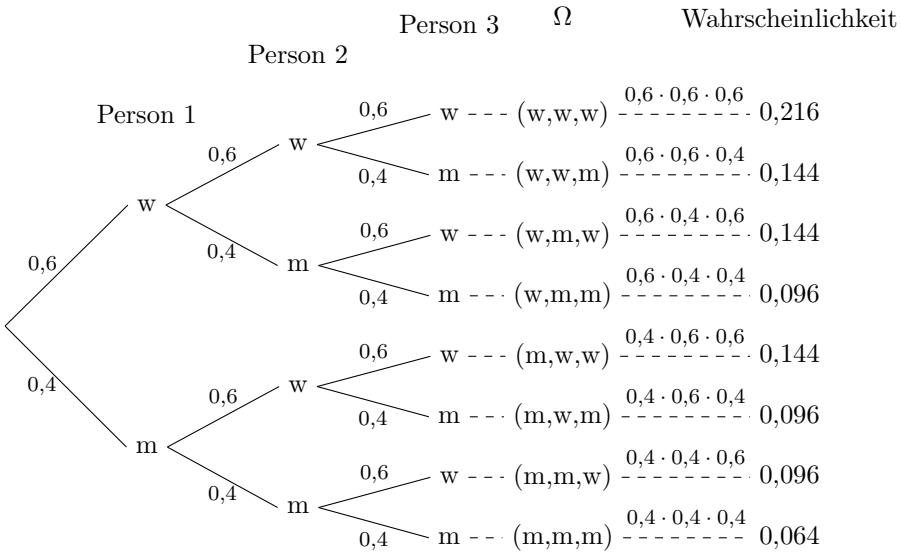
Beispiel Münzwurf

Der Wahrscheinlichkeitsbaum für das Zufallsexperiment zwei Münzen zu werfen ist:



Beispiel mit 3 Variablen

In einer Population gibt es 40% Männer und 60% Frauen. Der Wahrscheinlichkeitsbaum für das Ziehen einer zufälligen Stichprobe von drei Personen ist unten dargestellt.

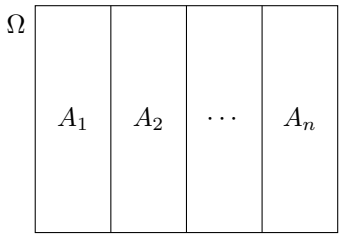


5.6 Satz von der vollständigen Wahrscheinlichkeit:

Definition "Aufteilung des Wahrscheinlichkeitsraums"

Eine Menge von Ereignissen $A_1, A_2, \dots, A_n$ aus demselben Ereignisraum Ω heißt eine *Aufteilung des Wahrscheinlichkeitsraums*, wenn sie die folgenden Bedingungen erfüllt:

1. Die Vereinigung der Ereignisse ist der Ereignisraum, das heißt, $A_1 \cup \dots \cup A_n = \Omega$
2. Alle Ereignisse sind paarweise ausschließend, das heißt $A_i \cap A_j = \emptyset \; \forall i \neq j$.



Üblicherweise ist es einfach, eine Aufteilung des Wahrscheinlichkeitsraums zu erhalten, indem man eine Population entsprechend einer kategorischen Variable aufteilt, wie beispielsweise Geschlecht oder Blutgruppe usw.

Wenn wir eine Aufteilung des Ereignisraums haben, können wir sie verwenden, um die Wahrscheinlichkeiten anderer Ereignisse in demselben Wahrscheinlichkeitsraum zu berechnen.

Definition "Gesamtwahrscheinlichkeit"

Gegeben eine Aufteilung $A_1, \dots, A_n$ eines Wahrscheinlichkeitsraums Ω, kann die Wahrscheinlichkeit jedes anderen Ereignisses B des selben Wahrscheinlichkeitsraums mit der folgenden Formel berechnet werden:

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(A_i) P(B|A_i).$$

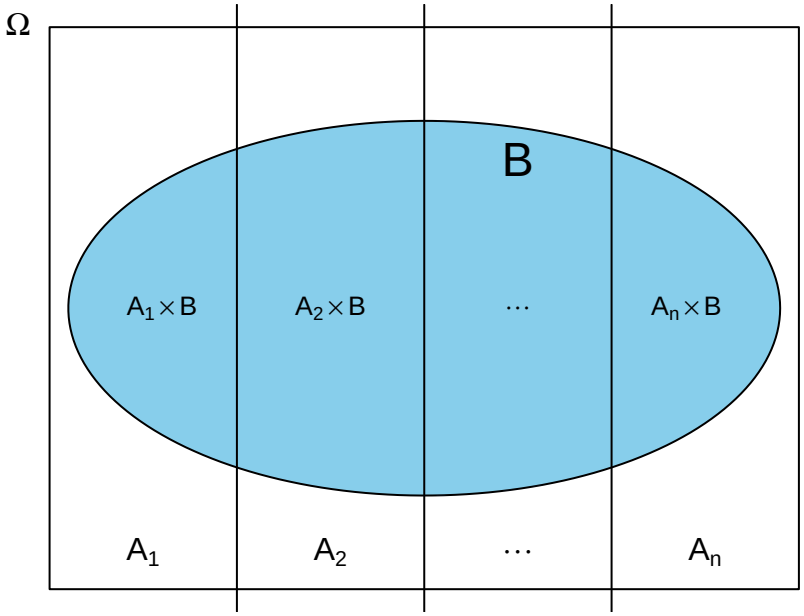
Beweis

Der Beweis dieses Satzes ist recht einfach. Da $A_1, \dots, A_n$ eine Aufteilung von Ω ist, haben wir:

$$B = B \cap \Omega = B \cap (A_1 \cup \dots \cup A_n) = (B \cap A_1) \cup \dots \cup (B \cap A_n).$$

Und alle Ereignisse dieser Vereinigung sind paarweise ausschließend, da $A_1, \dots, A_n$. Daher gilt:

$$P(B) = P((B \cap A_1) \cup \dots \cup (B \cap A_n)) = P(B \cap A_1) + \dots + P(B \cap A_n) =$$
$$= P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n) = \sum_{i=1}^{n} P(A_i)P(B|A_i).$$



Beispiel: Diagnose

Ein Symptom S kann durch eine Krankheit K verursacht werden, kann aber auch bei Personen ohne die Krankheit auftreten. In einer Population beträgt der Anteil der Menschen mit der Krankheit 0,2. Es ist bekannt, dass 90 % der Personen mit der Krankheit das Symptom aufweisen, während nur 40 % der Personen ohne die Krankheit es haben.
Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person aus der Bevölkerung das Symptom hat?

Um diese Frage zu beantworten, können wir den Satz von der vollständigen Wahrscheinlichkeit unter Verwendung der Aufteilung $\{K, \overline{K}\}$ anwenden:

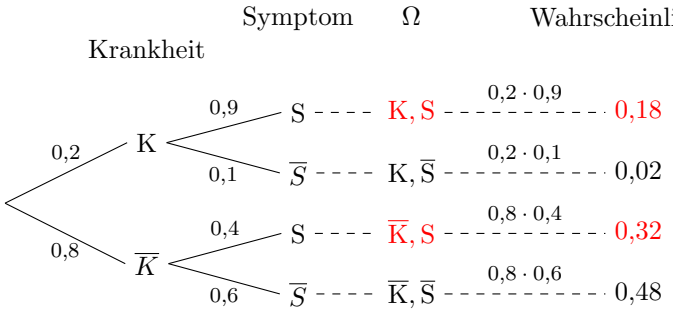
$$P(S) = P(K)P(S|K) + P(\overline{K})P(S|\overline{K}) = 0.2 \cdot 0.9 + 0.8 \cdot 0.4 = 0.5.$$

Das heißt, die Hälfte der Bevölkerung hat das Symptom.

Tatsächlich handelt es sich um einen gewichteten Mittelwert der Wahrscheinlichkeiten!

Beispiel Wahrscheinlichkeitsbaum

Das Ergebnis der vorherigen Frage wird durch den Wahrscheinlichkeitsbaum des Wahrscheinlichkeitsraums noch klarer:



$$P(S) = P(K \cap S) + P(\overline{K} \cap S) = P(K) \cdot P(S|K) + P(\overline{K}) \cdot P(S|\overline{K})$$
$$= 0.2 \cdot 0.9 + 0.8 \cdot 0.4 = 0.18 + 0.32 = 0.5.$$

5.7 Satz von Bayes

Eine Teilmenge eines Ereignisraums $A_1, \cdots, A_n$ kann auch als eine Menge von möglichen Hypothesen für ein Ereignis B interpretiert werden. In solchen Fällen kann es hilfreich sein, die posteriori Wahrscheinlichkeit $P(A_i|B)$ jeder Hypothese zu berechnen.

Satz von Bayes

Gegeben eine Partition $A_1, \cdots, A_n$ eines Ereignisraums Ω und ein weiteres Ereignis B desselben Mengenraums, kann die bedingte Wahrscheinlichkeit jedes Ereignisses A_i $i = 1, \ldots, n$ unter B mit der folgenden Formel berechnet werden:

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}.$$

In dem vorherigen Beispiel sollte die Diagnose für eine Person mit Symptomen gestellt werden.

In diesem Fall können wir K und $\overline{K}$ als die beiden möglichen Hypothesen für das Symptom S interpretieren. Die vorab Wahrscheinlichkeiten dafür sind $P(K) = 0.2$ und $P(\overline{K}) = 0.8$. Das bedeutet, dass wenn wir keine Informationen über das Symptom haben, die Diagnose nicht gestellt werden kann.

Allerdings ändert sich diese Unsicherheit bezüglich der Hypothese, wenn wir das Symptom beobachten. Dann müssen wir die posteriori Wahrscheinlichkeiten berechnen, um zu diagnostizieren, d.h.

$$P(K|S) \text{ und } P(\overline{K}|S)$$

Zur Berechnung nutzen wir den Satz von Bayes:

$$P(D|S) = \frac{P(D)P(S|D)}{P(D)P(S|D) + P(\overline{D})P(S|\overline{D})} = \frac{0.2 \cdot 0.9}{0.2 \cdot 0.9 + 0.8 \cdot 0.4} = \frac{0.18}{0.5} = 0.36,$$
$$P(\overline{D}|S) = \frac{P(\overline{D})P(S|\overline{D})}{P(D)P(S|D) + P(\overline{D})P(S|\overline{D})} = \frac{0.8 \cdot 0.4}{0.2 \cdot 0.9 + 0.8 \cdot 0.4} = \frac{0.32}{0.5} = 0.64.$$

Wie wir sehen können, hat sich die Wahrscheinlichkeit, die Krankheit zu haben, erhöht. Trotzdem ist die Wahrscheinlichkeit, sie nicht zu haben, immer noch größer als die Wahrscheinlichkeit, sie zu haben. Aus diesem Grund lautet die Diagnose, dass die Person die Krankheit nicht hat.

In diesem Fall wird gesagt, dass das Symptom S nicht entscheidend ist, um die Krankheit zu diagnostizieren.

5.8 Epidemiologie

Einzelne Zweige der Medizin, die eine intensive Nutzung von Wahrscheinlichkeiten vornehmen, sind Epidemiologie und Präventionsmedizin.

Die Epidemiologie untersucht die Häufigkeit und Ursachen von Krankheiten in Populationen, indem sie Risikofaktoren für Krankheiten und Ziele zur präventiven Gesundheitsversorgung identifiziert.

In der Epidemiologie ist man daran interessiert, wie häufig ein medizinisches Ereignis □ (typischerweise eine Krankheit wie Grippe, ein Risikofaktor wie Rauchen oder ein Schutzfaktor wie eine Impfung) auftritt, das als nominale Variable mit zwei Kategorien (das Eintreten oder Nicht-Eintreten des Ereignisses) gemessen wird.

Es gibt verschiedene Messgrößen zur Häufigkeit eines medizinischen Ereignisses. Die wichtigsten sind:

- Prävalenz
- Inzidenz
- Relatives Risiko
- Odds Ratio

5.8.1 Prävalenz

Definition "Prävalenz"

Die Prävalenz eines medizinischen Ereignisses K ist der Anteil einer bestimmten Bevölkerung, der von diesem medizinischen Ereignis betroffen ist.

$$\text{Prävalenz}(K) = \frac{\text{Anzahl von } K \text{ betroffener}}{\text{Größe der Bevölkerung}}$$

Die Prävalenz wird häufig anhand einer Stichprobe geschätzt, indem man den relativen Anteil der Menschen berechnet, die in der Stichprobe von dem Ereignis betroffen sind. Es ist auch üblich, diesen Anteil als Prozentangabe auszudrücken.

Beispiel

Um die Prävalenz von Grippe zu schätzen, wurde eine Stichprobe von 1.000 Personen untersucht und 150 davon hatten Grippe. Daher beträgt die Prävalenz der Grippe etwa $150/1000 = 0,15$ oder 15%.

5.8.2 Inzidenz

Die **Inzidenz** misst die Wahrscheinlichkeit des Auftretens eines medizinischen Ereignisses in einer Bevölkerung innerhalb eines bestimmten Zeitraums. Die Inzidenz kann als kumulative Anteilszahl oder als Rate gemessen werden.

⚠ Definition "kumulative Inzidenz"

Die kumulative Inzidenz eines medizinischen Ereignisses K ist der Anteil der Menschen, die das Ereignis in einem bestimmten Zeitraum erlitten haben, d.h. die Anzahl der neuen Fälle mit dem Ereignis in diesem Zeitraum geteilt durch die Größe der Risikobevölkerung. Anzahl neuer Fälle mit K

$$R(K) = \frac{\text{Anzahl Neuerkrankungen}}{\text{Größe der Risikobevölkerung}}$$

💡 Beispiel

Eine Bevölkerung enthält zu Beginn 1.000 Personen ohne Grippe und nach zweijähriger Beobachtungszeit hatten 160 davon die Grippe. Die Anteilszahl der Grippe-Inzidenz berechnet sich mit

$$R(K) = \frac{160}{1.000} = 0,16$$

Die kumulative Inzidenz liegt also beo 16% für zwei Jahre.

⚠ Definition "Inzidenzrate" (Absolutes Risiko)

Die Inzidenz*rate* oder das absolute Risiko eines medizinischen Ereignisse K ist die Anzahl neuer Erkrankungen geteilt durch die Größe der Risikobevölkerung und die Anzahl der Zeiteinheiten in einem bestimmten Zeitraum.

$$R(K) = \frac{\text{Anzahl Neuerkrankungen}}{\text{Größe der Risikobevölkerung} \times \text{Anzahl Zeiteinheiten}}$$

💡 Beispiel

Eine Bevölkerung enthält zu Beginn 1.000 Personen ohne Grippe und nach zweijähriger Beobachtungszeit hatten 160 davon die Grippe. Wenn man das Jahr als Zeiteinheit betrachtet, beträgt die Inzidenzrate der Grippe

$$R(K) = \frac{160}{1.000 \cdot 2} = 0,08$$

Die Inzidenzrate liegt also bei 8% Personen pro Jahr.

5.8.3 Prävlenz oder Inzidenz

Die Prävalenz zeigt an, wie verbreitet ein medizinisches Ereignis ist und ist eher ein Maß für die Belastung des Ereignisses für die Gesellschaft ohne Berücksichtigung der gefährdeten Zeit oder wann Personen möglicherweise einem möglichen Risikofaktor ausgesetzt waren

Die Inzidenz liefert Informationen über das Risiko, von dem Ereignis betroffen zu sein.

Die Prävalenz kann in Querschnittsstudien zu einem bestimmten Zeitpunkt gemessen werden, während zur Messung der Inzidenz eine Längsschnittstudie erforderlich ist, bei der die Individuen über einen bestimmten Zeitraum beobachtet werden.

Beim Verständnis der Ereignis-Ätiologie ist die Inzidenz nützlicher als die Prävalenz: wenn sich beispielsweise die Inzidenz einer Krankheit in einer Bevölkerung erhöht, dann gibt es einen Risikofaktor, der dies fördert.

Wenn die Inzidenz für die Dauer des Ereignisses ungefähr konstant ist, ist die Prävalenz ungefähr das Produkt aus Ereignis-Inzidenz und durchschnittlicher Ereignisdauer.

$$\text{Prävalenz} = \text{Inzidenz} \cdot \text{Dauer}$$

5.8.4 Risiko-Vergleich

Um zu bestimmen, ob ein Faktor oder eine Eigenschaft mit dem medizinischen Ereignis verbunden ist, müssen wir das Risiko des medizinischen Ereignisses in zwei Populationen vergleichen, von denen die eine dem Faktor ausgesetzt ist und die andere nicht. Die Gruppe von Menschen, die dem Faktor ausgesetzt sind, wird als *Interventionsgruppe* oder *experimentelle Gruppe* bezeichnet und die Gruppe von Menschen, die nicht ausgesetzt sind, als *Kontrollgruppe*.

In der Regel werden die beobachteten Fälle für jede Gruppe in einer 2x2-Tabelle wie der folgenden dargestellt.

	erkrankt K	gesund $\overline{K}$
Interventionsgruppe$(exponiert)$	a	b
Kontrollgruppe$(nicht\ exponiert)$	c	d

5.8.5 Zurechenbares Risiko

Definition “zurechendbares Risiko” (Risikodifferenz)

Das zurechenbare Risiko (*attributales Risiko* oder *Risikodifferenz*) eines medizinischen Ereignisses K für Personen, die einem Faktor (Exposition) ausgesetzt sind, ist der Unterschied zwischen den absoluten Risiken der Interventionsgruppe und der Kontrollgruppe.

$$AR(K) = R_{Intervention}(K) - R_{Kontrolle}(K) = \frac{a}{a+b} - \frac{c}{c+d}.$$

Das zurechenbare Risiko ist das Risiko eines Ereignisses, das spezifisch auf den Einflussfaktor von Interesse zurückzuführen ist. Beachten Sie, dass das zurechenbare Risiko *positiv* sein kann, wenn das Risiko der Interventionsgruppe *größer* als das Risiko der Kontrollgruppe ist, und umgekehrt.

Beispiel Impfung

Um die Wirksamkeit eines Impfstoffs gegen Grippe zu bestimmen, wurde zu Beginn des Jahres eine Stichprobe von 1.000 Personen ohne Grippe ausgewählt. Die Hälfte von ihnen wurde geimpft (Interventionsgruppe) und der Rest erhielt ein Placebo (Kontrollgruppe). Die folgende Tabelle fasst die Ergebnisse am Ende des Jahres zusammen.

	Grippe K	keine Grippe $\overline{K}$
Interventionsgruppe$(geimpft)$	20	480
Kontrollgruppe$(nicht\ geimpft)$	80	420

Für geimpfte Personen liegt das zurechenbare Risiko an Grippe zu erkranken bei

$$AR(K) = \frac{20}{20+480} - \frac{80}{80+420} = -0.12.$$

Das bedeutet, dass das Risiko an Grippe zu erkranken in der Gruppe der Geimpften 12% geringer ist als in der

Gruppe der Ungeimpften.

5.8.6 Relatives Risiko

Definition "relatives Risiko"

Das relative Risiko eines medizinischen Ereignisses K für Personen, die einer Exposition ausgesetzt sind, ist es der Quotient zwischen den Inzidenzen der Behandlungs- und Kontrollgruppe.

$$RR(K) = \frac{\text{Risiko der Interventionsgruppe}}{\text{Risiko der Kontrollgruppe}} = \frac{R_I(K)}{R_K(K)} = \frac{a/(a+b)}{c/(c+d)}$$

Das relative Risiko vergleicht das Risiko eines medizinischen Ereignisses zwischen der Behandlungs- und Kontrollgruppe.

- $RR = 1$ $\rightarrow$ Es gibt keine Verbindung zwischen dem Ereignis und der Exposition.
- $RR < 1$ $\rightarrow$ Die Exposition verringert das Risiko des Ereignisses.
- $RR > 1$ $\rightarrow$ Die Exposition erhöht das Risiko des Ereignisses.

Je weiter entfern RR von 1 ist, desto stärker die Verbindung.

Beispiel Impfung

Um die Wirksamkeit eines Impfstoffs gegen Grippe zu bestimmen, wurde zu Beginn des Jahres eine Stichprobe von 1.000 Personen ohne Grippe ausgewählt. Die Hälfte von ihnen wurde geimpft (Behandlungsgruppe) und der Rest erhielt ein Placebo (Kontrollgruppe). Die folgende Tabelle fasst die Ergebnisse am Ende des Jahres zusammen.

	Grippe K	keine Grippe $\overline{K}$
Interventionsgruppe$(geimpft)$	20	480
Kontrollgruppe$(nicht\ geimpft)$	80	420

Das relative Risiko für geimpfte Personen an Grippe zu erkranken ist:

$$RR(K) = \frac{20/(20+480)}{80/(80+420)} = 0,25.$$

Das bedeutet, dass geimpfte Personen im Vergleich zu ungeimpften nur ein Viertel des Risikos haben die Grippe bekommen. Das bedeutet, der Impfstoff verringerte das Risiko einer Grippe um 75%.

5.8.7 Odds

efinition "Odds"

Die Odds eines medizinischen Ereignisses K in einer Bevölkerung ist der Quotient zwischen den Personen, die das Ereignis erlitten haben und den Personen, die es nicht in einem bestimmten Zeitraum erlitten haben.

$$ODDS(K) = \frac{\text{Neue Fälle mit } K}{\text{Fälle ohne } K} = \frac{P(K)}{P(\overline{K})}$$

Im Gegensatz zur Inzidenz, die ein Anteil von weniger als oder gleich 1 ist, kann das Odds größer als 1 sein. Es ist jedoch möglich, das Odds in eine Wahrscheinlichkeit umzuwandeln, indem man folgende Formel verwendet:

$$P(K) = \frac{ODDS(K)}{ODDS(K) + 1}$$

Beispiel

Eine Bevölkerung enthält initial 1.000 Personen ohne Grippe und nach einem Jahr haben 160 von ihnen die Grippe bekommen. Das Odds der Grippe ist somit

$$ODDS(\text{Grippe}) = \frac{160}{840} = 0,1905.$$

Beachten Sie, dass die Inzidenz $160/1000 =0,16 $ beträgt.

5.8.8 Odds Ratio

Definition "Odds Ratio"

Die Odds Ratio eines medizinischen Ereignisses K für Personen, die einer Exposition ausgesetzt sind, ist der Quotient zwischen dem Odds des Ereignisses in der Behandlungs- und Kontrollgruppe.

$$OR(K) = \frac{\text{Odds in Interventionsgruppe}}{\text{Odds in Kontrollgruppe}} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

Die Odds Ratio vergleicht die Odds eines medizinischen Ereignisses zwischen der Behandlungs- und Kontrollgruppe. Die Interpretation ist ähnlich wie beim relativen Risiko.

- $OR = 1 \Rightarrow$ Es gibt keine Verbindung zwischen dem Ereignis und der Exposition.
- $OR < 1 \Rightarrow$ Die Exposition verringert das Risiko des Ereignisses.
- $OR > 1 \Rightarrow$ Die Exposition erhöht das Risiko des Ereignisses.

Je weiter OR weg von 1 ist, desto stärker ist die Verbindung.

Beispiel Impfung

Um die Wirksamkeit eines Impfstoffs gegen Grippe zu bestimmen, wurde zu Beginn des Jahres eine Stichprobe von 1.000 Personen ohne Grippe ausgewählt. Die Hälfte von ihnen wurde geimpft (Interventionsgruppe) und der Rest erhielt ein Placebo (Kontrollgruppe). Die folgende Tabelle fasst die Ergebnisse am Ende des Jahres zusammen.

	Grippe K	keine Grippe $\overline{K}$
Interventionsgruppe$(geimpft)$	20	480
Kontrollgruppe$(nicht\ geimpft)$	80	420

Die Odds Ratio der Grippe für geimpfte Personen ist:

$$OR(K) = \frac{20/480}{80/420} = 0,21875.$$

Das bedeutet, dass die Chancen, an Grippe zu erkranken (im Vergleich dazu, nicht zu erkranken), bei geimpften Personen fast ein Fünftel derer bei ungeimpften Personen betragen. Das heißt: Auf etwa 22 geimpfte Personen mit Grippe kommen etwa 100 ungeimpfte Personen mit Grippe.

5.8.9 Relatives Risiko vs Odds ratio

Relatives Risiko und Odds Ratio sind zwei Kennzahlen, aber ihre Interpretation unterscheidet sich leicht. Während das Relative Risiko einen Vergleich der Risiken zwischen den Behandlungs- und Kontrollgruppen ausdrückt, beschreibt die Odds Ratio einen Vergleich der *Chancen*, was nicht dasselbe wie das *Risiko* ist. Daher bedeutet ein Odds Ratio von 2 nicht, dass die Behandlungsgruppe das doppelte Risiko hat, das medizinische Ereignis zu erwerben.

Die Interpretation des Odds Ratio ist komplizierter, da sie kontrarfaktisch ist und uns angibt, wie viele Male das Ereignis in der Behandlungsgruppe häufiger ist im Vergleich zur Kontrollgruppe unter der Annahme, dass in der Kontrollgruppe das Ereignis so häufig wie das Nicht-Ereignis ist.

Der Vorteil des Odds Ratio besteht darin, dass sie nicht von der Prävalenz oder der Inzidenz des Ereignisses abhängt und notwendigerweise verwendet werden muss, wenn die Anzahl der Menschen mit dem medizinischen Ereignis in beiden Gruppen willkürlich ausgewählt wird, wie z.B. bei Fall-Kontroll-Studien.

Beispiel Rauchen und Lungenkrebs

Um die Assoziation zwischen Lungenkrebs und Rauchen zu bestimmen, wurden zwei Proben ausgewählt (die zweite mit der doppelten Anzahl von Nicht-Krebs-Individuen), wobei folgende Ergebnisse erzielt wurden:

	Krebs	kein Krebs
Raucher	60	80
Nichtraucher	40	320

$$RR(K) = \frac{60/(60+80)}{40/(40+320)} = 3.86.$$

$$OR(K) = \frac{60/80}{40/320} = 6.$$

	Krebs	kein Krebs
Raucher	60	160
Nichtraucher	40	640

$$RR(K) = \frac{60/(60+160)}{40/(40+640)} = 4.64.$$

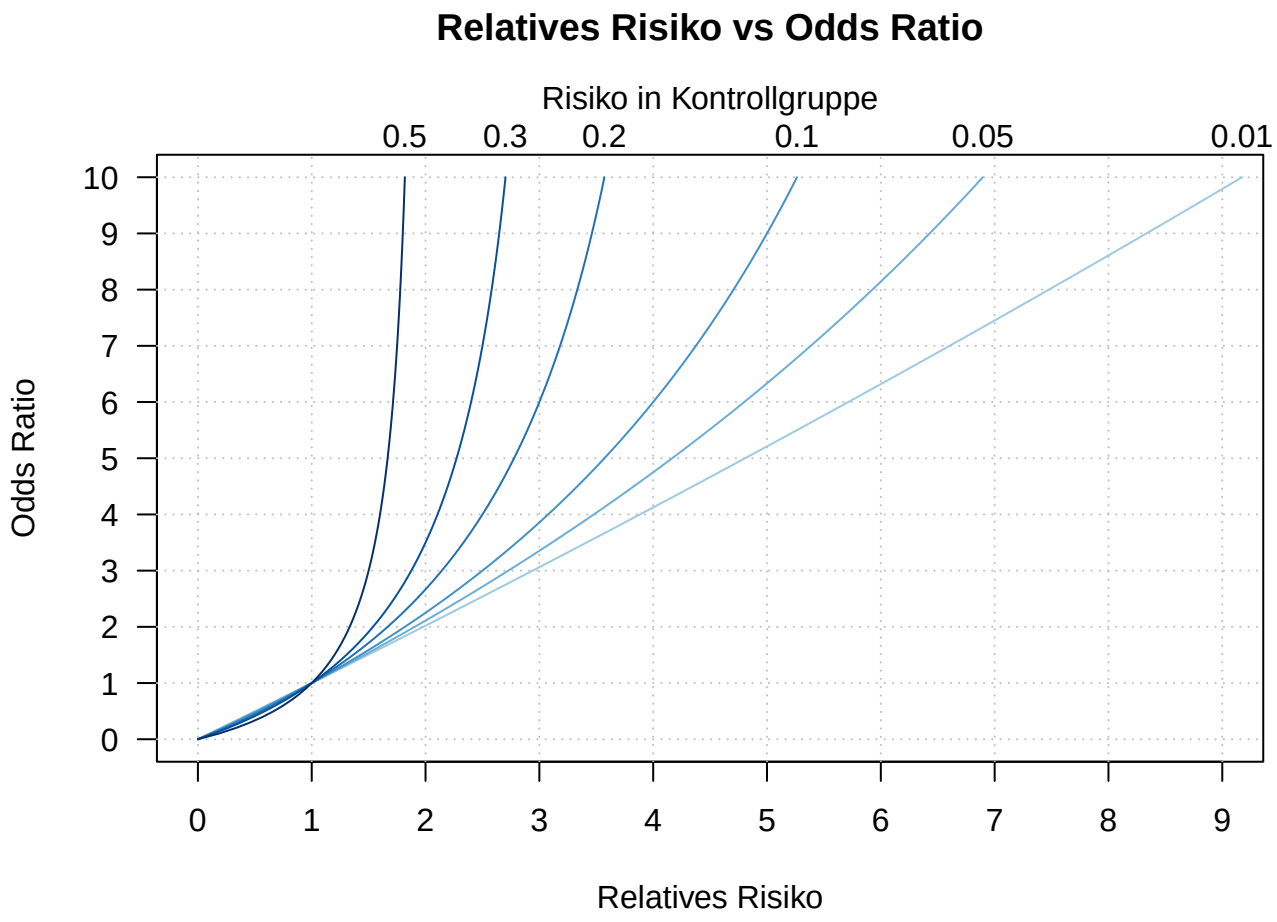
$$OR(K) = \frac{60/160}{40/640} = 6.$$

Das Relative Risiko verändert sich, wenn wir die Inzidenz oder Prävalenz des Ereignisses (Lungenkrebs) ändern, während sich die Odds Ratio nicht verändert.

Die Beziehung zwischen dem Relative Risiko und dem Odds Ratio wird durch die folgende Formel gegeben:

$$RR = \frac{OR}{1 - R_{Kontrolle} + R_{Kontrolle} \cdot OR} = OR \cdot \frac{1 - R_{Intervention}}{1 - R_{Kontrolle}}$$

Die Odds Ratio überschätzt stets das Relative Risiko, wenn es größer als 1 ist, und unterschätzt es, wenn es kleiner als 1 ist. Allerdings sind Relatives Risiko und Odds Ratio für seltene medizinische Ereignisse (mit sehr kleiner Prävalenz oder Inzidenz) fast gleich.



5.9 Diagnosetests

In der Epidemiologie ist es üblich, diagnostische Tests zur Diagnose von Krankheiten zu verwenden. Im Allgemeinen sind diese jedoch nicht vollständig zuverlässig und haben ein gewisses Risiko einer Fehldiagnose, wie in der folgenden Tabelle dargestellt:

	krank K	gesund $\overline{K}$
Testergebnis $positiv+$	richtig positiv (RP)	falsch positiv (FP)
Testergebnis $negativ-$	falsch negativ (FN)	richtig negativ (RN)

Die Leistung eines diagnostischen Tests hängt von den folgenden beiden Wahrscheinlichkeiten ab:

- Sensitivität und
- Spezifität

5.9.1 Sensitivität und Spezifität

Definition "Sensitivität"

Die Sensitivität eines diagnostischen Tests ist der Anteil positiver Ergebnisse bei erkrankten Personen.

$$P(+|K) = \frac{RP}{RP + FN}$$

Definition "Spezifität"

Die Spezifität eines diagnostischen Tests ist der Anteil negativer Ergebnisse bei gesunden Personen.

$$P(-|\overline{K}) = \frac{RN}{RN + FP}$$

Ein Test mit hoher Sensitivität wird die Krankheit bei den meisten erkrankten Personen erkennen, jedoch auch mehr falsche Positive produzieren als ein weniger empfindlicher Test. Daher ist ein positives Ergebnis eines Tests mit hoher Sensitivität nicht nützlich zur Bestätigung der Krankheit, aber ein negatives Ergebnis ist nützlich zum Ausschluss der Krankheit, da es selten bei erkrankten Personen zu falsch-negativen Ergebnissen kommt.

Andererseits wird ein Test mit hoher Spezifität die Krankheit bei den meisten gesunden Personen ausschließen, jedoch auch mehr falsche Negative produzieren als ein weniger spezifischer Test. Daher ist ein negatives Ergebnis eines Tests mit hoher Spezifität nicht nützlich zum Ausschluss der Krankheit. Aber ein positives Ergebnis ist nützlich zur Bestätigung der Krankheit, da es selten bei gesunden Personen zu falsch-positiven Ergebnissen kommt.

Die Entscheidung für einen Test mit größerer Sensitivität oder einen Test mit größerer Spezifität hängt von Art der Krankheit und dem Ziel des Tests ab. Im Allgemeinen werden wir einen sensitiven Test verwenden, wenn:

- die Krankheit ernst ist und es wichtig ist, sie zu erkennen.
- die Krankheit behandelbar ist.
- Falsch-positive Ergebnisse keine schweren Schäden verursachen.

Und wir werden einen spezifischen Test verwenden, wenn:

- die Krankheit ernst ist, aber schwierig oder unmöglich zu behandeln ist.
- falsch-positive Ergebnisse schwerwiegende Schäden verursachen.
- die Behandlung von falsch-positiven Fällen gefährliche Folgen haben kann.

5.9.2 prädiktive Werte

Der wichtigste Aspekt eines diagnostischen Tests ist jedoch seine Vorhersagekraft, die mit den folgenden beiden posterioren Wahrscheinlichkeiten gemessen wird:

- positiv prädiktiver Wert und
- negativ prädiktiver Wert

Definition "positiv prädiktiver Wert

Der positiv prädiktive Wert eines diagnostischen Tests ist der Anteil von Personen mit der Krankheit unter allen Personen mit einem positiven Ergebnis.

$$P(K|+) = \frac{RP}{RP + FP}$$

Definition "Negativ prädiktiver Wert"

Der negativ prädiktive Wert eines diagnostischen Tests ist der Anteil von Personen ohne die Krankheit unter allen Personen mit einem negativen Ergebnis.

$$P(\overline{K}|-) = \frac{RN}{RN + FN}$$

Positiv und negativ prädiktive Werte ermöglichen es, die Krankheit zu bestätigen oder auszuschließen, besonders, wenn sie mindestens einen Schwellenwert von 0,5 erreichen.

$PPW > 0.5 \Rightarrow$ Hinweis auf das Vorliegen der Krankheit

$PPV < 0.5 \Rightarrow$ Hinweis auf das Nichtvorliegen der Krankheit

Diese Wahrscheinlichkeiten hängen jedoch von der Häufigkeit der Krankheit $P(K)$ ab und können aus der Sensitivität und Spezifität des diagnostischen Tests unter Verwendung des Satzes von Bayes berechnet werden.

$$PPW = P(K|+) = \frac{P(K)P(+|K)}{P(K)P(+|K) + P(\overline{K})P(+|\overline{K})}$$
$$NPW = P(\overline{K}|-) = \frac{P(\overline{K})P(-|\overline{K})}{P(K)P(-|K) + P(\overline{K})P(-|\overline{K})}$$

Daher steigt der positiv prädiktive Wert bei häufigen Krankheiten und der negativ prädiktive Wert bei seltenen Krankheiten.

Beispiel

Bei 1.000 Personen wurde ein Grippetest durchgeführt. Die Ergebnisse sind in der Tabelle dargestellt.

	Grippe K	keine Grippe $\overline{K}$
Testergebnis+	95	90
Testergebnis−	5	810

- Die Prävalenz der Grippe ist $P(K) = \frac{95+5}{1000} = 0,1$.
- Die Sensitivität des Tests ist $P(+|K) = \frac{95}{95+5} = 0,95$.
- Die Spezifität des Tests ist $P(-|\overline{K}) = \frac{810}{90+810} = 0,9$.
- Der positiv prädiktive Wert ist $PPW = P(K|+) = \frac{95}{95+90} = 0,5135$. Da dieser Wert über 0,5 liegt, bedeutet dies, dass wir bei einem positiven Testergebnis wahrscheinlich wirklich Grippe gefunden haben. Allerdings wird das Vertrauen in diesesehr gering, da der Wert sehr nahe an 0,5 liegt.
- Der negativ prädiktive Wert ist $NPW = P(\overline{K}|-) = \frac{810}{5+810} = 0,9939$. Da dieser Wert fast 1 ist, bedeutet dies, dass es fast sicher ist, dass eine Person keine Grippe hat, wenn sie ein negatives Testergebnis erhält.

Daher ist dieser Test sehr leistungsfähig, um die Grippe auszuschließen, aber nicht so leistungsfähig, um sie zu bestätigen.

5.9.3 Likelihood Ratio

Die folgenden Maße werden normalerweise aus Sensitivität und Spezifität abgeleitet.

Definition “Positive Likelihood Ratio”

Die positive Likelihood Ratio eines diagnostischen Tests ist das Verhältnis zwischen der Wahrscheinlichkeit positiver Ergebnisse bei Personen mit der Krankheit und gesunden Personen.

$$LR+ = \frac{P(+|K)}{P(+|\overline{K})} = \frac{\text{Sensitivität}}{1 - \text{Spezifität}}$$

Definition “Negative Likelihood Ratio”

Die negative Likelihood Ratio eines diagnostischen Tests ist das Verhältnis zwischen der Wahrscheinlichkeit negativer Ergebnisse bei Personen mit der Krankheit und gesunden Personen.

$$LR- = \frac{P(-|K)}{P(-|\overline{K})} = \frac{1 - \text{Sensitivität}}{\text{Spezifität}}$$

Die positive Likelihood Ratio kann als die Anzahl interpretiert werden, wie oft ein positives Ergebnis bei Personen mit der Krankheit gegenüber Personen ohne sie wahrscheinlicher ist.

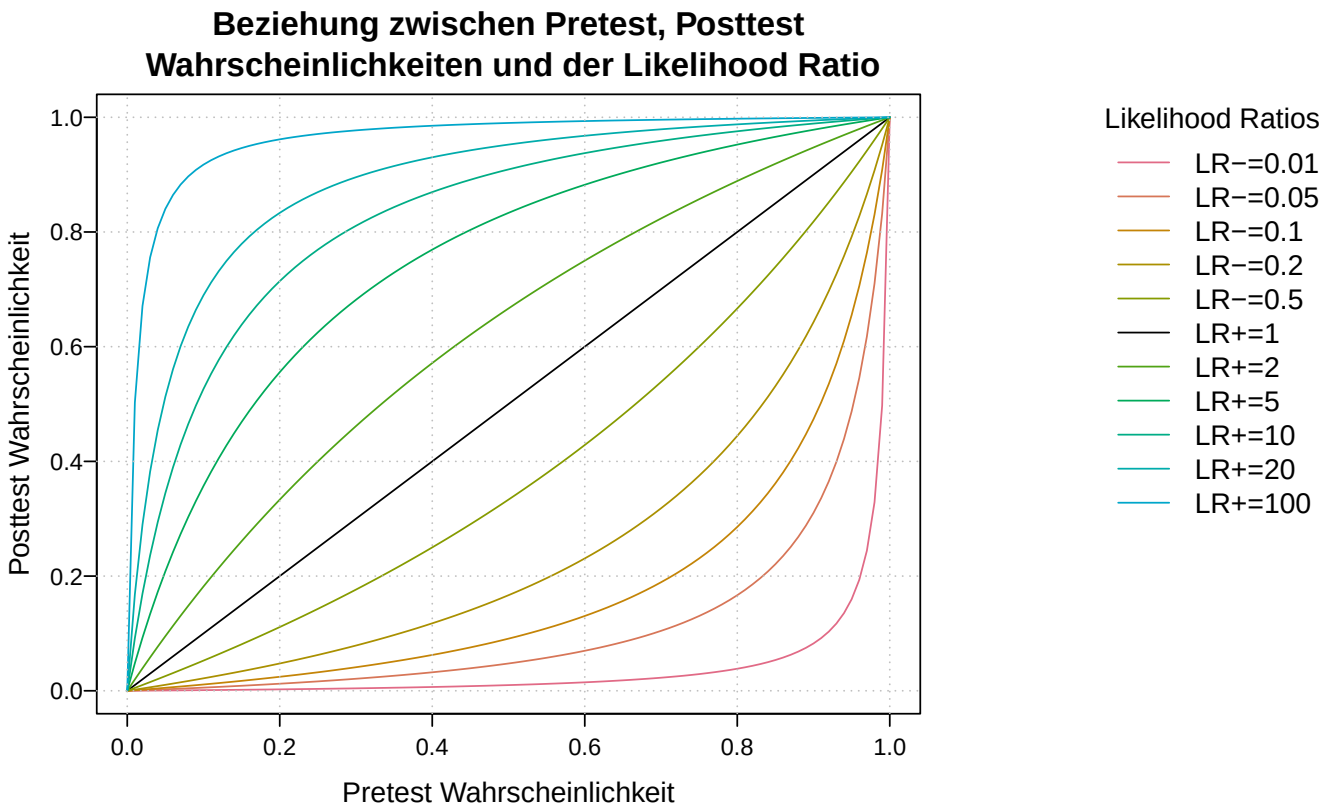
Andererseits kann die negative Likelihood Ratio als die Anzahl interpretiert werden, wie oft ein negatives Ergebnis bei Personen mit der Krankheit gegenüber Personen ohne sie wahrscheinlicher ist.

Post-Test-Wahrscheinlichkeiten können durch Likelihood Ratio aus Pre-Test-Wahrscheinlichkeiten berechnet werden.

$$P(K|+) = \frac{P(K)P(+|K)}{P(K)P(+|K) + P(\overline{K})P(+|\overline{K})} = \frac{P(K)LR+}{1 - P(K) + P(K)LR+}$$

Daher gilt:

- Eine Likelihood Ratio größer als 1 erhöht die Wahrscheinlichkeit der Erkrankung.
- Eine Likelihood Ratio kleiner als 1 verringert die Wahrscheinlichkeit der Erkrankung.
- Eine Likelihood Ratio von 1 ändert die Pre-Test-Wahrscheinlichkeit nicht.


Beziehung zwischen Pretest, Posttest Wahrscheinlichkeiten und der Likelihood Ratio
Posttest Wahrscheinlichkeit
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Pretest Wahrscheinlichkeit
Likelihood Ratios
LR−=0.01
LR−=0.05
LR−=0.1
LR−=0.2
LR−=0.5
LR+=1
LR+=2
LR+=5
LR+=10
LR+=20
LR+=100

6 Diskrete Zufallsvariablen

6.1 Zufallsvariable

Das Ziehen einer Probe auf zufällige Weise ist ein Zufallsexperiment und jede Variable, die in der Probe gemessen wird, ist eine *Zufallsvariable*, da die Werte, die die Variable bei den Individuen der Probe annimmt, dem Zufall unterliegen.

⚠ Definition "Zufallsvariable"

Eine Zufallsvariable X ist eine Funktion, die jedes Element einer Zufallsstichprobe auf eine reelle Zahl abbildet.

$$X : \Omega \to \mathbb{R}$$

Der Satz von Werten, den die Variable annehmen kann, wird Reichweite (*range*) genannt und durch $Ran(X)$ dargestellt.

Im Grunde genommen ist eine Zufallsvariable eine Variable, deren Werte aus einem Zufallsexperiment stammen und jeder Wert hat eine bestimmte Wahrscheinlichkeit für das Auftreten.

💡 Beispiel Würfel

Die Variable X, die das Ergebnis des Würfelns misst, ist eine Zufallsvariable und ihre Reichweite ist

$$\text{Ran}(X) = \{1, 2, 3, 4, 5, 6\}$$

6.1.1 Arten von Zufallsvariablen

Es gibt zwei Arten von Zufallsvariablen:

- **Diskret**: Sie nehmen isolierte Werte an und ihre Reichweite ist abzählbar, zum Beispiel Anzahl der Kinder einer Familie, Anzahl gerauchter Zigaretten, Anzahl bestandener Prüfungen usw.
- **Kontinuierlich**: Sie können jeden Wert in einem echten Intervall annehmen und ihre Reichweite ist nicht abzählbar, z.B. Beispiel Gewicht, Größe, Alter, Cholesterinspiegel usw.

Die Art der Modellierung ist für jede Art von Variable unterschiedlich. In diesem Kapitel werden wir untersuchen, wie *diskrete Zufallsvariablen* modelliert werden.

6.2 Wahrscheinlichkeitsverteilung diskreter Zufallsvariablen

Da die Werte einer diskreten Zufallsvariablen mit den elementaren Ereignissen eines Zufallsexperiments verknüpft sind, hat jeder Wert eine Wahrscheinlichkeit.

Definition "Wahrscheinlichkeitsfunktion"

Die Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen X ist die Funktion $f(x)$, die jeden Wert x_i der Variablen auf seine Wahrscheinlichkeit abbildet,

$$f(x_i) = P(X = x_i).$$

Wir können auch Wahrscheinlichkeiten auf die gleiche Weise kumulieren wie wir Häufigkeiten in der Stichprobe kumuliert haben.

Definition "Verteilungsfunktion"

Die Verteilungsfunktion einer diskreten Zufallsvariablen X ist die Funktion $F(x)$, die jeden Wert x_i der Variablen auf die Wahrscheinlichkeit abbildet, einen Wert kleiner oder gleich x_i zu haben,

$$F(x_i) = P(X \leq x_i) = f(x_1) + \cdots + f(x_i).$$

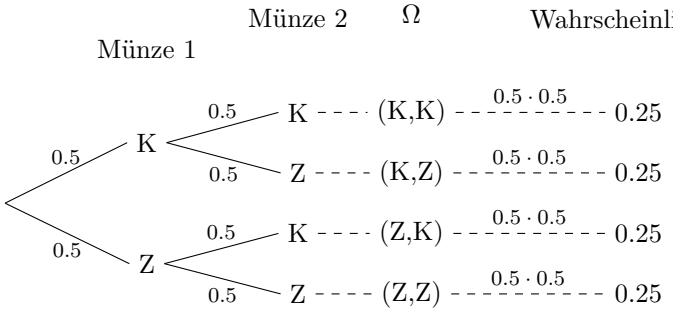
Der Bereich einer diskreten Zufallsvariablen und ihre Wahrscheinlichkeitsfunktion wird als Wahrscheinlichkeitsverteilung der Variablen bezeichnet und normalerweise in einer Tabelle dargestellt:

X	x_1	x_2	$\cdots$	x_n	$\sum$
$f(x)$	$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$	1
$F(x)$	$F(x_1)$	$F(x_2)$	$\cdots$	$F(x_n) = 1$	

Auf die gleiche Weise, wie die Tabelle der Häufigkeiten in der Stichprobe die Verteilung der Werte einer Variablen in der Stichprobe zeigt, zeigt die Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariablen die Verteilung der Werte in der Gesamtpopulation.

Beispiel Münzenwurf

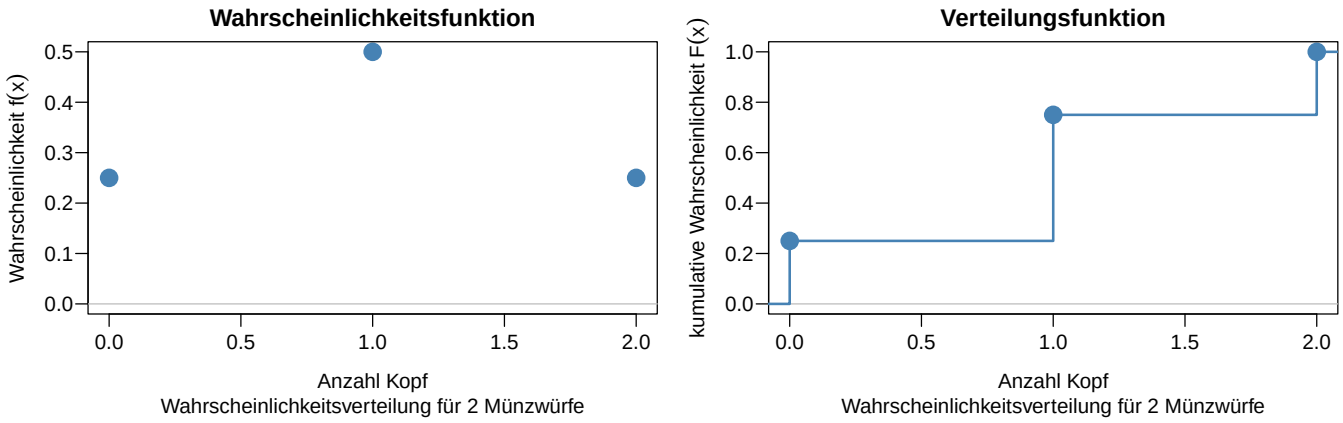
Sei X die diskrete Zufallsvariable, die die Anzahl der Köpfe nach dem Werfen von zwei Münzen misst. Der Wahrscheinlichkeitbaum des Wahrscheinlichkeitsraums des Zufallsexperiments ist wie folgt dargestellt.



Die Wahrscheinlichkeitsverteilung ist daher

X	0	1	2
$f(x)$	0.25	0.5	0.25
$F(x)$	0.25	0.75	1

$$F(x) = \begin{cases} 0 & \text{wenn } x < 0 \\ 0{,}25 & \text{wenn } 0 \leq x < 1 \\ 0{,}75 & \text{wenn } 1 \leq x < 2 \\ 1 & \text{wenn } x \geq 2 \end{cases}$$



6.2.1 Populationsstatistik

Auf die gleiche Weise, wie wir Stichprobestatistiken verwenden, um die Häufigkeitsverteilung einer Variablen in der Stichprobe zu beschreiben, verwenden wir Populationsstatistiken, um die Wahrscheinlichkeitsverteilung einer Zufallsvariablen in der Gesamtpopulation zu beschreiben.

Die Definition von Populationsstatistiken entspricht analog der Definition von Stichprobestatistiken, jedoch unter Verwendung von Wahrscheinlichkeiten anstelle von relativen Häufigkeiten.

Die wichtigsten sind:

- Mittelwert: $\mu = E(X) = \sum_{i=1}^{n} x_i f(x_i)$
- Varianz: $\sigma^2 = Var(X) = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2$
- Standardabweichung: $\sigma = +\sqrt{\sigma^2}$

Um sie besser von Stichprobenkennwerten zu unterscheiden, werden für die Populationsstatistiken griechische Buchstaben verwendet,

Beispiel Münzwurf

Aus einem Zufallsexperiment, bei dem 2 Münzen geworfen wurden, ergab sich folgende Wahrscheinlichkeitsverteilung

X	0	1	2
$f(x)$	0.25	0.5	0.25
$F(x)$	0.25	0.75	1

Daraus ergeben sich folgende Populationskennwerte:

$$\mu = \sum_{i=1}^{n} x_i f(x_i) = 0 \cdot 0.25 + 1 \cdot 0.5 + 2 \cdot 0.25 = 1 \text{ Kopf},$$

$$\sigma^2 = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2 = (0^0 \cdot 0.25 + 1^2 \cdot 0.5 + 2^2 \cdot 0.25) - 1^2 = 0.5 \text{ Kopf}^2,$$

$$\sigma = +\sqrt{0.5} = 0.71 \text{ Kopf}.$$

6.2.2 Diskrete Wahrscheinlichkeitsverteilungsmodelle

Je nach Art des Experiments, bei dem die Zufallsvariable gemessen wird, gibt es verschiedene Wahrscheinlichkeitsverteilungsmodelle. Die wichtigsten sind:

- Diskrete Gleichverteilung
- Binomialverteilung
- Poissonverteilung

6.3 Diskrete Gleichverteilung

Wenn alle Werte einer zufälligen Variablen X die gleiche Wahrscheinlichkeit haben, ist die Wahrscheinlichkeitsverteilung von X gleichförmig.

Definition 63 "Diskrete Gleichverteilung U(a, B)"

Eine diskrete Zufallsvariable X folgt einem diskreten Gleichverteilungsmodell mit Parametern a und b, notiert als $X \sim U(a, b)$, wenn ihr Bereich $\text{Ran}(X) = \{a, a+1, \dots, b\}$ ist und ihre Wahrscheinlichkeitsfunktion

$$f(x) = \frac{1}{b-a+1}.$$

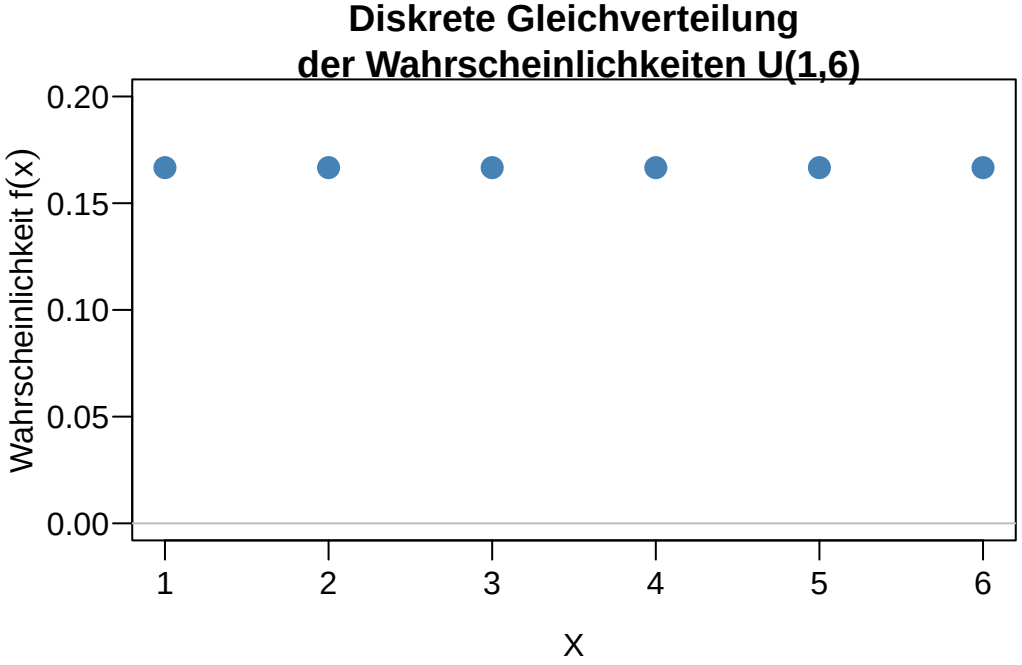
Beachten Sie, dass a und b das Minimum bzw. das Maximum des Bereichs sind.

Der Mittelwert (*Erwartungswert*) und die Varianz sind

$$\mu = \sum_{i=0}^{b-a} \frac{a+i}{b-a+1} = \frac{a+b}{2} \qquad \sigma^2 = \sum_{i=0}^{b-a} \frac{(a+i-\mu)^2}{b-a+1} = \frac{(b-a+1)^2-1}{12}$$

Beispiel Würfeln

Die Variable, die das Ergebnis beim Würfeln misst, folgt einem diskreten Gleichverteilungsmodell $U(1, 6)$.



6.4 Binomialverteilung

Die Binomialverteilung entspricht in der Regel einer Variablen, die in einem zufälligen Experiment mit den folgenden Merkmalen gemessen wird:

- Das Experiment besteht aus einer Sequenz von n Wiederholungen desselben Versuchs.
- Jeder Versuch wird unter identischen Bedingungen wiederholt und produziert zwei mögliche Ergebnisse namens *Erfolg* oder *Misserfolg*.
- Die Versuche sind unabhängig voneinander.
- Die Wahrscheinlichkeit für einen Erfolg ist in allen Versuchen gleich und beträgt $P(\text{Erfolg}) = p$.

Unter diesen Bedingungen folgt die diskrete Zufallsvariable X, die die Anzahl der Erfolge bei den n Versuchen misst, einem Binomialverteilungsmodell mit Parametern n und p.

Definition "Binomialverteilung $(B(n,p)$"

Eine diskrete Zufallsvariable X folgt einem binomialen Verteilungsmodell mit Parametern n und p, notiert als $X \sim B(n,p)$, wenn ihr Bereich $\text{Ran}(X) = \{0, 1, \dots, n\}$ ist und ihre Wahrscheinlichkeitsfunktion

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

ist.

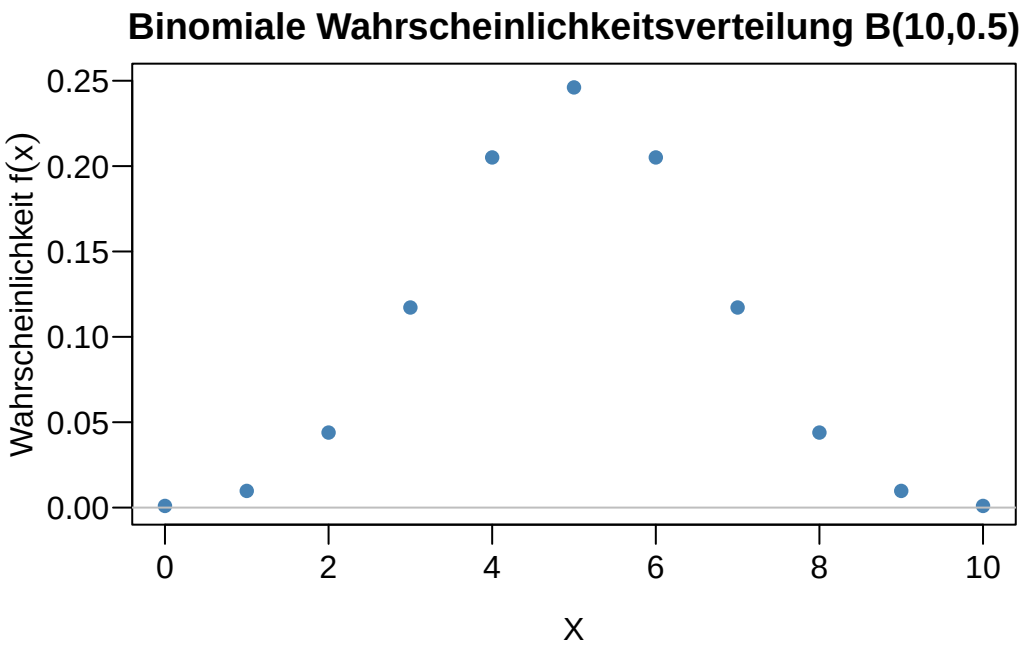
Beachten Sie, dass n als die Anzahl der Wiederholungen eines Versuchs bekannt ist und p als die Wahrscheinlichkeit für einen Erfolg bei jeder Wiederholung.

Der Erwartungswert und die Varianz sind:

$$\mu = n \cdot p \qquad \sigma^2 = n \cdot p \cdot (1-p).$$

Beispiel: Werfen von 10 Münzen:

Die Variable, die die Anzahl der Köpfe nach dem Werfen von 10 Münzen misst, folgt einem binomialen Verteilungsmodell $B(10,\ 0.5)$.



Wenn $X \sim B(10,\ 0.5)$ die zufällige Variable ist, die die Anzahl der Köpfe nach dem Werfen von 10 Münzen misst, dann ist:

- die Wahrscheinlichkeit, 4 mal Kopf zu werfen:

$$f(4) = \binom{10}{4} 0.5^4 (1-0.5)^{10-4} = \frac{10!}{4!6!} 0.5^4 0.5^6 = 210 \cdot 0.5^{10} = 0.2051.$$

- die Wahrscheinlichkeit, 2 mal oder weniger Kopf zu werfen:

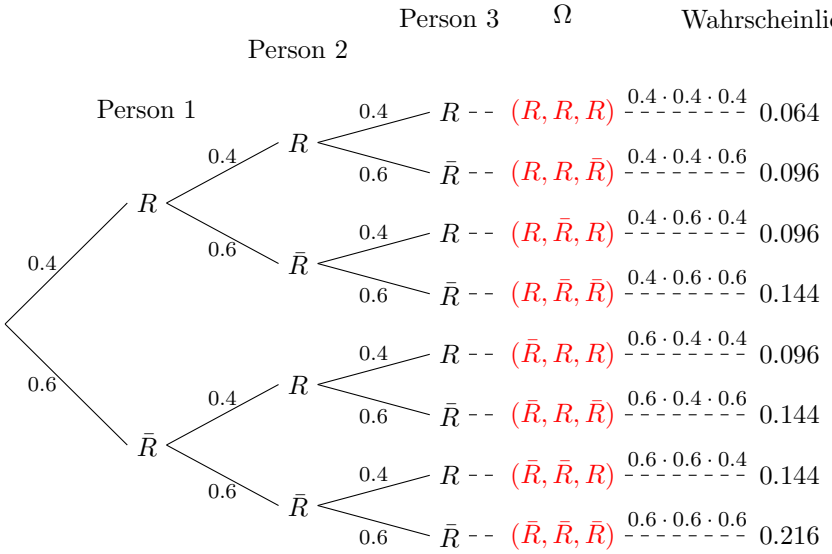
$$\begin{aligned} F(2) &= f(0) + f(1) + f(2) = \\ &= \binom{10}{0} 0.5^0 (1-0.5)^{10-0} + \binom{10}{1} 0.5^1 (1-0.5)^{10-1} + \binom{10}{2} 0.5^2 (1-0.5)^{10-2} = \\ &= 0.0547. \end{aligned}$$

- der erwartete Wert für die Anzahl an Kopfwürfen:

$$\mu = 10 \cdot 0.5 = 5 \text{ mal Kopf.}$$

Beispiel für eine zufällige Stichprobe mit Zurücklegen:

In einer Population gibt es 40% Raucher. Die Variable X, die die Anzahl der Raucher in einer zufälligen Stichprobe mit Zurücklegen von 3 Personen misst, folgt einem binomialen Verteilungsmodell $X \sim B(3,\ 0.4)$.



$f(0) = \binom{3}{0} 0.4^0 (1-0.4)^{3-0} = 0.6^3,$ $f(1) = \binom{3}{1} 0.4^1 (1-0.4)^{3-1} = 3 \cdot 0.4 \cdot 0.6^2,$
$f(2) = \binom{3}{2} 0.4^2 (1-0.4)^{3-2} = 3 \cdot 0.4^2 \cdot 0.6,$ $f(3) = \binom{3}{3} 0.4^3 (1-0.4)^{3-3} = 0.4^3.$

6.5 Poissonverteilung

Die Poisson-Verteilung entspricht in der Regel einer Variablen, die in einem zufälligen Experiment gemessen wird, für welches folgenden Merkmalen gelten:

- Das Experiment besteht darin, die Anzahl der Ereignisse zu beobachten, die in einem festgelegten Zeit- oder Raumintervall auftreten, zum Beispiel: Anzahl der Geburten in einem Monat, Anzahl der E-Mails in einer Stunde, Anzahl der roten Blutkörperchen in einem Blutvolumen usw.

- Die Ereignisse treten unabhängig voneinander auf.
- Das Experiment produziert dieselbe durchschnittliche Ereignisrate λ für jedes Intervall.

Unter diesen Bedingungen folgt die diskrete Zufallsvariable X, die die Anzahl der Ereignisse in einem Intervall misst, einem Poisson-Verteilungsmodell mit dem Parameter λ.

Definition "Poisson-Verteilung" $P(\lambda)$

Eine diskrete Zufallsvariable X folgt einem Poisson-Verteilungsmodell mit dem Parameter λ, notiert als $X \sim P(\lambda)$, wenn ihr Bereich $\text{Ran}(X) = \{0, 1, \dots, \infty\}$ ist und ihre Wahrscheinlichkeitsfunktion

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

ist.

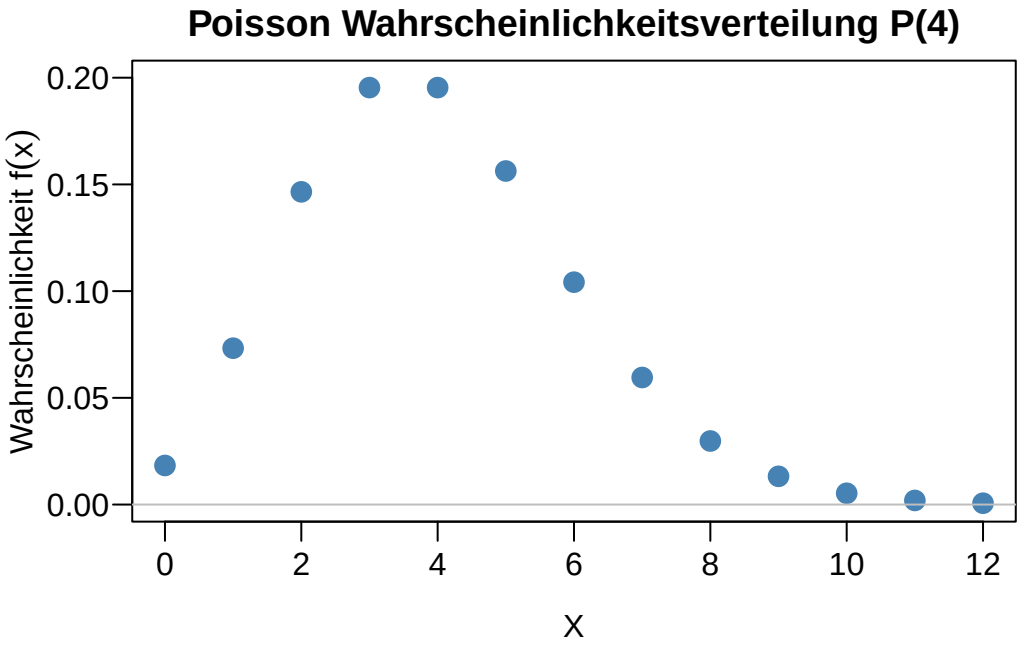
Beachten Sie, dass λ die durchschnittliche Ereignisrate für ein Intervall ist und sich ändern wird, wenn das Intervall verändert wird.

Der Erwartungswert und die Varianz sind:

$$\mu = \lambda \qquad \sigma^2 = \lambda.$$

Beispiel: Anzahl der Geburten in einer Stadt

In einer Stadt gibt es durchschnittlich 4 Geburten pro Tag. Die Zufallsvariable X, die die Anzahl der Geburten an einem Tag in der Stadt misst, folgt einem Poisson-Verteilungsmodell $X \sim P(4)$.



Wenn $X \sim P(4)$ die Zufallsvariable ist, die die Anzahl der Geburten in der Stadt misst, dann ist:

- die Wahrscheinlichkeit, dass es 5 Geburten an einem Tag gibt:

$$f(5) = e^{-4}\frac{4^5}{5!} = 0.1563.$$

- die Wahrscheinlichkeit, dass es weniger als 2 Geburten an einem Tag gibt:

$$F(1) = f(0) + f(1) = e^{-4}\frac{4^0}{0!} + e^{-4}\frac{4^1}{1!} = 5e^{-4} = 0.0916.$$

- die Wahrscheinlichkeit, dass es mehr als 1 Geburt an einem Tag gibt:

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = 1 - 0.0916 = 0.9084.$$

6.5.1 Annäherung an die Binomialverteilung durch die Poisson-Verteilung

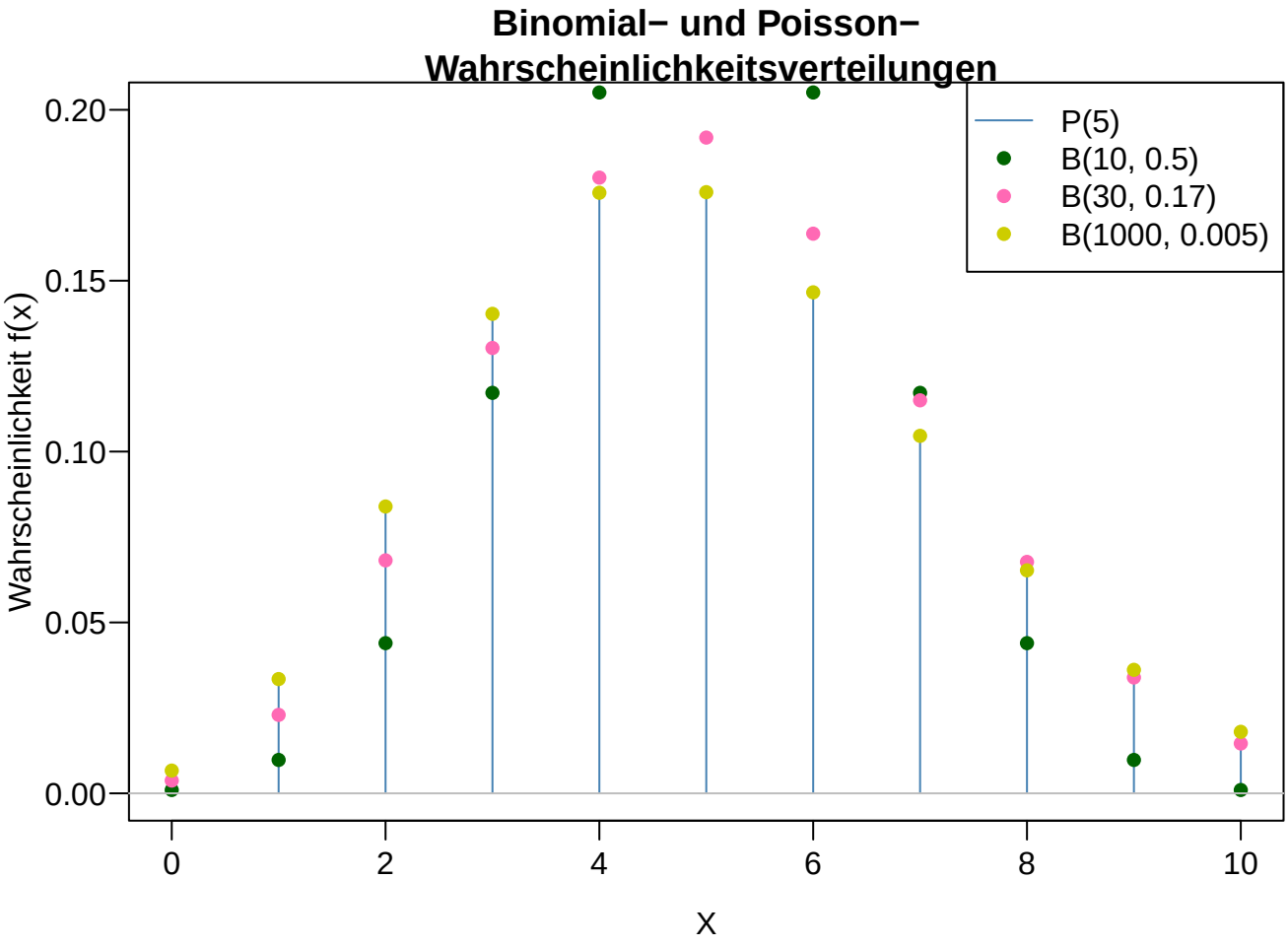
Die Poisson-Verteilung kann aus der Binomialverteilung erhalten werden, wenn die Anzahl der Versuche gegen unendlich geht und die Wahrscheinlichkeit für den Erfolg gegen Null tendiert.

Gesetz der seltenen Ereignisse

Die Binomialverteilung$X \sim B(n, p)$ tendiert gegen die Poisson-Verteilung $P(\lambda)$, mit $\lambda = n \cdot p$, wenn n gegen unendlich geht und p gegen Null geht, d.h.:

$$\lim_{n\to\infty, p\to 0} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda}\frac{\lambda^x}{x!}.$$

In der Praxis kann diese Annäherung für $n \geq 30$ und $p \leq 0.1$ verwendet werden.



Beispiel

Ein Impfstoff löst in 4% der Fälle eine unerwünschte Reaktion aus. Wenn 50 Personen geimpft werden, was ist die Wahrscheinlichkeit, dass mehr als 2 Personen eine unerwünschte Reaktion haben?
Die Zufallsvariable, die die Anzahl der Personen mit einer unerwünschten Reaktion im Sample misst, folgt einem Binomialverteilungsmodell $X \sim B(50, 0.04)$. Da aber $n = 50 > 30$ und $p = 0.04 < 0.1$ sind, können wir das Gesetz der seltenen Ereignisse anwenden und das Poisson-Verteilungsmodell $P(50 \cdot 0.04) = P(2)$ verwenden, um die Berechnungen durchzuführen.

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) = 1 - f(0) - f(1) - f(2) = \\ &= 1 - e^{-2}\frac{2^0}{0!} - e^{-2}\frac{2^1}{1!} - e^{-2}\frac{2^2}{2!} = \\ &= 1 - 5e^{-2} = 0.3233. \end{aligned}$$

7 Kontinuierliche Zufallsvariablen

Kontinuierliche Zufallsvariablen können im Gegensatz zu diskreten Zufallsvariablen jeden Wert in einem echten Intervall annehmen. Daher ist der Bereich einer kontinuierlichen Zufallsvariablen unendlich und nicht abzählbar. Diese Vielzahl von Werten macht es unmöglich, die Wahrscheinlichkeit für jeden einzelnen zu berechnen, wodurch es nicht möglich ist, ein Wahrscheinlichkeitsmodell über eine Wahrscheinlichkeitsfunktion wie bei diskreten Zufallsvariablen zu definieren.

Darüber hinaus wird die Messung einer kontinuierlichen Zufallsvariablen in der Regel durch die Genauigkeit des Messinstruments begrenzt. Zum Beispiel beträgt die wahre Größe einer Person, für die unser Messinstrument 1,68 Meter groß sagt, nicht genau 1,68 Meter, da die Genauigkeit des Messinstruments nur auf zwei Dezimalstellen cm (Zentimeter) beträgt. Dies bedeutet, dass die wahre Größe dieser Person zwischen 1,675 und 1,685 Metern liegt.

Daher ist es für kontinuierliche Variablen unsinnig, die Wahrscheinlichkeit eines einzelnen Werts zu berechnen, und wir werden Wahrscheinlichkeiten für Intervalle berechnen.

7.1 Verteilungsfunktion einer kontinuierlichen Zufallsvariablen

Um die Verteilungsfunktion einer kontinuierlichen Zufallsvariablen zu modellieren, wird eine Wahrscheinlichkeitsdichtefunktion verwendet.

⚠ Definition “Wahrscheinlichkeitsdichtefunktion”

Die Wahrscheinlichkeitsdichtefunktion einer kontinuierlichen Zufallsvariablen X ist eine Funktion $f(x)$, die folgende Bedingungen erfüllt:

- Sie ist nicht-negativ: $f(x) \geq 0 \; \forall x \in \mathbb{R}$.
- Der von der Kurve der Dichtefunktion und der x-Achse begrenzte Bereich beträgt gleich 1, d.h.:

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

- Die Wahrscheinlichkeit, dass X einen Wert zwischen a und b annimmt, entspricht dem von der Dichtefunktion begrenzten Bereich unter a und b, d.h.:

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)\, dx$$

Die Wahrscheinlichkeitsdichtefunktion misst die relative Wahrscheinlichkeit jedes Werts, jedoch ist $f(x)$ nicht die Wahrscheinlichkeit von x, da $P(X = x) = 0$ für jeden Wert x gilt.

7.1.1 Verteilungsfunktion

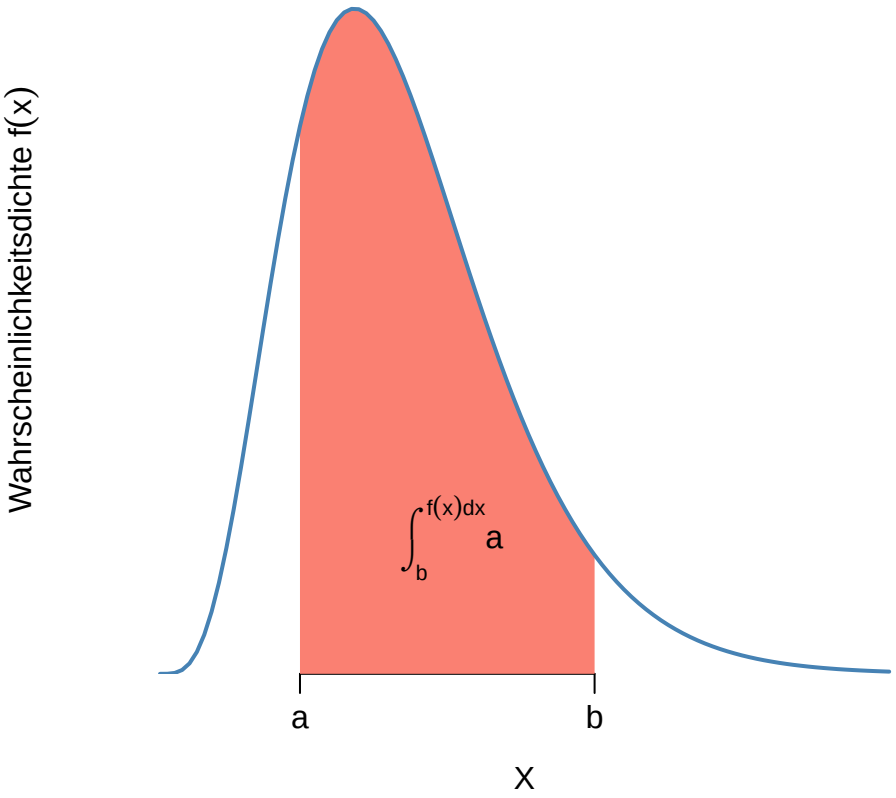
Genauso wie bei diskreten Zufallsvariablen macht es auch bei kontinuierlichen Zufallsvariablen Sinn, kumulative Wahrscheinlichkeiten zu berechnen.

Definition "Verteilungsfunktion"

Die Verteilungsfunktion einer kontinuierlichen Zufallsvariablen X ist eine Funktion $F(x)$, die jeden Wert a auf die Wahrscheinlichkeit abbildet, dass X einen Wert kleiner oder gleich a annimmt, d.h.:

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)\, dx.$$

7.1.2 Wahrscheinlichkeitsbereiche



$$P(a \leq X \leq b) = \int_{a}^{b} f(x)\, dx = F(b) - F(a)$$

Beispiel

Gegeben ist die folgende Funktion:

$$f(x) = \begin{cases} 0 & \text{bei } x < 0 \\ e^{-x} & \text{bei } x \geq 0 \end{cases}$$

Überprüfen wir, ob es sich um eine Dichtefunktion handelt.
Da diese Funktion klarerweise nicht-negativ ist, müssen wir prüfen, dass der Gesamtbereich, der durch die Kurve und die x-Achse begrenzt wird, gleich 1 beträgt:

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} f(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{\infty} e^{-x}\,dx = \\ = [-e^{-x}]_0^{\infty} = -e^{-\infty} + e^0 = 1.$$

Nun berechnen wir die Wahrscheinlichkeit, dass X einen Wert zwischen 0 und 2 annimmt:

$$P(0 \leq X \leq 2) = \int_0^2 f(x)\,dx = \int_0^2 e^{-x}\,dx = [-e^{-x}]_0^2 = -e^{-2} + e^0 = 0.8646.$$

7.1.3 Populationsstatistik

Die Berechnung der Populationsstatistik erfolgt ähnlich wie im Fall diskreter Variabler, jedoch mit der Dichtefunktion anstelle der Wahrscheinlichkeitsfunktion und durch Ersetzen der diskreten Summe durch das Integral.

Die wichtigsten sind:

- Mittelwert: $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$
- Varianz: $\sigma^2 = Var(X) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$
- Standardabweichung: $\sigma = +\sqrt{\sigma^2}$

Beispiel

Gegeben ist die Variable X mit der folgenden Wahrscheinlichkeitsdichtefunktion:

$$f(x) = \begin{cases} 0 & \text{bei } x < 0 \\ e^{-x} & \text{bei } x \geq 0 \end{cases}$$

Der Mittelwert beträgt:

$$\mu = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{-\infty}^{0} xf(x)\,dx + \int_{0}^{\infty} xf(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{\infty} xe^{-x}\,dx = \\ = [-e^{-x}(1+x)]_0^{\infty} = 1.$$

Die Varianz beträgt:

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2 = \int_{-\infty}^{0} x^2 f(x)\,dx + \int_{0}^{\infty} x^2 f(x)\,dx - \mu^2 = \\ = \int_{-\infty}^{0} 0\,dx + \int_{0}^{\infty} x^2 e^{-x}\,dx - \mu^2 = [-e^{-x}(x^2+2x+2)]_0^{\infty} - 1^2 = 2e^0 - 1 = 1.$$

Die wichtigsten kontinuierlichen Wahrscheinlichkeitsverteilungsmodelle sind:

- Gleichverteilung
- Normalverteilung
- Student's t-Verteilung

- Chi-Quadrat-Verteilung

7.2 stetige Gleichverteilung

Wenn alle Werte einer Zufallsvariablen X dieselbe Wahrscheinlichkeit haben, dann ist die Wahrscheinlichkeitsverteilung von X gleichverteilt.

Definition "Kontinuierliche gleichverteilte Verteilung" $U(a,b)$)

Eine kontinuierliche Zufallsvariable X folgt dem Wahrscheinlichkeitsverteilungsmodell der Gleichverteilung mit den Parametern a und b, notiert als $X \sim U(a,b)$, wenn ihr Bereich $\text{Ran}(X) = [a,b]$ ist und ihre Dichtefunktion

$$f(x) = \frac{1}{b-a} \quad \forall x \in [a,b]$$

ist.

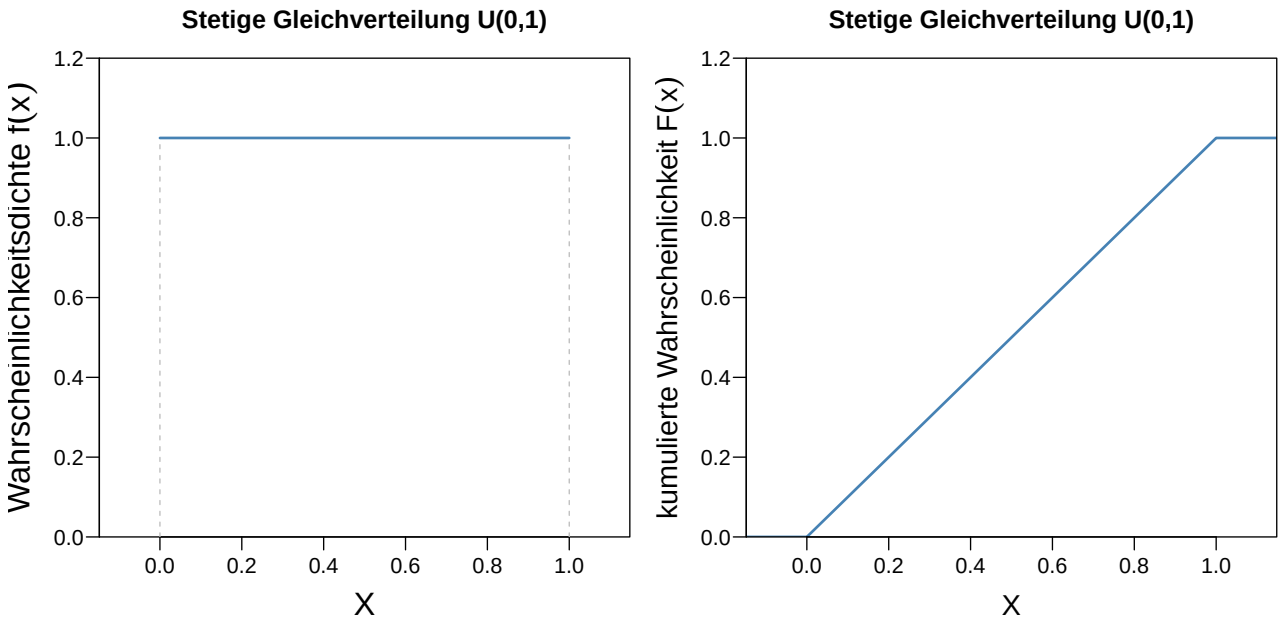
Dabei sind a und b der minimale bzw. maximale Wert des Bereichs, und die Dichtefunktion ist konstant.

Der Erwartungswert und die Varianz sind:

$$\mu = \frac{a+b}{2} \qquad \sigma^2 = \frac{(b-a)^2}{12}.$$

Beispiel

Die Generierung einer Zufallszahl zwischen 0 und 1 folgt der kontinuierlichen gleichverteilten Verteilung $U(0,1)$.

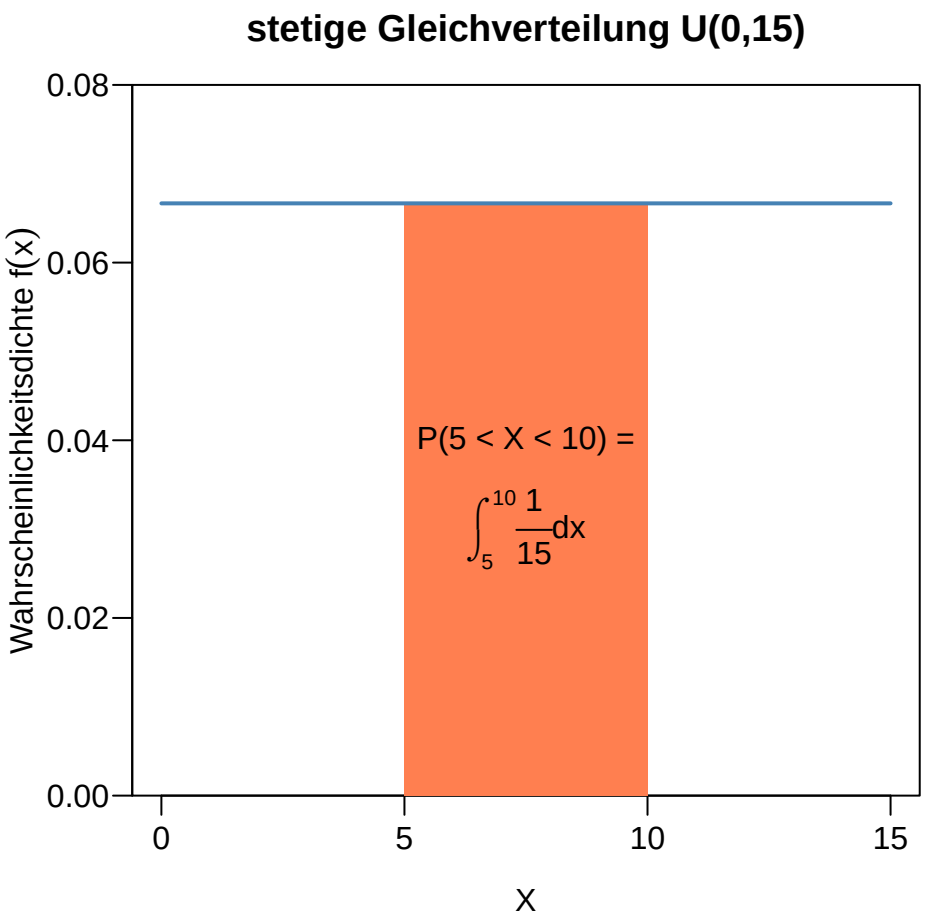


Da die Dichtefunktion konstant ist, wächst die Verteilungsfunktion linear.

Beispiel: Wartezeit für einen Bus

Ein Bus hat eine Frequenz von 15 Minuten. Angenommen, eine Person kann zu jeder Zeit an der Bushaltestelle ankommen. Wie groß ist die Wahrscheinlichkeit, dass man auf den Bus zwischen 5 und 10 Minuten wartet?
In diesem Fall folgt die Variable X, welche die Wartezeit misst, einer kontinuierlichen gleichverteilten Verteilung $U(0, 15)$, da jede Wartezeit zwischen 0 und 15 gleich wahrscheinlich ist.
Die Wahrscheinlichkeit, zwischen 5 und 10 Minuten auf den Bus zu warten, beträgt:

$$P(5 \leq X \leq 10) = \int_5^{10} \frac{1}{15}\, dx = \left[\frac{x}{15}\right]_5^{10} = \\ = \frac{10}{15} - \frac{5}{15} = \frac{1}{3}.$$



Die erwarete (durchschnittliche) Wartezeit ist $\mu = \frac{0+15}{2} = 7.5$ Minuten.

7.3 Normalverteilung

Das Normalverteilungsmodell ist ohne Zweifel das wichtigste Modell für kontinuierliche Verteilungen und das häufigste in der Natur.

Definition "Normalverteilung" $N(\mu, \sigma)$

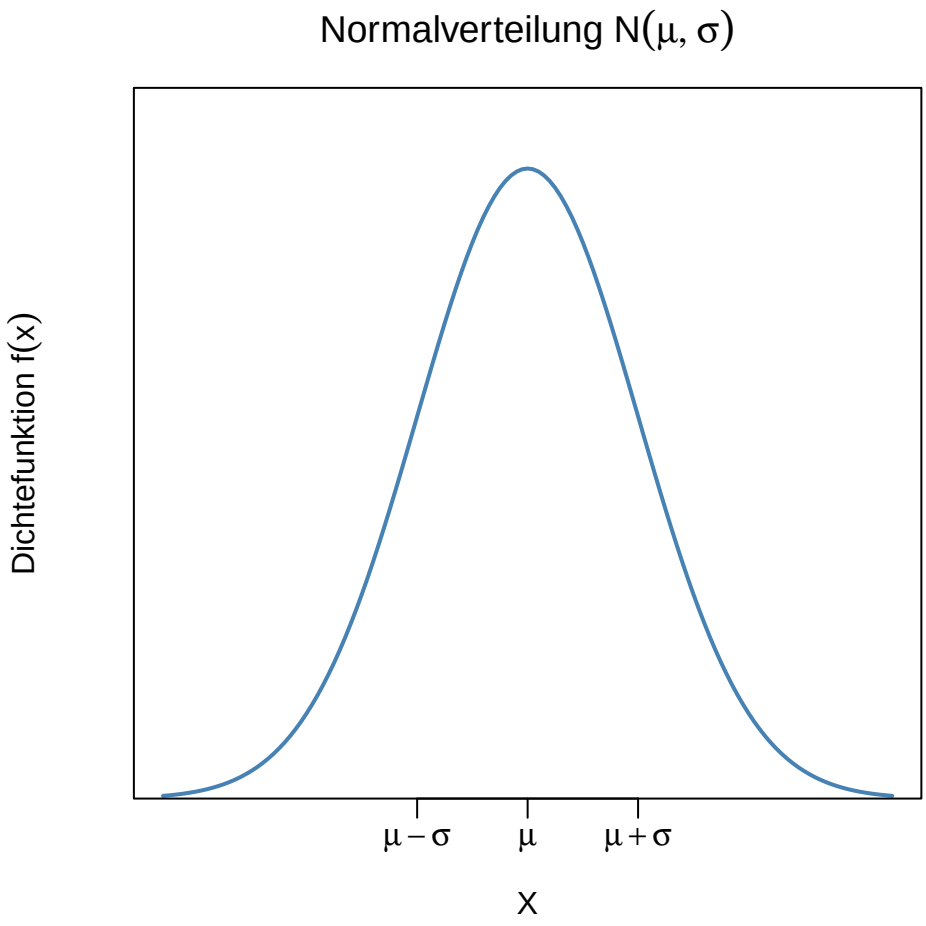
Eine stetige Zufallsvariable X folgt dem Normalverteilungsmodell $N(\mu, \sigma)$, wenn ihr Wertebereich $\text{Ran}(X) = (-\infty, \infty)$ ist und ihre Dichtefunktion

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Die beiden Parameter μ und σ sind der Mittelwert (Erwartungswert) bzw. die Standardabweichung der Population.

7.3.1 Dichtefunktion der Normalverteilung

Der Graph der Wahrscheinlichkeitsdichtefunktion einer Normalverteilung $N(\mu, \sigma)$ ist glockenförmig und wird auch als *Gauß-Glocke* bezeichnet.

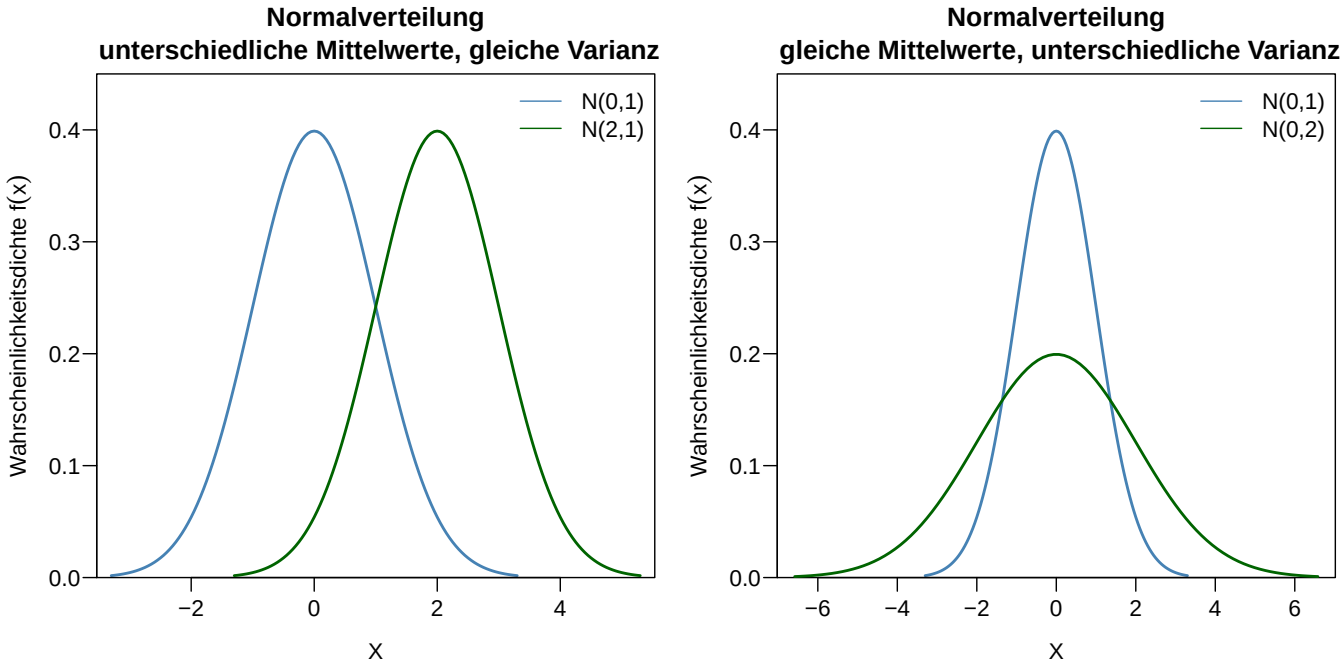


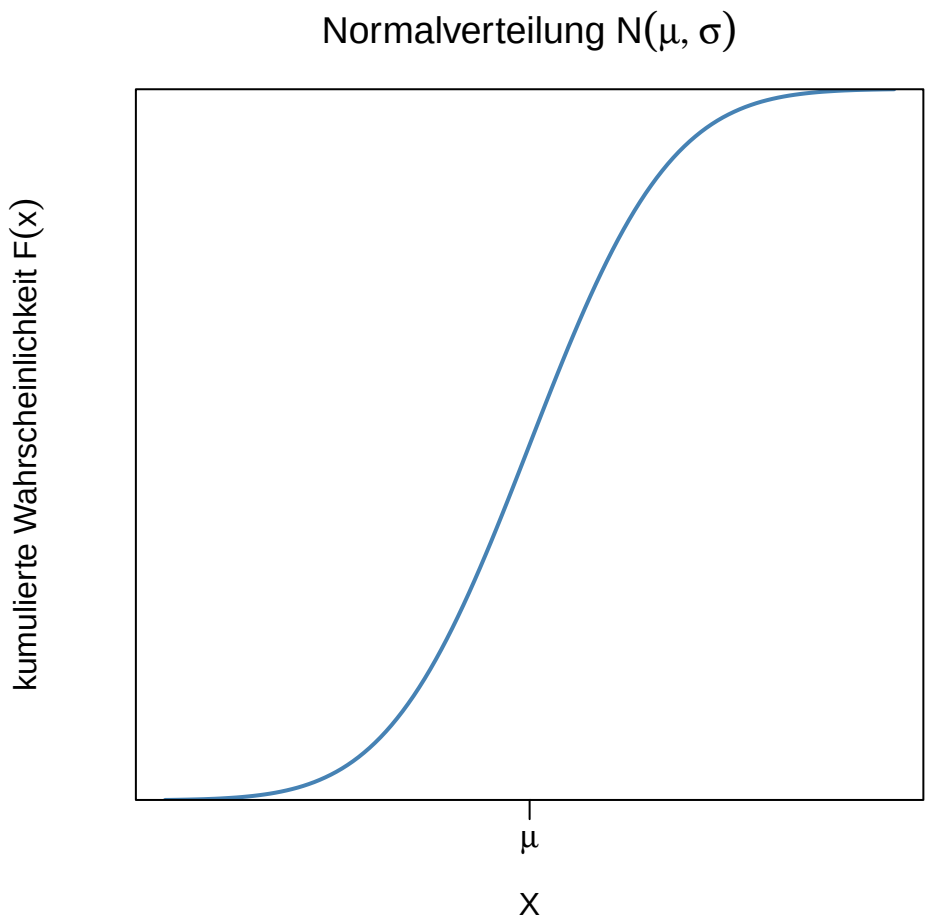
Die Glockenform hängt vom Mittelwert μ und der Standardabweichung σ ab.

- Der Mittelwert μ legt die Mitte der Glocke fest.
- Die Standardabweichung σ legt die Breite der Glocke fest.

7.3.2 Verteilungsfunktion

Die Verteilungsfunktion der Normalverteilung ist S-förmig

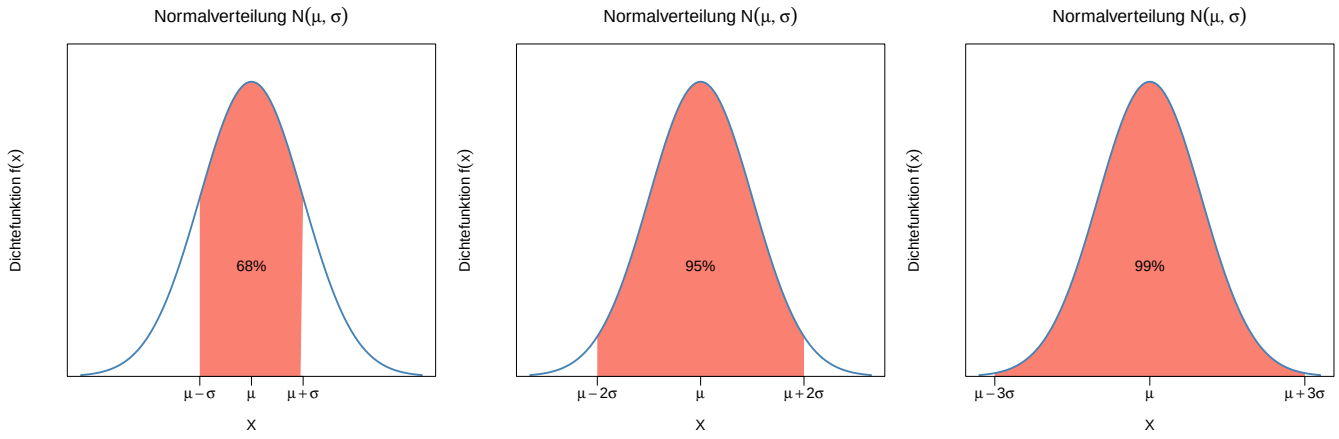




7.3.3 Eigenschaften der Normalverteilung

- Sie ist symmetrisch bezüglich der Mitte, sodass Schiefe null ist: $g_1 = 0$.

- Sie ist mesokurtisch, da die Dichtefunktion glockenförmig ist und der Kurtosiskoeffizient daher null beträgt: $g_2 = 0$.
- Mittelwert, Median und Modalwert sind gleich: $\mu = Me = Mo$.
- Sie nähert sich asymptotisch Null an, wenn x gegen $\pm\infty$ geht.



$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.68$$
$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95$$
$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.99$$

Beispiel Cholesterin

Es ist bekannt, dass die Cholesterinwerte von Frauen im Alter zwischen 40 und 50 Jahren einer Normalverteilung mit $\mu = 210$mg/dl und $\sigma = 20$mg/dl folgen.
Durch die Eigenschaften der Glockenkurve wissen wir, dass

- 68% der Frauen einen Cholesterinwert von 210 ± 20 mg/dl aufweisen, also Werte zwischen 190 und 230 mg/dl.
- 95% der Frauen einen Cholesterinwert von $210 \pm 2 \cdot 20$ mg/dl aufweisen, also Werte zwischen 170 und 250 mg/dl.
- 99% der Frauen einen Cholesterinwert von $210 \pm 3 \cdot 20$ mg/dl aufweisen, also Werte zwischen 150 und 270 mg/dl.

Beispiel Blutanalyse

Bei der Blutanalyse wird häufig das Intervall $□±2□verwendet, um mögliche Pathologien zu erkennen. Im Fall des Cholesterins liegt dieses Intervall bei $[170$ mg/dl, 250 mg/dl$]$.
Somit wird eine Frau im Alter zwischen 40 und 50 Jahren mit einem Cholesteringehalt außerhalb dieses Intervalls häufig auf eine mögliche Erkrankung untersucht. Allerdings könnte diese Person gesund sein, obwohl die Wahrscheinlichkeit dafür nur 5 % beträgt.

7.3.4 Der Zentrale Grenzwertsatz

Dieses Verhalten ist bei vielen physikalischen und biologischen Variablen in der Natur zu beobachten.

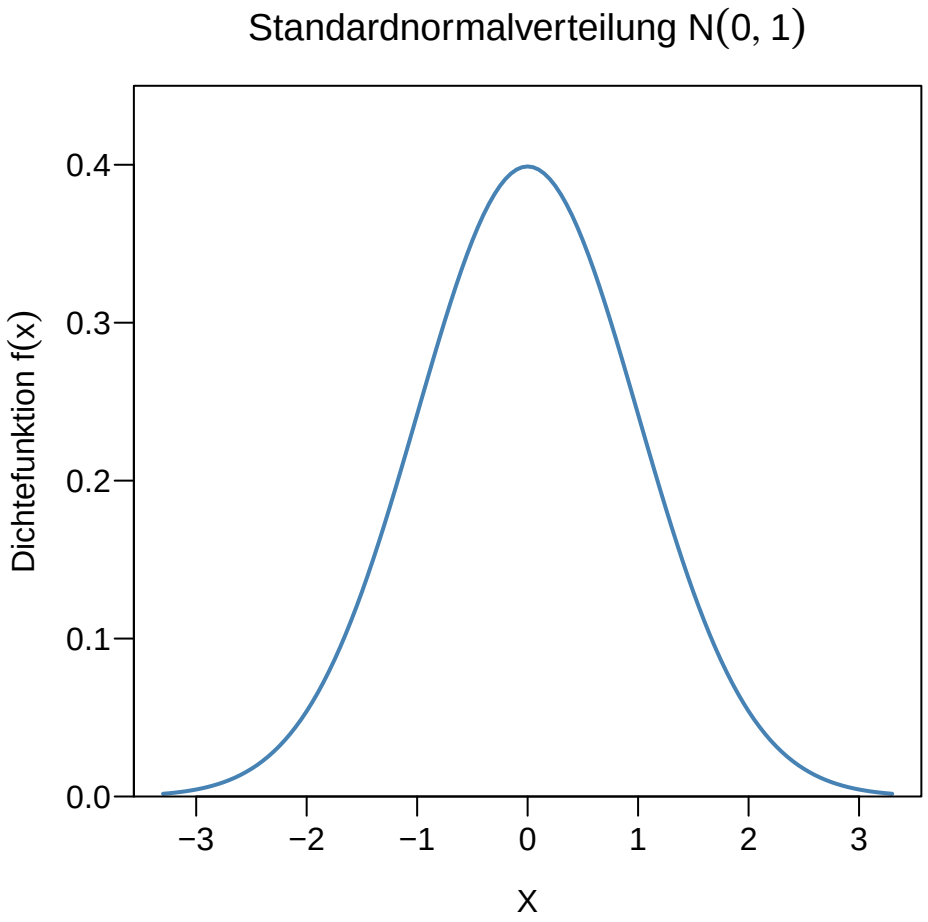
Wenn man zum Beispiel an die Verteilung der Körpergröße denkt, kann man feststellen, dass die meisten Menschen in der Bevölkerung eine Körpergröße um den Mittelwert haben. Allerdings gibt es immer weniger Menschen mit Körpergrößen, je weiter sie vom Mittelwert entfernt sind - sowohl darunter als auch darüber. Die Erklärung für dieses

Verhalten ist der *Zentrale Grenzwertsatz*, auf den wir im nächsten Kapitel zurückkommen werden. Er besagt, dass eine stetige Zufallsvariable, deren Werte von einer großen Anzahl unabhängiger Faktoren abhängen, die ihre Effekte addieren, immer einer Normalverteilung folgt.

7.3.5 Standardnormalverteilung

Die wichtigste Normalverteilung hat einen Mittelwert $\mu = 0$ und eine Standardabweichung $\sigma = 1$.

Sie wird als **Standardnormalverteilung** bezeichnet und üblicherweise mit $Z \sim N(0, 1)$ dargestellt.



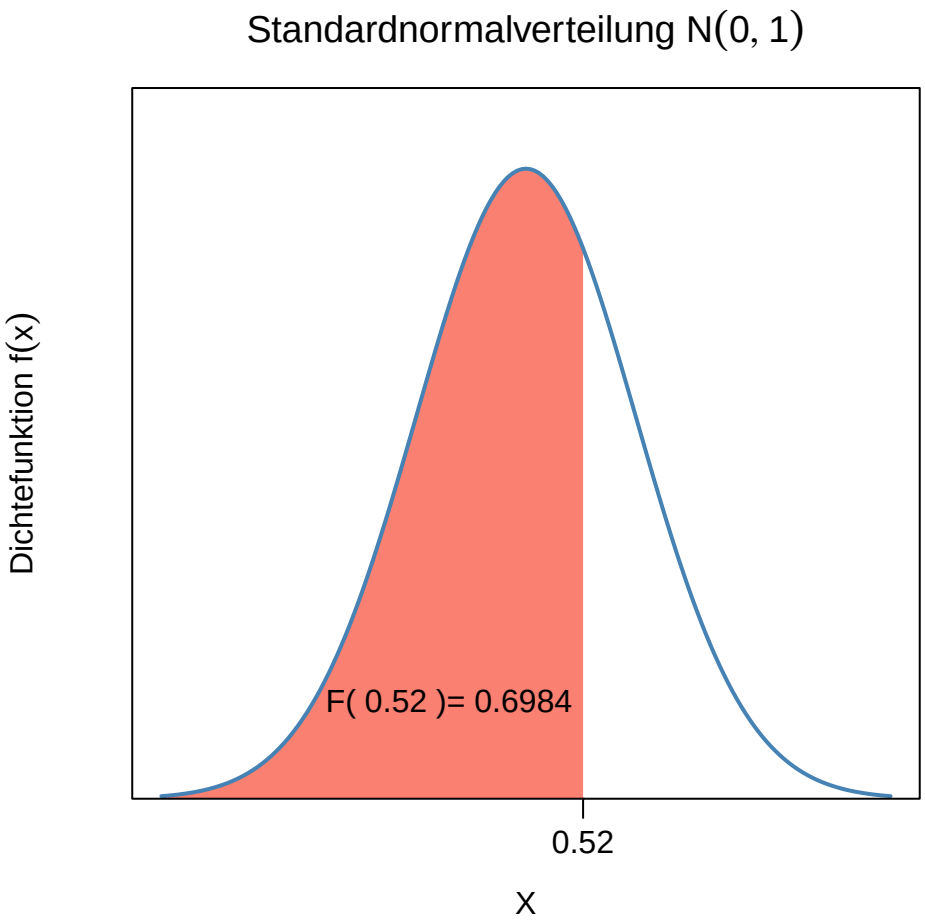
7.3.6 Berechnung von Wahrscheinlichkeiten

Um die Integration der normalen Dichtefunktion zu vermeiden, ist es üblich, die Verteilungsfunktion zu verwenden, die in Tabellenform wie unten angegeben ist.

Beispiel: Werte aus Tabellen ablesen

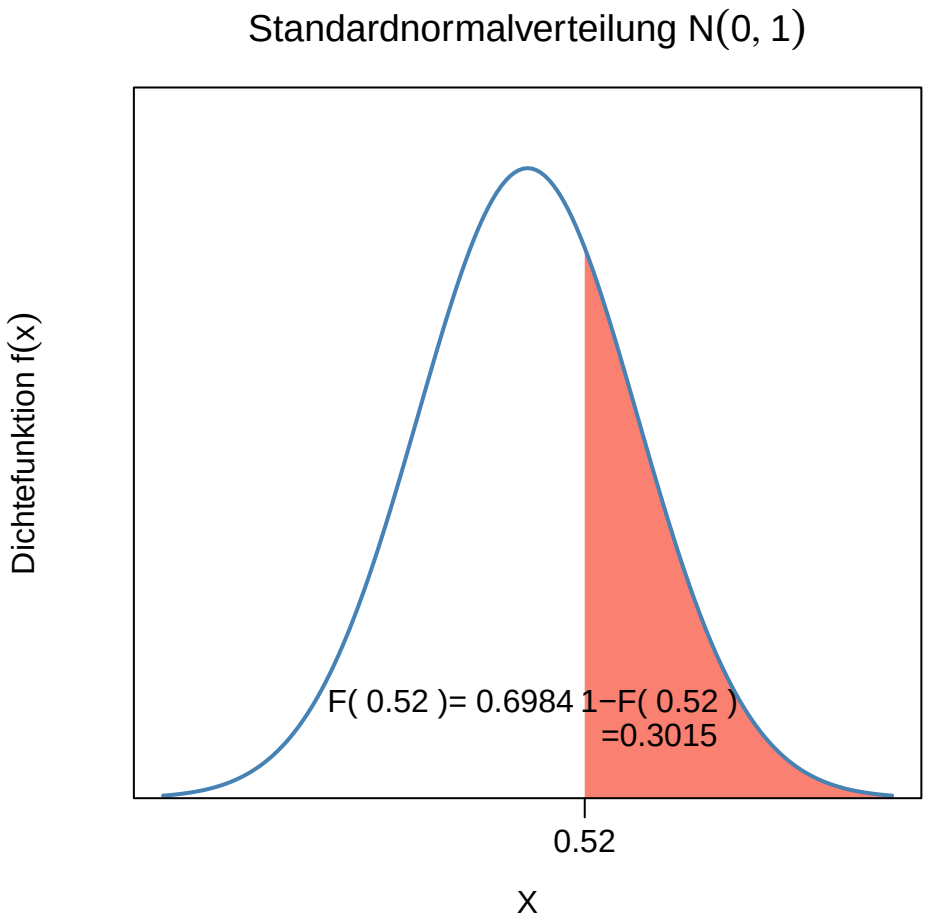
Bestimmung von $P(Z \leq 0.52)$
$\rightarrow$ Reihe 0.5 + Spalte 0.02 $\rightarrow$ **0,6985** ist der gesuchte Wert.

z	0.00	0.01	0.02	⋯
0.0	0.5000	0.5040	0.5080	⋯
0.1	0.5398	0.5438	0.5478	⋯
0.2	0.5793	0.5832	0.5871	⋯
0.3	0.6179	0.6217	0.6255	⋯
0.4	0.6554	0.6591	0.6628	⋯
0.5	0.6915	0.6950	**0.6985**	⋯
⋮	⋮	⋮	⋮	⋱



Um kumulative Wahrscheinlichkeiten für Werte rechts von einem Wert zu berechnen, kann die Regel für komplementäre Ereignisse angewendet werden, beispielsweise:

$$P(Z > 0.52) = 1 - P(Z \leq 0.52) = 1 - F(0.52) = 1 - 0.6985 = 0.3015.$$



7.3.7 Standardisierung

Wir haben gelernt, wie wir die Tabelle der Standardnormalverteilungsfunktion verwenden können, um Wahrscheinlichkeiten zu berechnen. Aber was tun wir, wenn es sich nicht um die Standardnormalverteilung handelt?

In diesem Fall können wir *Standardisierung* anwenden, um jede Normalverteilung in eine Standardnormalverteilung umzuwandeln.

Definition "Standardisierung"

Wenn X eine stetige Zufallsvariable ist, die einer Normalverteilungsfunktion mit Mittelwert μ und Standardabweichung σ folgt, $X \sim N(\mu, \sigma)$, dann folgt die Variable, die durch Subtrahieren von μ und Dividieren durch σ erhalten wird, einer Standardnormalverteilung:

$$X \sim N(\mu, \sigma) \Rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Um daher Wahrscheinlichkeiten mit einer nicht-standardmäßigen Normalverteilung zu berechnen, müssen wir die Variable zuerst standardisieren, bevor wir die Tabelle der Standardnormalverteilungsfunktion verwenden.

Beispiel

Angenommen, die Note X eines Examens folgt einer normalen Wahrscheinlichkeitsverteilungsfunktion $N(\mu = 6, \sigma = 1.5)$. Welcher Prozentsatz der Schüler hat das Examen nicht bestanden?
Da X einer nicht-standardmäßigen Normalverteilung folgt, müssen wir zuerst standardisieren,

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 6}{1.5}$$

$$P(X < 5) = P\left(\frac{X - 6}{1.5} < \frac{5 - 6}{1.5}\right) = P(Z < -0.67).$$

Jetzt können wir die Tabelle der Standardnormalverteilungsfunktion verwenden:

$$P(Z < -0.67) = F(-0.67) = 0.2514.$$

Demnach haben 25,14% der Schüler das Examen nicht bestanden.

7.4 Chi-Quadrat-Verteilung

Definition "Chi-Quadrat-Verteilung $\mathrm{X}^2(n)$

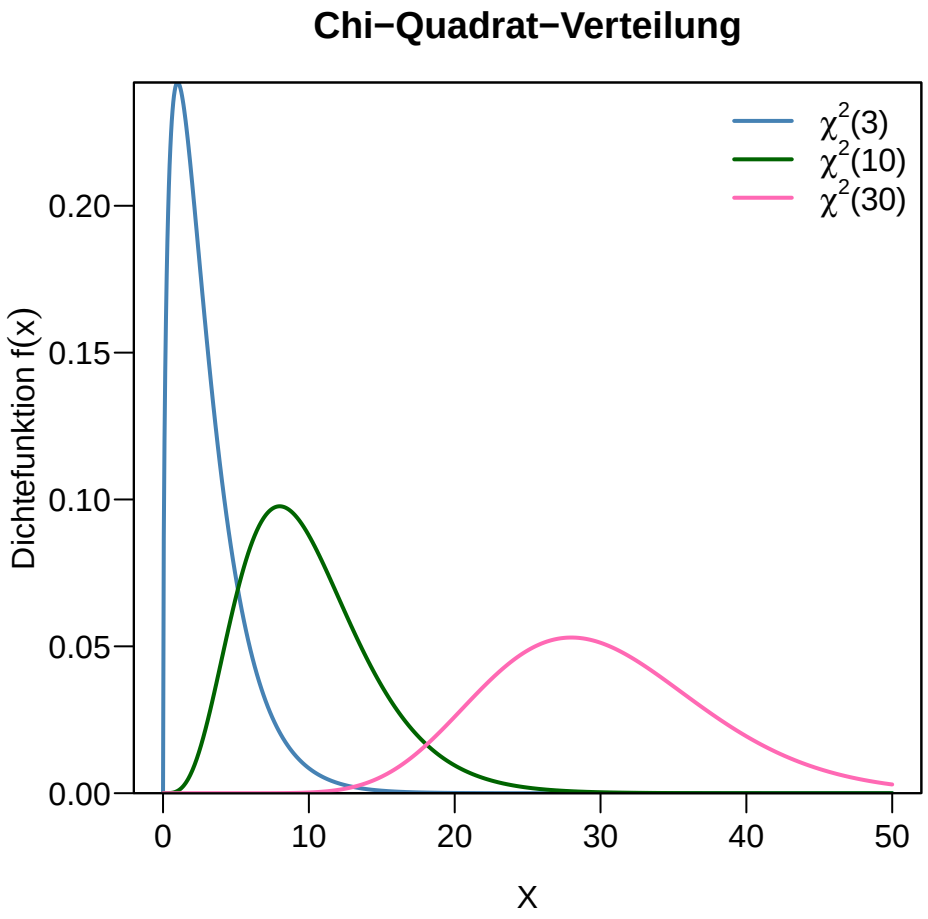
Gegeben n unabhängige Zufallsvariablen $Z_1, \dots, Z_n$, die alle einer Standardnormalverteilung folgen, folgt die Variable

$$\chi^2(n) = Z_1^2 + \cdots + Z_n^2,$$

einer Chi-Quadrat-Verteilung mit n Freiheitsgraden. Ihr Bereich ist $\mathbb{R}^+$, und ihr Mittelwert und ihre Varianz sind:

$$\mu = n, \qquad \sigma^2 = 2n.$$

Die Chi-Quadrat-Verteilung spielt eine wichtige Rolle bei der Schätzung der Populationsvarianz und beim Studium von Beziehungen zwischen qualitativen Variablen.



Eigenschaften der Chi-Quadrat-Verteilung

- Die Spannweite ist nicht-negativ.
- Wenn $X \sim \chi^2(n)$ and $Y \sim \chi^2(m)$, dann gilt $X + Y \sim \chi^2(n + m)$.
- Sie nähert sich asymptotisch einer normalen Verteilung an, wenn die Anzahl an Freiheitsgraden zunimmt.

7.5 Student's t-Verteilung

Definition "Student's t-Verteilung" $T(n)$

Gegeben eine Variable Z, die einer Standardnormalverteilung folgt, $Z \sim N(0, 1)$, und eine Variable X, die einer Chi-Quadrat-Verteilung mit n Freiheitsgraden folgt,

$$X \sim \chi^2(n)$$

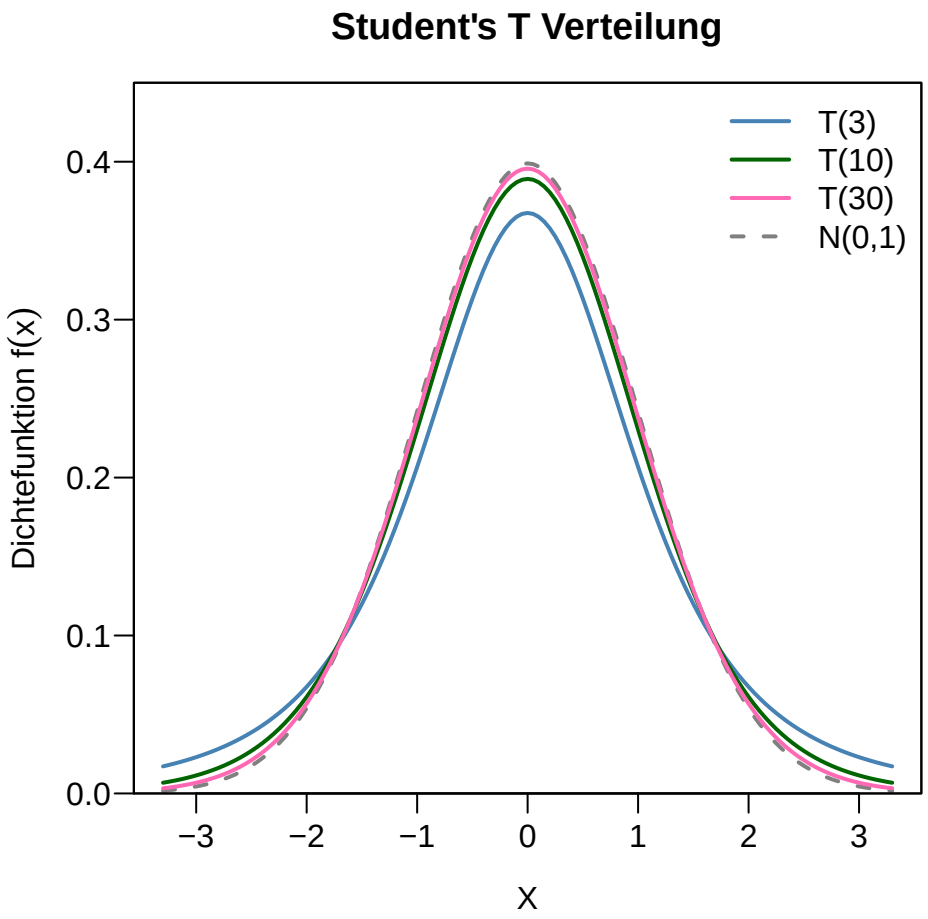
, dann folgt die Variable

$$T = \frac{Z}{\sqrt{X/n}},$$

einer Student's t-Verteilung mit n Freiheitsgraden. Ihr Bereich ist $\mathbb{R}$, und ihr Mittelwert und ihre Varianz sind:

$$\mu = 0, \qquad \sigma^2 = \frac{n}{n-2} \text{ wenn } n > 2.$$

Die t-Verteilung spielt eine wichtige Rolle bei der Schätzung des Populationsmittelwerts.



Eigenschaften der Student's t-Verteilung:

- Der Mittelwert, der Median und der Modus sind gleich, $\mu = Me = Mo$.
- Sie ist symmetrisch, $g_1 = 0$.
- Sie nähert sich asymptotisch einer Standardnormalverteilung an, wenn die Anzahl der Freiheitsgrade abnimmt. In der Praxis gilt für n ≥ 30, dass beide Verteilungen annähernd gleich sind.

$$T(n) \overset{n\to\infty}{\approx} N(0,1).$$

7.6 Fisher-Snedecor-Verteilung

Definition "Fisher-Snedecor-Verteilung" $F(m,n)$

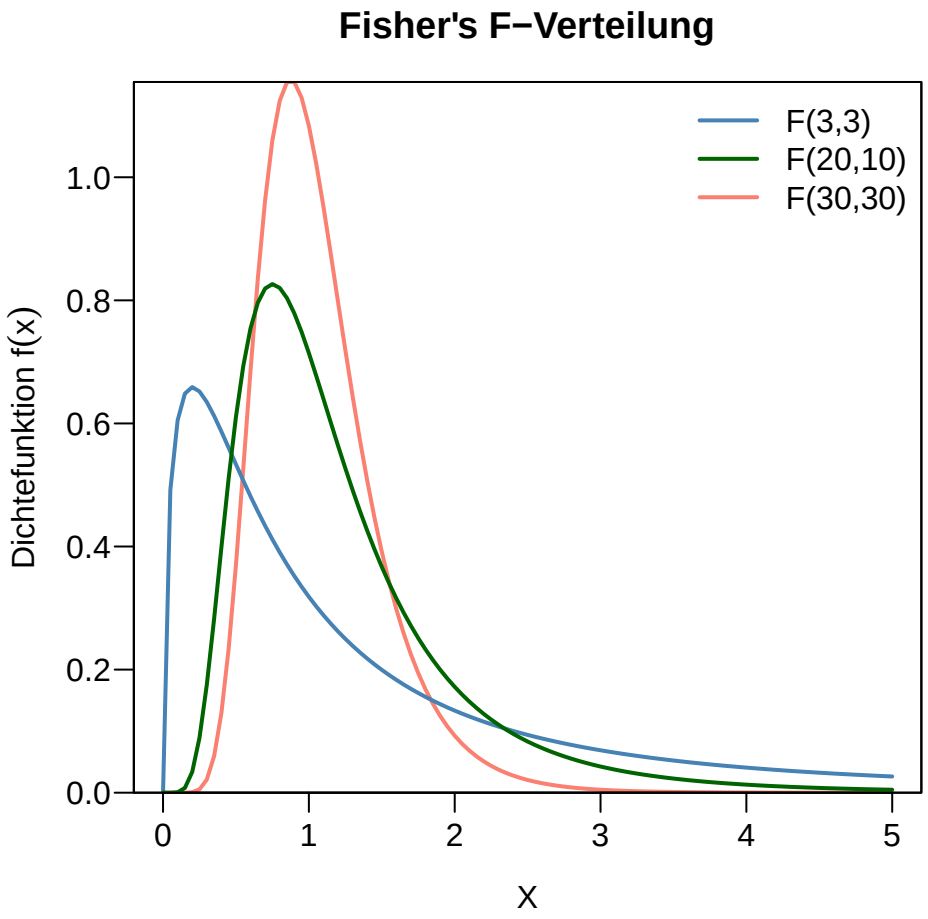
Gegeben zwei unabhängige Variablen X und Y, die jeweils einer Chi-Quadrat-Verteilung mit m bzw. n Freiheitsgraden folgen, $X \sim \chi^2(m)$ and $Y \sim \chi^2(n)$, folgt die Variable

$$F = \frac{X/m}{Y/n},$$

einer Fisher-Snedecor-Verteilung mit m und n Freiheitsgraden. Ihr Bereich ist $\mathbb{R}^+$, und ihr Mittelwert und ihre Varianz sind:

$$\mu = \frac{n}{n-2}, \qquad \sigma^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \text{ wenn } n > 4.$$

Die Fisher-Snedecor-Verteilung spielt eine wichtige Rolle beim Vergleich von Populationsvarianzen und bei der ANOVA-Analyse (Analysis of Variance).



Eigenschaften der Fisher-Snedecor-Verteilung:

- Der Bereich ist nicht-negativ.
- Sie erfüllt die Beziehung

$$F(m,n) = \frac{1}{F(n,m)}.$$

Dies bedeutet, dass wenn wir den Wert $f(m,n)_p$ definieren als den Wert, der die Gleichung $P(F(m,n) \leq f(m,n)_p) = p$ erfüllt, dann gilt:

$$f(m,n)_p = \frac{1}{f(n,m)_{1-p}}$$

Dies ist hilfreich beim Berechnen von Wahrscheinlichkeiten mithilfe der Verteilungsfunktionstabelle.

8 Schätzen der Populationsparameter

Die Wahrscheinlichkeitsverteilungsmodelle, die im vorherigen Abschnitt behandelt wurden, erklären das Verhalten von Zufallsvariablen. Hierfür muss bekannt sein, welchem Verteilungsmodell eine bestimmte Variable folgt. Dies ist der erste Schritt der *inferenziellen Statistik*.

Um das Verteilungsmodell einer Variable genau zu bestimmen, müssen die charakteristischen Merkmale aller Individuen in der Population bekannt sein, was in den meisten Fällen nicht möglich ist (wirtschaftliche, körperliche, zeitliche Unmöglichkeit usw.).

Um diese Nachteile zu vermeiden, wird auf die Untersuchung einer Stichprobe zurückgegriffen, mit dem Ziel, das Verteilungsmodell der Variable in der Population ungefähr zu bestimmen.

Die Untersuchung einer kleineren Anzahl von Individuen aus einer Stichprobe statt der gesamten Population (Vollerhebung) hat eindeutige Vorteile:

- geringere Kosten.
- größere Schnelligkeit.
- größere Leichtigkeit.

Es hat jedoch auch einige Nachteile:

- Notwendigkeit, eine *repräsentative* (siehe hierzu von der Lippe (2002, 2010, 2011; 2002)) Stichprobe zu erhalten.
- Gefahr, Verzerrungen zu erhalten.

Im Idealfall können diese Nachteile überwunden werden:

- Die Repräsentativität der Stichprobe wird durch angemessene Samplingmethoden erreicht (Zufallsstichprobe);
- Fehler können nicht vermieden werden, aber es wird versucht, sie so weit wie möglich zu reduzieren und einzuschränken.

8.1 Stichprobenverteilungen

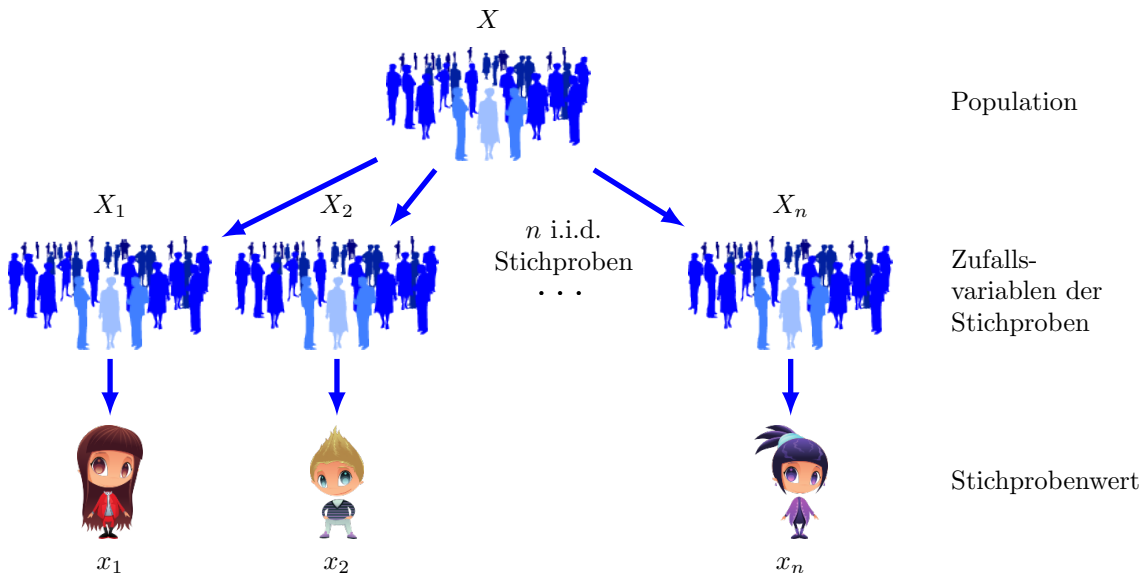
Eine Stichprobe der Größe n aus einer Zufallsvariable X kann als Realisation einer n-dimensionalen Zufallsvariablen $X = (X_1, \dots, X_n)$ aufgefasst werden.

⚠ Definition “Stichprobe unabhängiger identisch verteilter Zufallsvariablen” (i.i.d.)

Eine *Stichprobe unabhängiger identisch verteilter Zufallsvariablen* (abgekürzt **i.i.d.**, von *independent and identically distributed*) einer Zufallsvariable X, die in einer Population untersucht wird, ist eine Sammlung von n Zufallsvariablen $X_1, \dots, X_n$, die folgende Eigenschaften erfüllen:

- Jede der Zufallsvariablen X_i folgt derselben Wahrscheinlichkeitsverteilung wie die Variable X in der Population.
- Alle Zufallsvariablen X_i sind voneinander unabhängig.

Die möglichen Werte dieser n-dimensionalen Zufallsvariablen entsprechen allen denkbaren Stichproben der Größe n, die aus der Population gezogen werden können.



Die drei grundlegenden Merkmale der Stichproben-Zufallsvariable sind:

1. Homogenität: Die Variablen, die die Stichproben-Zufallsvariable bilden, folgen der gleichen Verteilung.
2. Unabhängigkeit: Die Variablen sind voneinander unabhängig.
3. Verteilungsmodell: Das Verteilungsmodell, dem die n Variablen folgen.

Die ersten beiden Punkte können gelöst werden, wenn eine einfache Zufallsstichprobe zur Erhebung der Daten verwendet wird. Beim dritten Punkt müssen wiederum zwei Fragen beantwortet werden:

1. Welches Verteilungsmodell passt am besten zu unserem Datensatz? Dies wird teilweise durch den Einsatz nichtparametrischer Methoden geklärt.
2. Sobald das passendste Verteilungsmodell ausgewählt wurde: Welcher Parameter des Modells ist von Interesse, und wie kann sein Wert bestimmt werden? Damit befasst sich der Teil der statistischen Inferenz, der als Parameterschätzung bekannt ist.

In diesem Abschnitt wird die zweite Frage behandelt, d. h., unter der Annahme, dass das Verteilungsmodell einer Grundgesamtheit bekannt ist, werden die wichtigsten Parameter geschätzt, die es beschreiben. Zum Beispiel sind die zentralen Parameter, die die im vorherigen Thema behandelten Verteilungen definieren:

Verteilung	Parameter
Binomial	n, p
Poisson	λ
Uniform	a, b
Normal	μ, σ
Chi-Quadrat	n
T-Student	n
F-Fisher	m, n

Die Wahrscheinlichkeitsverteilung der Werte der Stichprobenvariablen hängt eindeutig von der Wahrscheinlichkeitsverteilung der Werte in der Grundgesamtheit ab.

Beispiel: Kinder in Familien

Betrachten wir eine Grundgesamtheit, in der ein Viertel der Familien keine Kinder hat, die Hälfte der Familien ein Kind besitzt und der Rest zwei Kinder hat.

Verteilung in der Population

X	$P(x)$
0	0,25
1	0,50
2	0,25

Stichproben vom Umfang 2 →

Verteilung in den Stichproben

(X_1, X_2)	$P(x_1, x_2)$
$(0, 0)$	0,0625
$(0, 1)$	0,1250
$(0, 2)$	0,0625
$(1, 0)$	0,1250
$(1, 1)$	0,2500
$(1, 2)$	0,1250
$(2, 0)$	0,0625
$(2, 1)$	0,1250
$(2, 2)$	0,0625

Da ein Statistikergebnis eine Funktion einer Zufallsvariable ist, stellt es selbst ebenfalls eine Zufallsvariable dar. Somit hängt seine Wahrscheinlichkeitsverteilung ebenfalls von der Verteilung der Grundgesamtheit und den sie bestimmenden Parametern ab (μ, σ, p, …).

Beispiel: Mittelwert

Wird der Stichprobenmittelwert $\bar{X}$ aus Stichproben der Größe 2 (gemäß vorherigem Beispiel) berechnet, ergibt sich folgende Wahrscheinlichkeitsverteilung:

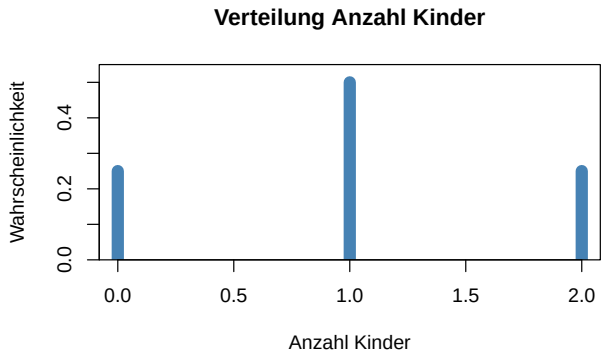
Verteilung in den Stichproben

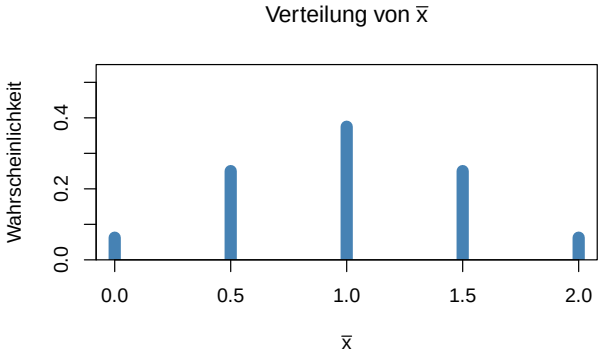
(X_1, X_2)	$P(x_1, x_2)$
$(0, 0)$	0,0625
$(0, 1)$	0,1250
$(0, 2)$	0,0625
$(1, 0)$	0,1225
$(1, 1)$	0,2500
$(1, 2)$	0,1225
$(2, 0)$	0,0625
$(2, 1)$	0,1250
$(2, 2)$	0,0625

$\overline{x}=\frac{X_1+X_2}{2}$ →

Verteilung von $\overline{x}$

$\overline{X}$	$P(x)$
0	0,0625
0,5	0,2500
1	0,3750
1,5	0,2500
2	0,0625





Wie hoch ist die Wahrscheinlichkeit, einen Stichprobenmittelwert zu erhalten, der den wahren Populationsmittelwert mit einem maximalen Fehler von 0,5 approximiert?

Wie wir gesehen haben, ist für die Bestimmung der Verteilung einer Stichprobenkennfunktion die Kenntnis der Populationsverteilung erforderlich, was jedoch nicht immer möglich ist. Glücklicherweise lässt sich für große Stichproben

die Verteilung bestimmter Kennfunktionen wie des Mittelwerts näherungsweise bestimmen, dank des folgenden Theorems:

Zentraler Grenzwertsatz

Wenn $X_1, \dots, X_n$ unabhängige Zufallsvariablen ($n \geq 30$) mit Erwartungswerten $\mu_i = E(X_i)$ und Varianzen $\sigma_i^2 = Var(X_i)$ (für $i = 1, \dots, n$) sind, dann folgt die Zufallsvariable $X = X_1 + \cdots + X_n$ annähernd einer Normalverteilung.

$$X = X_1 + \cdots + X_n \overset{n\geq 30}{\sim} N\left(\sum_{i=1}^{n} \mu_i, \sqrt{\sum_{i=1}^{n} \sigma_i^2}\right)$$

Der zentrale Grenzwertsatz erklärt auch, warum die meisten biologischen Variablen einer Normalverteilung folgen: Sie entstehen typischerweise durch das Zusammenspiel zahlreicher Faktoren, deren Effekte sich unabhängig voneinander addieren.

8.1.1 Verteilung des Stichprobenmittels für große Stichproben ($n \geq 30$)

Das Stichprobenmittel einer Zufallsstichprobe vom Umfang n ist die Summe von n unabhängigen, identisch verteilten Zufallsvariablen:

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{X_1}{n} + \cdots + \frac{X_n}{n}$$

Gemäß den Eigenschaften linearer Transformationen sind Erwartungswert und Varianz jeder dieser Variablen:

$$E\left(\frac{X_i}{n}\right) = \frac{\mu}{n} \quad \text{und} \quad Var\left(\frac{X_i}{n}\right) = \frac{\sigma^2}{n^2}$$

wobei μ und σ^2 den Erwartungswert und die Varianz der zugrundeliegenden Population darstellen.

Somit folgt für große Stichprobenumfänge ($n \geq 30$) nach dem zentralen Grenzwertsatz, dass die Verteilung des Stichprobenmittels normal ist:

$$\bar{X} \sim N\left(\sum_{i=1}^{n} \frac{\mu}{n}, \sqrt{\sum_{i=1}^{n} \frac{\sigma^2}{n^2}}\right) = N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Beispiel für große Stichproben (n ≥ 30)

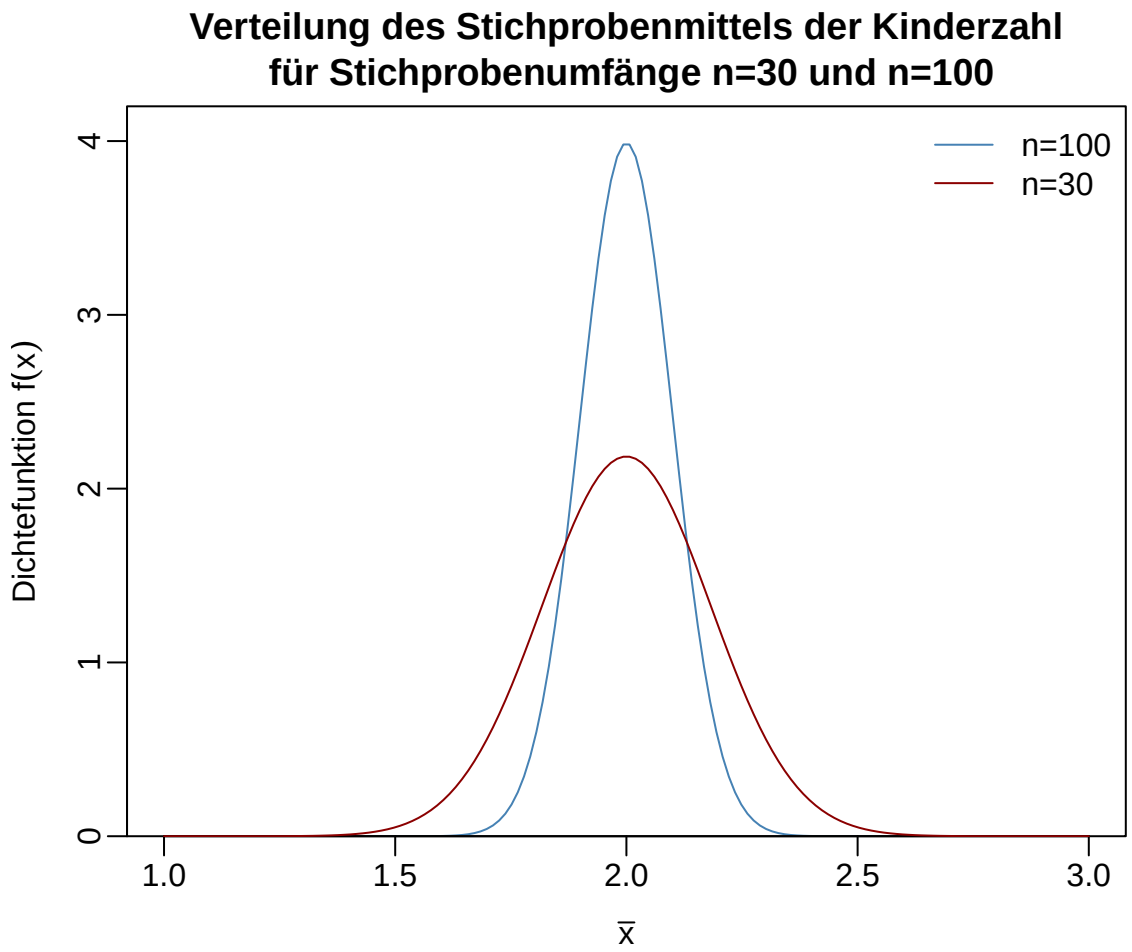
Angenommen, man möchte die durchschnittliche Kinderzahl in einer Population schätzen, wobei der wahre Mittelwert $\mu = 2$ Kinder und die Standardabweichung $\sigma = 1$ Kind beträgt.
Wie hoch ist die Wahrscheinlichkeit, μ mittels $\bar{x}$ mit einem Fehler von weniger als 0,2 zu schätzen?
Gemäß dem zentralen Grenzwertsatz gilt:

1. Für $n = 30$ ist $\bar{x} \sim N(2, 1/\sqrt{30})$ und

$$P(1.8 < \bar{x} < 2.2) = 0.7267.$$

2. Für $n = 100$ ist $\bar{x} \sim N(2, 1/\sqrt{100})$ und

$$P(1.8 < \bar{x} < 2.2) = 0.9545.$$



8.1.2 Verteilung einer Stichprobenproportion für große Stichproben (n 30)

Eine Populationsproportion p kann als Mittelwert einer dichotomen $(0, 1)$-Variable berechnet werden. Diese Variable wird als *Bernoulli-Variable* $B(p)$ bezeichnet, ein Spezialfall der Binomialverteilung für $n = 1$.

Für eine Zufallsstichprobe vom Umfang n lässt sich die Stichprobenproportion $\hat{p}$ somit als Summe von n unabhängigen, identisch verteilten Zufallsvariablen darstellen:

$$\hat{p} = \bar{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{X_1}{n} + \cdots + \frac{X_n}{n}, \text{ mit } X_i \sim B(p)$$

mit Erwartungswert und Varianz

$$E\left(\frac{X_i}{n}\right) = \frac{p}{n} \quad \text{und} \quad Var\left(\frac{X_i}{n}\right) = \frac{p(1-p)}{n^2}$$

Für große Stichprobenumfänge ($n \geq 30$) folgt nach dem zentralen Grenzwertsatz, dass die Verteilung der Stichprobenproportion ebenfalls normalverteilt ist:

$$\hat{p} \sim N\left(\sum_{i=1}^{n} \frac{p}{n}, \sqrt{\sum_{i=1}^{n} \frac{p(1-p)}{n^2}}\right) = N\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

8.2 Schätzer

Stichprobenstatistiken können zur Approximation von Populationsparametern verwendet werden. Wenn ein Kennwert zu diesem Zweck eingesetzt wird, bezeichnet man ihn als *Schätzer* für den Parameter.

Definition "Schätzer und Schätzwert"

Ein Schätzer ist eine Funktion der Stichproben-Zufallsvariablen:

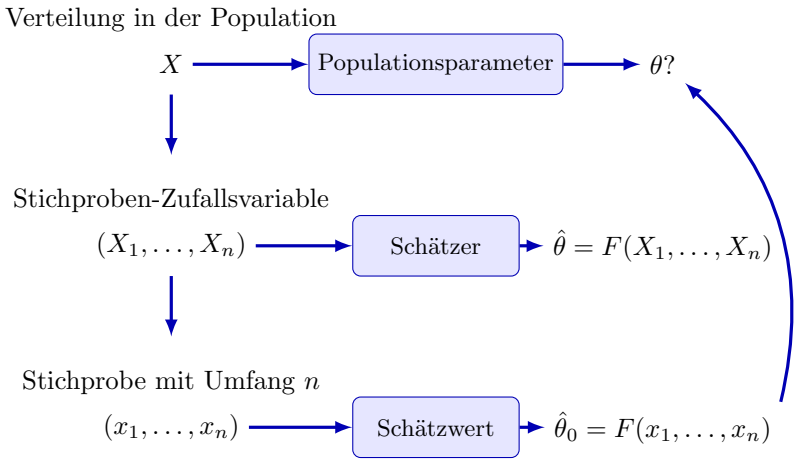
$$\hat{\theta} = F(X_1, \dots, X_n).$$

Für eine konkrete Stichprobe $(x_1, \dots, x_n)$ wird der Wert des Schätzers, angewendet auf diese Stichprobe, als *Schätzwert* bezeichnet:

$$\hat{\theta}_0 = F(x_1, \dots, x_n).$$

Da es sich um eine Funktion der Stichproben-Zufallsvariablen handelt, ist ein Schätzer selbst wieder eine Zufallsvariable, deren Verteilung von der zugrundeliegenden Population abhängt.

Während der *Schätzer* als Funktion eindeutig ist, ist der *Schätzwert* nicht eindeutig, sondern hängt von der gezogenen Stichprobe ab.



Beispiel Raucher

Angenommen, man möchte den Anteil p der Raucher in einer Stadt ermitteln. In diesem Fall folgt die dichotome Variable, die misst, ob eine Person raucht (1) oder nicht (0), einer Bernoulli-Verteilung $B(p)$.
Wenn eine Zufallsstichprobe vom Umfang 5, $(X_1, X_2, X_3, X_4, X_5)$, aus dieser Grundgesamtheit gezogen wird, kann der Anteil der Raucher in der Stichprobe als Schätzer für den Anteil der Raucher in der Grundgesamtheit verwendet werden:

$$\hat{p} = \frac{\sum_{i=1}^{5} X_i}{5}$$

Dieser Schätzer ist eine Zufallsvariable, die folgendermaßen verteilt ist:

$$\hat{p} \sim \frac{1}{n} B\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

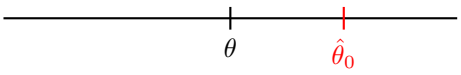
Wenn verschiedene Stichproben gezogen werden, erhält man unterschiedliche Schätzwerte:

Stichprobe	Schätzwert
$(1,0,0,1,1)$	$3/5$
$(1,0,0,0,0)$	$1/5$
$(0,1,0,0,1)$	$2/5$
...	...

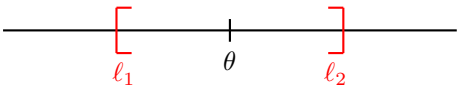
Die Parameterschätzung kann auf zwei Arten durchgeführt werden:

1. **Punktschätzung:** Es wird ein einzelner Schätzer verwendet, der einen Wert oder eine Näherung für den Parameter liefert. Der Hauptnachteil dieser Schätzmethode besteht darin, dass die Genauigkeit der Schätzung nicht angegeben wird.
2. **Intervallschätzung:** Hier werden zwei Schätzer verwendet, die die Grenzen eines Intervalls liefern, in dem der wahre Wert des Parameters mit einer bestimmten Sicherheit vermutet wird. Diese Schätzmethode ermöglicht es, den Schätzfehler zu kontrollieren.

Punkt-Schätzwert



Intervallschätzung



8.3 Punktschätzung

Die Punktschätzung verwendet einen einzelnen Schätzer, um den Wert eines unbekannten Parameters der Grundgesamtheit zu schätzen.

Theoretisch können verschiedene Schätzer für denselben Parameter verwendet werden. Zum Beispiel hätte man zur Schätzung des Raucheranteils in einer Stadt neben dem Stichprobenanteil auch andere mögliche Schätzer verwenden können, wie:

$$\hat{\theta}_1 = \sqrt[5]{X_1 X_2 X_3 X_4 X_5}\ \hat{\theta}_2 \quad = \frac{X_1 + X_5}{2}\ \hat{\theta}_3 = X_1 \cdots$$

Welcher Schätzer ist der beste?

Die Antwort hängt von den Eigenschaften der einzelnen Schätzer ab.

Obwohl die Punktschätzung keine direkte Aussage über die Genauigkeit der Schätzung liefert, gibt es verschiedene Eigenschaften, die diese Qualität sicherstellen.

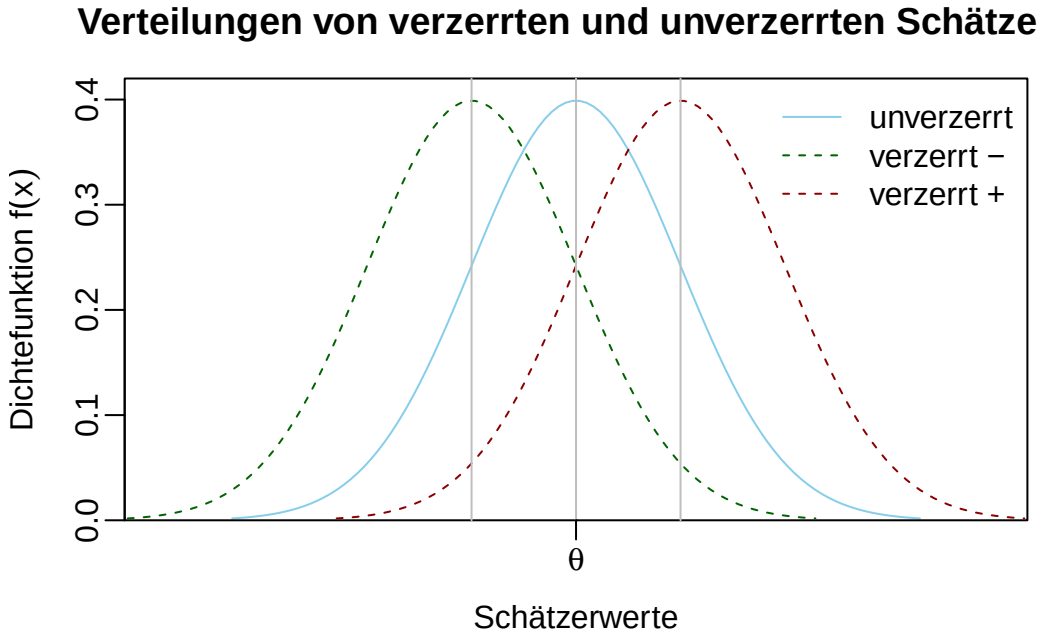
Die wünschenswertesten Eigenschaften eines Schätzers sind:

- Erwartungstreue (Unverzerrtheit)
- Effizienz
- Konsistenz
- Asymptotische Normalverteilung
- Vollständigkeit

⚠ Definition “Erwartungstreuer Schätzer” (unverzerrter Schätzer)

Ein Schätzer $\hat{\theta}$ heißt erwartungstreu (oder unverzerrt) für einen Parameter θ, wenn sein Erwartungswert genau θ entspricht, d. h.,

$$E(\hat{\theta}) = \theta.$$



Wenn ein Schätzer nicht erwartungstreu ist, wird die Differenz zwischen seinem Erwartungswert und dem wahren Parameterwert θ als Verzerrung (Bias) bezeichnet:

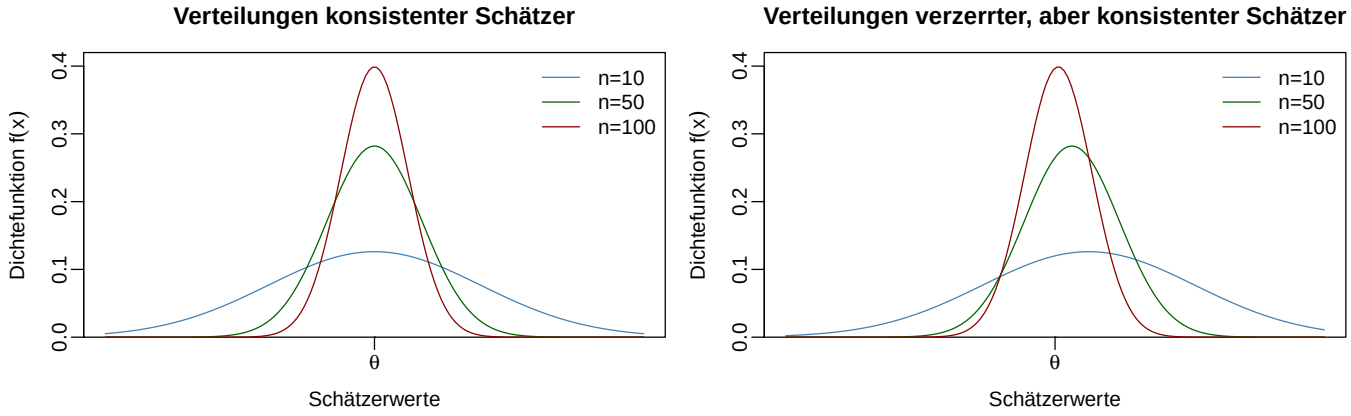
$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Je kleiner die Verzerrung eines Schätzers ist, desto näher liegen seine Schätzwerte am tatsächlichen Parameterwert.

Definition "Konsistenter Schätzer"

Ein Schätzer $\hat{\theta}_n$ für Stichproben des Umfangs n heißt **konsistent** für einen Parameter θ, wenn für jedes beliebige $\epsilon > 0$ gilt:

$$\lim_{n \to \infty} P(|\hat{\theta}_n - \theta| < \epsilon) = 1.$$



Ein Schätzer ist konsistent, wenn folgende Bedingungen erfüllt sind:

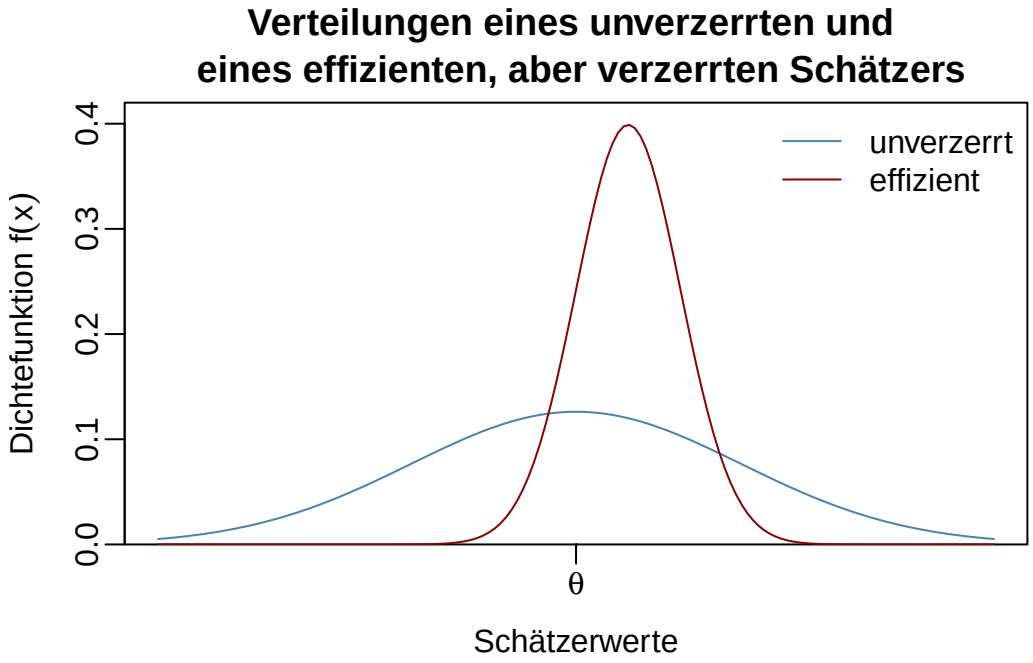
- $\text{Bias}(\hat{\theta}_n) = 0$ oder $\lim_{n\to\infty} \text{Bias}(\hat{\theta}_n) = 0$
- $\lim_{n\to\infty} \text{Var}(\hat{\theta}_n) = 0$

Wenn also sowohl die Varianz als auch die Verzerrung mit wachsendem Stichprobenumfang abnehmen, ist der Schätzer konsistent.

⚠ Definition "Effizienter Schätzer"

Ein Schätzer $\hat{\theta}$ für einen Parameter θ heißt *effizient*, wenn er den mittleren quadratischen Fehler (MSE) minimiert:

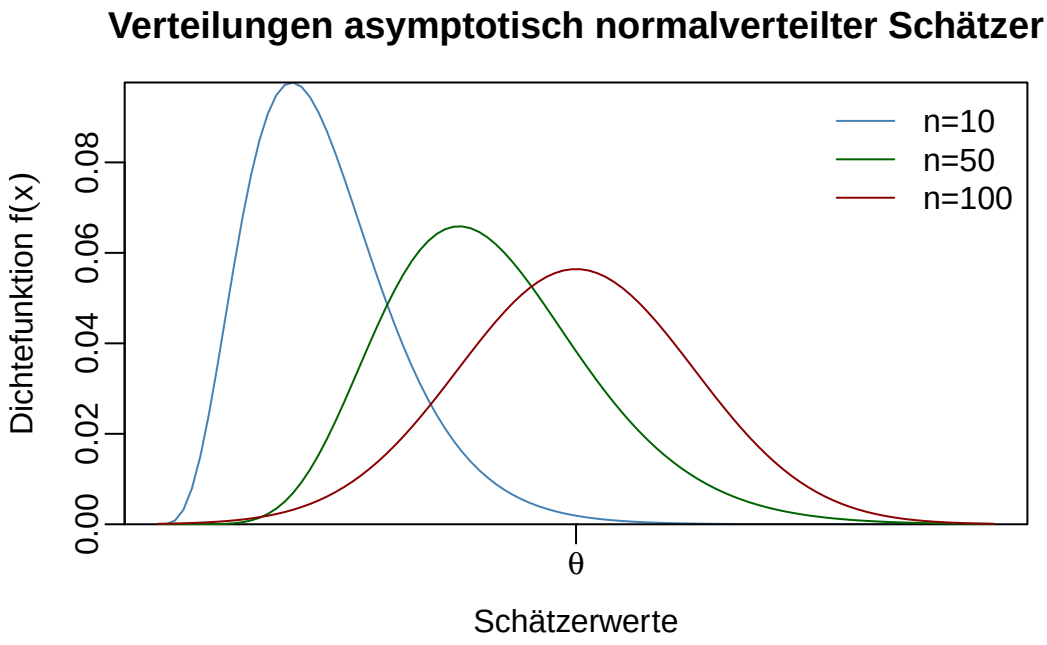
$$\text{MSE}(\hat{\theta}) = \text{Bias}(\hat{\theta})^2 + Var(\theta).$$



⚠ Definition "Asymptotisch normalverteilter Schätzer"

Ein Schätzer $\hat{\theta}$ heißt asymptotisch normalverteilt, wenn – unabhängig von der Verteilung der Zufallsvariablen in der Stichprobe – seine Verteilung bei hinreichend großem Stichprobenumfang einer Normalverteilung folgt.

Wie wir später sehen werden, ist diese Eigenschaft besonders wichtig für die Parameterschätzung mittels Konfidenzintervallen.



Definition "Suffizienter Schätzer"

Ein Schätzer $\hat{\theta}$ heißt suffizient (erschöpfend) für einen Parameter θ, wenn die bedingte Verteilung der Stichprobenvariablen gegeben den Schätzwert $\hat{\theta} = \hat{\theta}_0$ nicht mehr von θ abhängt.

Dies bedeutet, dass nach Erhalt der Schätzung alle weiteren Informationen über die Stichprobe für θ irrelevant werden.

Der übliche Schätzer für den Erwartungswert ist das Stichprobenmittel:

Für Stichproben vom Umfang n ergibt sich die Zufallsvariable:

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

Wenn die Grundgesamtheit den Mittelwert μ und Varianz σ^2 besitzt, gilt:

$$E(\bar{X}) = \mu \quad \text{und} \quad Var(\bar{X}) = \frac{\sigma^2}{n}$$

Somit ist das Stichprobenmittel:

- ein erwartungstreuer Schätzer
- konsistent (da die Varianz mit wachsendem n gegen 0 geht)
- effizient

Die Stichprobenvarianz:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$$

ist jedoch ein *verzerrter* Schätzer für die Populationsvarianz, denn:

$$E(S^2) = \frac{n-1}{n}\sigma^2.$$

Diese Verzerrung lässt sich leicht korrigieren, um einen erwartungstreuen Schätzer zu erhalten.

⚠ Definition “Stichproben-Quasivarianz”

Für eine Stichprobe vom Umfang n einer Zufallsvariablen X wird die Stichproben-Quasivarianz definiert als:

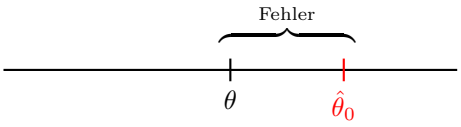
$$\hat{S}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} = \frac{n}{n-1}S^2.$$

8.4 Intervallschätzer

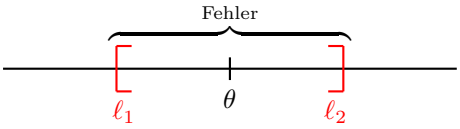
Das Hauptproblem der Punktschätzung besteht darin, dass nach Auswahl der Stichprobe und Berechnung der Schätzung der tatsächliche Fehler unbekannt bleibt.

Um den Schätzfehler kontrollieren zu können, ist die Intervallschätzung die bessere Methode.

Punkt-Schätzwert



Intervallschätzung



Bei der Intervallschätzung wird versucht, ausgehend von der Stichprobe ein Intervall zu konstruieren, in dem der zu schätzende Parameter mit einer bestimmten Konfidenzwahrscheinlichkeit vermutet wird. Hierzu werden zwei Schätzer verwendet - einer für die untere und einer für die obere Intervallgrenze.

⚠ Definition “Konfidenzintervall”

Gegeben zwei Schätzer $\hat{l}_i(X_1, \dots, X_n)$ und $\hat{l}_s(X_1, \dots, X_n)$ sowie deren jeweilige Schätzwerte l_1 und l_2 für eine konkrete Stichprobe, heißt das Intervall $I = [l_1, l_2]$ Konfidenzintervall für einen Populationsparameter θ zum Konfidenzniveau $1-\alpha$ (oder Signifikanzniveau α), wenn gilt:

$$P(\hat{l}_i(X_1, \dots, X_n) \leq \theta \leq \hat{l}_s(X_1, \dots, X_n)) = 1 - \alpha.$$

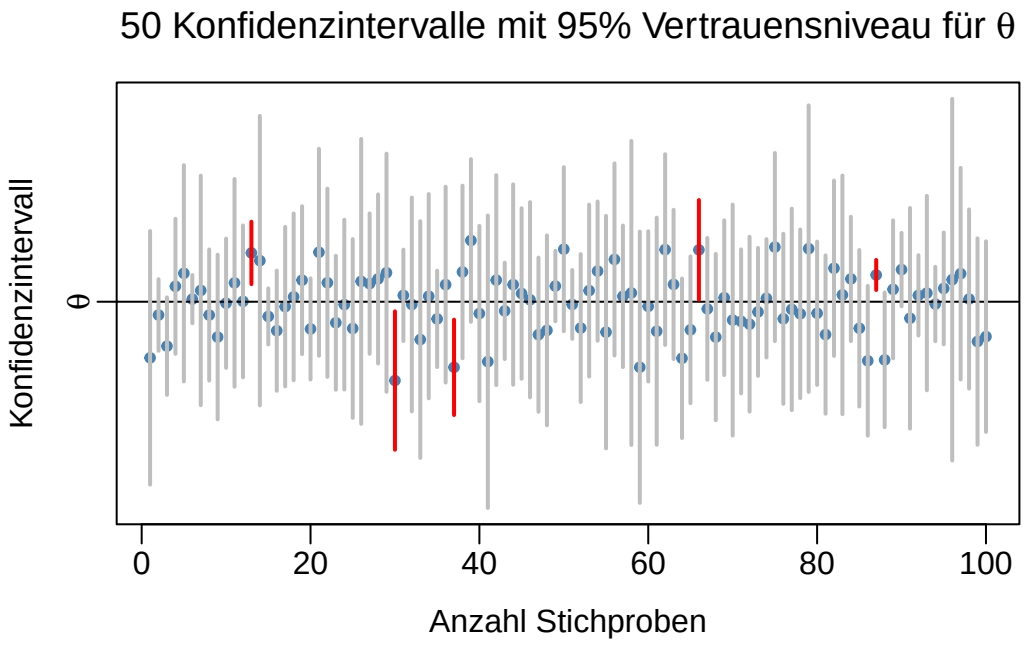
Wichtige Anmerkungen zu Konfidenzintervallen:

- Ein Konfidenzintervall garantiert niemals mit absoluter Sicherheit, dass der Parameter darin enthalten ist.
- Man kann nicht sagen, die Wahrscheinlichkeit, dass der Parameter im Intervall liegt, betrage $1-\alpha$ - denn nach Berechnung des Intervalls haben die Zufallsvariablen feste Werte angenommen.
- Die korrekte Interpretation ist: $(1-\alpha)$ aller möglichen Stichprobenintervalle enthalten den wahren Parameter.
- Daher spricht man von Konfidenz (Vertrauen), nicht von Wahrscheinlichkeit.

Typische Konfidenzniveaus in der Praxis:

$$\begin{aligned} 1-\alpha &= 0.90 \quad (\alpha = 0.10) \\ 1-\alpha &= 0.95 \quad (\alpha = 0.05) \quad \text{(am häufigsten verwendet)} \\ 1-\alpha &= 0.99 \quad (\alpha = 0.01) \quad \text{(für kritische Anwendungen)} \end{aligned}$$

Theoretisch enthalten bei einem Konfidenzniveau von $1-\alpha = 0,95$ etwa 95 von 100 berechneten Intervallen den wahren Parameter θ, während etwa 5 Intervalle ihn nicht enthalten.



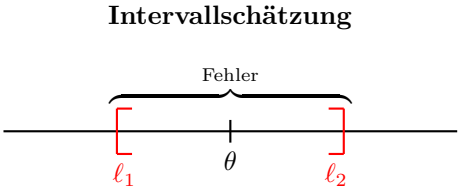
8.4.1 Schätzfehler

Ein weiterer zentraler Aspekt von Konfidenzintervallen ist ihre Fehlerspanne.

Definition "Fehlerspanne"

Die Fehlerspanne oder Ungenauigkeit eines Konfidenzintervalls $[l_i, l_s]$ wird durch seine Breite bestimmt:

$$A = l_s - l_i.$$



Damit ein Intervall praktisch nutzbar ist, darf diese Ungenauigkeit nicht zu groß sein.

Grundsätzlich hängt die Präzision eines Intervalls von drei Faktoren ab:

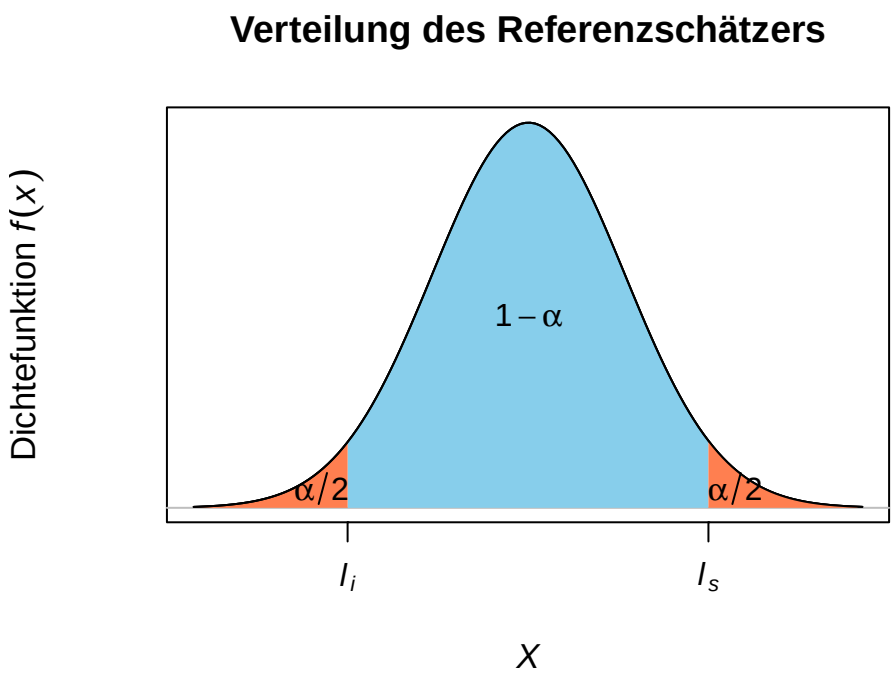
1. Streuung der Grundgesamtheit: Je größer die Streuung, desto ungenauer das Intervall.
2. Konfidenzniveau: Höhere Konfidenzniveaus führen zu weniger präzisen Intervallen.
3. Stichprobenumfang: Größere Stichproben erhöhen die Präzision.

   Wenn Konfidenz und Präzision im Konflikt stehen - wie kann man dann Präzision gewinnen ohne Konfidenz zu verlieren?

In der Praxis geht man typischerweise wie folgt vor:

- Man beginnt mit einem Punktschätzer, dessen Stichprobenverteilung bekannt ist.
- Auf Basis dieser Verteilung bestimmt man die Intervallgrenzen so, dass sie eine Wahrscheinlichkeit von $1 - \alpha$ einschließen.
- Üblicherweise wählt man die Grenzen symmetrisch:

- Die untere Grenze schließt $\alpha/2$ der Wahrscheinlichkeit aus
- Die obere Grenze schließt ebenfalls $\alpha/2$ aus



8.5 Konfidenzintervalle für eine Grundgesamtheit

Im Folgenden werden Konfidenzintervalle zur Schätzung von Parametern einer Grundgesamtheit dargestellt:

- Konfidenzintervall für den Mittelwert einer normalverteilten Grundgesamtheit mit bekannter Varianz
- Konfidenzintervall für den Mittelwert einer normalverteilten Grundgesamtheit mit unbekannter Varianz
- Konfidenzintervall für den Mittelwert einer Grundgesamtheit mit unbekannter Varianz bei großen Stichproben
- Konfidenzintervall für die Varianz einer normalverteilten Grundgesamtheit
- Konfidenzintervall für einen Anteilswert einer Grundgesamtheit

8.5.1 Konfidenzintervall für den Mittelwert einer normalverteilten Grundgesamtheit mit bekannter Varianz

Sei X eine Zufallsvariable, die folgende Annahmen erfüllt:

- Sie ist normalverteilt: $X \sim N(\mu, \sigma)$
- Der Mittelwert μ ist unbekannt, aber die Varianz σ^2 ist bekannt

Unter diesen Annahmen folgt der Stichprobenmittelwert für Stichproben des Umfangs n ebenfalls einer Normalverteilung:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

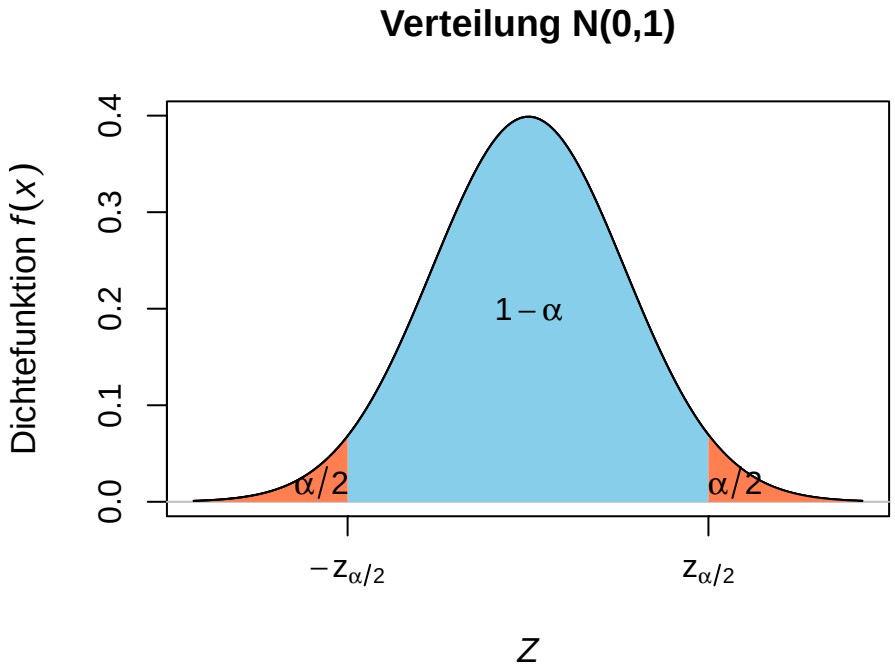
Durch Standardisierung der Variable erhält man:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Auf dieser Verteilung lassen sich einfach die Werte z_i und z_s berechnen, so dass gilt:

$$P(z_i \leq Z \leq z_s) = 1 - \alpha.$$

Da die Standardnormalverteilung symmetrisch um 0 ist, wählt man am besten entgegengesetzte Werte $-z_{\alpha/2}$ und $z_{\alpha/2}$, die jeweils $\alpha/2$ der Wahrscheinlichkeit in den Rändern lassen.



Ausgehend von der Standardisierung lässt sich durch Umformung einfach zu den Schätzern für die Grenzen des Konfidenzintervalls gelangen:

$$\begin{aligned}
1 - \alpha = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = \\
= P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = \\
= P\left(-\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = \\
= P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).
\end{aligned}$$

⚠ Konfidenzintervall für den Mittelwert einer Normalverteilung mit bekannter Varianz

Falls $X \sim N(\mu, \sigma)$ mit bekannter Standardabweichung σ, dann lautet das Konfidenzintervall für den Mittelwert μ zum Konfidenzniveau $1 - \alpha$:

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

alternativ geschrieben als:

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Aus der Intervallformel

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

ergeben sich folgende Charakteristika:

1. Das Intervall ist zentriert um den Stichprobenmittelwert $\bar{X}$, den besten Schätzer für den Populationsmittelwert.
2. Die Breite (Ungenauigkeit) des Intervalls beträgt:

$$A = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Diese hängt ab von:

- σ : Je größer die Populationsvarianz, desto ungenauer das Intervall
- $z_{\alpha/2:}$ Abhängig vom Konfidenzniveau - höhere Konfidenzniveaus $(1-\alpha)$ führen zu größeren Intervallen
- n : Größere Stichproben verringern die Ungenauigkeit

Die einzige Möglichkeit, die Intervallgenauigkeit bei gleichbleibendem Konfidenzniveau zu verbessern, besteht folglich in der Erhöhung des Stichprobenumfangs.

8.5.1.1 Berechnung des Stichprobenumfangs zur Schätzung des Mittelwerts einer Normalverteilung mit bekannter Varianz

Unter Berücksichtigung der Breite (Ungenauigkeit) des Konfidenzintervalls für den Mittelwert einer Normalverteilung mit bekannter Varianz:

$$A = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

kann der erforderliche Stichprobenumfang zur Erzielung einer Intervallbreite A mit Konfidenzniveau $1-\alpha$ leicht berechnet werden:

$$A = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \Leftrightarrow \sqrt{n} = 2z_{\alpha/2}\frac{\sigma}{A},$$

woraus folgt:

$$n = 4z_{\alpha/2}^2\frac{\sigma^2}{A^2}$$

Beispiel

Betrachten wir eine Studentenpopulation, deren Prüfungsergebnisse einer Normalverteilung $X \sim N(\mu, \sigma = 1.5)$ folgen.

Zur Schätzung der Durchschnittsnote μ wird eine Stichprobe von 10 Studenten gezogen:

$$4 - 6 - 8 - 7 - 7 - 6 - 5 - 2 - 5 - 3$$

Daraus berechnen wir das 95%-Konfidenzintervall ($\alpha = 0.05$) für μ:

- $\bar{X} = \frac{4+\cdots+3}{10} = \frac{53}{10} = 5.3$ Punkte
- $z_{\alpha/2} = z_{0.025} \approx 1.96$ (Standardnormalverteilung)

Einsetzen in die Intervallformel ergibt:

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 5.3 \pm 1.96\frac{1.5}{\sqrt{10}} = 5.3 \pm 0.93 = [4.37,\ 6.23]\,.$$

Mit 95%iger Sicherheit liegt μ somit zwischen 4.37 und 6.23 Punkten.

Beispiel

Die Ungenauigkeit dieses Intervalls beträgt ± 0.93 Punkte.
Für eine höhere Präzision von ± 0.5 Punkten berechnet sich der erforderliche Stichprobenumfang zu:

$$n = 4z_{\alpha/2}^2\frac{\sigma^2}{A^2} = 4 \cdot 1.96^2\frac{1.5^2}{(2 \cdot 0.5)^2} = 34.57.$$

Es wird daher eine Stichprobe von mindestens 35 Studenten benötigt, um ein 95%-Konfidenzintervall mit $\pm 0,5$ Punkten Genauigkeit zu erhalten.

8.5.2 Konfidenzintervall für den Mittelwert einer Normalverteilung mit unbekannter Varianz

Gegeben sei eine Zufallsvariable X, die folgende Annahmen erfüllt:

- Normalverteilung:$X \sim N(\mu, \sigma)$
- Sowohl der Mittelwert μ als auch die Varianz σ^2 sind unbekannt

Wenn die Populationsvarianz unbekannt ist, wird sie typischerweise durch die korrigierte Stichprobenvarianz $\hat{S}^2$ geschätzt. Dadurch folgt der Referenzschätzer nicht mehr einer Normalverteilung (wie bei bekannter Varianz), sondern einer t-Verteilung mit $n-1$ Freiheitsgraden:

$$\left.\begin{array}{l} \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \\ \dfrac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2(n-1) \end{array}\right\} \Rightarrow \frac{\bar{X}-\mu}{\hat{S}/\sqrt{n}} \sim T(n-1),$$

Da die t-Verteilung (wie die Normalverteilung) symmetrisch um 0 ist, können zwei entgegengesetzte Werte $-t_{\alpha/2}^{n-1}$ und $t_{\alpha/2}^{n-1}$ gewählt werden, sodass gilt:

$$
\begin{aligned}
1-\alpha &= P\left(-t_{\alpha/2}^{n-1} \leq \frac{\bar{X}-\mu}{\hat{S}/\sqrt{n}} \leq t_{\alpha/2}^{n-1}\right) \\
&= P\left(-t_{\alpha/2}^{n-1}\frac{\hat{S}}{\sqrt{n}} \leq \bar{X}-\mu \leq t_{\alpha/2}^{n-1}\frac{\hat{S}}{\sqrt{n}}\right) \\
&= P\left(\bar{X}-t_{\alpha/2}^{n-1}\frac{\hat{S}}{\sqrt{n}} \leq \mu \leq \bar{X}t_{\alpha/2}^{n-1}\frac{\hat{S}}{\sqrt{n}}\right)
\end{aligned}
$$

⚠ Konfidenzintervall für den Mittelwert einer Normalverteilung mit unbekannter Varianz

Falls $X \sim N(\mu, \sigma)$ mit unbekannter Standardabweichung σ, dann lautet das Konfidenzintervall für den Mittelwert μ zum Konfidenzniveau $1-\alpha$:

$$
\left[\bar{X}-t_{\alpha/2}^{n-1}\frac{\hat{S}}{\sqrt{n}}, \bar{X}+t_{\alpha/2}^{n-1}\frac{\hat{S}}{\sqrt{n}}\right]
$$

alternativ geschrieben als:

$$
\bar{X} \pm t_{\alpha/2}^{n-1}\frac{\hat{S}}{\sqrt{n}}
$$

8.5.2.1 Berechnung des Stichprobenumfangs zur Schätzung des Mittelwerts einer Normalverteilung mit unbekannter Varianz

Analog zum vorherigen Fall ergibt sich die Breite (Ungenauigkeit) des Konfidenzintervalls für den Mittelwert bei unbekannter Varianz zu:

$$
A = 2t_{\alpha/2}^{n-1}\frac{\hat{S}}{\sqrt{n}}
$$

Daraus lässt sich der benötigte Stichprobenumfang für eine gewünschte Intervallbreite A mit Konfidenzniveau $1-\alpha$ ableiten:

$$
A = 2t_{\alpha/2}^{n-1}\frac{\hat{S}}{\sqrt{n}} \Leftrightarrow \sqrt{n} = 2t_{\alpha/2}^{n-1}\frac{\hat{S}}{A},
$$

was zu folgender Formel führt:

$$
n = 4(t_{\alpha/2}^{n-1})^2\frac{\hat{S}^2}{A^2}
$$

Der entscheidende Unterschied zum Fall mit bekannter Varianz besteht darin, dass $\hat{S}$ bekannt sein muss. Daher wird typischerweise wie folgt vorgegangen:

- Eine kleine Vorstichprobe wird gezogen, um $\hat{S}$ zu schätzen
- Der t-Wert kann für größere Stichproben durch den z-Wert der Normalverteilung approximiert werden: $t_{\alpha/2}^{n-1} \approx z_{\alpha/2}$

Beispiel

Angenommen, im vorherigen Beispiel sei die Populationsvarianz der Prüfungsergebnisse unbekannt.
Mit derselben Stichprobe von 10 Studenten:

$$4 - 6 - 8 - 7 - 7 - 6 - 5 - 2 - 5 - 3$$

ergibt sich das 95%-Konfidenzintervall ($\alpha = 0.05$) für μ wie folgt:

- $\bar{X} = \frac{53}{10} = 5.3$ Punkte
- $\hat{S}^2 = \frac{\sum(x_i - \bar{x})^2}{9} = 3.5667$ und $\hat{S} = 1.8886$ Punkte
- $t_{\alpha/2}^{n-1} = t_{0.025}^{9} = 2.2622$ (t-Verteilung mit 9 Freiheitsgraden)

Das Konfidenzintervall berechnet sich zu:

$$\bar{X} \pm t_{\alpha/2}^{n-1} \frac{\hat{S}}{\sqrt{n}} = 5.3 \pm 2.2622 \frac{1.8886}{\sqrt{10}} = 5.3 \pm 1.351 = [3.949,\ 6.651]\,.$$

Beispiel

Die Ungenauigkeit des Intervalls ($\pm 1,8886$ Punkte) ist deutlich größer als bei bekannter Varianz - dies erklärt sich durch die zusätzliche Unsicherheit bei der Schätzung der Populationsvarianz.
Der benötigte Stichprobenumfang für eine Genauigkeit von $\pm 0,5$ Punkten beträgt:

$$n = 4(z_{\alpha/2})^2 \frac{\hat{S}^2}{A^2} = 4 \cdot 1.96^2 \frac{3.5667}{(2 \cdot 0.5)^2} = 54.81.$$

Somit sind mindestens 55 Studenten nötig, um bei unbekannter Varianz ein 95%-Konfidenzintervall mit $\pm 0,5$ Punkten Genauigkeit zu erhalten.

8.5.3 Konfidenzintervall für den Mittelwert einer nicht-normalverteilten Grundgesamtheit

Gegeben sei eine Zufallsvariable X mit folgenden Eigenschaften:

- Ihre Verteilung ist nicht normal
- Sowohl der Mittelwert μ als auch die Varianz σ^2 sind unbekannt

Bei nicht-normalverteilten Grundgesamtheiten ändern sich die Verteilungen der Referenzschätzer, sodass die bisherigen Intervalle nicht anwendbar sind.

Für große Stichproben ($n \geq 30$) gilt jedoch nach dem zentralen Grenzwertsatz, dass sich die Verteilung des Stichprobenmittelwerts einer Normalverteilung annähert:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Somit bleibt das zuvor beschriebene Intervall gültig.

Konfidenzintervall für den Mittelwert einer nicht-normalverteilten Grundgesamtheit bei großen Stichproben

Für eine nicht-normalverteilte Variable X mit $n \geq 30$ lautet das $(1 - \alpha)$-Konfidenzintervall für μ:

$$\bar{X} \pm t_{\alpha/2}^{n-1} \frac{\hat{S}}{\sqrt{n}}$$

8.5.4 Konfidenzintervall für die Varianz einer normalverteilten Grundgesamtheit

Voraussetzungen:

- X ist normalverteilt: $X \sim N(\mu, \sigma)$
- Sowohl μ als auch σ^2 sind unbekannt

Als Ausgangspunkt dient der Referenzschätzer:

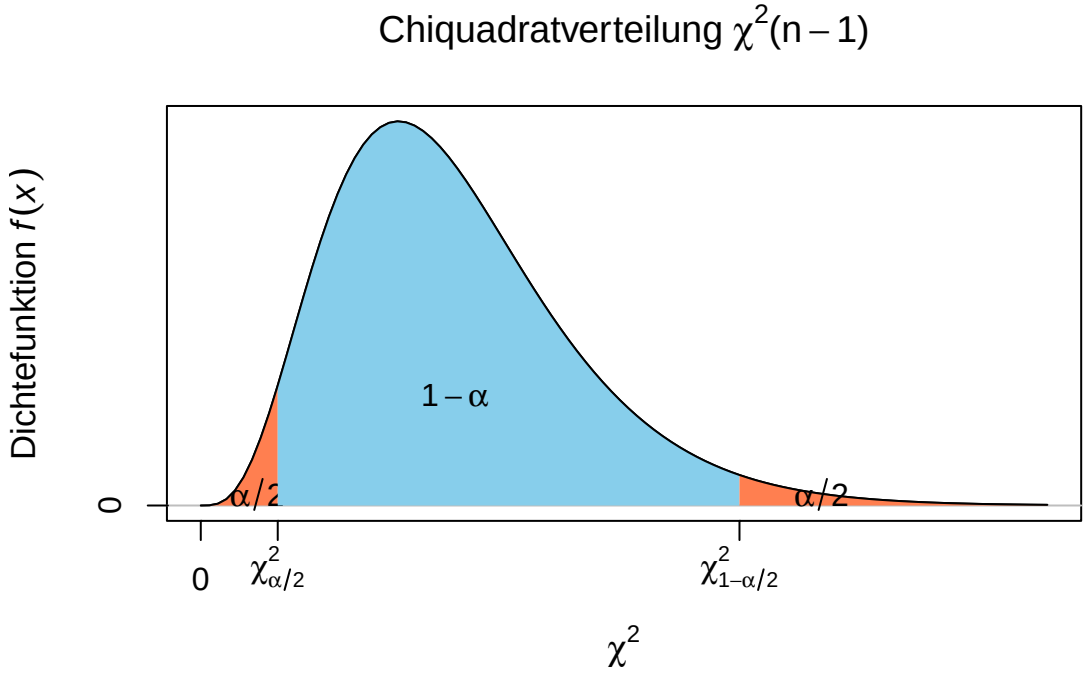
$$\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2(n-1),$$

der einer Chi-Quadrat-Verteilung mit $n-1$ Freiheitsgraden folgt.

Es müssen die Werte χ_i und χ_s bestimmt werden, für die gilt:

$$P(\chi_i \leq \chi^2(n-1) \leq \chi_s) = 1 - \alpha.$$

Da die Chi-Quadrat-Verteilung nicht symmetrisch ist, werden die Quantile $\chi_{\alpha/2}^{n-1}$ und $\chi_{1-\alpha/2}^{n-1}$ verwendet.



Somit ergibt sich

$$1-\alpha = P\left(\chi^{n-1}_{\alpha/2} \leq \frac{nS^2}{\sigma^2} \leq \chi^{n-1}_{1-\alpha/2}\right) = P\left(\frac{1}{\chi^{n-1}_{\alpha/2}} \geq \frac{\sigma^2}{nS^2} \geq \frac{1}{\chi^{n-1}_{1-\alpha/2}}\right) =$$
$$= P\left(\frac{1}{\chi^{n-1}_{1-\alpha/2}} \leq \frac{\sigma^2}{nS^2} \leq \frac{1}{\chi^{n-1}_{\alpha/2}}\right) = P\left(\frac{nS^2}{\chi^{n-1}_{1-\alpha/2}} \leq \sigma^2 \leq \frac{nS^2}{\chi^{n-1}_{\alpha/2}}\right).$$

Daher lautet das Konfidenzintervall für die Varianz einer normalverteilten Population:

Konfidenzintervall für die Varianz einer normalverteilten Grundgesamtheit

Für $X \sim N(\mu, \sigma)$ lautet das $(1-\alpha)$-Konfidenzintervall für σ :

$$\left[\frac{nS^2}{\chi^{n-1}_{1-\alpha/2}}, \frac{nS^2}{\chi^{n-1}_{\alpha/2}}\right]$$

Beispiel

Fortsetzung des Prüfungsbeispiels mit der Stichprobe:

$$4-6-8-7-7-6-5-2-5-3$$

Das Konfidenzintervall für σ^2 bei einem Vertrauensniveau von 95% ($\alpha = 0,05$) lautet:

- $S^2 = \frac{(4-5.3)^2+\cdots+(3-5.3)^2}{10} = 3.21$ Punkte^2.
- $\chi^{n-1}_{\alpha/2} = \chi^9_{0.025}$ ist der Wert der Chi-Quadrat-Verteilung mit 9 Freiheitsgraden, der eine kumulative Wahrscheinlichkeit von $0,025$ unterschreitet, und beträgt $2,7$.
- $\chi^{n-1}_{1-\alpha/2} = \chi^9_{0.975}$ ist der Wert der Chi-Quadrat-Verteilung mit 9 Freiheitsgraden, der eine kumulative Wahrscheinlichkeit von $0,975$ unterschreitet, und beträgt 19.

Das Intervall berechnet sich zu:

$$\left[\frac{nS^2}{\chi^{n-1}_{1-\alpha/2}}, \frac{nS^2}{\chi^{n-1}_{\alpha/2}}\right] = \left[\frac{10 \cdot 3.21}{19}, \frac{10 \cdot 3.21}{2.7}\right] = [1.69,\ 11.89] \text{ Punkte}^2.$$

8.5.5 Konfidenzintervall für einen Anteilswert

Zur Schätzung des Anteils p einer Grundgesamtheit mit einer bestimmten Eigenschaft geht man von der Zufallsvariablen X aus, die die Anzahl der Individuen mit dieser Eigenschaft in einer Stichprobe vom Umfang n zählt. Diese Variable folgt einer Binomialverteilung:

$$X \sim B(n, p)$$

Für ausreichend große Stichproben (genauer: wenn $np \geq 5$ und $n(1-p) \geq 5$) gilt) besagt der zentrale Grenzwertsatz, dass X approximativ normalverteilt ist:

$$X \sim N(np, \sqrt{np(1-p)}).$$

Folglich ist auch der Stichprobenanteil $\hat{p} = X/n$ approximativ normalverteilt:

$$\hat{p} = \frac{X}{n} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right),$$

Dies dient als Referenzschätzer. Durch Standardisierung erhält man:

$$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Nach der Standardisierung lassen sich – genauso wie zuvor – leicht die Werte $-z_{\alpha/2}$ und $z_{\alpha/2}$ finden, die erfüllen:

$$P\left(-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \leq z_{\alpha/2}\right) = 1-\alpha.$$

Somit erhält man, indem man die Standardisierung rückgängig macht und analog zum vorherigen Vorgehen argumentiert:

$$\begin{aligned}
1-\alpha &= P\left(-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \leq z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2}\frac{\sqrt{p(1-p)}}{n} \leq \hat{p}-p \leq z_{\alpha/2}\frac{\sqrt{p(1-p)}}{n}\right) \\
&= P\left(\hat{p}-z_{\alpha/2}\frac{\sqrt{p(1-p)}}{n} \leq p \leq \hat{p}+z_{\alpha/2}\frac{\sqrt{p(1-p)}}{n}\right)
\end{aligned}$$

Durch Umformung ergibt sich das Konfidenzintervall:

Konfidenzintervall für einen Anteilswert

Für $X \sim B(n,p)$ mit $np \geq 5$ und $n(1-p) \geq 5$ lautet das $(1-\alpha)$-Konfidenzintervall für p:

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

oder alternativ:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

8.5.5.1 Berechnung des Stichprobenumfangs zur Schätzung eines Anteils

Die Breite oder Ungenauigkeit des Konfidenzintervalls für einen Anteil in einer Population beträgt

$$A = 2z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Daher lässt sich der erforderliche Stichprobenumfang leicht berechnen, um ein Intervall mit der Breite A und einem Konfidenzniveau von $1-\alpha$ zu erhalten:

$$A = 2z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Leftrightarrow A^2 = 4z^2_{\alpha/2}\frac{\hat{p}(1-\hat{p})}{n},$$

woraus sich ableiten lässt:

$$n = 4z^2_{\alpha/2}\frac{\hat{p}(1-\hat{p})}{A^2}$$

Für die Berechnung wird eine Schätzung des Anteils $\hat{p}$ benötigt. Daher wird üblicherweise eine kleine Vorstichprobe gezogen, um diesen Wert zu ermitteln. Im schlimmsten Fall, wenn keine Vorstichprobe verfügbar ist, kann $\hat{p} = 0.5$ angenommen werden.

Beispiel

Angenommen, es soll der Anteil der Raucher in einer bestimmten Population geschätzt werden. Dazu wird eine Stichprobe von 20 Personen gezogen, und es wird erfasst, ob sie rauchen (1) oder nicht (0):

$$0-1-1-0-0-0-1-0-0-1-0-0-0-1-1-0-1-1-0-0$$

Daraus ergibt sich:

- $\hat{p} = \frac{8}{20} = 0.4$. Somit gilt $np = 20 \cdot 0.4 = 8 \geq 5$ und $n(1-p) = 20 \cdot 0.6 = 12 \geq 5$.
- $z_{\alpha/2} = z_{0.025}$ ist der Wert der Standardnormalverteilung, der eine obere kumulative Wahrscheinlichkeit von 0.025 aufweist und ungefähr 1.96 beträgt.

Durch Einsetzen dieser Werte in die Formel für das Konfidenzintervall ergibt sich:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.4 \pm 1.96\sqrt{\frac{0.4 \cdot 0.6}{10}} = 0.4 \pm 0.3 = [0.1,\ 0.7]\,.$$

Das bedeutet, mit einer 95%igen Konfidenz liegt der wahre Anteil p zwischen 0.1 und 0.7.

Beispiel

Wie zu erkennen ist, beträgt die Ungenauigkeit des obigen Intervalls ± 0.215, was angesichts eines Anteils sehr groß ist.
Um präzise Konfidenzintervalle für Anteilsschätzungen zu erhalten, sind relativ große Stichprobenumfänge erforderlich. Wenn beispielsweise eine Genauigkeit von $\pm 0,05$ gewünscht wird, wäre der erforderliche Stichprobenumfang:

$$n = 4z^2_{\alpha/2}\frac{\hat{p}(1-\hat{p})}{A^2} = 4 \cdot 1.96^2\frac{0.4 \cdot 0.6}{(2 \cdot 0.05)^2} = 368.79.$$

Das heißt, es wären mindestens 369 Personen notwendig, um ein 95%-Konfidenzintervall für den Anteil mit einer Genauigkeit von $\pm 0,05$ zu erhalten.

8.6 Konfidenzintervalle für den Vergleich zweier Populationen

In vielen Studien besteht das eigentliche Ziel nicht darin, den Wert eines Parameters zu ermitteln, sondern ihn mit dem einer anderen Population zu vergleichen.

Beispielsweise könnte man untersuchen, ob ein bestimmter Parameter in der männlichen und der weiblichen Population denselben Wert aufweist.

In solchen Fällen geht es nicht primär um die separate Schätzung der beiden Parameter, sondern um eine Schätzung, die ihren Vergleich ermöglicht.

Es werden drei Fälle betrachtet:

- Vergleich von Mittelwerten: Schätzung der Differenz der Mittelwerte $\mu_1 - \mu_2$.
- Vergleich von Varianzen: Schätzung des Verhältnisses der Varianzen $\frac{\sigma_1^2}{\sigma_2^2}$.
- Vergleich von Anteilen: Schätzung der Differenz der Anteile $\hat{p}_1 - \hat{p}_2$.

Im Folgenden werden die folgenden Konfidenzintervalle für den Vergleich zweier Populationen vorgestellt:

- Intervall für die Differenz der Mittelwerte zweier normalverteilter Populationen mit bekannten Varianzen.
- Intervall für die Differenz der Mittelwerte zweier normalverteilter Populationen mit unbekannten, aber gleichen Varianzen.
- Intervall für die Differenz der Mittelwerte zweier normalverteilter Populationen mit unbekannten und ungleichen Varianzen.
- Intervall für das Verhältnis der Varianzen zweier normalverteilter Populationen.
- Intervall für die Differenz der Anteile zweier Populationen.

8.6.1 Konfidenzintervall für die Differenz der Mittelwerte normalverteilter Populationen mit bekannten Varianzen

Seien X_1 und X_2 zwei Zufallsvariablen, die folgende Annahmen erfüllen:

- Sie sind normalverteilt: $X_1 \sim N(\mu_1, \sigma_1)$ und $X_2 \sim N(\mu_2, \sigma_2)$.
- Ihre Mittelwerte μ_1 und μ_2 sind unbekannt, aber ihre Varianzen σ_1^2 und σ_2^2 sind bekannt.

Unter diesen Annahmen folgt die Differenz der Stichprobenmittelwerte aus zwei unabhängigen Stichproben der Größen n_1 bzw. n_2 einer Normalverteilung:

$$\left.\begin{array}{l} \bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1}{\sqrt{n_1}}\right) \\ \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2}{\sqrt{n_2}}\right) \end{array}\right\} \Rightarrow \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

Durch Standardisierung ergeben sich die Werte der Standardnormalverteilung $-z_{\alpha/2}$ und $z_{\alpha/2}$, die folgende Bedingung erfüllen:

$$P\left(-z_{\alpha/2} \leq \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq z_{\alpha/2}\right) = 1 - \alpha.$$

Durch Rückgängigmachen der Standardisierung erhält man:

$$\begin{aligned} 1 - \alpha &= P\left(-z_{\alpha/2} \leq \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq z_{\alpha/2}\right) \\ &= P\left(-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq (\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2) \leq z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \\ &= P\left(\bar{X}_1 - \bar{X}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \end{aligned}$$

Somit lautet das Konfidenzintervall für die Differenz der Mittelwerte:

Konfidenzintervall für die Differenz der Mittelwerte normalverteilter Populationen mit bekannten Varianzen

Falls $X_1 \sim N(\mu_1, \sigma_1)$ und $X_2 \sim N(\mu_2, \sigma_2)$ mit bekannten σ_1 und σ_2, dann ist das Konfidenzintervall für die Differenz der Mittelwerte $\mu_1 - \mu_2$ zum Konfidenzniveau $1 - \alpha$ gegeben durch:

$$\left[\bar{X}_1 - \bar{X}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{X}_1 - \bar{X}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

oder alternativ:

$$\bar{X}_1 - \bar{X}_2 \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

8.6.2 Konfidenzintervall für die Differenz der Mittelwerte zweier normalverteilter Populationen mit unbekannten, aber gleichen Varianzen

Seien X_1 und X_2 zwei Zufallsvariablen, die folgende Annahmen erfüllen:

- Sie sind normalverteilt: $X_1 \sim N(\mu_1, \sigma_1)$ und $X_2 \sim N(\mu_2, \sigma_2)$
- Ihre Mittelwerte μ_1 und μ_2 sind unbekannt und ihre Varianzen ebenfalls, aber sie sind gleich: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Wenn die Populationsvarianz unbekannt ist, kann sie aus Stichproben der Größen n_1 und n_2 beider Populationen mittels der gewichteten Stichprobenvarianz (gepoolte Varianz) geschätzt werden:

$$\hat{S}_p^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}.$$

Die Teststatistik folgt in diesem Fall einer t-Verteilung:

$$\left.\begin{array}{l} \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sigma\sqrt{\frac{n_1+n_2}{n_1 n_2}}\right) \\ \frac{n_1 S_1^2 + n_2 S_2^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2) \end{array}\right\} \Rightarrow \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\hat{S}_p\sqrt{\frac{n_1+n_2}{n_1 n_2}}} \sim T(n_1 + n_2 - 2).$$

Daraus lassen sich die kritischen Werte $-t^{n_1+n_2-2}\alpha/2$ und $t^{n_1+n_2-2}\alpha/2$ der t-Verteilung bestimmen, die folgende Bedingung erfüllen:

$$P\left(-t_{\alpha/2}^{n_1+n_2-2} \leq \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\hat{S}_p\sqrt{\frac{n_1+n_2}{n_1 n_2}}} \leq t_{\alpha/2}^{n_1+n_2-2}\right) = 1 - \alpha.$$

Und durch Rücktransformation ergibt sich

$$
\begin{aligned}
1-\alpha &= P\left(-t_{\alpha/2}^{n_1+n_2-2} \leq \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\hat{S}_p\sqrt{\frac{n_1+n_2}{n_1 n_2}}} \leq t_{\alpha/2}^{n_1+n_2-2}\right) \\
&= P\left(-t_{\alpha/2}^{n_1+n_2-2}\hat{S}_p\sqrt{\frac{n_1+n_2}{n_1 n_2}} \leq (\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2) \leq t_{\alpha/2}^{n_1+n_2-2}\hat{S}_p\sqrt{\frac{n_1+n_2}{n_1 n_2}}\right) \\
&= P\left(\bar{X}_1-\bar{X}_2-t_{\alpha/2}^{n_1+n_2-2}\hat{S}_p\sqrt{\frac{n_1+n_2}{n_1 n_2}} \leq \mu_1-\mu_2 \leq \bar{X}_1-\bar{X}_2+t_{\alpha/2}^{n_1+n_2-2}\hat{S}_p\sqrt{\frac{n_1+n_2}{n_1 n_2}}\right).
\end{aligned}
$$

Durch weitere Umformung erhält man das Konfidenzintervall:

⚠ Konfidenzintervall für die Differenz der Mittelwerte normalverteilter Populationen mit unbekannten, aber gleichen Varianzen

Wenn $X_1 \sim N(\mu_1, \sigma_1)$ und $X_2 \sim N(\mu_2, \sigma_2)$ mit unbekannten, aber gleichen Varianzen $\sigma_1 = \sigma_2$, dann ist das Konfidenzintervall für die Mittelwertdifferenz $\mu_1 - \mu_2$ zum Konfidenzniveau $1-\alpha$ gegeben durch:

$$
\left[\bar{X}_1-\bar{X}_2-t_{\alpha/2}^{n_1+n_2-2}\hat{S}_p\sqrt{\frac{n_1+n_2}{n_1 n_2}}, \bar{X}_1-\bar{X}_2+t_{\alpha/2}^{n_1+n_2-2}\hat{S}_p\sqrt{\frac{n_1+n_2}{n_1 n_2}}\right]
$$

oder alternativ:

$$
\bar{X}_1-\bar{X}_2 \pm t_{\alpha/2}^{n_1+n_2-2}\hat{S}_p\sqrt{\frac{n_1+n_2}{n_1 n_2}}
$$

Wenn $[l_i, l_s]$ ein Konfidenzintervall mit dem Niveau $1-\alpha$ für die Differenz der Mittelwerte $\mu_1 - \mu_2$ ist, dann gilt

$$
\mu_1 - \mu_2 \in [l_i, l_s]
$$

Interpretation des Konfidenzintervalls $[l_i, l_s]$ für $\mu_1 - \mu_2$:

- Wenn das Intervall vollständig negativ ist ($l_s < 0$), dann gilt mit $(1-\alpha)$ Konfidenz, dass $\mu_1 < \mu_2$
- Wenn das Intervall vollständig positiv ist ($l_i > 0$), dann gilt mit $(1-\alpha)$ Konfidenz, dass $\mu_1 > \mu_2$
- Wenn das Intervall die Null enthält, gibt es keinen statistisch signifikanten Unterschied zwischen den Mittelwerten

In den ersten beiden Fällen spricht man von statistisch signifikanten Unterschieden zwischen den Mittelwerten.

Beispiel

Vergleich der akademischen Leistungen zweier Schülergruppen (10 bzw. 12 Schüler) mit unterschiedlichen Lehrmethoden:

$$
\begin{aligned}
X_1 &: 4-6-8-7-7-6-5-2-5-3 \\
X_2 &: 8-9-5-3-8-7-8-6-8-7-5-7
\end{aligned}
$$

Unter der Annahme, dass beide Variablen dieselbe Varianz besitzen, ergibt sich

- $\bar{X}_1 = \frac{4+\cdots+3}{10} = 5.3$ und $\bar{X}_2 = \frac{8+\cdots+7}{12} = 6.75$ Punkte.
- $S_1^2 = \frac{4^2+\cdots+3^2}{10} - 5.3^2 = 3.21$ und $S_2^2 = \frac{8^2+\cdots+3^2}{12} - 6.75^2 = 2.6875$ Punkte^2.
- $\hat{S}_p^2 = \frac{10\cdot 3.21+12\cdot 2.6875}{10+12-2} = 3.2175$ Punkte^2, und $\hat{S}_p = 1.7937$.
- $t_{\alpha/2}^{n_1+n_2-2} = t_{0.025}^{20}$ ist der Wert der Student-t-Verteilung mit 20 Freiheitsgraden, der eine obere kumulierte Wahrscheinlichkeit von $0,025$ abbildet, und beträgt ungefähr $2,09$.

Das 95%-Konfidenzintervall für die Leistungsdifferenz beträgt:

$$5.3 - 6.75 \pm 2.086 \cdot 1.7937\sqrt{\frac{10+12}{10 \cdot 12}} = -1.45 \pm 1.6021 = [-3.0521,\ 0.1521] \text{ Punkte.}$$

Das heißt, die Differenz der durchschnittlichen Werte $\mu_1 - \mu_2$ liegt mit einer Konfidenz von 95% zwischen -3.0521 und 0.1521 Punkten.

Angesichts dieses Intervalls kann man folgern, dass, da das Intervall sowohl positive als auch negative Werte enthält und somit auch die 0 einschließt, nicht festgestellt werden kann, dass eine der Mittelwerte größer ist als die andere. Daher wird angenommen, dass sie gleich sind, und es kann nicht gesagt werden, dass es signifikante Unterschiede zwischen den Gruppen gibt.

8.6.3 Konfidenzintervall für die Differenz der Mittelwerte zweier normalverteilter Populationen mit unbekannten und ungleichen Varianzen

Seien X_1 und X_2 zwei Zufallsvariablen, die folgende Annahmen erfüllen:

- Ihre Verteilung ist normal: $X_1 \sim N(\mu_1, \sigma_1)$ und $X_2 \sim N(\mu_2, \sigma_2)$.
- Ihre Mittelwerte μ_1, μ_2 und Varianzen σ_1^2, σ_2^2 sind unbekannt, aber es gilt $\sigma_1^2 \neq \sigma_2^2$.

In diesem Fall folgt die Teststatistik einer t-Verteilung:

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}} \sim T(g),$$

wobei die Anzahl der Freiheitsgrade $g = n_1 + n_2 - 2 - \Delta$ beträgt, mit

$$\Delta = \frac{(\frac{n_2-1}{n_1}\hat{S}_1^2 - \frac{n_1-1}{n_2}\hat{S}_2^2)^2}{\frac{n_2-1}{n_1^2}\hat{S}_1^4 + \frac{n_1-1}{n_2^2}\hat{S}_2^4}.$$

Daraus lassen sich die kritischen Werte der t-Verteilung $-t^g\alpha/2$ und $t^g\alpha/2$ bestimmen, die folgende Bedingung erfüllen:

$$P\left(-t_{\alpha/2}^g \leq \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}} \leq t_{\alpha/2}^g\right) = 1 - \alpha.$$

Durch Umstellung erhält man:

$$
\begin{aligned}
1-\alpha &= P\left(-t^g_{\alpha/2} \leq \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}}} \leq t^g_{\alpha/2}\right) \\
&= P\left(-t^g_{\alpha/2}\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}} \leq (\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2) \leq t^g_{\alpha/2}\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}}\right) \\
&= P\left(\bar{X}_1-\bar{X}_2-t^g_{\alpha/2}\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}} \leq \mu_1-\mu_2 \leq \bar{X}_1-\bar{X}_2+t^g_{\alpha/2}\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}}\right)
\end{aligned}
$$

Somit ergibt sich das Konfidenzintervall für die Differenz der Mittelwerte:

⚠ Konfidenzintervall für die Differenz der Mittelwerte normalverteilter Populationen mit unbekannten und ungleichen Varianzen

Wenn $X_1 \sim N(\mu_1, \sigma_1)$ und $X_2 \sim N(\mu_2, \sigma_2)$ mit unbekannten und ungleichen Varianzen $\sigma_1 \neq \sigma_2$, dann ist das Konfidenzintervall für die Mittelwertdifferenz $\mu_1 - \mu_2$ zum Konfidenzniveau $1-\alpha$ gegeben durch:

$$
\left[\bar{X}_1-\bar{X}_2-t^g_{\alpha/2}\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}}, \bar{X}_1-\bar{X}_2-t^g_{\alpha/2}\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}}\right]
$$

oder alternativ:

$$
\bar{X}_1-\bar{X}_2 \pm t^g_{\alpha/2}\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}}
$$

Wie wir gesehen haben, gibt es zwei mögliche Konfidenzintervalle zur Schätzung der Mittelwertdifferenz: eines für den Fall gleicher Populationsvarianzen und eines für ungleiche Varianzen.

Doch wenn die Populationsvarianzen unbekannt sind,

> wie lässt sich entscheiden, welches Intervall verwendet werden soll?

Die Antwort liefert das nächste Konfidenzintervall, das wir betrachten werden. Es ermöglicht die Schätzung des Varianzverhältnisses $\frac{\sigma_2^2}{\sigma_1^2}$ und somit deren Vergleich.

Daher ist es notwendig, bevor man das Konfidenzintervall für den Mittelwertvergleich berechnet (wenn die Populationsvarianzen unbekannt sind), zuerst das Konfidenzintervall für das Varianzverhältnis zu bestimmen. Auf Grundlage dieses Intervalls kann dann die geeignete Methode für den Mittelwertvergleich ausgewählt werden.

8.6.4 Konfidenzintervall für das Verhältnis von Varianzen

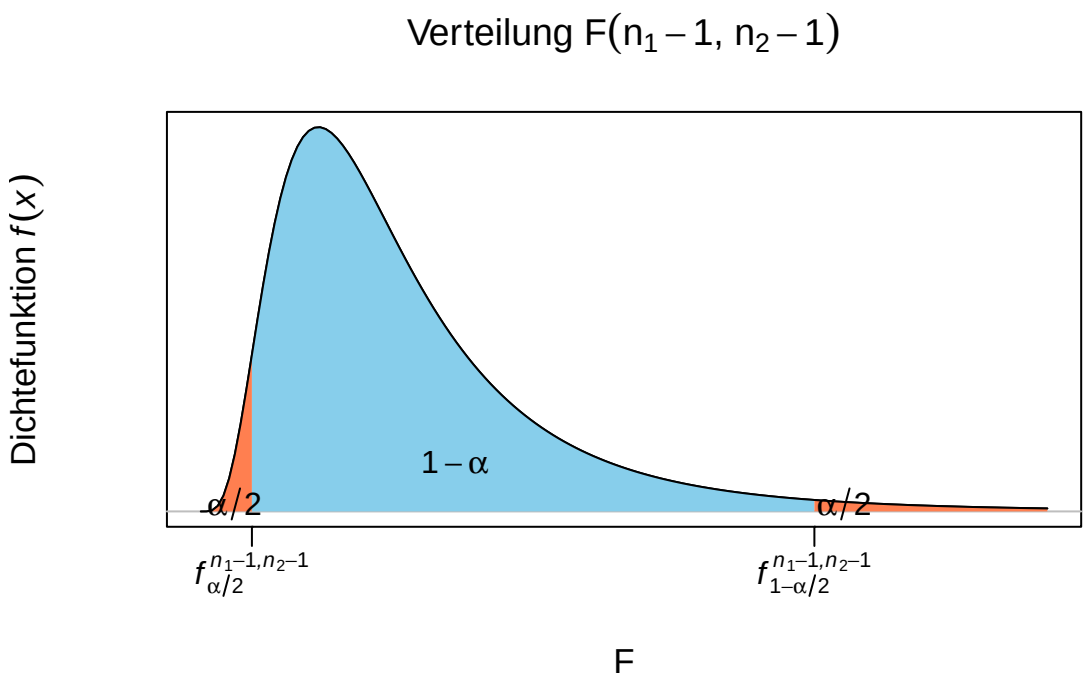
Seien X_1 und X_2 zwei Zufallsvariablen, die folgende Annahmen erfüllen:

- Ihre Verteilung ist normal: $X_1 \sim N(\mu_1, \sigma_1)$ und $X_2 \sim N(\mu_2, \sigma_2)$.
- Ihre Mittelwerte μ_1, μ_2 und Varianzen σ_1^2, σ_2^2 sind unbekannt.

In diesem Fall folgt die Teststatistik für Stichproben der Größen n_1 und n_2 einer F-Verteilung nach Fisher-Snedecor:

$$\left.\begin{array}{l}\frac{(n_1-1)\hat{S}_1^2}{\sigma_1^2} \sim \chi^2(n_1-1) \\ \frac{(n_2-1)\hat{S}_2^2}{\sigma_2^2} \sim \chi^2(n_2-1)\end{array}\right\} \Rightarrow \frac{\frac{(n_2-1)\hat{S}_2^2}{\sigma_2^2}}{\frac{(n_1-1)\hat{S}_1^2}{\sigma_1^2}} = \frac{\sigma_1^2}{\sigma_2^2}\frac{\hat{S}_2^2}{\hat{S}_1^2} \sim F(n_2-1, n_1-1).$$

Da die F-Verteilung nicht symmetrisch ist, werden zwei kritische Werte $f^{n_2-1,n_1-1}\alpha/2$ und $f^{n_2-1,n_1-1}1-\alpha/2$ verwendet, die jeweils untere Wahrscheinlichkeiten von $\alpha/2$ und $1-\alpha/2$ begrenzen.



Somit ergibt sich:

$$1-\alpha = P\left(f_{\alpha/2}^{n_2-1,n_1-1} \leq \frac{\sigma_1^2}{\sigma_2^2}\frac{\hat{S}_2^2}{\hat{S}_1^2} \leq f_{1-\alpha/2}^{n_2-1,n_1-1}\right) =$$
$$= P\left(f_{\alpha/2}^{n_2-1,n_1-1}\frac{\hat{S}_1^2}{\hat{S}_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq f_{1-\alpha/2}^{n_2-1,n_1-1}\frac{\hat{S}_1^2}{\hat{S}_2^2}\right)$$

Daher lautet das Konfidenzintervall für den Vergleich von Varianzen zweier normalverteilter Populationen:

Konfidenzintervall für das Verhältnis von Varianzen normalverteilter Populationen

Wenn $X_1 \sim N(\mu_1, \sigma_1)$ und $X_2 \sim N(\mu_2, \sigma_2)$, dann ist das Konfidenzintervall für das Varianzverhältnis σ_1/σ_2 zum Konfidenzniveau $1-\alpha$ gegeben durch:

$$\left[f_{\alpha/2}^{n_2-1,n_1-1}\frac{\hat{S}_1^2}{\hat{S}_2^2}, f_{1-\alpha/2}^{n_2-1,n_1-1}\frac{\hat{S}_1^2}{\hat{S}_2^2}\right]$$

Das Verhältnis der Varianzen $\frac{\sigma_1^2}{\sigma_2^2}$ liegt mit einer Konfidenz von $1-\alpha$ im Intervall $[l_i, l_s]$.

$$\frac{\sigma_1^2}{\sigma_2^2} \in [l_i, l_s]$$

Interpretation des Konfidenzintervalls $[l_i, l_s]$ für $\frac{\sigma_1^2}{\sigma_2^2}$:

- Wenn alle Werte des Intervalls kleiner als 1 sind $(l_s < 1)$, kann geschlossen werden, dass $\sigma_1^2 < \sigma_2^2$
- Wenn alle Werte des Intervalls größer als 1 sind $(l_i > 1)$, kann geschlossen werden, dass $\sigma_1^2 > \sigma_2^2$
- Wenn das Intervall sowohl Werte unter als auch über 1 enthält (also 1 im Intervall liegt), kann keine Aussage über die Größenverhältnisse der Varianzen gemacht werden. In diesem Fall wird normalerweise die Hypothese gleicher Varianzen $\sigma_1^2 = \sigma_2^2$ angenommen.

Beispiel

Fortsetzung des Beispiels mit den Punktzahlen zweier Gruppen:

$$X_1 : 4 - 6 - 8 - 7 - 7 - 6 - 5 - 2 - 5 - 3$$
$$X_2 : 8 - 9 - 5 - 3 - 8 - 7 - 8 - 6 - 8 - 7 - 5 - 7$$

Zur Berechnung des Konfidenzintervalls für das Varianzverhältnis bei einem Konfidenzniveau von 95% gilt:

- $\bar{X}_1 = 5.3$ Punkte und $\bar{X}_2 = 6.75$ Punkte
- $\hat{S}_1^2 = 3.5667$ Punkte^2 und $\hat{S}_2^2 = 2.9318$ Punkte^2
- $f_{0.025}^{11,9} \approx 0.2787$ (unterer kritischer F-Wert)
- $f_{0.975}^{11,9} \approx 3.9121$ (oberer kritischer F-Wert)

Einsetzen in die Formel ergibt das Intervall:

$$\left[0.2787\frac{3.5667}{2.9318},\ 3.9121\frac{3.5667}{2.9318}\right] = [0.3391,\ 4.7591] \text{ Punkte}^2.$$

Das bedeutet, das Varianzverhältnis $\frac{\sigma_1^2}{\sigma_2^2}$ liegt mit 95% Konfidenz zwischen $0,3391$ und $4,7591$.

Da das Intervall sowohl Werte unter als auch über 1 enthält, kann keine signifikante Unterschiedlichkeit der Varianzen angenommen werden. Für den Mittelwertvergleich wäre daher das Verfahren mit der Annahme gleicher Varianzen zu verwenden - genau die Methode, die wir bereits zuvor angewendet haben.

8.6.5 Konfidenzintervall für die Differenz von Proportionen

Um die Proportionen p_1 und p_2 von Individuen, die eine bestimmte Eigenschaft in zwei unabhängigen Populationen aufweisen, zu vergleichen, wird ihre Differenz $p_1 - p_2$ geschätzt.

Wenn aus jeder Population eine Stichprobe mit den Umfängen n_1 und n_2 gezogen wird, folgen die Variablen, die die Anzahl der Individuen mit der Eigenschaft in jeder Stichprobe messen, den Verteilungen

$$X_1 \sim B(n_1, p_1) \quad \text{y} \quad X_2 \sim B(n_2, p_2)$$

Wenn die Stichprobenumfänge groß sind (tatsächlich reicht es aus, dass $n_1 p_1 \geq 5$, $n_1(1-p_1) \geq 5$, $n_2 p_2 \geq 5$ und $n_2(1-p_2) \geq 5$ erfüllt sind), sichert der zentrale Grenzwertsatz, dass X_1 und X_2 Normalverteilungen folgen

$$X_1 \sim N(n_1 p_1, \sqrt{n_1 p_1 (1-p_1)}) \quad \text{und} \quad X_2 \sim N(n_2 p_2, \sqrt{n_2 p_2 (1-p_2)}),$$

und die Stichprobenproportionen

$$\hat{p}_1 = \frac{X_1}{n_1} \sim N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}}\right) \quad \text{und} \quad \hat{p}_2 = \frac{X_2}{n_2} \sim N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{n_2}}\right)$$

Aus den Stichprobenproportionen wird der Referenzschätzer konstruiert

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right).$$

Durch Standardisierung werden Werte $-z_{\alpha/2}$ und $z_{\alpha/2}$ gesucht, die erfüllen

$$P\left(-z_{\alpha/2} \leq \frac{(\hat{p}_1 - \hat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \leq z_{\alpha/2}\right) = 1 - \alpha.$$

Und durch Rückgängigmachen der Standardisierung ergibt sich

$$\begin{aligned}
1 - \alpha &= P\left(-z_{\alpha/2} \leq \frac{(\hat{p}_1 - \hat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \leq z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \leq (\hat{p}_1 - \hat{p_2}) - (p_1 - p_2) \leq z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right) \\
&= P\left(\hat{p}_1 - \hat{p_2} - z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \leq \hat{p}_1 - \hat{p_2} + p_1 - p_2 \leq z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)
\end{aligned}$$

Somit ist das Konfidenzintervall für die Differenz der Proportionen

Konfidenzintervall für die Differenz von Proportionen

Wenn $X_1 \sim B(n_1, p_1)$ und $X_2 \sim B(n_2, p_2)$, mit $n_1 p_1 \geq 5$, $n_1(1-p_1) \geq 5$, $n_2 p_2 \geq 5$ und $n_2(1-p_2) \geq 5$, dann ist das Konfidenzintervall für die Differenz der Proportionen $p_1 - p_2$ mit dem Konfidenzniveau $1 - \alpha$

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Beispiel

Angenommen, es sollen die Proportionen oder Prozentsätze der bestandenen Prüfungen in zwei Gruppen verglichen werden, die unterschiedliche Methoden verfolgt haben. In der ersten Gruppe haben 24 von 40 Schülern bestanden, während in der zweiten Gruppe 48 von 60 Schülern bestanden haben.
Um das Konfidenzintervall für die Differenz der Proportionen mit einem Konfidenzniveau von 95% zu berechnen, gilt:

- $\hat{p}_1 = 24/40 = 0.6$ und
- $\hat{p}_2 = 48/60 = 0.8$, sodass die Hypothesen
  - $n_1\hat{p}_1 = 40 \cdot 0.6 = 24 \geq 5$,

  - $n_1(1-\hat{p}_1) = 40(1-0.6) = 26 \geq 5$,
  - $n_2\hat{p}_2 = 60 \cdot 0.8 = 48 \geq 5$ und
  - $n_2(1-\hat{p}_2) = 60(1-0.8) = 12 \geq 5$ erfüllt sind.
- $z_{\alpha/2} = z_{0.025} = 1.96$.

Durch Einsetzen in die Formel des Intervalls ergibt sich

$$0.6 - 0.8 \pm 1.96\sqrt{\frac{0.6(1-0.6)}{40} + \frac{0.8(1-0.8)}{60}} = -0.2 \pm 0.17 = [-0.37,\ -0.03].$$

Da das Intervall negativ ist, gilt $p_1 - p_2 < 0 \Rightarrow p_1 < p_2$, und es kann geschlossen werden, dass es signifikante Unterschiede in den Bestehensquoten gibt.

9 Parametrische Hypothesentests

9.1 Statistische Hypothese und Arten von Tests

In vielen statistischen Studien besteht das Ziel eher darin, die Richtigkeit einer über die untersuchte Population aufgestellten Hypothese zu überprüfen, als den Wert eines unbekannten Parameters in der Population zu schätzen.

Der Forscher hat aufgrund seiner Erfahrung oder früherer Studien oft Vermutungen über die untersuchte Population, die er in Form von Hypothesen formuliert.

⚠ Definition "Statistische Hypothese"

Eine *statistische* Hypothese ist jede Aussage oder Vermutung, die die Verteilung einer oder mehrerer Variablen in der Population ganz oder teilweise bestimmt.

💡 Beispiel

Um die akademische Leistung einer Studierendengruppe in einem bestimmten Fach zu überprüfen, könnten wir die Hypothese aufstellen, ob die Bestehensquote über 50% liegt.

9.1.1 Hypothesentest

Im Allgemeinen wird man nie mit absoluter Sicherheit wissen, ob eine statistische Hypothese wahr oder falsch ist, da hierfür alle Individuen der Population untersucht werden müssten.

Um die Richtigkeit oder Falschheit dieser Hypothesen zu überprüfen, müssen sie mit den empirischen Ergebnissen der Stichproben verglichen werden. Wenn die beobachteten Ergebnisse der Stichproben – innerhalb der durch den Zufall bedingten Fehlertoleranz – mit dem übereinstimmen, was bei Gültigkeit der Hypothese zu erwarten wäre, wird die Hypothese als wahr akzeptiert. Andernfalls wird sie als falsch verworfen, und es werden neue Hypothesen gesucht, die die beobachteten Daten erklären können.

Da Stichproben zufällig gezogen werden, wird die *Entscheidung*, eine statistische Hypothese anzunehmen oder abzulehnen, auf probabilistischer Basis getroffen.

Die Methodik, die sich mit der Überprüfung der Richtigkeit statistischer Hypothesen befasst, wird als *Hypothesentest* bezeichnet.

9.1.2 Arten von Hypothesentests

- **Anpassungstests** (Goodness-of-Fit-Tests): Ziel ist die Überprüfung einer Hypothese über die Verteilungsform der Population. Beispiel: Testen, ob die Noten einer Schülergruppe einer Normalverteilung folgen.
- **Verträglichkeitstests** (Konformitätstests): Ziel ist die Überprüfung einer Hypothese über einen Parameter der Population. Beispiel: Testen, ob der Notendurchschnitt einer Schülergruppe gleich 2 ist.
- **Homogenitätstests**: Ziel ist der Vergleich zweier Populationen hinsichtlich bestimmter Parameter. Beispiel: Testen, ob die Leistung zweier Schülergruppen gleich ist, indem ihre Durchschnittsnoten verglichen werden.

- **Unabhängigkeitstests**: Ziel ist die Überprüfung, ob ein Zusammenhang zwischen zwei Variablen der Population besteht. Beispiel: Testen, ob ein Zusammenhang zwischen den Noten zweier unterschiedlicher Fächer besteht.

Wenn Hypothesen über Parameter der Population aufgestellt werden, spricht man auch von **parametrischen** Tests.

9.1.3 Nullhypothese und Alternativhypothese

In den meisten Fällen geht es bei einem Hypothesentest darum, eine Entscheidung zwischen zwei gegensätzlichen Hypothesen zu treffen:

- **Nullhypothese**: Dies ist die konservative Hypothese, die beibehalten wird, solange die Stichprobendaten ihre Falschheit nicht eindeutig belegen. Sie wird als H_0 bezeichnet.
- **Alternativhypothese**: Sie stellt die Verneinung der Nullhypothese dar und entspricht in der Regel der Aussage, die nachgewiesen werden soll. Sie wird als H_1 bezeichnet.

Die Wahl beider Hypothesen folgt dem Prinzip der wissenschaftlichen Einfachheit (Ockhams Rasiermesser):

> "Ein einfaches Modell sollte nur dann durch ein komplexeres ersetzt werden, wenn es starke Beweise für das komplexere Modell gibt."

Ein Beispiel aus der Rechtsprechung: Wenn ein Richter entscheiden muss, ob ein Angeklagter schuldig oder unschuldig ist, sollten die Hypothesen wie folgt lauten:

$$H_0 : \text{Unschuldig}$$
$$H_1 : \text{Schuldig}$$

Denn Unschuld wird angenommen, während Schuld bewiesen werden muss.

Entsprechend würde der Richter die Alternativhypothese nur dann akzeptieren, wenn es signifikante Beweise für die Schuld des Angeklagten gibt.

Der Forscher übernimmt in diesem Vergleich die Rolle des Anklägers, da sein Ziel darin besteht, die Nullhypothese zu widerlegen – also die Schuld des Angeklagten nachzuweisen.

> **Achtung**
>
> Diese Methodik begünstigt immer die Nullhypothese!

9.1.4 Parametrische Hypothesentests

Bei vielen Tests, insbesondere bei Konformitäts- und Homogenitätstests, beziehen sich die Hypothesen auf unbekannte Parameter der Grundgesamtheit wie den Mittelwert, die Varianz oder einen Anteilswert.

In diesem Fall weist die Nullhypothese dem Parameter stets einen konkreten Wert zu, während die Alternativhypothese meist eine offene Aussage darstellt, die zwar der Nullhypothese widerspricht, aber keinen festen Wert für den Parameter vorgibt.

Daraus ergeben sich drei Arten von Tests:

Zweiseitig	Einsitig (links)	Einseitig (rechts)
$H_0 : \theta = \theta_0$ $H_1 : \theta \neq \theta_0$	$H_0 : \theta = \theta_0$ $H_1 : \theta < \theta_0$	$H_0 : \theta = \theta_0$ $H_1 : \theta > \theta_0$

Beispiel

Angenommen, es besteht der Verdacht, dass in einer Population weniger Männer als Frauen vorhanden sind. Welche Art von Test sollte durchgeführt werden, um diesen Verdacht zu bestätigen oder zu widerlegen?

- Der Verdacht bezieht sich auf den Anteilswert p der Männer in der Population, es handelt sich also um einen parametrischen Test.
- Das Ziel ist die Überprüfung des Werts von p, daher handelt es sich um einen Konformitätstest. In der Nullhypothese wird p auf 0.5 festgesetzt, da gemäß den Gesetzen der Genetik in der Population gleich viele Männer wie Frauen zu erwarten wären.
- Schließlich besteht der Verdacht, dass der Männeranteil geringer ist als der Frauenanteil, daher lautet die Alternativhypothese $p < 0.5$.

Somit sollte folgender Test durchgeführt werden:

$$H_0 : p = 0,5$$
$$H_1 : p < 0,5$$

9.2 Methodik zur Durchführung eines Hypothesentests

9.2.1 Teststatistik

Die Annahme oder Ablehnung der Nullhypothese hängt letztlich von den Beobachtungen in der Stichprobe ab.

Die Entscheidung wird basierend auf dem Wert einer Stichprobenstatistik getroffen, die mit dem zu testenden Parameter oder Merkmal in Zusammenhang steht. Die Wahrscheinlichkeitsverteilung dieser Statistik muss unter der Annahme der Nullhypothese und bei festgelegtem Stichprobenumfang bekannt sein. Diese Statistik wird als **Teststatistik** bezeichnet.

Für jede Stichprobe liefert die Teststatistik eine Schätzung, auf deren Grundlage die Entscheidung getroffen wird: Wenn die Schätzung zu stark vom unter H_0 erwarteten Wert abweicht, wird die Nullhypothese verworfen, andernfalls wird sie beibehalten.

Die zugrundeliegende Logik folgt dem Prinzip, die Nullhypothese beizubehalten, es sei denn, die Stichprobe liefert eindeutige Beweise gegen sie. In Analogie zum Gerichtsverfahren würde dies bedeuten, die Unschuldsvermutung aufrechtzuerhalten, solange keine klaren Schuldbeweise vorliegen.

Beispiel

Kehren wir zum Beispiel des Tests über den Männeranteil in einer Population zurück:

$$H_0 : p = 0,5$$
$$H_1 : p < 0,5$$

Wenn für den Test eine Zufallsstichprobe von 10 Personen gezogen wird, könnte die Anzahl der Männer in der Stichprobe X als Teststatistik gewählt werden.
Unter der Annahme, dass die Nullhypothese zutrifft, folgt die Teststatistik einer Binomialverteilung $X \sim B(10,\ 0.5)$, sodass die erwartete Anzahl von Männern in der Stichprobe 5 beträgt.
Daher ist es logisch, die Nullhypothese zu akzeptieren, wenn die Stichprobe eine Anzahl von Männern in der Nähe von 5 ergibt, und sie zu verwerfen, wenn die Anzahl deutlich unter 5 liegt.
Doch wo genau sollte die Grenze zwischen den X-Werten gezogen werden, die zur Annahme bzw. zur Ablehnung führen?

9.2.2 Annahme- und Ablehnungsbereich

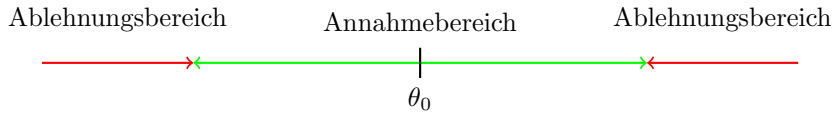
Nach Auswahl der Teststatistik gilt es festzulegen, für welche Werte dieser Statistik die Nullhypothese angenommen bzw. verworfen wird. Dadurch wird der Wertebereich der Statistik in zwei Regionen unterteilt:

- **Annahmebereich**: Die Menge der Werte der Teststatistik, für die die Nullhypothese angenommen wird.
- **Ablehnungsbereich**: Die Menge der Werte der Teststatistik, für die die Nullhypothese verworfen und folglich die Alternativhypothese angenommen wird.

Abhängig von der Art des Tests (einseitig/zweiseitig) befindet sich der Ablehnungsbereich entweder links, rechts oder auf beiden Seiten des unter der Nullhypothese erwarteten Werts der Teststatistik.

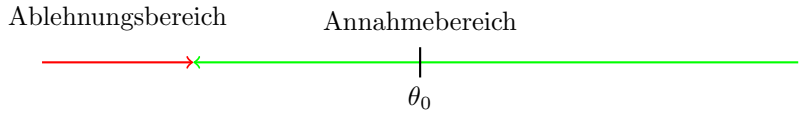
- Zweiseitiger Hypothesentest

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$



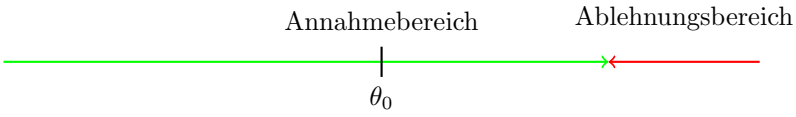
- Einseitiger Test (linksseitig)

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta < \theta_0$$



- Einseitiger Test (rechtsseitig)

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta > \theta_0$$

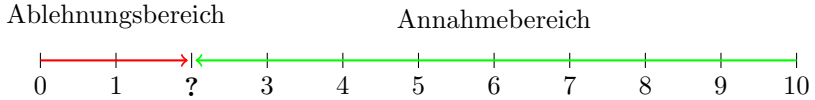


Beispiel

Im fortgesetzten Beispiel des Hypothesentests zum Bevölkerungsanteil von Männern

$$H_0 : p = 0,5$$
$$H_1 : p < 0,5$$

folgt die Teststatistik bei gültiger Nullhypothese einer $B(10, 0.5)$-Verteilung. Ihr Wertebereich reicht somit von 0 bis 10 bei einem Erwartungswert von 5. Da ein linksseitiger Test vorliegt ($H_1 : \ p < 0,5$), konzentriert sich der Ablehnungsbereich auf die Werte deutlich unter 5. Die entscheidende Frage bleibt jedoch:
An welchem kritischen Wert soll die Trennlinie zwischen Annahme und Verwerfung der Nullhypothese festgelegt werden?



9.2.3 Fehler bei einem Hypothesentest

Wir haben gesehen, dass ein Hypothesentest auf einer Entscheidungsregel basiert, die je nach Wert der Teststatistik die Annahme oder Ablehnung der Nullhypothese ermöglicht.

Letztlich führt der Test zu einer Entscheidung gemäß dieser Regel. Das Problem ist, dass man niemals mit absoluter Sicherheit weiß, ob eine Hypothese wahr oder falsch ist. Daher besteht bei jeder Entscheidung - ob Annahme oder Ablehnung - die Möglichkeit eines Fehlers.

Bei einem Hypothesentest können zwei Arten von Fehlern auftreten:

- **Fehler 1. Art** (α-Fehler): Tritt auf, wenn die Nullhypothese abgelehnt wird, obwohl sie wahr ist.
- **Fehler 2. Art** (β-Fehler): Tritt auf, wenn die Nullhypothese angenommen wird, obwohl sie falsch ist.

Entscheidung	H_0 gilt	H_1 gilt
H_0 annehmen	korrekte Entscheidung	Fehler 2. Art (β)
H_0 ablehnen	Fehler 1. Art (α)	korrekte Entscheidung

9.2.4 Fehlerrisiko von Hypothesentests

Die Risiken der beiden Fehlerarten werden durch Wahrscheinlichkeiten quantifiziert:

Definition "Alpha- und Betarisiko"

Bei einem Hypothesentest wird das α-Risiko definiert als die maximale Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, d.h.

$$P(\text{Ablehnung } H_0 | H_0) \leq \alpha,$$

und das β-Risiko als die maximale Wahrscheinlichkeit, einen Fehler 2. Art zu begehen, also

$$P(\text{Annahme } H_0 | H_1) \leq \beta.$$

Vorsicht

Da diese Methodik grundsätzlich die Nullhypothese begünstigt, gilt der Fehler 1. Art als gravierender als der Fehler 2. Art. Daher wird das α-Risiko üblicherweise auf niedrige Werte von $0,1$, $0,05$ oder $0,01$ festgelegt, wobei $0,05$ der gebräuchlichste Wert ist.

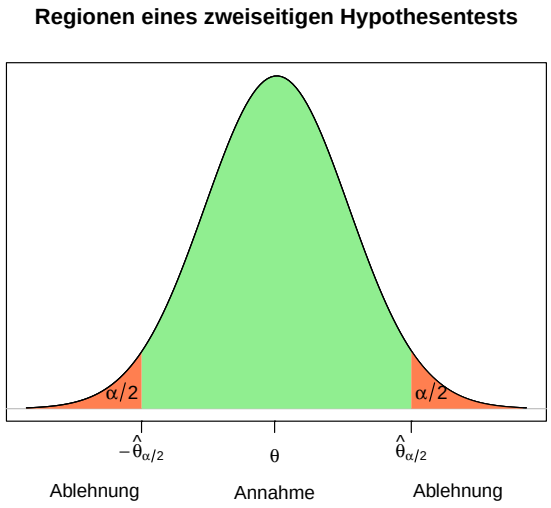
Bei der Interpretation des α-Risikos ist Vorsicht geboten, da es sich um eine bedingte Wahrscheinlichkeit unter der Annahme handelt, dass die Nullhypothese wahr ist. Wenn man die Nullhypothese mit einem α-Risiko von $0,05$ verwirft, ist die Aussage "in 5 von 100 Fällen liegen wir falsch" nur dann korrekt, wenn die Nullhypothese tatsächlich immer wahr wäre.

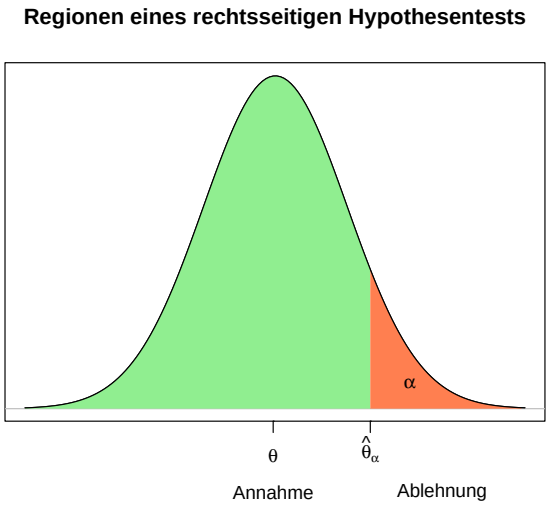
Ebenso wenig sinnvoll ist es, nach einer konkreten Testentscheidung von der "Wahrscheinlichkeit eines Fehlers" zu sprechen - denn für die getroffene Entscheidung gilt:

- Entweder man hat richtig entschieden (Fehlerwahrscheinlichkeit 0)
- oder falsch (Fehlerwahrscheinlichkeit 1).

9.2.5 Festlegung von Annahme- und Ablehnungsbereich basierend auf dem α-Risiko

Nach Festlegung des tolerierbaren α-Risikos können die Regionen für Annahme und Ablehnung der Nullhypothese so bestimmt werden, dass die kumulierte Wahrscheinlichkeit des Ablehnungsbereichs unter der Annahme der Nullhypothese genau α beträgt.





Beispiel: Fortsetzung des Tests zum Männeranteil in einer Population

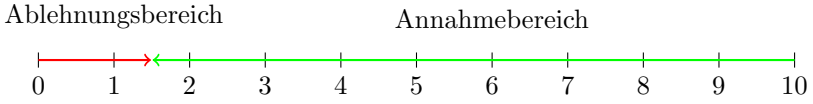
Wird die Nullhypothese verworfen, wenn höchstens 2 Männer in der Stichprobe sind, ergibt sich unter $X \sim B(10, 0.5)$ eine Fehlerwahrscheinlichkeit 1. Art von:

$$P(X \leq 2) = f(0) + f(1) + f(2) = 0.0010 + 0.0098 + 0.0439 = 0.0547.$$

Wenn die maximal tolerierbare Wahrscheinlichkeit für einen Fehler 1. Art auf $\alpha = 0,05$ festgelegt ist, für welche Werte der Teststatistik darf dann die Nullhypothese verworfen werden?

$$P(X \leq 1) = f(0) + f(1) = 0.0010 + 0.0098 = 0.0107.$$

Das bedeutet, die Nullhypothese könnte nur verworfen werden, wenn die Stichprobe 0 oder 1 Mann enthält.



9.2.6 Betarisiko und Effektgröße

Obwohl der Fehler 2. Art weniger gravierend erscheinen mag, ist es dennoch wünschenswert, das β-Risiko gering zu halten. Andernfalls wird es schwierig sein, die Nullhypothese zu verwerfen (was meist das Ziel ist), selbst wenn es deutliche Anzeichen für ihre Falschheit gibt.

Das Problem bei parametrischen Tests besteht darin, dass die Alternativhypothese eine offene Hypothese ist, die keinen festen Wert für den zu testenden Parameter vorgibt. Um das β-Risiko berechnen zu können, muss daher ein spezifischer Parameterwert angenommen werden.

Üblicherweise wird der Parameterwert auf die kleinste praktisch oder klinisch relevante Differenz festgelegt. Diese minimal als bedeutsam erachtete Differenz wird als **Effektgröße** bezeichnet und mit δ (Delta) symbolisiert.

9.2.7 Teststärke (Power) eines Hypothesentests

Da das Forschungsziel meist in der Ablehnung der Nullhypothese besteht, ist oft besonders interessant, wie gut ein Test in der Lage ist, die Falschheit der Nullhypothese zu erkennen, wenn tatsächlich eine Differenz größer als δ zwischen dem wahren Parameterwert und dem unter der Nullhypothese angenommenen Wert besteht.

Definition "Teststärke (Power)"

Die Teststärke (Power) eines Hypothesentests ist definiert als

$$\text{Teststärke} = P(\text{Ablehnung } H_0|H_1) = 1 - P(\text{Annahme } H_0|H_1) = 1 - \beta.$$

Somit führt eine Verringerung des β-Risikos zu einer Erhöhung der Teststärke.

Ein Test mit geringer Power ist meist wenig aussagekräftig, da er die Nullhypothese selbst dann nicht verwerfen kann, wenn es deutliche gegen sie sprechende Evidenzen gibt.

9.2.8 Berechnung des Betarisikos und der Teststärke $1 - \beta$

Angenommen, beim Test des Männeranteils wird eine Abweichung von weniger als 10% vom Wert der Nullhypothese als nicht signifikant betrachtet, d.h. $\delta = 0.1$.

Dies ermöglicht die Festlegung der Alternativhypothese:

$$H_1 : \ p = 0,5 - 0,1 = 0,4.$$

Unter der Annahme, dass diese Hypothese zutrifft, folgt die Teststatistik einer Binomialverteilung $X \sim B(10, , 0.4)$.

In diesem Fall beträgt das β-Risiko für die zuvor definierten Annahme- und Ablehnungsbereiche:

$$\beta = P(\text{Annahme } H_0|H_1) = P(X \geq 2) = 1 - P(X < 2) = 1 - 0,0464 = 0,9536.$$

Wie zu erkennen ist, handelt es sich um ein sehr hohes β-Risiko, sodass die Teststärke nur:

$$1 - \beta = 1 - 0,9536 = 0,0464$$

beträgt. Dies zeigt, dass dieser Test ungeeignet wäre, um Unterschiede von 10% im Parameterwert zu erkennen.

9.2.9 Zusammenhang zwischen β-Risiko und Effektgröße δ

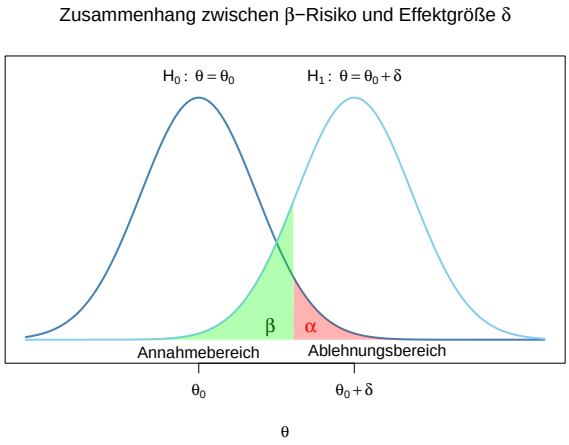
Das β-Risiko hängt direkt von der minimalen Differenz δ ab, die im Vergleich zum unter der Nullhypothese angenommenen Parameterwert erkannt werden soll.

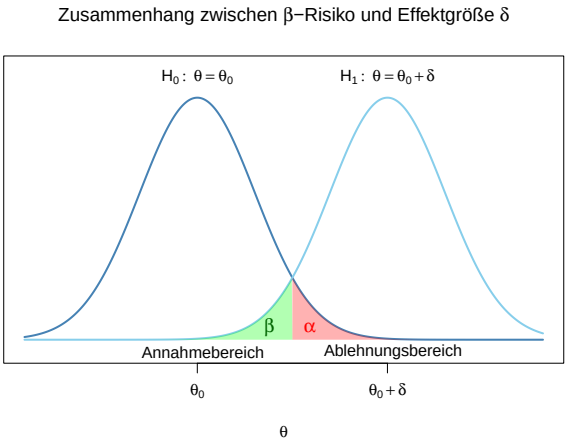
Beispiel

Wenn beim Test des Männeranteils eine Abweichung von mindestens 20% vom Wert der Nullhypothese erkannt werden soll, also $\delta = 0.2$, dann würde die Alternativhypothese auf

$$H_1 : \ p = 0.5 - 0.2 = 0.3$$

festgelegt werden.





Unter dieser Annahme folgt die Teststatistik einer Binomialverteilung $X \sim B(10, , 0.3)$.
In diesem Fall beträgt das β-Risiko für die zuvor definierten Annahme- und Ablehnungsbereiche:

$$\beta = P(\text{Annahme } H_0 | H_1) = P(X \geq 2) = 1 - P(X < 2) = 1 - 0,1493 = 0,8507$$

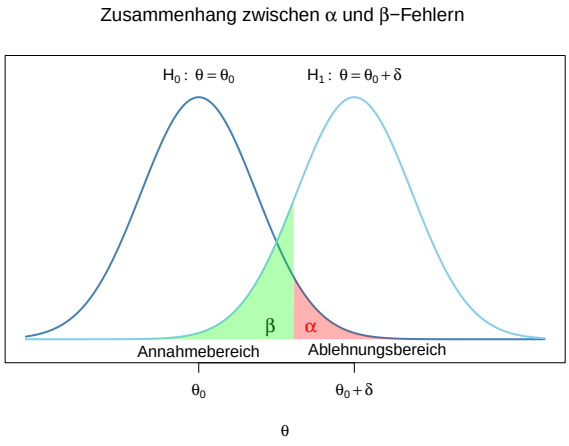
sodass das β-Risiko sinkt und die Teststärke steigt:

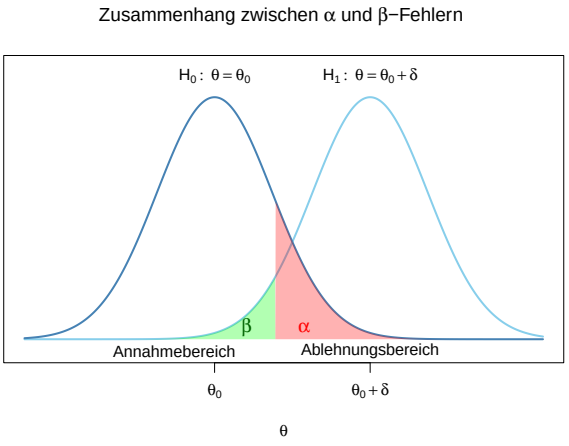
$$1 - \beta = 1 - 0,8507 = 0,1493$$

wobei der Test damit immer noch eine geringe Power aufweist.

9.2.10 Beziehung zwischen α- und β-Fehlern

Die Fehlerwahrscheinlichkeiten α und β stehen in einem Spannungsverhältnis: Wenn die eine steigt, sinkt in der Regel die andere – und umgekehrt.





Beispiel

Wenn beim Test des Männeranteils ein Risiko von $\alpha = 0.1$ festgelegt wird, dann wäre der Ablehnungsbereich $X \leq 2$, da unter der Annahme der Nullhypothese $X \sim B(10, 0.5)$ gilt und

$$P(X \leq 2) = 0.0547 \leq 0.1 = \alpha.$$

Für eine minimale Differenz von $\delta = 0.1$ und unter Annahme der Alternativhypothese $X \sim B(10, 0.4)$ beträgt das β-Risiko dann:

$$\beta = P(\text{Annahme } H_0 | H_1) = P(X \geq 3) = 1 - P(X < 3) = 1 - 0,1673 = 0,8327$$

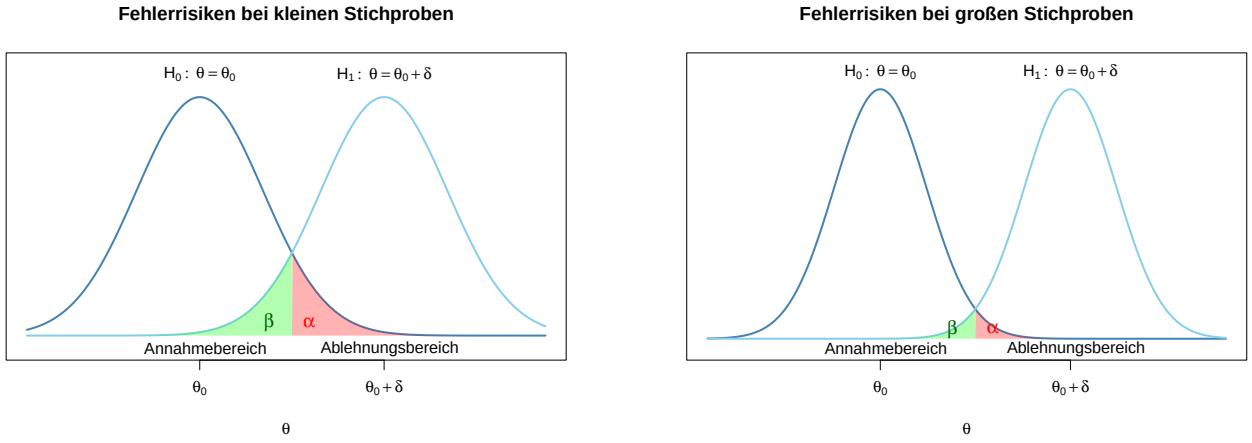
sodass die Teststärke nun auf

$$1 - \beta = 1 - 0,8327 = 0,1673$$

angestiegen ist.

9.2.11 Zusammenhang zwischen Fehlerrisiken und Stichprobenumfang

Die Fehlerrisiken hängen ebenfalls vom Stichprobenumfang ab, denn mit zunehmender Stichprobengröße verringert sich die Streuung der Teststatistik und damit auch die Fehlerrisiken.



Beispiel

Wenn für den Test des Männeranteils eine Stichprobe des Umfangs 100 (statt 10) verwendet worden wäre, würde die Teststatistik unter Gültigkeit der Nullhypothese einer Binomialverteilung $B(100, 0.5)$ folgen. In diesem Fall wäre der Ablehnungsbereich $X \leq 41$, da

$$P(X \leq 41) = 0,0443 \leq 0,05 = \alpha.$$

Für $\delta = 0.1$ und unter Annahme der Alternativhypothese $X \sim B(100, , 0.4)$ würde das β-Risiko dann

$$\beta = P(\text{Annahme } H_0 | H_1) = P(X \geq 42) = 0.3775$$

betragen, und die Teststärke hätte sich deutlich erhöht auf

$$1 - \beta = 1 - 0,3775 = 0,6225.$$

Dieser Test wäre wesentlich besser geeignet, um eine Abweichung von mindestens 10% vom unter der Nullhypothese angenommenen Parameterwert zu erkennen.

9.3 Power-Kurve

Die Teststärke (Power) eines Hypothesentests hängt vom unter der Alternativhypothese angenommenen Parameterwert ab und ist somit eine Funktion dieses Parameters:

$$\text{Power}(x) = P(\text{Ablehnung } H_0 | \theta = x).$$

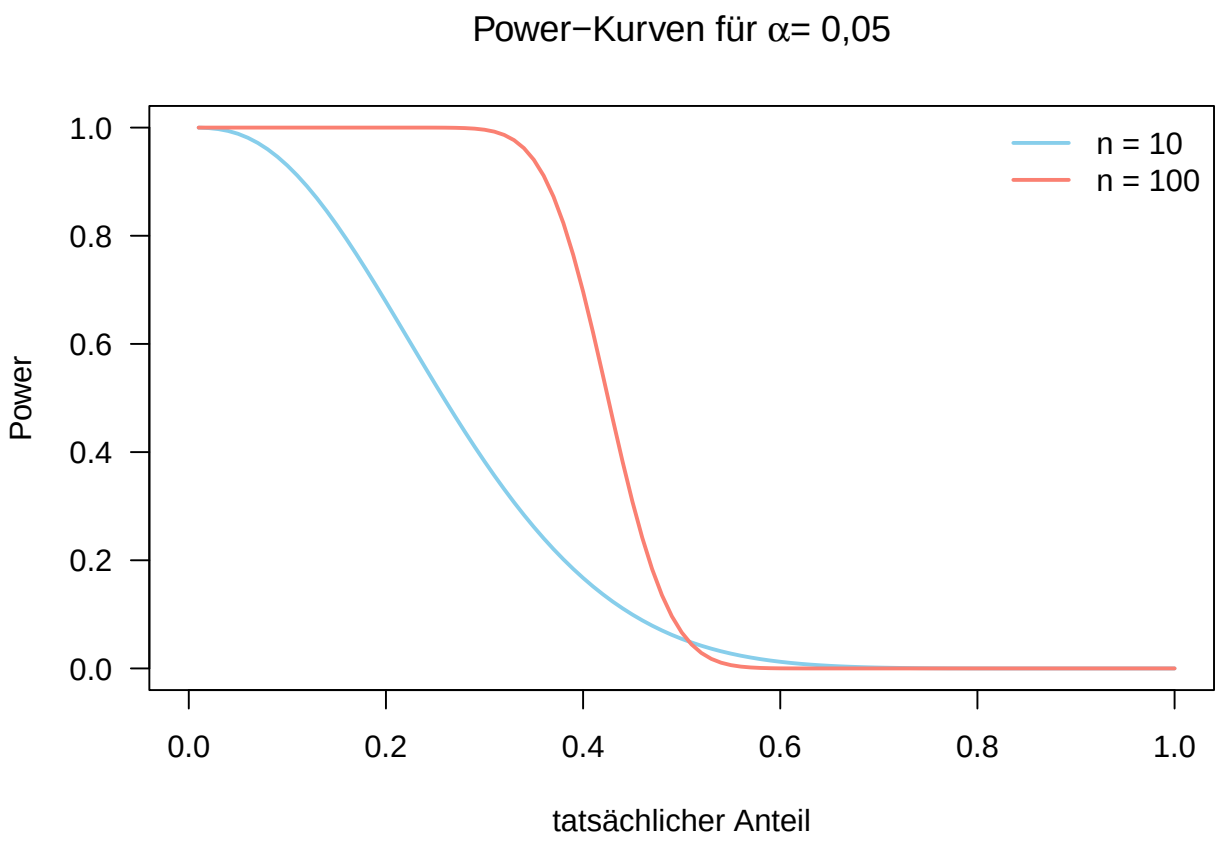
Diese Funktion gibt die Wahrscheinlichkeit an, die Nullhypothese für jeden möglichen Parameterwert zu verwerfen und wird als **Power-Kurve** oder **Trennschärfekurve** bezeichnet.

Wenn der genaue Parameterwert der Alternativhypothese nicht festgelegt werden kann, ist die Darstellung dieser Kurve besonders nützlich, um die Güte des Tests bei Nicht-Ablehnung der Nullhypothese zu beurteilen. Ebenso hilfreich ist sie, wenn nur eine begrenzte Stichprobengröße verfügbar ist, um die Sinnhaftigkeit der Untersuchung einzuschätzen.

Die Qualität eines Tests ist umso höher, je größer die Fläche unter der Power-Kurve ist.

Beispiel

Die zugehörige Power-Kurve des Tests zur Prüfung des Männeranteils in der Bevölkerung ist die folgende:



9.3.1 Der p-Wert eines Hypothesentests

Grundsätzlich wird die Nullhypothese verworfen, wenn die Schätzung der Teststatistik in den Ablehnungsbereich fällt. Jedoch haben wir deutlich mehr Vertrauen in die Ablehnung, wenn die Schätzung weit vom Annahmebereich entfernt liegt, als wenn sie nahe an der Grenze zwischen Annahme- und Ablehnungsbereich liegt.

Aus diesem Grund wird bei der Durchführung eines Tests zusätzlich die Wahrscheinlichkeit berechnet, mit der eine Abweichung auftritt, die mindestens so groß ist wie die beobachtete Differenz zwischen der Schätzung der Teststatistik und ihrem erwarteten Wert gemäß der Nullhypothese.

Definition "p-Wert"

In einem Hypothesentest wird für jede Schätzung x_0 der Teststatistik X - abhängig von der Art des Tests - der p-Wert definiert als:

$$\begin{array}{ll} \text{zweiseitiger Test :} & 2P(X \geq x_0|H_0) \\ \text{einseitiger Test (links) :} & P(X \leq x_0|H_0) \\ \text{einseitiger Test (links) :} & P(X \geq x_0|H_0) \end{array}$$

In gewisser Weise drückt der p-Wert das Vertrauen in die Entscheidung aus, die Nullhypothese abzulehnen. Je näher der p-Wert bei 1 liegt, desto größer ist das Vertrauen in die Annahme der Nullhypothese, und je näher er bei 0 liegt, desto größer ist das Vertrauen in ihre Ablehnung.

9.3.2 Entscheidungsregel eines Tests

Nach Festlegung des Risikos α kann die Entscheidungsregel für einen Test auch wie folgt formuliert werden:

Entscheidungsregel

$$\begin{array}{lcl} \text{Wenn } p \leq \alpha & \rightarrow & \text{Ablehnung von } H_0 \\ \text{Wenn } p > \alpha & \rightarrow & \text{Annehmen von } H_0. \end{array}$$

Somit gibt der p-Wert Auskunft darüber, bei welchen Signifikanzniveaus die Nullhypothese verworfen werden kann und bei welchen nicht.

Beispiel

Wenn beim Test des Männeranteils eine Stichprobe der Größe 10 gezogen wird und 1 Mann beobachtet wird, dann beträgt der p-Wert unter Annahme der Nullhypothese $X \sim B(10, , 0.5)$:

$$p = P(X \leq 1) = 0,0107$$

während bei einer Beobachtung von 0 Männern der p-Wert:

$$p = P(X \leq 0) = 0,001$$

beträgt. Im ersten Fall würde die Nullhypothese für ein Risiko $\alpha = 0.05$ verworfen werden, aber nicht für $\alpha = 0.01$, während sie im zweiten Fall auch für $\alpha = 0.01$ verworfen würde. Offensichtlich würde im zweiten Fall die Entscheidung zur Ablehnung der Nullhypothese mit größerer Sicherheit getroffen werden.

9.3.3 Vorgehensweise bei der Durchführung eines Hypothesentests

1. Formulierung der Nullhypothese H_0 und Alternativhypothese H_1
2. Festlegung der gewünschten Risiken α und β
3. Auswahl der geeigneten Teststatistik
4. Bestimmung der klinisch relevanten Mindestdifferenz (Effektgröße) δ
5. Berechnung des erforderlichen Stichprobenumfangs n
6. Abgrenzung von Annahme- und Ablehnungsbereich

7. Ziehung einer Stichprobe des Umfangs n
8. Berechnung der Teststatistik für die Stichprobe
9. Ablehnung der Nullhypothese, wenn der Schätzwert in den Ablehnungsbereich fällt oder der p-Wert kleiner als α ist; andernfalls Beibehaltung der Nullhypothese

9.4 Wichtigste parametrische Tests

Anpassungstests (Goodness-of-Fit-Tests):

- Test für den Mittelwert einer Normalverteilung mit bekannter Varianz
- Test für den Mittelwert einer Normalverteilung mit unbekannter Varianz
- Test für den Mittelwert einer Population mit unbekannter Varianz bei großen Stichproben
- Test für die Varianz einer normalverteilten Population
- Test für einen Anteilswert in einer Population

Homogenitätstests:

- Vergleichstest für Mittelwerte zweier Normalverteilungen mit bekannten Varianzen
- Vergleichstest für Mittelwerte zweier Normalverteilungen mit unbekannten, aber gleichen Varianzen
- Vergleichstest für Mittelwerte zweier Normalverteilungen mit unbekannten und ungleichen Varianzen
- Vergleichstest für Varianzen zweier Normalverteilungen
- Vergleichstest für Anteilswerte zweier Populationen

9.5 Test für den Mittelwert einer Normalverteilung mit bekannter Varianz

Gegeben sei eine Zufallsvariable X mit folgenden Eigenschaften:

- Normalverteilung $X \sim N(\mu, \sigma)$
- Unbekannter Mittelwert μ, aber bekannte Varianz σ^2

Testproblem:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

Teststatistik:

$$\bar{x} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1).$$

Annahmebereich: $z_{\alpha/2} < Z < z_{1-\alpha/2}$

Ablehnungsbereich: $Z \leq z_{\alpha/2}$ oder $Z \geq z_{1-\alpha/2}$

9.6 Test für den Mittelwert einer normalverteilten Grundgesamtheit mit unbekannter Varianz

Sei X eine Zufallsvariable, die die folgenden Bedingungen erfüllt:

- Ihre Verteilung ist normal $X \sim N(\mu, \sigma)$.
- Sowohl ihr Mittelwert μ als auch ihre Varianz σ^2 sind unbekannt.

Testproblem:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

Teststatistik: Unter Verwendung der Stichprobenvarianz als Schätzer der Populationsvarianz ergibt sich

$$\bar{x} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow T = \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}} \sim T(n-1).$$

Annahmebereich: $t^{n-1}\alpha/2 < T < t^{n-1}1 - \alpha/2$.

Ablehnungsbereich: $T \leq t^{n-1}\alpha/2$ und $T \geq t^{n-1}1 - \alpha/2$.

Beispiel

In einer Gruppe von Studierenden soll getestet werden, ob die durchschnittliche Note in Statistik höher als 5 Punkte ist. Dazu wird die folgende Stichprobe entnommen:

$$6,3 - 5,4 - 4,1 - 5,0 - 8,2 - 7,6 - 6,4 - 5,6 - 4,3 - 5,2$$

Der formulierte Test lautet:

$$H_0 : \mu = 5 \quad H_1 : \mu > 5$$

Für die Durchführung des Test ergibt sich:

- $\bar{x} = \frac{6.3+\cdots+5.2}{10} = \frac{58.1}{10} = 5.81$ Punkte.
- $\hat{s}^2 = \frac{(6.3-5.81)^2+\cdots+(5.2-5.81)^2}{9} = \frac{15.949}{9} = 1.7721$ Punkte^2, und $\hat{s} = 1.3312$ Punkte.

Die Teststatistik beträgt

$$T = \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}} = \frac{5,81 - 5}{1,3312/\sqrt{10}} = 1,9246.$$

Der p-Wert des Tests ist $P(T(9) \geq 1,9246) = 0,04323$, was bedeutet, dass die Nullhypothese für $\alpha = 0.05$ abgelehnt würde.
Der Ablehnungsbereich ist

$$T = \frac{\bar{x} - 5}{1,3312/\sqrt{10}} \geq t^9_{0,95} = 1,8331 \Leftrightarrow \bar{x} \geq 5 + 1,8331\frac{1,3312}{\sqrt{10}} = 5,7717$$

sodass die Nullhypothese abgelehnt wird, wenn der Stichprobenmittelwert größer als $5,7717$ ist, andernfalls wird sie angenommen.
Unter der Annahme, dass in der Praxis ein minimaler wichtiger Unterschied in der Durchschnittsnote ein Punkt $\delta = 1$ wäre, dann unter der Alternativhypothese $H_1 : \mu = 6$, wenn die Nullhypothese abgelehnt würde, wäre das Risiko β

$$\beta = P\left(T(9) \leq \frac{5,7717 - 6}{1,3312\sqrt{10}}\right) = P(T(9) \leq -0,5424) = 0,3004$$

sodass die Teststärke zur Erkennung einer Differenz von $\delta = 1$ Punkt $1 - \beta = 1 - 0,3004 = 0,6996$ beträgt.

9.6.1 Bestimmung des Stichprobenumfangs für einen Test des Mittelwerts

Es wurde gezeigt, dass für ein Risiko α der Ablehnungsbereich

$$T = \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}} \geq t_{1-\alpha}^{n-1} \approx z_{1-\alpha} \quad \text{für } n \geq 30.$$

oder äquivalent

$$\bar{x} \geq \mu_0 + z_{1-\alpha}\frac{\hat{s}}{\sqrt{n}}.$$

Wenn die Effektgröße δ beträgt, ergibt sich für eine Alternativhypothese $H_1 : \mu = \mu_0 + \delta$ das Risiko β zu

$$\beta = P\left(Z < \frac{\mu_0 + z_{1-\alpha}\frac{\hat{s}}{\sqrt{n}} - (\mu_0 + \delta)}{\frac{\hat{s}}{\sqrt{n}}}\right) = P\left(Z < \frac{z_{1-\alpha}\frac{\hat{s}}{\sqrt{n}} - \delta}{\frac{\hat{s}}{\sqrt{n}}}\right).$$

sodass

$$z_\beta = \frac{z_{1-\alpha}\frac{\hat{s}}{\sqrt{n}} - \delta}{\frac{\hat{s}}{\sqrt{n}}} \Leftrightarrow \delta = (z_{1-\alpha} - z_\beta)\frac{\hat{s}}{\sqrt{n}} \Leftrightarrow n = (z_{1-\alpha} - z_\beta)^2\frac{\hat{s}^2}{\delta^2} = (z_\alpha + z_\beta)^2\frac{\hat{s}^2}{\delta^2}.$$

Beispiel

Im vorherigen Beispiel wurde gezeigt, dass die Teststärke zur Erkennung einer Differenz in der Durchschnittsnote von 1 Punkt bei $69,96$ lag.
Um die Teststärke auf 90 zu erhöhen, wie viele Schüler müssten in die Stichprobe aufgenommen werden?
Da eine Teststärke von $1 - \beta = 0,9$ gewünscht ist, beträgt das Risiko $\beta = 0.1$ und aus der Tabelle der Standardnormalverteilung ergibt sich $z_\beta = z_{0,1} = 1,2816$.
Unter Anwendung der obigen Formel zur Bestimmung des erforderlichen Stichprobenumfangs ergibt sich

$$n = (z_\alpha + z_\beta)^2\frac{\hat{s}^2}{\delta^2} = (1,6449 + 1,2816)^2\frac{1,7721}{1^2} = 15,18,$$

sodass mindestens 16 Schüler hätten untersucht werden müssen.

9.7 Test für den Mittelwert einer Grundgesamtheit mit unbekannter Varianz und großen Stichproben

Sei X eine Zufallsvariable, die die folgenden Bedingungen erfüllt:

- Ihre Verteilung kann beliebig sein.
- Sowohl ihr Mittelwert μ als auch ihre Varianz σ^2 sind unbekannt.

Testproblem:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

Teststatistik: Unter Verwendung der Stichprobenvarianz als Schätzer der Populationsvarianz und dank des zentralen Grenzwertsatzes (da große Stichproben $n \geq 30$ vorliegen) gilt:

$$\bar{x} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow Z = \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}} \sim N(0,1).$$

Annahmebereich: $-z_{\alpha/2} < Z < z_{\alpha/2}$.

Ablehnungsbereich: $Z \leq -z_{\alpha/2}$ y $Z \geq z_{\alpha/2}$.

9.8 Test für die Varianz einer normalverteilten Grundgesamtheit

Sei X eine Zufallsvariable, die folgende Annahmen erfüllt:

- Ihre Verteilung ist normal
- Sowohl ihr Mittelwert μ als auch ihre Varianz σ^2 sind unbekannt

Testproblem:

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 \neq \sigma_0^2$$

Teststatistik: Ausgehend von der Stichprobenvarianz als Schätzer für die Populationsvarianz gilt:

$$J = \frac{nS^2}{\sigma_0^2} = \frac{(n-1)\hat{S}^2}{\sigma_0^2} \sim \chi^2(n-1),$$

die einer Chi-Quadrat-Verteilung mit $n-1$ Freiheitsgraden folgt.

Annahmebereich: $\chi^{n-1}_{\alpha/2} < J < \chi^{n-1}_{1-\alpha/2}$

Ablehnungsbereich: $J \leq \chi^{n-1}_{\alpha/2}$ oder $J \geq \chi^{n-1}_{1-\alpha/2}$

Beispiel

In einer Schülergruppe soll getestet werden, ob die Standardabweichung der Noten größer als 1 Punkt ist. Dazu wird folgende Stichprobe erhoben:

$$6,3 - 5,4 - 4,1 - 5,0 - 8,2 - 7,6 - 6,4 - 5,6 - 4,3 - 5,2$$

Es wird folgender Test durchgeführt:

$$H_0 : \sigma = 1 \quad H_1 : \sigma > 1$$

Für die Testdurchführung ergibt sich:

- $\bar{x} = \frac{6.3+\cdots+5,2}{10} = \frac{58,1}{10} = 5,81$ Punkte
- $\hat{s}^2 = \frac{(6,3-5,81)^2+\cdots+(5,2-5,81)^2}{9} = \frac{15,949}{9} = 1,7721$ Punkte^2

Die Teststatistik beträgt:

$$J = \frac{(n-1)\hat{S}^2}{\sigma_0^2} = \frac{9 \cdot 1,7721}{1^2} = 15,949,$$

und der p-Wert des Tests ist $P(\chi(9) \geq 15,949) = 0,068$, sodass die Nullhypothese für $\alpha = 0.05$ nicht verworfen werden kann.

9.9 Test für einen Anteilswert in einer Grundgesamtheit

Sei p der Anteil von Individuen in einer Population, die eine bestimmte Eigenschaft aufweisen.

Testproblem:

$$H_0 : p = p_0$$
$$H_1 : p \neq p_0$$

Teststatistik: Die Zufallsvariable, die die Anzahl der Individuen mit der Eigenschaft in einer Zufallsstichprobe des Umfangs n zählt, folgt einer Binomialverteilung $X \sim B(n, p_0)$. Gemäß dem zentralen Grenzwertsatz gilt für große Stichproben ($np \geq 5$ und $n(1-p) \geq 5$) die Approximation $X \sim N(np_0, \sqrt{np_0(1-p_0)})$, und somit:

$$\hat{p} = \frac{X}{n} \sim N\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right) \Rightarrow Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1).$$

Annahmebereich: $z_{\alpha/2} < Z < z_{1-\alpha/2}$.

Ablehnungsbereich: $Z \leq z_{\alpha/2}$ oder $Z \geq z_{1-\alpha/2}$.

Beispiel

In einer Gruppe von Studierenden soll untersucht werden, ob die Durchfallquote über 50% liegt. Dazu wird eine Stichprobe von 80 Studierenden gezogen, von denen 50 bestanden haben.
Das Testproblem lautet:

$$H_0 : p = 0.5$$
$$H_1 : p > 0.5$$

Für die Testdurchführung ergibt sich $\hat{p} = 50/80 = 0,625$. Da die Approximationsbedingungen $n\hat{p} = 80 \cdot 0,625 = 50 \geq 5$ und $n(1-\hat{p}) = 80(1-0,625) = 30 \geq 5$ erfüllt sind, berechnet sich die Teststatistik zu:

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0,625 - 0.5}{\sqrt{0,5(1-0,5)/80}} = 2,2361.$$

Der p-Wert des Tests beträgt $P(Z \geq 2,2361) = 0,0127$, sodass die Nullhypothese für $\alpha = 0,05$ verworfen werden kann. Es lässt sich schließen, dass die Bestehensquote signifikant über 50% liegt.

9.10 Vergleichstest für Mittelwerte zweier Normalverteilungen mit bekannten Varianzen

Seien X_1 und X_2 zwei Zufallsvariablen, die folgende Bedingungen erfüllen:

- Ihre Verteilung ist normal: $X_1 \sim N(\mu_1, \sigma_1)$ und $X_2 \sim N(\mu_2, \sigma_2)$
- Ihre Mittelwerte μ_1 und μ_2 sind unbekannt, aber ihre Varianzen σ_1^2 und σ_2^2 sind bekannt.

Testproblem:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Teststatistik:

$$\left.\begin{array}{l} \bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1}{\sqrt{n_1}}\right) \\ \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2}{\sqrt{n_2}}\right) \end{array}\right\} \Rightarrow$$

$$\Rightarrow \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \Rightarrow Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1).$$

Annahmebereich: $-z_{\alpha/2} < Z < z_{\alpha/2}$

Ablehnungsbereich: $Z \leq -z_{\alpha/2}$ oder $Z \geq z_{\alpha/2}$

9.11 Vergleichstest für Mittelwerte zweier Normalverteilungen mit unbekannten aber gleichen Varianzen

Seien X_1 und X_2 zwei Zufallsvariablen, die folgende Bedingungen erfüllen:

- Ihre Verteilung ist normal: $X_1 \sim N(\mu_1, \sigma_1)$ und $X_2 \sim N(\mu_2, \sigma_2)$
- Ihre Mittelwerte μ_1 und μ_2 sind unbekannt und ihre Varianzen ebenfalls, aber sie sind gleich: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Testproblem:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Teststatistik:

$$\left.\begin{array}{l} \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sigma\sqrt{\frac{n_1+n_2}{n_1 n_2}}\right) \\ \frac{n_1 S_1^2 + n_2 S_2^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2) \end{array}\right\} \Rightarrow T = \frac{\bar{X}_1 - \bar{X}_2}{\hat{S}_p\sqrt{\frac{n_1+n_2}{n_1 n_2}}} \sim T(n_1 + n_2 - 2).$$

Annahmebereich: $-t_{\alpha/2}^{n_1+n_2-2} < T < t_{\alpha/2}^{n_1+n_2-2}$

Ablehnungsbereich: $T \leq -t_{\alpha/2}^{n_1+n_2-2}$ oder $T \geq t_{\alpha/2}^{n_1+n_2-2}$

Beispiel

Es soll die akademische Leistung zweier Schülergruppen (10 bzw. 12 Schüler) mit unterschiedlichen Lehrmethoden verglichen werden. Dazu werden folgende Prüfungsergebnisse erzielt:

$$X_1 : 4 - 6 - 8 - 7 - 7 - 6 - 5 - 2 - 5 - 3$$
$$X_2 : 8 - 9 - 5 - 3 - 8 - 7 - 8 - 6 - 8 - 7 - 5 - 7$$

Das Testproblem lautet:

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$$

Für die Testdurchführung ergibt sich:

- $\bar{X}_1 = \frac{4+\cdots+3}{10} = 5.3$ Punkte und $\bar{X}_2 = \frac{8+\cdots+7}{12} = 6.75$ Punkte
- $S_1^2 = \frac{4^2+\cdots+3^2}{10} - 5.3^2 = 3.21$ Punkte^2 und $S_2^2 = \frac{8^2+\cdots+3^2}{12} - 6.75^2 = 2.69$ Punkte^2
- $\hat{S}_p^2 = \frac{10 \cdot 3.21 + 12 \cdot 2.6875}{10+12-2} = 3.2175$ Punkte^2, und $\hat{S}_p = 1.7937$

Unter Annahme gleicher Varianzen beträgt die Teststatistik:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\hat{S}_p \sqrt{\frac{n_1+n_2}{n_1 n_2}}} = \frac{5,3 - 6,75}{1,7937 \sqrt{\frac{10+12}{10 \cdot 12}}} = -1,8879$$

Der p-Wert des Tests ist $2P(T(20) \leq -1,8879) = 0,0736$, sodass die Nullhypothese nicht verworfen werden kann. Es gibt keine signifikanten Unterschiede zwischen den Durchschnittsnoten der Gruppen.

9.12 Vergleichstest für Mittelwerte zweier Normalverteilungen mit unbekannten und ungleichen Varianzen

Seien X_1 und X_2 zwei Zufallsvariablen, die folgende Bedingungen erfüllen:

- Ihre Verteilung ist normal: $X_1 \sim N(\mu_1, \sigma_1)$ und $X_2 \sim N(\mu_2, \sigma_2)$
- Ihre Mittelwerte μ_1, μ_2 und Varianzen σ_1^2, σ_2^2 sind unbekannt, wobei $\sigma_1^2 \neq \sigma_2^2$

Testproblem:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Teststatistik:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}} \sim T(g),$$

mit $g = n_1 + n_2 - 2 - \Delta$ und

$$\Delta = \frac{(\frac{n_2-1}{n_1}\hat{S}_1^2 - \frac{n_1-1}{n_2}\hat{S}_2^2)^2}{\frac{n_2-1}{n_1^2}\hat{S}_1^4 + \frac{n_1-1}{n_2^2}\hat{S}_2^4}.$$

Annahmebereich: $-t_{\alpha/2}^g < T < t_{\alpha/2}^g$

Ablehnungsbereich: $T \leq -t_{\alpha/2}^g$ oder $T \geq t_{\alpha/2}^g$

9.13 Vergleichstest für Varianzen zweier Normalverteilungen

Seien X_1 und X_2 zwei Zufallsvariablen mit:

- Normalverteilung: $X_1 \sim N(\mu_1, \sigma_1)$ und $X_2 \sim N(\mu_2, \sigma_2)$
- Unbekannten Mittelwerten μ_1, μ_2 und Varianzen σ_1^2, σ_2^2

Testproblem:

$$H_0 : \sigma_1 = \sigma_2$$
$$H_1 : \sigma_1 \neq \sigma_2$$

Teststatistik:

$$\left.\begin{array}{l} \frac{(n_1-1)\hat{S}_1^2}{\sigma_1^2} \sim \chi^2(n_1-1) \\ \frac{(n_2-1)\hat{S}_2^2}{\sigma_2^2} \sim \chi^2(n_2-1) \end{array}\right\} \Rightarrow F = \frac{\frac{\frac{(n_1-1)\hat{S}_1^2}{\sigma_1^2}}{n_1-1}}{\frac{\frac{(n_2-1)\hat{S}_2^2}{\sigma_2^2}}{n_2-1}} = \frac{\sigma_2^2}{\sigma_1^2}\frac{\hat{S}_1^2}{\hat{S}_2^2} \sim F(n_1-1, n_2-1).$$

Annahmebereich: $F_{\alpha/2}^{n_1-1,n_2-1} < F < F_{1-\alpha/2}^{n_1-1,n_2-1}$

Ablehnungsbereich: $F \leq F_{\alpha/2}^{n_1-1,n_2-1}$ oder $F \geq F_{1-\alpha/2}^{n_1-1,n_2-1}$

Beispiel

Fortsetzung des Beispiels mit den Punktzahlen zweier Gruppen:

$$X_1 : 4-6-8-7-7-6-5-2-5-3$$
$$X_2 : 8-9-5-3-8-7-8-6-8-7-5-7$$

Für den Varianzvergleich lautet das Testproblem:

$$H_0 : \sigma_1 = \sigma_2 \qquad H_1 : \sigma_1 \neq \sigma_2$$

Berechnungen:

- $\bar{X}_1 = \frac{4+\cdots+3}{10} = 5,3$ Punkte
- $\bar{X}_2 = \frac{8+\cdots+7}{12} = 6,75$ Punkte
- $\hat{S}_1^2 = \frac{(4-5,3)^2+\cdots+(3-5,3)^2}{9} = 3,5667$ Punkte^2
- $\hat{S}_2^2 = \frac{(8-6,75)^2+\cdots+(3-6,75)^2}{11} = 2,9318$ Punkte^2

Teststatistik:

$$F = \frac{\hat{S}_1^2}{\hat{S}_2^2} = \frac{3,5667}{2,9318} = 1,2165$$

Der p-Wert beträgt $2P(F(9,11) \leq 1.2165) = 0.7468$, sodass die Nullhypothese gleicher Varianzen beibehalten wird.

9.14 Vergleichstest von Anteilen zweier Populationen

Seien p_1 und p_2 die jeweiligen Anteile von Individuen, die ein bestimmtes Merkmal in zwei Populationen aufweisen.

Testproblem:

$$H_0 : p_1 = p_2 \qquad H_1 : p_1 \neq p_2$$

Teststatistik: Die Zufallsvariablen, die die Anzahl der Individuen mit dem Merkmal in zwei unabhängigen Stichproben der Größen n_1 und n_2 messen, folgen binomialverteilten Zufallsvariablen $X_1 \sim B(n_1, p_1)$ und $X_2 \sim B(n_2, p_2)$. Wenn die Stichproben groß sind ($n_i p_i \geq 5$ und $n_i(1-p_i) \geq 5$), folgt gemäß dem zentralen Grenzwertsatz:

$$X_1 \sim N(np_1, \sqrt{np_1(1-p_1)}) \quad \text{und} \quad X_2 \sim N(np_2, \sqrt{np_2(1-p_2)})$$

und es gilt:

$$\left.\begin{aligned} \hat{p}_1 = \tfrac{X_1}{n_1} &\sim N\left(p_1, \sqrt{\tfrac{p_1(1-p_1)}{n_1}}\right) \\ \hat{p}_2 = \tfrac{X_2}{n_2} &\sim N\left(p_2, \sqrt{\tfrac{p_2(1-p_2)}{n_2}}\right) \end{aligned}\right\} \Rightarrow Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1)$$

Annahmebereich: $z_{\alpha/2} < Z < z_{1-\alpha/2}$

Ablehnungsbereich: $z \leq z_{\alpha/2}$ und $z \geq z_{1-\alpha/2}$

Beispiel

Es soll geprüft werden, ob sich die Bestehensquoten zweier Gruppen unterscheiden, die unterschiedliche Lehrmethoden durchlaufen haben. In der ersten Gruppe haben 24 von insgesamt 40 Schülern bestanden, in der zweiten Gruppe 48 von 60.
Der formulierte Test lautet:

$$H_0 : \ p_1 = p_2 \qquad H_1 : \ p_1 \neq p_2$$

Für den Test ergibt sich: $\hat{p}_1 = 24/40 = 0.6$ und $\hat{p}_2 = 48/60 = 0.8$. Damit sind die Voraussetzungen erfüllt:

- $n_1\hat{p}_1 = 40 \cdot 0.6 = 24 \geq 5$,
- $n_1(1-\hat{p}_1) = 40(1-0.6) = 16 \geq 5$,
- $n_2\hat{p}_2 = 60 \cdot 0.8 = 48 \geq 5$ und
- $n_2(1-\hat{p}_2) = 60(1-0.8) = 12 \geq 5$.

Die Teststatistik ergibt:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} = \frac{0,6 - 0,8}{\sqrt{\frac{0,6(1-0,6)}{40} + \frac{0,8(1-0,8)}{60}}} = -2,1483$$

und der p-Wert des Tests ist $2P(Z \leq -2,1483) = 0,0317$, sodass die Nullhypothese bei einem Signifikanzniveau von $\alpha = 0,05$ verworfen wird. Es wird also geschlossen, dass es einen Unterschied zwischen den Gruppen gibt.

9.15 Durchführung von Tests mittels Konfidenzintervallen

Eine interessante Alternative zur Durchführung eines Hypothesentests

$$H_0 : \theta = \theta_0 \qquad H_1 : \theta \neq \theta_0$$

mit einem Signifikanzniveau α ist die Berechnung eines Konfidenzintervalls für θ mit einem Vertrauensniveau von $1-\alpha$. Dieses Intervall kann als die Menge akzeptabler Hypothesen für θ interpretiert werden. Befindet sich θ_0 außerhalb des Intervalls, erscheint die Nullhypothese wenig glaubwürdig und kann verworfen werden. Befindet sich θ_0 hingegen innerhalb des Intervalls, wird die Hypothese als plausibel betrachtet und beibehalten.

Bei einem einseitigen Test mit der Alternativhypothese "kleiner" wird θ_0 mit der oberen Grenze eines Konfidenzintervalls für θ mit einem Vertrauensniveau von $1 - 2\alpha$ verglichen. Ist der Test einseitig mit der Alternativhypothese "größer", so wird θ_0 mit der unteren Grenze des Intervalls verglichen.

Testart	**Konfidenzintervall**	**Entscheidung**
Zweiseitig	$[l_i, l_s]$ mit Vertrauensniveau $1-\alpha$	Verwerfe H_0 wenn $\theta_0 \notin [l_i, l_s]$
Einseitig (kleiner)	$[-\infty, l_s]$ mit Vertrauensniveau $1-2\alpha$	Verwerfe H_0 wenn $\theta_0 \geq l_s$
Einseitig (größer)	$[l_i, \infty]$ mit Vertrauensniveau $1-2\alpha$	Verwerfe H_0 wenn $\theta_0 \leq l_i$

Beispiel

Zurück zum Test, der den schulischen Erfolg zweier Schülergruppen vergleicht. Diese haben folgende Punktzahlen erreicht:

$$X_1 : 4-6-8-7-7-6-5-2-5-3$$
$$X_2 : 8-9-5-3-8-7-8-6-8-7-5-7$$

Der getestete Hypothesensatz lautete:

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$$

Da es sich um einen zweiseitigen Test handelt, ergibt sich das Konfidenzintervall für die Differenz der Mittelwerte $\mu_1 - \mu_2$ bei einem Vertrauensniveau von $1 - \alpha = 0,95$ (unter der Annahme gleicher Varianzen) zu $[-3{,}0521 \mid 0{,}1521]$ Punkten.
Da unter der Nullhypothese $\mu_1 - \mu_2 = 0$ gilt und 0 innerhalb dieses Intervalls liegt, wird die Nullhypothese beibehalten.
Der Vorteil eines Konfidenzintervalls besteht darin, dass es uns nicht nur erlaubt, einen Test durchzuführen, sondern auch eine Vorstellung von der Größenordnung der Differenz zwischen den Gruppenmittelwerten vermittelt.

10 Varianzanalysen

10.1 Varianzanalyse mit einem Faktor

Die *Varianzanalyse mit einem Faktor* (im Englischen ANOVA für *Analysis of Variance*) ist eine statistische Hypothesentest-Methode, die dazu dient, die Mittelwerte einer quantitativen Variablen – üblicherweise als *abhängige Variable* oder *Antwortvariable* bezeichnet – zwischen verschiedenen Gruppen oder Stichproben zu vergleichen. Diese Gruppen werden durch eine qualitative Variable definiert, die als *unabhängige Variable* oder *Faktor* bezeichnet wird. Die verschiedenen Ausprägungen des Faktors, die die zu vergleichenden Gruppen darstellen, nennt man *Stufen* oder *Behandlungen* des Faktors.

Es handelt sich dabei um eine Verallgemeinerung des *t-Tests zum Vergleich der Mittelwerte zweier unabhängiger Stichproben* für Versuchsanordnungen mit mehr als zwei Gruppen. Der Hauptunterschied zu einer einfachen Regressionsanalyse – bei der sowohl die abhängige als auch die unabhängige Variable quantitativ sind – besteht darin, dass beim einfaktoriellen Varianzanalysedesign die unabhängige Variable bzw. der Faktor eine qualitative Variable ist. Wie wir später bei den Regressionskontrasten sehen werden, lässt sich ein ANOVA-Test jedoch auch als Spezialfall eines linearen Regressionsmodells formulieren.

Ein typisches Beispiel für die Anwendung dieser Technik wäre der Vergleich des durchschnittlichen Cholesterinspiegels je nach Blutgruppe. In diesem Fall ist die unabhängige Variable oder der *Faktor* die Blutgruppe mit vier Stufen (`A`, `B`, `0`, `AB`), während die *Antwortvariable* der Cholesterinspiegel ist.

Zum Vergleich der Mittelwerte der Antwortvariablen in den verschiedenen Stufen des Faktors wird ein Hypothesentest aufgestellt, bei dem die Nullhypothese H_0 lautet, dass die Antwortvariable in allen Gruppen denselben Mittelwert hat. Die Alternativhypothese H_1 besagt, dass es statistisch signifikante Unterschiede zwischen mindestens zwei der Mittelwerte gibt. Dieser Test basiert auf der Zerlegung der Gesamtvarianz der Antwortvariablen – daher der Name dieser Methode.

10.1.1 Der ANOVA-Test

Die übliche Notation in der ANOVA lautet:

- k: ist die Anzahl der Stufen (Niveaus) des Faktors.
- n_i: ist der Stichprobenumfang, der zur i-ten Stufe des Faktors gehört.
- $n = \sum_{i=1}^{k} n_i$: ist die Gesamtanzahl der Beobachtungen.
- X_{ij} $(i = 1, ..., k;\ j = 1, ..., n_i)$: ist eine Zufallsvariable, die die Antwort des j-ten Individuums auf die i-te Stufe des Faktors angibt.
- x_{ij}: ist der konkrete Wert von X_{ij} in einer gegebenen Stichprobe.

Faktorstufen			
1	2	$\cdots$	k
X_{11}	X_{21}	$\cdots$	X_{k1}
X_{12}	X_{22}	$\cdots$	X_{k2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_{1n_1}	X_{2n_2}	$\cdots$	X_{kn_k}

- μ_i: ist der Mittelwert der Grundgesamtheit der i-ten Stufe.

- $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$: ist der Stichprobenmittelwert der i-ten Stufe und Schätzer für μ_i.
- $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i$: ist der konkrete Wert dieses Mittelwerts in einer gegebenen Stichprobe.
- μ: ist der Gesamtmittelwert der Grundgesamtheit (über alle Stufen hinweg).
- $\bar{X} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij}/n$: ist der Gesamtstichprobenmittelwert, Schätzer für μ.
- $\bar{x} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}/n$: ist der konkrete Wert dieses Mittelwerts in einer gegebenen Stichprobe.

Mit dieser Notation lässt sich die Antwortvariable durch ein mathematisches Modell ausdrücken, das sie in Komponenten zerlegt, die verschiedenen Ursachen zugeordnet werden:

$$X_{ij} = \mu + (\mu_i - \mu) + (X_{ij} - \mu_i),$$

Das heißt, die Antwort der j-ten Beobachtung in der i-ten Stufe setzt sich zusammen aus dem Gesamtmittelwert, einer Abweichung, die dem Einfluss des i-ten Faktors zugeschrieben wird, und einer weiteren Abweichung aufgrund zufälliger Schwankungen innerhalb der Stufe.

Auf Basis dieses Modells formulieren wir die Nullhypothese: Die Mittelwerte aller Stufen sind gleich. Die Alternativhypothese lautet: Mindestens zwei Mittelwerte unterscheiden sich.

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \cdots = \mu_k \\ H_1 &: \mu_i \neq \mu_j \quad \text{für einige } i \neq j \end{aligned}$$

Damit dieser Test gültig ist, müssen folgende Modellannahmen erfüllt sein:

- **Unabhängigkeit**: Die k Stichproben, die den k Stufen entsprechen, sind unabhängige Zufallsstichproben aus k Populationen mit unbekannten Mittelwerten $\mu_1 = \mu_2 = \cdots = \mu_k$.
- **Normalverteilung**: Jede der k Populationen ist normalverteilt.
- **Homoskedastizität**: Alle Populationen haben die gleiche Varianz σ^2.

Wird in obigem Modell der Populationsmittelwert durch seinen Stichprobenschätzer ersetzt, ergibt sich:

$$X_{ij} = \bar{X} + (\bar{X}_i - \bar{X}) + (X_{ij} - \bar{X}_i),$$

oder umgestellt:

$$X_{ij} - \bar{X} = (\bar{X}_i - \bar{X}) + (X_{ij} - \bar{X}_i).$$

Durch Quadrieren und unter Ausnutzung der Summeneigenschaften ergibt sich die sogenannte *Identität der Quadratsummen*:

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 = \sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2,$$

wobei:

- $\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$: ist die *Gesamtquadratsumme* (SCT) – misst die gesamte Streuung der Daten.
- $\sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2$: ist die *Zwischengruppen-Quadratsumme* ($SCInter$) – misst die Streuung aufgrund der Unterschiede zwischen den Gruppenmitteln.
- $\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$: ist die *Intragruppen-Quadratsumme* ($SCIntra$) – misst die Streuung innerhalb der Gruppen.

Somit gilt:

$$SCT = SCInter + SCIntra$$

Um zur Teststatistik zu gelangen, definiert man die *mittleren Quadratsummen* (mean squares), indem man jede Quadratsumme durch die entsprechenden Freiheitsgrade teilt:

- Für SCT: $n - 1$
- Für $SCInter$: $k - 1$
- Für $SCIntra$: $n - k$

Daraus folgt:

$$\begin{aligned} CMT &= \frac{SCT}{n-1} \\ CMInter &= \frac{SCInter}{k-1} \\ CMIntra &= \frac{SCIntra}{n-k} \end{aligned}$$

Man kann zeigen, dass unter der Annahme, dass H_0 sowie die Modellvoraussetzungen zutreffen, der Quotient

$$\frac{CMInter}{CMIntra}$$

einer F-Verteilung von Fisher mit $k - 1$ und $n - k$ Freiheitsgraden folgt.

Wenn also H_0 gilt, liegt dieser Quotient für gegebene Stichproben nahe bei 1 (aber stets > 0). Ist H_0 falsch, dann nimmt die Streuung zwischen den Gruppen zu – der Wert des Quotienten steigt. Daher handelt es sich um einen einseitigen Test (rechtsseitig) mit Hilfe der F-Verteilung. Wir berechnen den p-Wert der beobachteten F-Statistik und entscheiden, ob wir H_0 beibehalten oder verwerfen – abhängig vom festgelegten Signifikanzniveau.

10.1.1.1 ANOVA-Tabelle

Alle im vorherigen Abschnitt eingeführten Statistiken werden in einer sogenannten **ANOVA-Tabelle** zusammengefasst. In dieser Tabelle werden die Schätzwerte dieser Statistiken für die konkret untersuchten Stichproben dargestellt. Solche Tabellen sind auch das Standardausgabeformat statistischer Softwareprogramme bei der Durchführung einer ANOVA. Am Ende der Tabelle wird üblicherweise der p-Wert des berechneten F-Statistikums angegeben, anhand dessen entschieden werden kann, ob die Nullhypothese – dass die Mittelwerte aller Faktorstufen gleich sind – akzeptiert oder verworfen wird.

	Quadratsumme	Freiheitsgrade	Mittlere Quadrate	F-Statistik	p-Wert
Zwischengruppen	$SCInter$	$k-1$	$CMInter = \frac{SCInter}{k-1}$	$f = \frac{CMInter}{CMIntra}$	$P(F > f)$
Intragruppen	$SCIntra$	$n-k$	$CMIntra = \frac{SCIntra}{n-k}$		
Gesamt	SCT	$n-1$			

10.1.2 Tests für multiple und paarweise Vergleiche

Nachdem eine einfaktorielle ANOVA durchgeführt wurde, um die k Mittelwerte der k Faktorstufen bzw. Behandlungen zu vergleichen, kann man im Ergebnis entweder die Nullhypothese beibehalten – in diesem Fall endet die Analyse hinsichtlich signifikanter Unterschiede – oder sie verwerfen. Wird die Nullhypothese verworfen, ist es sinnvoll, die Analyse fortzusetzen, um präzise zu bestimmen, **zwischen welchen Gruppen** signifikante Unterschiede bestehen.

Für diesen zweiten Fall existieren verschiedene Verfahren, die als **Tests für multiple Vergleiche** bezeichnet werden. Diese lassen sich weiter unterteilen in:

- **Tests für paarweise Vergleiche**: Ziel ist es, alle möglichen Paare von Mittelwerten zwischen den Faktorstufen einzeln zu vergleichen. Die Ergebnisse werden meist in Tabellenform dargestellt, mit den Differenzen zwischen allen möglichen Mittelwertpaaren und den zugehörigen Konfidenzintervallen. Dabei wird angegeben, ob die Unterschiede signifikant sind oder nicht. Wichtig ist, dass diese Intervalle **nicht** dieselben wären wie bei einem isolierten Vergleich zweier Gruppen, da das Verwerfen von H_0 im Gesamt-ANOVA-Test bedeutet, dass gleichzeitig mehrere Einzelvergleiche impliziert sind. Um das vorgegebene Signifikanzniveau α aufrechterhalten zu können, muss deshalb ein deutlich kleineres korrigiertes Niveau α' bei den Einzeltests verwendet werden.
- **Tests mit multiplen Rängen**: Ziel ist es, homogene Gruppen von Mittelwerten zu identifizieren, zwischen denen **keine signifikanten Unterschiede** bestehen.

Für paarweise Vergleiche kann der **Bonferroni-Test** verwendet werden; für Ranggruppierungen der **Duncan-Test**. Für beide Arten zusammen eignen sich die Verfahren **Tukey-HSD** und **Scheffé-Test**.

10.2 ANOVA mit zwei oder mehr Faktoren

In vielen Fragestellungen liegen nicht nur ein Faktor mit k Stufen vor, sondern es treten zwei oder mehr Faktoren auf, mit denen die Stichprobe nach verschiedenen Kriterien in Gruppen eingeordnet werden kann. Ziel ist es zu prüfen, ob signifikante Unterschiede zwischen den Mittelwerten der abhängigen Variable bestehen.

Zur Behandlung solcher Fragestellungen existiert das **ANOVA mit zwei oder mehr Faktoren** (auch **Mehrfaktorielle ANOVA**), eine Verallgemeinerung der einfaktoriellen ANOVA. Sie ermöglicht neben der Analyse der Haupteffekte jedes Faktors auch die Untersuchung ihrer **Interaktion**.

Außerdem gibt es Situationen, in denen pro Versuchsperson mehrere Messungen einer quantitativen Variable erfolgen. Bei **zwei Messungen** verwendet man den gepaarten Student-Test (oder bei nichtparametrischen Voraussetzungen den Wilcoxon-Test). Wird **drei oder mehr Messwerte** pro Subjekt erhoben, kommt die **ANOVA für Messwiederholungen** zum Einsatz.

Es kann auch vorkommen, dass eine Variable mehrfach pro Individuum gemessen wird **und** gleichzeitig zwei oder mehr Klassifikationsfaktoren vorliegen. Dann kombiniert man repetierte Messungen mit einer mehrfaktoriellen ANOVA.

Die komplexeste Variante ist die **ANCOVA (Analyse der Kovarianz)**, die in einem Design mit Messwiederholungen, mehreren Faktoren und zusätzlich **Kovariablen** (quantitative Einflussgrößen) durchgeführt wird. Die ANCOVA untersucht die Faktoren- und Messwiederholungseffekte, **bereinigt um den Einfluss der Kovariablen**.

10.2.1 Zweifaktorielle ANOVA mit zwei Faktorstufen

Um eine zweifaktorielle oder mehrfaktorielle ANOVA zu verstehen, ist es hilfreich, mit einem einfachen Fall mit zwei Faktoren und jeweils zwei Stufen zu beginnen. Beispielsweise könnte man ein Experiment durchführen, bei dem Personen entweder eine Diät einhalten (erster Faktor: Diät, mit zwei Stufen: ja und nein) und zusätzlich ein bestimmtes Medikament einnehmen (zweiter Faktor: Medikament, mit zwei Stufen: ja und nein), um ihr Körpergewicht zu reduzieren (numerische Zielvariable: Gewichtsreduktion in kg).

In diesem Fall ergeben sich vier verschiedene Gruppen:

1. Keine Diät und kein Medikament (Nein-Nein)
2. Keine Diät, aber Medikament (Nein-Ja)
3. Diät, aber kein Medikament (Ja-Nein)
4. Diät und Medikament (Ja-Ja)

Es können drei verschiedene Effekte untersucht werden:

1. **Diät-Effekt**: Untersucht, ob es signifikante Unterschiede in der Gewichtsabnahme zwischen Personen mit und ohne Diät gibt.
2. **Medikamenten-Effekt**: Untersucht, ob es signifikante Unterschiede in der Gewichtsabnahme zwischen Personen mit und ohne Medikament gibt.
3. **Interaktionseffekt**: Untersucht, ob die kombinierte Wirkung von Diät und Medikament unterschiedlich ist zur Summe ihrer Einzeleffekte. Bei einer Interaktion kann diese in zwei Richtungen wirken:

- Synergie: Die Kombination führt zu einer stärkeren Gewichtsabnahme als die Summe der Einzeleffekte.
- Antagonismus: Die Kombination führt zu einer schwächeren Gewichtsabnahme als die Summe der Einzeleffekte.

Beispiel

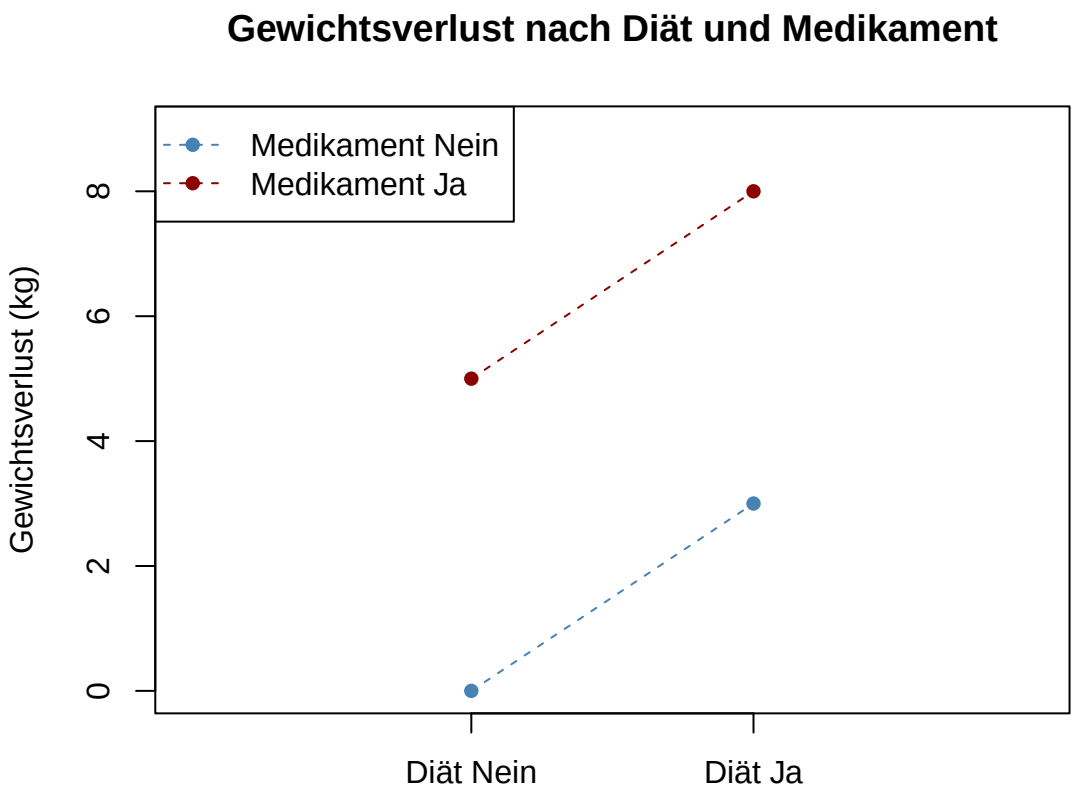
Angenommen, die folgende Tabelle zeigt die durchschnittliche Gewichtsabnahme (in kg) für jede Gruppe (ohne Berücksichtigung der individuellen Variabilität, die in einer echten ANOVA berücksichtigt würde):

	Medikament Nein	Medikament Ja
Diät Nein	0	5
Diät Ja	3	8

In diesem Fall gibt es keine Interaktion zwischen Diät und Medikament, denn:

- Der Effekt des Medikaments allein (ohne Diät) beträgt +5 kg.
- Der Effekt der Diät allein (ohne Medikament) beträgt +3 kg.
- Die kombinierte Wirkung (Diät + Medikament) beträgt +8 kg, was exakt der Summe der Einzeleffekte (5 + 3) entspricht.

Grafische Darstellung: In einem Interaktionsplot wären die Linien für die Gruppen parallel (innerhalb eines gewissen Variabilitätsbereichs), was auf das Fehlen einer Interaktion hindeutet.



Im Gegensatz dazu könnte auch eine Tabelle entstehen, in der die Summe der Einzeleffekte geringer ist als der kombinierte Effekt von Diät und Medikament:

	Medikament Nein	Medikament Ja
Diät Nein	0	5
Diät Ja	3	12

In diesem Fall (wenn wir die Variabilität innerhalb jeder Gruppe vernachlässigen und annehmen, dass sie klein genug ist, damit die Unterschiede signifikant sind), wären die 8 kg durchschnittlicher Gewichtsabnahme, die sich aus der Summe der Einzeleffekte von Diät und Medikament ergeben, geringer als die 12 kg, die Personen im Durchschnitt verloren haben, die sowohl das Medikament einnahmen als auch die Diät befolgten.

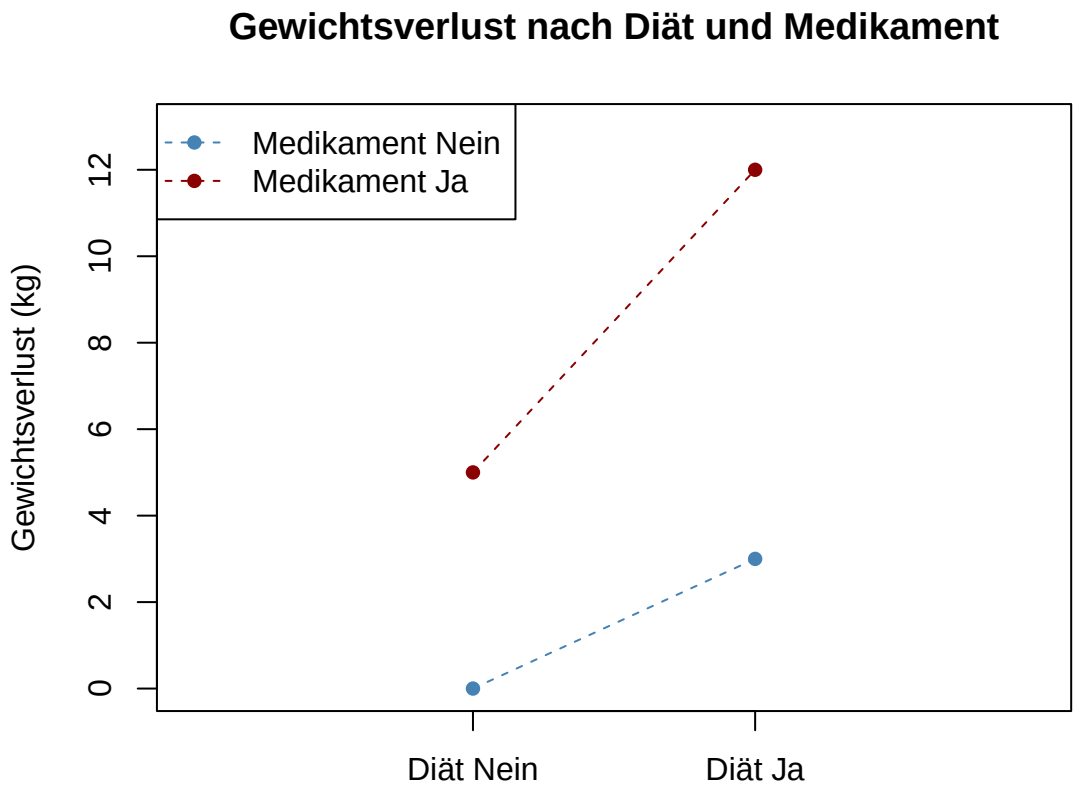
Interpretation:

- Es liegt eine Interaktion der beiden Faktoren vor, die ihre Einzeleffekte verstärkt hat.

- Um das Endergebnis zu erklären, müsste man einen zusätzlichen Interaktionsterm in die Berechnung aufnehmen, der 4 kg zusätzlichen Gewichtsverlust beisteuert (über die Summe der Einzeleffekte hinaus).
- Da dieser Term die Wirkung beider Faktoren verstärkt, handelt es sich um eine synergistische Interaktion.

Grafische Darstellung:

In einem Interaktionsplot wären die Linien nicht parallel, sondern würden sich annähern oder entfernen, was auf eine Interaktion hindeutet.



Zusammenfassung der Effekte:

- Medikament allein: +5 kg
- Diät allein: +3 kg
- Erwarteter kombinierter Effekt (Summe): 5 + 3 = 8 kg
- Tatsächlicher kombinierter Effekt: 12 kg
- Interaktionsterm: 12 - 8 = +4 kg (Synergie)

Schließlich könnte sich auch eine Tabelle ergeben, bei der die Summe der Einzeleffekte größer ist als der kombinierte Effekt beider Faktoren:

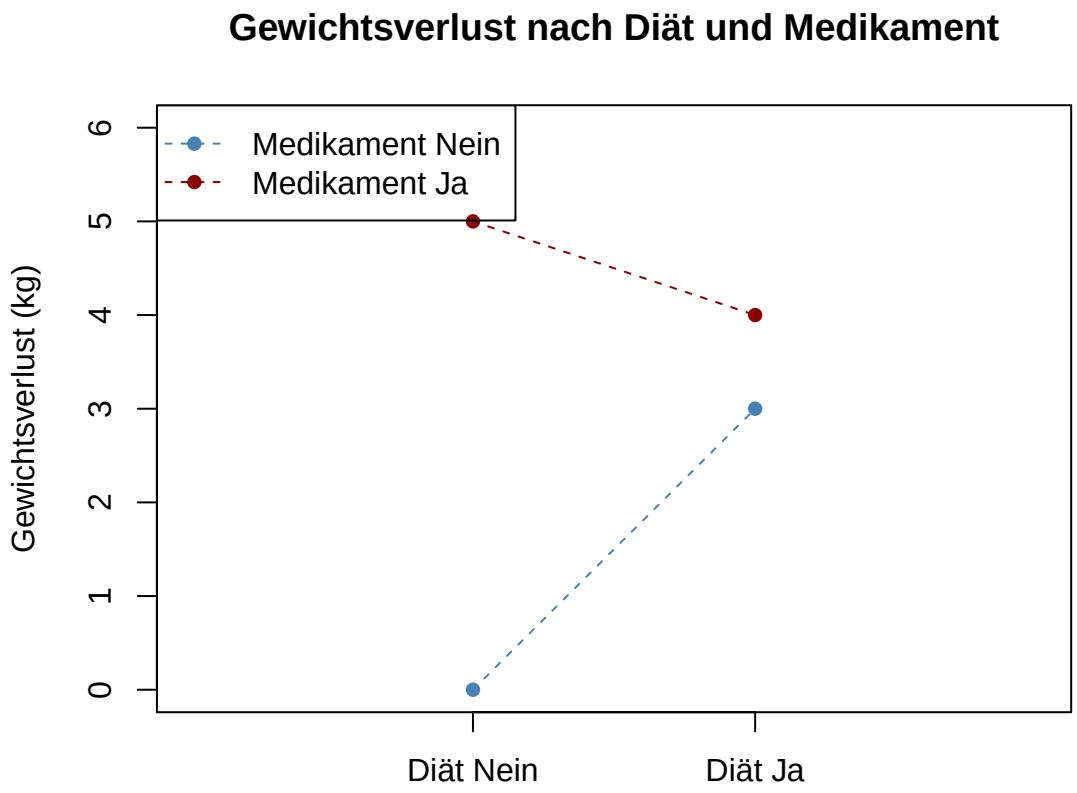
	Medikament Nein	Medikament Ja
Diät Nein	0	5
Diät Ja	3	4

In diesem Beispiel würden sich durch die Summe der Einzeleffekte (5 + 3 = 8 kg Gewichtsverlust) ein höherer Wert ergeben als die tatsächlich beobachteten 4 kg bei den Personen, die sowohl Diät hielten als auch das Medikament einnahmen.

Schlussfolgerungen:

- Es liegt eine antagonistische Interaktion vor
- Zur Erklärung der Ergebnisse muss ein negativer Interaktionsterm von -4 kg eingeführt werden (8 kg erwartet - 4 kg Interaktionseffekt = 4 kg beobachtet)
- Die kombinierte Wirkung ist schwächer als die Summe der Einzeleffekte

Der Interaktionsterm wäre in diesem Fall statistisch signifikant negativ. Grafisch würden sich die Linien in einem Interaktionsplot kreuzen oder divergieren. Dies deutet darauf hin, dass die Faktoren sich gegenseitig in ihrer Wirkung abschwächen.



Zusammenfassung der Effekte:

Effekt	Gewichtsverlust
Medikament allein	$+5$kg
Diät allein	$+3$kg
Erwartete Kombination	$+8$kg
Beobachtete Kombination	$+4$kg
Interaktionseffekt	-4kg

Tatsächlich kann eine Interaktion sowohl synergistisch mit einem Faktor als auch antagonistisch mit dem anderen wirken, insbesondere wenn die Faktoren entgegengesetzte Effekte haben. Ein Beispiel wäre eine Studie, in der:

- Eine kalorienreiche Diät (erwarteter Effekt: Gewichtszunahme)
- Mit einem Gewichtsreduktionsmedikament (erwarteter Effekt: Gewichtsabnahme)kombiniert werden.

Wie anhand der Tabellen und Grafiken ersichtlich, zeigt eine Interaktion an, dass die Mittelwertdifferenzen zwischen den Gruppen nicht konsistent sind. Beispiel:

- In der zweiten Tabelle betrug die Differenz zwischen Medikament und kein Medikament:
  - Ohne Diät: 5 - 0 = 5 kg
  - Mit Diät: 12 - 3 = 9 kg
- Grafisch äußert sich dies in unterschiedlichen Steigungen der Verbindungslinien

Es werden drei zentrale Hypothesen geprüft:

1. Haupteffekt Diät

- $H_0 : \mu_{\text{Diät}} = \mu_{\text{keine Diät}}$
- $H_1 : \mu_{\text{Diät}} \neq \mu_{\text{keine Diät}}$

2. Haupteffekt Medikament

- $H_0 : \mu_{\text{Medikament}} = \mu_{\text{kein Medikament}}$
- $H_1 : \mu_{\text{Medikament}} \neq \mu_{\text{kein Medikament}}$

3. Interaktionseffekt (zwei äquivalente Formulierungen):

   a) Unterschiedliche Medikamentenwirkung in Diät-Gruppen:

   - $H_0 : (\mu_{\text{Med}} - \mu_{\text{kein Med}})_{\text{keine Diät}} = (\mu\text{Med} - \mu_{\text{kein Med}})\text{Diät}$
   - $H_1 : (\mu\text{Med} - \mu_{\text{kein Med}})_{\text{keine Diät}} \neq (\mu\text{Med} - \mu_{\text{kein Med}})_{\text{Diät}}$

   b) Unterschiedliche Diätwirkung in Medikamentengruppen:

   - $H_0 : (\mu_{\text{Diät}} - \mu_{\text{keine Diät}})_{\text{kein Med}} = (\mu\text{Diät} - \mu_{\text{keine Diät}})\text{Med}$
   - $H_1 : (\mu\text{Diät} - \mu_{\text{keine Diät}})_{\text{kein Med}} \neq (\mu\text{Diät} - \mu_{\text{keine Diät}})_{\text{Med}}$

Die Gesamtvarianz (Sum of Squares Total, SST) wird in vier Komponenten zerlegt:

- Faktor A (Diät): SS_{Dit}
- Faktor B (Medikament): SS_{Med}
- Interaktion: $SS_{Interaktion}$
- Residualvarianz: $SS_{Residual}$

Die ANOVA-Tabelle strukturiert diese Analyse:

Quelle	Quadratsumme	Freiheitsgrade	Mittelquadrate	F-Wert	p-Wert
Faktor A	SS_A	$k_1 - 1$	$MS_A = \frac{SS_A}{df_A}$	$F_A = \frac{MS_A}{MS_{Res}}$	$P(F > F_A)$
Faktor B	SS_B	$k_2 - 1$	$MS_B = \frac{SS_B}{df_B}$	$F_B = \frac{MS_B}{MS_{Res}}$	$P(F > F_B)$
Interaktion	SS_{AB}	$(k_1-1)(k_2-1)$	$MS_{AB} = \frac{SS_{AB}}{df_{AB}}$	$F_{AB} = \frac{MS_{AB}}{MS_{Res}}$	$P(F > F_{AB})$
Residual	SS_{Res}	$n - k_1 k_2$	$MS_{Res} = \frac{SS_{Res}}{df_{Res}}$		
Total	SS_{Total}	$n - 1$			

Nach der Erstellung der ANOVA-Tabelle – üblicherweise mithilfe eines Statistikprogramms, um die umfangreichen Berechnungen zu vermeiden – folgt die Interpretation der erhaltenen p-Werte für die einzelnen Faktoren sowie für die Interaktion. Dabei kommt dem p-Wert der Interaktion eine zentrale Bedeutung zu, da er die gesamte Analyse maßgeblich beeinflusst:

- **Keine signifikante Interaktion** Liegt der p-Wert der Interaktion oberhalb des festgelegten Signifikanzniveaus (meist 0,05), kann die Wirkung der beiden Faktoren unabhängig voneinander betrachtet werden. Das bedeutet, man prüft für jeden Faktor separat, ob es signifikante Unterschiede zwischen dessen Stufen gibt. Beispiel: Bei der Analyse von Gewichtsverlusten in Abhängigkeit von Diät und Medikament zeigt sich keine signifikante Interaktion. In diesem Fall untersucht man den Effekt der Diät anhand des entsprechenden p-Werts. Ist dieser kleiner als 0,05, so ist der Faktor Diät statistisch signifikant, was bedeutet, dass sich die Gewichtsverluste der Personen mit und ohne Diät signifikant unterscheiden – unabhängig davon, ob das Medikament eingenommen wurde oder nicht. Analog wird der Effekt des Medikaments geprüft.
- **Signifikante Interaktion** Ist der p-Wert der Interaktion hingegen kleiner als das Signifikanzniveau, können die Faktoren nicht unabhängig voneinander betrachtet werden. Die Wirkung eines Faktors hängt vom jeweiligen Niveau des anderen Faktors ab. Deshalb ist es notwendig, die Effekte innerhalb der Stufen des jeweils anderen Faktors getrennt zu analysieren. Beispiel: Bei signifikantem Interaktionseffekt zwischen Diät und Medikament bedeutet dies, dass der Einfluss des Medikaments auf den Gewichtsverlust sich je nach Diätstatus unterscheidet – und umgekehrt. Es existiert somit kein allgemeingültiger Haupteffekt eines Faktors, sondern der Effekt ist kontextabhängig.

Wichtiger Hinweis:

Ein zweifaktorielles ANOVA mit je zwei Stufen pro Faktor ist nicht mit zwei unabhängigen T-Tests zu vergleichen. Selbst wenn keine Interaktion vorliegt, entsprechen die p-Werte aus dem ANOVA nicht zwangsläufig denen, die man durch separate T-Tests erhält. Das ANOVA berücksichtigt die gemeinsamen Einflüsse beider Faktoren und eliminiert den Anteil der Variabilität, der durch den jeweils anderen Faktor verursacht wird. Dadurch liefert es eine multivariate Betrachtung der Effekte. Im Beispiel bedeutet dies, dass der Einfluss der Diät auf den Gewichtsverlust unter Berücksichtigung der Wirkung des Medikaments (und gegebenenfalls deren Interaktion) bewertet wird – was bei einfachen T-Tests unberücksichtigt bliebe. Ebenso entspricht die Analyse der Interaktion nicht der Durchführung eines einfaktoriellen ANOVA mit einer Faktorstufe, die alle Kombinationen zusammenfasst, da die multivariate Struktur des zweifaktoriellen ANOVA eine differenziertere Betrachtung ermöglicht.

10.2.2 Zweifaktorielles ANOVA mit drei oder mehr Stufen in einem Faktor

Das Vorgehen bei einem zweifaktoriellen ANOVA, bei dem mindestens ein Faktor drei oder mehr Stufen aufweist, ähnelt dem bereits beschriebenen Fall mit je zwei Stufen pro Faktor. Lediglich die Nullhypothesen werden erweitert, indem die Gleichheit aller Mittelwerte der jeweiligen Stufen eines Faktors geprüft wird, und die Alternativhypothese besagt, dass mindestens ein Mittelwert sich unterscheidet. Bei den Interaktionen werden ebenfalls Mittelwertunterschiede betrachtet, wobei aufgrund der höheren Anzahl von Stufen mehr mögliche Unterschiede vorliegen.

Zur Interpretation der ANOVA-Ergebnistabelle gilt: Liegt keine signifikante Interaktion vor, aber signifikante Unterschiede in einem der Faktoren mit drei oder mehr Stufen, so folgt eine genauere Untersuchung, zwischen welchen Stufen diese Unterschiede bestehen. Beispiel: Wenn der Faktor 1 drei Stufen hat, keine Interaktion vorliegt und die Nullhypothese der Mittelwertgleichheit abgelehnt wird, prüft man, ob die Unterschiede beispielsweise zwischen Stufe 1 und 2, 1 und 3 oder 2 und 3 bestehen – und zwar unabhängig vom Faktor 2. Um die Unterschiede zwischen den Stufen zu identifizieren, führt man Multiple Paarvergleiche durch, etwa den Bonferroni-Test oder andere aus dem einfaktoriellen ANOVA bekannte Verfahren. Bei signifikanter Interaktion wird analog vorgegangen, jedoch werden die Unterschiede innerhalb jeder Stufe des jeweils anderen Faktors separat betrachtet.

Wie bereits beim zweifaktoriellen ANOVA mit zwei Stufen und den unabhängigen T-Tests erläutert, entspricht ein zweifaktorielles ANOVA mit drei oder mehr Stufen nicht der Durchführung zweier separater einfaktorieller ANOVAs. Die erhaltenen p-Werte unterscheiden sich – selbst wenn keine Interaktion vorliegt –, da die gemeinsame Varianzstruktur und Einflüsse berücksichtigt werden.

10.2.3 ANOVA mit drei oder mehr Faktoren

Die Grundprinzipien des ANOVA mit drei oder mehr Faktoren sind ähnlich denen des zweifaktoriellen ANOVA, und die Ergebnistabellen weisen vergleichbare Strukturen auf. Die Interpretation wird jedoch deutlich komplexer. Ein dreifaktorielles ANOVA enthält die Haupteffekte der drei Faktoren, die drei möglichen zweifachen Interaktionen (Faktor 1 mit 2, 1 mit 3, 2 mit 3) sowie optional die dreifache Interaktion aller Faktoren (je nach Software und Einstellungen können Interaktionen beliebiger Ordnung berücksichtigt werden).

Ist die dreifache Interaktion signifikant, kann kein allgemeiner Haupteffekt eines Faktors betrachtet werden. Stattdessen muss der Effekt dieses Faktors innerhalb jeder Stufe der beiden anderen Faktoren separat analysiert werden. Liegt keine signifikante dreifache Interaktion vor, wohl aber eine zweifache (beispielsweise zwischen Faktor 1 und 2), so wird der Effekt von Faktor 1 innerhalb der Stufen von Faktor 2 untersucht, unabhängig von Faktor 3.

Dies führt zu einer Vielzahl möglicher Analysen und multipler Vergleichstests. Daher liegt es in der Verantwortung des Forschers, die Anzahl der durchgeführten Tests sinnvoll zu begrenzen, den Versuchsaufbau klar zu definieren und ggf. höhere Interaktionen (dreifach oder mehr) nur dann zu berücksichtigen, wenn deren Interpretation auch tatsächlich möglich und sinnvoll ist.

Abschließend gilt auch hier: Ein ANOVA mit drei oder mehr Faktoren entspricht nicht der Durchführung mehrerer einfaktorieller ANOVAs für jeden Faktor einzeln.

10.2.4 Feste Faktoren und Zufallsfaktoren

Beim mehrfaktoriellen ANOVA unterscheiden sich der Umgang mit der durch jeden Faktor verursachten Variabilität und die daraus folgenden Schlussfolgerungen je nachdem, ob die Faktoren als feste oder zufällige Faktoren behandelt werden.

Ein fester Faktor (oder Faktor mit festen Effekten) ist ein Faktor, dessen Stufen vom Untersucher vorab festgelegt oder bestimmt werden (zum Beispiel festgelegte Dosierungen eines Medikaments oder vorgegebene Zeitpunkte) oder die sich aus der Natur des Faktors ergeben (z. B. Geschlecht oder Diät). Die Variabilität solcher Faktoren ist leichter kontrollierbar, und ihre Behandlung in den Berechnungen zum ANOVA-Endergebnis ist einfacher. Ein Nachteil besteht jedoch darin, dass nur für die konkret untersuchten Stufen Aussagen getroffen werden können. Das heißt, die Ergebnisse gelten nur für genau diese festen Stufen, und eine Verallgemeinerung auf andere, nicht untersuchte Stufen ist nicht zulässig.

Im Gegensatz dazu ist ein Zufallsfaktor (oder Faktor mit zufälligen Effekten) ein Faktor, dessen Stufen zufällig aus einer größeren Grundgesamtheit von möglichen Stufen ausgewählt wurden (beispielsweise eine Medikamentendosierung mit den Stufen 23 mg, 132 mg und 245 mg, die zufällig aus dem Bereich 0 bis 250 mg gezogen wurden). Die Analyse solcher Faktoren ist komplexer, da diese Stufen eine Stichprobe aller möglichen Stufen darstellen. Ziel ist es, Rückschlüsse auf die gesamte Population aller möglichen Stufen zu ziehen.

10.2.4.1 Modellannahmen des mehrfaktoriellen ANOVA

Wie beim einfaktoriellen ANOVA handelt es sich auch beim ANOVA mit zwei oder mehr Faktoren um einen parametrischen Test mit folgenden Voraussetzungen:

- Die Daten innerhalb jeder Kategorie – definiert als Kombination aller Stufen aller Faktoren – folgen einer Normalverteilung. Beispiel: Bei einem zweifaktoriellen ANOVA mit je drei Stufen pro Faktor gibt es $3^2 = 9$ Kategorien.
- Die Varianzen dieser Normalverteilungen sind gleich (Varianzhomogenität bzw. Homoskedastizität).

Werden diese Voraussetzungen nicht erfüllt und liegen zudem kleine Stichproben vor, sollte ein mehrfaktorieller ANOVA nicht angewendet werden. Ein zusätzliches Problem ist, dass es für diesen Fall keinen entsprechenden nicht-parametrischen Ersatztest gibt. Zwar können mit nicht-parametrischen Tests (wie dem Kruskal-Wallis-Test) die Einflüsse der einzelnen Faktoren separat geprüft werden, die wichtige Untersuchung der Interaktion zwischen Faktoren ist jedoch nicht möglich.

10.3 ANOVA mit Messwiederholungen

In vielen Fragestellungen wird eine abhängige Variable mehrfach am selben Individuum gemessen (z. B. bei einer Gruppe, die eine bestimmte Diät verfolgt, wird das verlorene Gewicht nach einem, zwei und drei Monaten erfasst). Ziel ist es, den Mittelwert der Variable über die verschiedenen Messzeitpunkte zu vergleichen, also zu prüfen, ob sich die Variable über die Zeit verändert (im Beispiel: Entwicklung des Gewichtsverlusts). Konzeptuell entspricht dies der Situation bei paarweise gepaarten Messwerten, wie sie etwa mit dem gepaarten t-Test oder dem nichtparametrischen Wilcoxon-Test verglichen werden, jedoch liegen hier mehr als zwei gepaarte Messungen pro Individuum vor. In solchen Fällen verwendet man die ANOVA mit Messwiederholungen.

Diese ANOVA hat wie alle Tests mit gepaarten Daten den Vorteil, dass die Vergleiche innerhalb desselben Individuums (intraindividuell) erfolgen. Dadurch wird die Variabilität, die zwischen verschiedenen Individuen auftritt, reduziert. Zum Beispiel könnte man das Gewicht in drei verschiedenen Gruppen messen, die alle dieselbe Diät machen, aber mit unterschiedlichen Alters- oder Geschlechtsverteilungen – das würde die Ergebnisse beeinflussen. Mit Messwiederholungen werden alle Messungen innerhalb desselben Individuums verglichen, sodass z.B. Geschlecht, Alter oder sportliche Aktivität konstant bleiben. Dadurch werden kleine Unterschiede besser erkennbar.

10.3.1 ANOVA mit Messwiederholungen als zweifaktorieller ANOVA ohne Interaktion

Die ANOVA mit Messwiederholungen kann als zweifaktorieller ANOVA ohne Interaktion berechnet werden, wenn die Daten entsprechend aufbereitet werden.

Angenommen, es liegen k gepaarte Messungen einer numerischen abhängigen Variable bei n Individuen vor, dann können die Daten wie folgt dargestellt werden:

Individuum	VarDep 1	VarDep 2	...	VarDep k
Individuum 1	$x_{1,1}$	$x_{1,2}$	...	$x_{1,k}$
Individuum 2	$x_{2,1}$	$x_{2,2}$	...	$x_{2,k}$
...	...	...	...	...
Individuum n	$x_{n,1}$	$x_{n,2}$	...	$x_{n,k}$

Diese Daten können auch in einem “langen” Format angeordnet werden, das sich gut für eine zweifaktorielle ANOVA eignet:

Var Dep	Individuum	Messzeitpunkt
$x_{1,1}$	1	1
$x_{2,1}$	2	1
...	...	...
$x_{n,1}$	n	1
$x_{1,2}$	1	2
$x_{2,2}$	2	2
...	...	...
$x_{n,2}$	n	2

Var Dep	Individuum	Messzeitpunkt
...	...	...
$x_{1,k}$	1	k
$x_{2,k}$	2	k
...	...	...
$x_{n,k}$	n	k

Hier sind *Individuum* und *Messzeitpunkt* kategoriale Variablen, die die Gesamtheit der Daten ($n \times k$) in Gruppen einteilen: n Gruppen für Individuen und k Gruppen für Messzeitpunkte. Die Kreuzung der beiden Variablen erzeugt $n \times k$ Gruppen mit jeweils genau einem Messwert.

Die Variabilität der abhängigen Variablen wird aufgeteilt in drei Quellen:

- Variabilität durch Messzeitpunkt
- Variabilität durch Individuum
- Residuale Variabilität

Eine Interaktion zwischen Messzeitpunkt und Individuum kann hier nicht analysiert werden, da jede Gruppe nur einen Wert enthält und somit keine Streuung berechnet werden kann. Daher wird eine zweifaktorielle ANOVA **ohne Interaktion** durchgeführt, die folgende Tabelle liefert:

Quelle	Quadratsummen	Freiheitsgrade	Mittelquadrat	F-Wert	p-Wert
F1 = Messzeitpunkt	$SF1$	$GF1 = k - 1$	$CF1 = \frac{SF1}{GF1}$	$f1 = \frac{CF1}{CR}$	$P(F > f1)$
F2 = Individuum	$SF2$	$GF2 = n - 1$	$CF2 = \frac{SF2}{GF2}$	$f2 = \frac{CF2}{CR}$	$P(F > f2)$
Residuale	SR	$GR = (n \cdot k) - 1 - GF1 - GF2$	$CR = \frac{SR}{GR}$		
Total	ST	$GT = (n \cdot k) - 1$			

Die Hypothesen, die mit diesem Modell getestet werden, sind:

1. **Für den Faktor Messzeitpunkt:**

   $$H_0 : \mu_{\text{Messzeitpunkt 1}} = \mu_{\text{Messzeitpunkt 2}} = \cdots = \mu_{\text{Messzeitpunkt k}}$$
   $$H_1 : \text{Mindestens ein Mittelwert ist unterschiedlich.}$$

   Ist der p-Wert kleiner als das Signifikanzniveau, so gibt es einen signifikanten Effekt der Messzeitpunkte, d.h., die Variable ändert sich über die Zeit innerhalb der Individuen.

2. **Für den Faktor Individuum:**

   $$H_0 : \mu_{\text{Individuum 1}} = \mu_{\text{Individuum 2}} = \cdots = \mu_{\text{Individuum n}}$$
   $$H_1 : \text{Mindestens ein Mittelwert ist unterschiedlich.}$$

   Ein signifikanter Effekt hier zeigt, dass Individuen sich im Mittel unterschiedlich verhalten. Dies ist jedoch meist erwartbar und weniger interessant im Kontext der Messwiederholungen.

Falls mindestens eine der Nullhypothesen verworfen wird, empfiehlt sich als nächster Schritt ein Multiple-Comparison-Test, z. B. der Bonferroni-Test, um herauszufinden, welche Mittelwerte sich unterscheiden – insbesondere bei den Messzeitpunkten.

10.3.1.1 Voraussetzungen der ANOVA mit Messwiederholungen

Wie bei jeder anderen ANOVA gilt auch für die ANOVA mit Messwiederholungen:

- Die Daten der abhängigen Variablen sollten innerhalb jeder Gruppe normalverteilt sein, egal ob die Gruppen durch die Variable *Messzeitpunkt* oder durch die Variable *Individuum* gebildet werden. Da der wichtigste Test bei der Variable *Messzeitpunkt* durchgeführt wird, ist es besonders wichtig, dass die Verteilungen aller Messzeitpunkte normal sind.
- Alle Normalverteilungen sollten die gleiche Varianz aufweisen (Homoskedastizität), besonders die der verschiedenen Messzeitpunkte.

Wenn bei einer ANOVA mit Messwiederholungen sowohl Normalität als auch Homoskedastizität erfüllt sind, spricht man von *Sphärizität* der Daten. Es gibt spezielle statistische Tests, um die Sphärizität zu überprüfen, beispielsweise den *Mauchly-Test*.

Werden die genannten Bedingungen nicht erfüllt und sind die Stichproben zudem klein, sollte die ANOVA mit Messwiederholungen nicht angewendet werden. Allerdings existiert zumindest ein nichtparametrischer Test, der überprüft, ob es signifikante Unterschiede zwischen den verschiedenen Stufen der Variable *Messzeitpunkt* gibt: der *Friedman-Test*.

10.3.2 ANOVA mit Messwiederholungen + ANOVA mit einem oder mehreren Faktoren

Es gibt viele Fälle, in denen neben der Analyse des intraindividuellen Effekts auf eine mehrfach gemessene quantitative abhängige Variable (ANOVA mit Messwiederholungen) auch qualitative Variablen berücksichtigt werden sollen, die vermutlich mit der abhängigen Variable zusammenhängen. Diese qualitativen Variablen führen zu Effekten, die üblicherweise als interindividuelle Effekte klassifiziert werden, die man aber auch als *intergruppale* Effekte verstehen kann, da sie Gruppen definieren, zwischen denen eine ANOVA mit einem oder mehreren Faktoren möglich ist.

Beispielsweise könnte man den Gewichtsverlust bei Individuen nach einem, zwei und drei Monaten (ANOVA mit Messwiederholungen) untersuchen, wobei die Individuen in sechs Gruppen aufgeteilt sind, die sich aus der Kreuzung zweier Faktoren ergeben: *Diät* (a, b, c) und *Bewegung* (niedrig, hoch). Um den Einfluss dieser beiden interindividuellen Faktoren zu untersuchen, wäre eine zweifaktorielle ANOVA mit Interaktion notwendig.

Obwohl die Daten ähnlich wie bei einer zweifaktoriellen ANOVA mit den Faktoren *Messzeitpunkt* und *Individuum* strukturiert werden könnten, um zwei weitere Faktoren (Diät und Bewegung) hinzuzufügen, ist es unpraktisch, für dasselbe Individuum mehrere Zeilen (eine pro Messzeitpunkt) einzufügen.

Deshalb erlauben bestimmte Statistikprogramme, wie z. B. PASW (SPSS), ANOVAs mit Messwiederholungen auf Basis des klassischen Formats (eine Zeile pro Individuum, eine Variable pro Messzeitpunkt). Dabei werden intraindividuelle Faktoren definiert, die alle Messzeitpunkte umfassen, und es können interindividuelle Faktoren (Kategorien) hinzugefügt werden, die die Antwortvariablen beeinflussen können. Zudem können Interaktionen zwischen den interindividuellen Faktoren sowie zwischen inter- und intraindividuellen Faktoren getestet werden.

Diese Verfahren kombinieren somit eine ANOVA mit Messwiederholungen und eine ANOVA mit einem oder mehreren Faktoren, wobei die Daten weiterhin klassisch eingegeben werden können: eine Zeile pro Individuum.

Die Ergebnisse ähneln den zuvor beschriebenen: Es werden ANOVA-Tabellen erzeugt, in denen für jeden Faktor (intraindividuell bzw. Messwiederholung, interindividuell bzw. Kategorien) sowie für die Interaktionen p-Werte berechnet werden.

10.4 Kovarianzanalyse: ANCOVA

Die Kovarianzanalyse (ANCOVA) ist eine Erweiterung der ANOVA (sei es einfaktorielle, mehrfaktorielle oder mit Messwiederholungen), die es erlaubt, den Einfluss aller kategorialen unabhängigen Variablen (Faktoren) und der Messwiederholungen auf die quantitative abhängige Variable zu analysieren, dabei aber zusätzlich den Effekt anderer quantitativer unabhängiger Variablen auf die Antwortvariable zu eliminieren. Diese unabhängigen Variablen, deren Effekt kontrolliert oder herausgerechnet werden soll, nennt man *Kovariablen* oder *Kovariate*, da man erwartet, dass sie mit der abhängigen Variable kovariieren, also korreliert sind.

Die detaillierte Durchführung einer ANCOVA geht über den Rahmen dieser Übung hinaus, die Grundidee ist aber einfach: Man kann eine Regressionsanalyse der abhängigen Variable in Abhängigkeit von der Kovariablen (oder mehreren Kovariablen) durchführen und den Teil der Variabilität der abhängigen Variable, der durch die Kovariable erklärt werden kann, eliminieren. Dies geschieht, indem man mit den Residuen des Regressionsmodells arbeitet anstatt mit den Originaldaten. Anschließend wird eine ANOVA mit einem oder mehreren Faktoren, auch mit Messwiederholungen, auf den Residuen durchgeführt.

Das Endergebnis der ANCOVA ist eine Tabelle, die der der ANOVA sehr ähnlich ist, aber mit zusätzlichen Zeilen für jede Kovariable. Diese Zeilen zeigen, wie viel Variabilität durch jede Kovariable erklärt wird, sowie deren zugehörigen p-Wert. Dieser p-Wert beantwortet die Frage, ob die Kovariable notwendig ist, um die abhängige Variable zu erklären (technisch ausgedrückt: ob die Steigung der Regressionsfunktion der unabhängigen Variable in Abhängigkeit von der Kovariable gleich Null sein kann oder nicht). In der ANCOVA-Tabelle gibt es keine zusätzliche Zeile für die mögliche Interaktion zwischen Kovariable und den verschiedenen interindividuellen Faktoren, weil ein ANCOVA-Modell bei vorhandener Interaktion nicht sinnvoll ist: Der Effekt des Faktors könnte nicht geschätzt werden, da er vom konkreten Wert der Kovariablen abhängt, welche als kontinuierliche Variable unendlich viele Werte annimmt. Dadurch gäbe es unendlich viele verschiedene Effekte des Faktors und kein eindeutiger p-Wert.

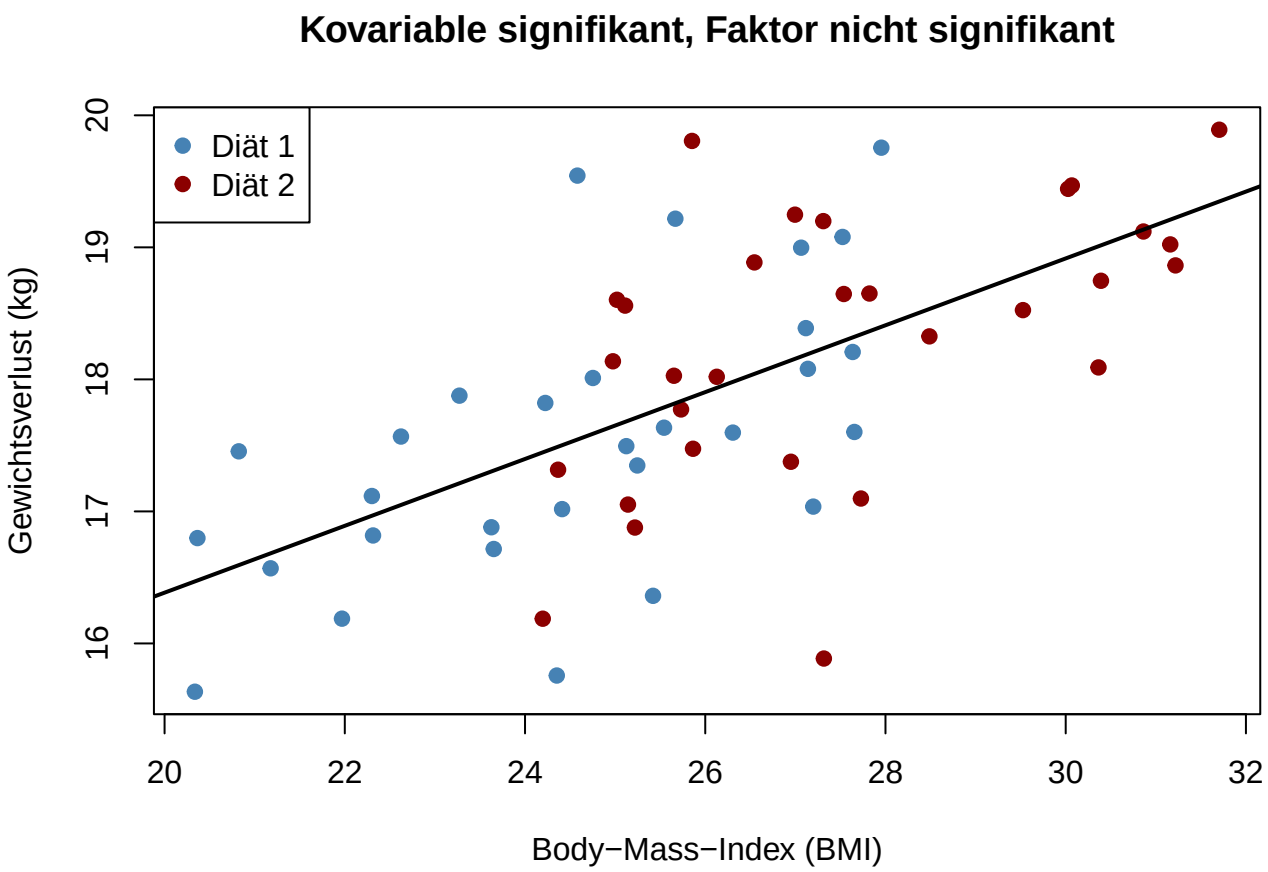
Allerdings enthält die Tabelle eine Zeile für die Interaktion jedes intraindividuellen Faktors mit jeder Kovariable, da jeder intraindividuelle Faktor aus mehreren quantitativen Variablen besteht, die in der Regression unterschiedliche Steigungen in Abhängigkeit von der Kovariable aufweisen können.

Wenn bei einer ANOVA zur Überprüfung des Einflusses verschiedener Faktoren auf eine quantitative Antwortvariable meist ein *Mittelwertdiagramm* verwendet wird, zeigt sich bei der ANCOVA der Effekt der Kovariablen auf die Antwortvariable durch ein *Punktwolken-Diagramm* der Antwortvariable in Abhängigkeit von der Kovariablen. Dieses wird je nach Grad der linearen Korrelation zwischen beiden mehr oder weniger linear aussehen. Zudem lässt sich erahnen, ob ein bestimmter Faktor nach Herausrechnung des Einflusses der Kovariablen noch einen Einfluss auf die Antwortvariable hat:

- Wenn die Punktwolke durch eine einzige Gerade mit Null-Steigung gut beschrieben werden kann, unabhängig von den Faktorstufen, bedeutet dies, dass weder die Kovariable noch der Faktor signifikant sind, um die Antwortvariable zu erklären.
- Wenn die Punktwolke durch eine einzige Gerade mit nicht null Steigung gut beschrieben werden kann, unabhängig von den Faktorstufen, bedeutet dies, dass die Kovariable signifikant ist, der Faktor aber nicht. Denn nach Herausrechnung des Einflusses der Kovariablen (also wenn man die Residuen der Anfangsregression betrachtet) gibt es keine Unterschiede zwischen den Faktorstufen (die Punkte der verschiedenen Kategorien liegen auf gleicher Höhe).

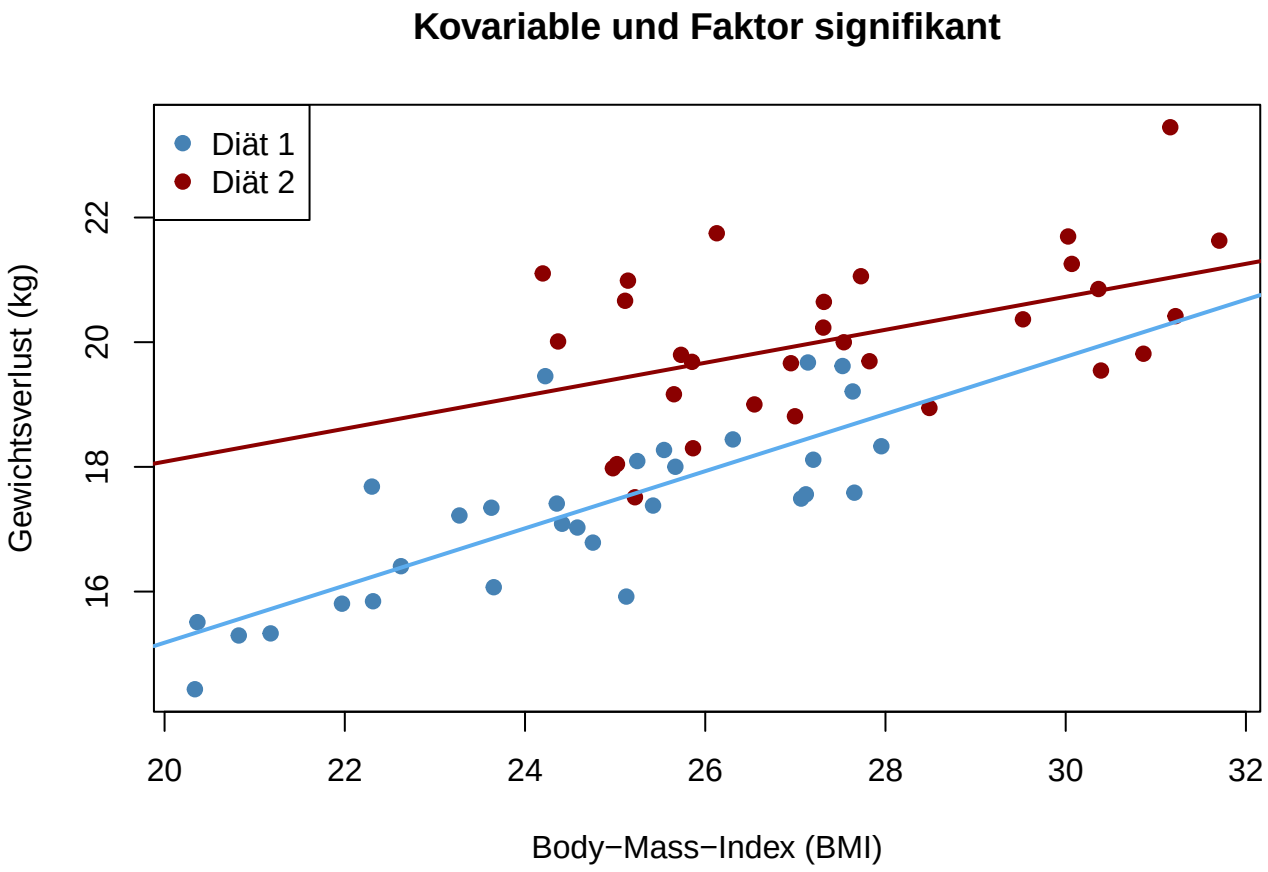
  Zum Beispiel zeigt die folgende Abbildung das Ergebnis eines Experiments, bei dem die verlorenen Kilogramm von Personen gemessen wurden, die zwei verschiedene Diäten befolgten. Als Kovariable wurde der Body-Mass-Index (BMI) berücksichtigt, der ebenfalls Einfluss auf den Gewichtsverlust haben könnte, aber bei der Gruppeneinteilung nicht kontrolliert wurde. Die Gruppe mit Diät 2 hat im Mittel einen höheren BMI als die Gruppe mit Diät 1. Laut Abbildung scheint es signifikante Unterschiede im Gewichtsverlust zwischen den Diäten zu geben (Diät 2 führt zu mehr Gewichtsverlust), aber tatsächlich ist der Unterschied auf die Kovariable zurückzuführen. Wird der Effekt der Kovariablen eliminiert (also die Steigung der Geraden auf 0 gesetzt), hätten beide Gruppen

etwa gleich viel Gewicht verloren. Das bedeutet, dass bei gleichem BMI Diät 2 nicht zu mehr Gewichtsverlust führt.



- Wenn die Punktwolke durch mehrere Geraden mit Null-Steigung, eine pro Faktorstufe, gut beschrieben wird, ist die Kovariable nicht signifikant, aber der Faktor schon.
- Wenn die Punktwolke durch Geraden mit gleicher, nicht null Steigung, eine pro Faktorstufe, beschrieben wird und mindestens eine Gerade sich von den anderen unterscheidet (mindestens eine Faktorstufe ist verschoben), sind sowohl Kovariable als auch Faktor signifikant bei der Erklärung der abhängigen Variable.

  Zum Beispiel zeigt die folgende Abbildung das Ergebnis eines ähnlichen Experiments wie oben: Gewichtsverlust in Abhängigkeit von Diät und Kovariable BMI, die bei der Gruppeneinteilung nicht kontrolliert wurde (die Gruppe mit Diät 2 hat einen höheren BMI). Die Grafik deutet sogar auf signifikante Unterschiede im Gewichtsverlust hin, sodass die Diät 2 mehr Gewichtsverlust verursachen würde, was aber alles auf die Kovariable zurückzuführen ist. Nach Eliminierung des Kovariablen-Effekts ist der Gewichtsverlust der Diät-1-Gruppe größer (wenn die Steigung der Geraden entfernt wird, liegen die Punkte der Diät 1 höher).



- Wenn die Punktwolke durch unterschiedliche Geraden, eine pro Faktorstufe, mit unterschiedlichen, nicht null Steigungen beschrieben wird, deutet dies auf eine Interaktion zwischen Kovariabler und Faktor hin, und es sollte kein ANCOVA-Modell verwendet werden.

Literatur

Grabinger, B. (2024). *Fit fürs Studium – Statistik* (3. Aufl.). Rheinwerk Computing. https://www.rheinwerk-verlag.de/fit-fuers-studium-statistik/

große Schlarmann, J. (2024). *Angewandte Übungen in R*. Hochschule Niederrhein. https://github.com/produnis/angewandte_uebungen_in_R

große Schlarmann, J. (2025a). *Statistik mit R und RStudio - Ein Nachschlagewerk für Gesundheitsberufe*. Hochschule Niederrhein. https://www.produnis.de/R

große Schlarmann, J. (2025b). *table traineR. Workouts für {data.table}*. Hochschule Niederrhein. https://www.produnis.de/tabletrainer/

große Schlarmann, J. (2025c). *trainingslageR. Ein Übungsbuch für R-Einsteiger*innen und Fortgeschrittene*. Hochschule Niederrhein. https://www.produnis.de/trainingslager

Koller, M. M. (2017). *Statistik für Pflege- und andere Gesundheitsberufe* (2. Aufl.). Facultas.

Müller, M. (2011). *Statistik für die Pflege: Handbuch für Pflegeforschung und -wissenschaft* (1. Aufl.). Hogrefe AG.

Sánchez Alberca, A., Conesa Egea, J., Garro, J. C., López Ramírez, E., Oncala Mesa, R., & Zaragoza de Lorite, A. (2023). *Elementary Statistics Manual*. https://github.com/asalber/statistics-manual

von der Lippe, P. (2002). Repräsentativität, convenience samples und Signifikanztests. Wie das Konzept "Repräsentativität" zu Fehlanwendungen der Statistik verleitet. *Marketing*, *24*, 227–238. http://www.von-der-lippe.org/dokumente/Repräsentativität_4.pdf

von der Lippe, P. (2010). Kapitel 10 - Stichprobentheorie. In *Induktive Statistik, Formeln, Aufgaben, Klausurtraining*. http://www.von-der-lippe.org/dokumente/INKAP10L.pdf

von der Lippe, P. (2011). *Wie groß muss meine Stichprobe sein, damit sie repräsentativ ist?* Universität Duisburg-Essen, Diskussionsbeitrag Nr. 187. http://www.von-der-lippe.org/dokumente/Wieviele.pdf

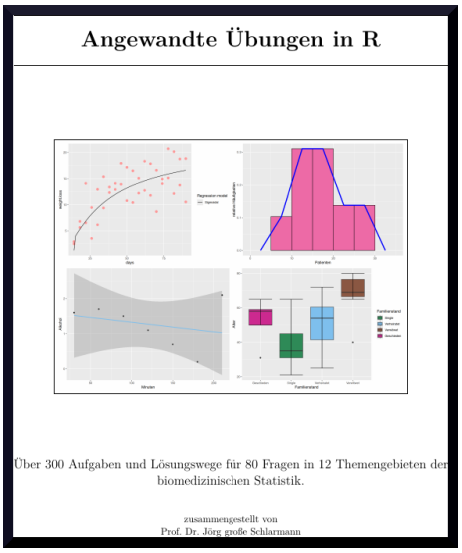
von der Lippe, P., & Kladroba, A. (2002). Repräsentativität von Stichproben. *Marketing*, *24*. http://www.von-der-lippe.org/dokumente/Repraesentativitaet.pdf

Bildquellen

- **?@fig-windmuehle1**: Alfredo Sánchez Alberca

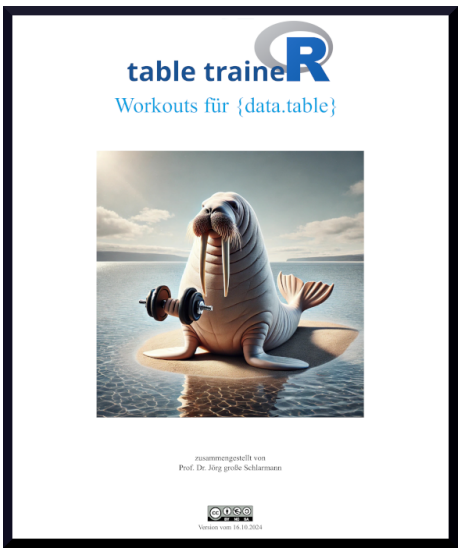
Credits

(a) große Schlarmann (2025a)



(a) große Schlarmann (2024)

(a) große Schlarmann (2025c)



(a) große Schlarmann (2025b)

Prof. Dr. Jörg große Schlarmann, BScN, MScN, RN

Hochschule Niederrhein, Krefeld

joerg.grosseschlarmann@hs-niederrhein.de

https://github.com/produnis/trainingslageR