

# Statistik 1

Daniel J. F. Gerber

21 October, 2024



# Inhaltsverzeichnis

<b>Vorwort</b>	<b>5</b>
<b>1 Einleitung</b>	<b>7</b>
1.1 Worum geht es? . . . . .	7
1.2 Inhaltlicher Aufbau . . . . .	7
1.3 Wie soll ich dieses Buch lesen? . . . . .	8
1.4 Formeln . . . . .	9
1.5 Software . . . . .	9
<b>I Eine intervallskaliertes Merkmal</b>	<b>11</b>
<b>2 Intervallskalierte Merkmale</b>	<b>13</b>
2.1 Was ist ein intervallskaliertes Merkmal? . . . . .	13
2.2 Wie kann ein intervallskaliertes Merkmal beschrieben werden? . .	14
2.3 Übungen . . . . .	16
<b>3 Stichprobenziehung</b>	<b>19</b>
3.1 Was ist das Problem der Stichprobenziehung? . . . . .	20
3.2 Wie kann man Aussagen über die Grundgesamtheit machen? . .	22
3.3 Übungen . . . . .	23
<b>4 Durchschnitt und Standardabweichung schätzen</b>	<b>25</b>
4.1 Wo liegt der Durchschnitt der Grundgesamtheit? . . . . .	25
4.2 Übungen . . . . .	33

<b>5</b>	<b>Durchschnitt testen</b>	<b>37</b>
5.1	Entspricht der Durchschnitt der Grundgesamtheit einem gewissen Wert? . . . . .	37
5.2	Weicht der gefundene Durchschnitt stark vom hypothetischen Wert ab? . . . . .	37
5.3	Übungen . . . . .	37
<b>II</b>	<b>Gruppenvergleich einer intervallskalierten Variable</b>	<b>39</b>
<b>6</b>	<b>Gruppenvergleich einer intervallskalierten Variable</b>	<b>41</b>
6.1	Zwei Gruppen vergleichen . . . . .	41
6.2	Was ist das Problem der Stichprobenziehung? . . . . .	41
6.3	Wie kann man Aussagen über die Grundgesamtheit machen? . .	41
6.4	Übungen . . . . .	41
<b>7</b>	<b>Welch-Test</b>	<b>43</b>
7.1	Zwei Gruppen vergleichen . . . . .	43
7.2	Sind die Durchschnitte der beiden Gruppen in der Grundgesamtheit gleich? . . . . .	43
7.3	Wie stark unterscheiden sich die Durchschnitte? . . . . .	43
7.4	Übungen . . . . .	43
	<b>Begriffsverzeichnis</b>	<b>45</b>

# Vorwort

Dieses Buch ist im Rahmen meiner Lehrtätigkeit an der FHNW entstanden und frei verfügbar.



# Kapitel 1

## Einleitung

### 1.1 Worum geht es?

### 1.2 Inhaltlicher Aufbau

Dieses Buch umfasst die untenstehenden Inhalte. Die Inhalte wurden hier nach Zwecken sortiert angeordnet:

Stichprobe beschreiben (**deskriptive Statistik**):

- Arithmetisches Mittel
- Median
- Quantile
- Anteil
- Odds Ratio
- Relatives Risiko

Population beschreiben (**Wahrscheinlichkeitslehre**):

- Zufallsvariable
- Erwartungswert
- Standardabweichung
- Varianz
- Wahrscheinlichkeitsdichte
- Wahrscheinlichkeitsverteilung
- Verteilungen

Populationsparameter aus Stichproben schätzen (**Konfidenzintervalle** + Stichprobengröße):

- Mittelwert
- Standardabweichung
- Anteil
- Berichten
- Darstellen

Aussagen auf die Population aufgrund von Stichproben machen (Test-Theorie):

- Effektstärke
- Berichten
- T-Test (1 Stichprobe)
- T-Test (2 Stichproben), Welch-Test
- Welch Test
- U-Test
- Korrelation absichern gegen 0
- Vierfelder/Mehrfeldertest

Zusammenhänge beschreiben (Zusammenhangsmasse):

- Pearsons  $r$
- Spearmans  $\rho$
- Vierfelderkorrelation /  $\Phi$
- Punktbiseriale Korrelation
- Kontingenzkoeffizient
- Cramér's  $V$

Die Inhalte nach Zweck zu gruppieren ist eine Option, die andere ist die Verfahren der Skalierung der Variablen folgend aufzubauen. Bei dieser Gruppierung ist der Zweck nicht direkt ersichtlich, dafür ist einfacher zu begreifen welches Verfahren für welche Ausgangslage geeignet ist. Diese Gruppierung wurde für die Präsentation der Inhalte in diesem Buch gewählt.

### 1.3 Wie soll ich dieses Buch lesen?

Dieses Buch enthält zu jedem Thema eine kurze Beschreibung der Theorie, Beispiele und Übungen. Das selbstständige Lösen der Übungen ist unerlässlich für das Verständnis und die Emanzipation im korrekten Umgang mit Daten. Ohne Übungen fehlt die Auseinandersetzung mit dem Unterrichtsstoff und ohne diese fällt es den allermeisten schwer sogar einfachste Zusammenhänge zu begreifen. Es wird deshalb empfohlen, dass die Übungen zum jeweiligen Thema zeitnah zur Theorie gelöst werden. Damit überprüft werden kann, ob die Übungen richtig gelöst wurden, ist zu jeder Übung eine kurze Lösung hinterlegt. Wer beim ersten selbstständigen Versuch der Übungslösung scheitert - was garantiert



den meisten Lesenden hier ein oder mehrmals passieren wird -, kann die Übung mit Hilfe der Lösung lösen und zu einem späteren Zeitpunkt die Übung selbstständig nochmal machen ohne Lösung. Für die Statistik ist es also *nicht* genug den Stoff einmal auswendig zu lernen, Übung ist unerlässlich.

## 1.4 Formeln

Die Statistik bedient sich der universellen Sprache der Formeln. Es ist deshalb unerlässlich einige Formeln zu verstehen. Das Verständnis von Formeln ist für ungeübte Lesende verwirrend und schwierig. Deshalb wird dieses Verständnis in diesem Buch nach und nach aufgebaut. Dazu werden Teilformeln isoliert und erklärt und die Einflüsse der verschiedenen Kenngrößen in der Formel exploriert.

## 1.5 Software

Für die Lösung der Übungen wird oft die freie Software Jamovi verwendet. Dem Leser wird deshalb empfohlen diese Software zu installieren. Für die Erstellung dieses Buches wurden ferner die folgenden Softwareprodukte verwendet:

- Jamovi software (Version 2.3.21.0)
- Jamovi R-package (Selker et al., 2024)
- R (R Core Team, 2024)
- Tidyverse (Wickham et al., 2019)
- Bookdown (Xie, 2016)



## Teil I

# Eine intervallskaliertes Merkmal



## Kapitel 2

# Intervallskalierte Merkmale

### 2.1 Was ist ein intervallskaliertes Merkmal?

Ein Merkmal ist dann **intervallskaliert**, wenn die einzelnen Beobachtungen in eine natürliche Reihenfolge gebracht werden können und zwischen dem tiefsten und höchsten möglichen Wert, alle erdenklichen Zwischenwerte möglich sind.

Ein Beispiel für ein intervallskaliertes Merkmal ist die Körpertemperatur. Beobachtungen der Körpertemperatur einer lebenden Person sind Werte zwischen ungefähr  $10^{\circ}\text{C}$  und  $42^{\circ}\text{C}$ . Es ist möglich zu sagen, dass eine Person mit  $40^{\circ}\text{C}$  Körpertemperatur eine höhere Temperatur hat als eine mit  $38^{\circ}\text{C}$  Körpertemperatur. Ausserdem sind alle erdenklichen Zwischenwerte möglich, so auch dass bei einer Person eine Körpertemperatur von  $37.821239^{\circ}\text{C}$  gemessen wird.

Ein weiteres Beispiel für ein intervallskaliertes Merkmal ist der Intelligenzquotient  $IQ$ . Der  $IQ$  bewegt sich normalerweise zwischen 50 und 150, eine Person mit einem  $IQ$  von 105 hat einen höheren  $IQ$  als eine Person mit einem  $IQ$  von 103. Ausserdem sind  $IQ$ -Werte von 103.12 oder 118.9182 durchaus möglich.

Klicke hier, falls dir verhältnisskalierte Merkmale bekannt sind

Die folgende Diskussion ist auch auf verhältnisskalierte Merkmale anwendbar. Letztere sind intervallskalierte Merkmale, welche einen absoluten Nullpunkt aufweisen.

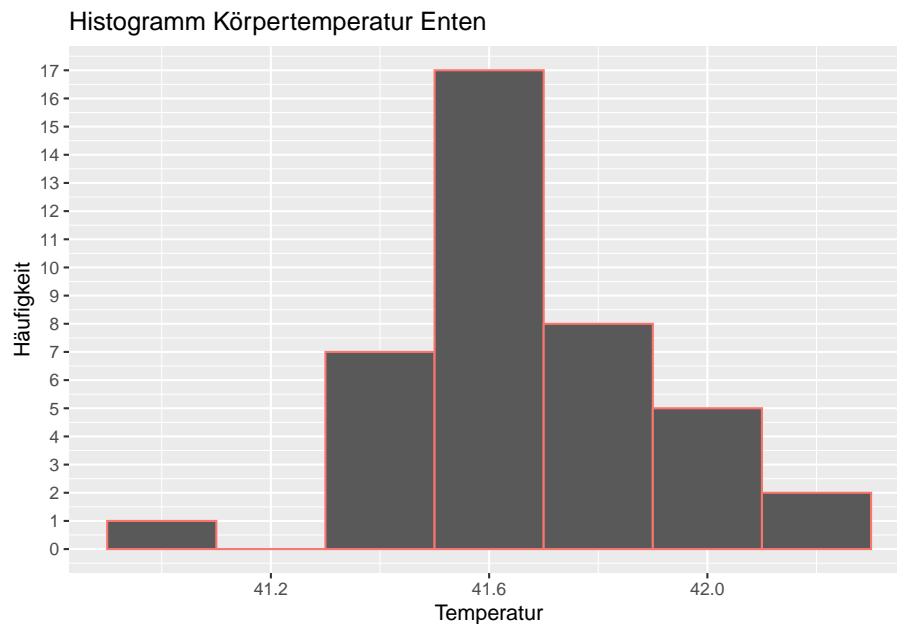
## 2.2 Wie kann ein intervallskaliertes Merkmal beschrieben werden?

Eine Veterinärin möchte herausfinden, welche Körpertemperatur Enten aufweisen. Dazu untersucht sie 40 Enten und misst die Körpertemperaturen 42.01, 41.72, 41.51, 41.52, 41.5, 41.6, 41.46, 41.81, 42.14, 41.82, 42.06, 41.53, 41.66, 41.65, 41.46, 41.48, 41.92, 41.58, 41.32, 41.58, 41.81, 41.7, 41.62, 41.52, 41.89, 41.53, 41.67, 41.43, 42.18, 41.52, 41.82, 41.96, 41.8, 41.54, 41.88, 41.69, 41.92, 41.35, 41.07, 41.67.

Für einen Menschen ist es schwierig direkt aus der Sichtung dieser Zahlen zu begreifen, welche Körpertemperatur Enten haben. Ein Mensch kann sich jedoch helfen, indem er die Zahlen zusammenfasst.

### 2.2.1 Verteilung

Um die Zahlen zusammenzufassen, kann die Veterinärin zum Beispiel Temperaturabschnitte von  $0.2^{\circ}\text{C}$  betrachten und zählen wie viele Beobachtungen sie in den jeweiligen Abschnitten gemacht hat. Diese Zählzeiten können tabellarisch oder grafisch mit einem Balkendiagramm dargestellt werden. Letzteres wird ein **Histogramm** genannt.



Aufgrund dieser Darstellung kann die Veterinärin nun sehen, wie häufig welche Körpertemperaturen sind. Dies wird die **Verteilung** des Merkmals genannt.

## 2.2. WIE KANN EIN INTERVALLSKALIERTES MERKMAL BESCHRIEBEN WERDEN?15

Sie bemerkt zum Beispiel, dass Beobachtungen der Körpertemperatur rund um 41.6°C am häufigsten sind und tiefere und höhere Temperaturen seltener vorkommen. Auf einen Blick sieht sie auch, dass die Temperatur aller Enten zwischen 41°C und 42.2°C war.

Die Verteilung eines Merkmals zu kennen ist hilfreich, jedoch in vielen Situationen (z. B. in der Kommunikation) noch zu komplex. Einfacher ist es die Komplexität einer Verteilung auf zwei Faktoren herunterzubrechen: Die Zentralität und die Variabilität eines Merkmals.

### 2.2.2 Zentralität

Mit der Zentralität ist ein Wert gemeint, welcher die zentrale Tendenz des Merkmals abbildet. Um die Zentralität zu messen, gibt es drei Möglichkeiten:

- Der **Modus** ist der am häufigsten vorkommende Wert. Im Beispiel ist das der Wert 41.52, welcher 3 mal und damit am häufigsten vorkommt.
- Wenn die Werte des Merkmals aufsteigend sortiert werden und der Wert betrachtet wird, welcher die Beobachtungen in eine tiefere und eine höhere Hälfte teilt, dann wird dieser Wert als **Median** (abgekürzt *Mdn*, Symbol  $\tilde{x}$ ) bezeichnet. Bei einer geraden Anzahl Beobachtungen, wird in der Regel der Durchschnittswert der beiden mittigsten Beobachtungen verwendet. Im Beispiel haben wir 40 Beobachtungen. Der Median entspricht also dem Durchschnittswert zwischen dem 20. und dem 21. der aufsteigend sortierten Werte 41.07, 41.32, 41.35, 41.43, 41.46, 41.46, 41.48, 41.5, 41.51, 41.52, 41.52, 41.52, 41.53, 41.53, 41.54, 41.58, 41.58, 41.6, 41.62, 41.65, 41.66, 41.67, 41.67, 41.69, 41.7, 41.72, 41.8, 41.81, 41.81, 41.82, 41.82, 41.88, 41.89, 41.92, 41.92, 41.96, 42.01, 42.06, 42.14, 42.18, also 41.655.
- Das **arithmetische Mittel** (abgekürzt *M*, Symbol  $\bar{x}$ ) bezeichnet, was gemeinhin mit Durchschnitt gemeint ist. Wenn wir die erste von insgesamt  $n$  Beobachtung mit  $x_1$  und die letzte Beobachtung mit  $x_n$  bezeichnen, so ist das arithmetische Mittel

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.1)$$

Im Beispiel ist das arithmetische Mittel der Körpertemperaturen 41.6725.

**Achtung**

*Hinweis. Erklärung der Formel:* Hier wird zum ersten Mal eine Formel verwendet.  $\sum$  steht für die Summe von allen Beobachtungen  $x_i$ , wenn der Index  $i$  in 1-Schritten von der Zahl unter dem Summenzeichen  $i = 1$  bis zu der Zahl oben am Summenzeichen  $i = n$  läuft. In unserem Beispiel ist  $n = 40$ , also ist  $i = 1, 2, 3, 4, \dots, 39, 40$ . Der Teil  $\sum_{i=1}^n x_i$  bedeutet also nichts anderes als  $x_1 + x_2 + \dots + x_{39} + x_{40}$ , also die Summe aller Beobachtungen.  $\frac{1}{n}$  bedeutet, dass wir diese Summe jetzt noch durch die Anzahl Beobachtungen teilen.

*Welchen Einfluss haben die verschiedenen Einflussgrößen:* Dies wird in Übung 2.2 erklärt.

Jedes dieser Masse für die Zentralität hat Vor- und Nachteile und sie werden dementsprechend in unterschiedlichen Situationen eingesetzt, siehe Übungen.

### 2.2.3 Variabilität

TODO: Quantile TODO: Standardabweichung

## 2.3 Übungen

### Übung 2.1.

- (a) Versuch selbst ein Histogramm der Daten oben (*Enten\_n40.sav*) mit Jamovi zu erstellen und begründe, weshalb es nicht gleich aussieht wie das Histogramm oben.
- (b) Berechne zusätzlich das arithmetische Mittel und die Standardabweichung des Merkmals.

*Lösung.*

- (a) Das Histogramm, siehe Abbildung 2.1 sieht nicht gleich aus, da Jamovi die Temperaturabschnitte kürzer gewählt hat nämlich bei  $0.125^\circ\text{C}$  statt  $0.2^\circ\text{C}$  wie oben im Text. In Jamovi gibt es aktuell keine Möglichkeit die Abschnittsweite anzupassen. Ein Histogramm sieht immer anders aus je nach ausgewählter Abschnittsweite.
- (b) TODO



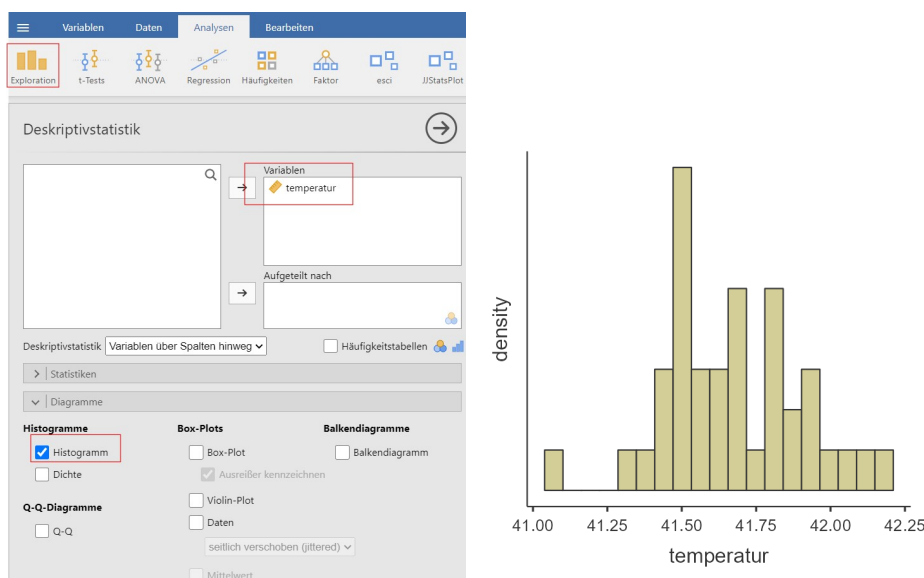


Abbildung 2.1: Links: Jamovi-Anleitung zur Erstellung des Histogramms; rechts: Histogramm der Temperatur.

**Übung 2.2.** In einem psychologischen Test machen 5 Probandinnen die Werte 18, 21, 20, 19, 22. Um mit einer Zahl zu sagen, wo die Testresultate liegen, wird ein zentraler Wert berechnet.

- Wie gross ist das arithmetische Mittel und der Median dieser Werte?
- Nehme an, der Testleiter hat den Wert der ersten Probandin falsch in seine Tabelle übertragen - statt 18 hat er 81 geschrieben. Wie gross ist das arithmetische Mittel und der Median dieser Werte in diesem Fall?
- Was sagt dies über den Median und das arithmetische Mittel aus?

*Lösung.* Die Aufgabe kann im Kopf gelöst werden, oder mithilfe eines Taschenrechners, oder indem die Zahlen manuell bei Jamovi eingegeben werden.

- (a) Wir haben hier  $n = 5$  Beobachtungen, nämlich  $x_1 = 18, x_2 = 21, x_3 = 20, x_4 = 19, x_5 = 22$ . Wird dies in die Formel (2.1) eingesetzt, so gibt dies das arithmetische Mittel

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + x_2 + x_3 + x_4 + x_5) = \frac{1}{5} (18 + 21 + 20 + 19 + 22) = 20.$$

Um den Median zu berechnen, werden die Werte zuerst aufsteigend sortiert 18, 19, 20, 21, 22. Der Wert, welcher die Werte in eine grössere und eine kleinere Hälfte teilt, ist hier 20, was dem Median entspricht.

- (b) Die Beobachtungen sind jetzt  $x_1 = 81, x_2 = 21, x_3 = 20, x_4 = 19, x_5 = 22$ . Analog wie in (a) kann demnach das arithmetische Mittel als  $\bar{x} = 32.6$  bestimmt werden. Die aufsteigend sortierten Beobachtungen sind nun 19, 20, 21, 22, 81. Der Median ist also 21.
- (c) Durch die fälschliche Übertragung eines Wertes, ist das arithmetische Mittel sehr stark und der Median fast gar nicht beeinflusst worden. Wenn die Daten wenige fehlerhafte Beobachtungen enthalten, ist der Median das bessere Mass für den zentralen Wert als das arithmetische Mittel. Wenn die Daten keine Fehler enthalten, ist das arithmetische Mittel gleich gut geeignet wie der Median.

## Kapitel 3

# Stichprobenziehung

**Beispiel 3.1** (Angst). Forschende haben ein Messinstrument STAI (State-Trait Anxiety Inventory), welches Angst misst (Spielberger et al., 1983). Sie unterscheiden dabei zwischen Zustandesangst und dem Persönlichkeitszug Ängslichkeit. Hier interessiert uns nur die Zustandesangst, welche fortan Angst genannt wird und misst wie grosse Angst man aktuell empfindet. Die so gemessene Angst entspricht einem Wert zwischen 20 und 80. A priori haben die Forschenden keine Ahnung, wie viel Angst eine Person im Durchschnitt hat und ob die ganze Skala der Werte genutzt wird. Die Forschenden machen deshalb eine kleine Befragung mit  $n = 30$  zufällig ausgewählten Studierenden. Die Forschenden finden die zusammenfassenden Werte  $M = 43.34$ ,  $s = 9.72$ ,  $n = 30$  für die Angst in ihren Beobachtungen.

Zufällig ausgewählte Beobachtungen eines Merkmals werden als **Stichprobe** bezeichnet. Nach der Auswahl der Stichprobe ist die **Stichprobenziehung**. Ist mit diesen Beobachtungen die Aussage beschränkt auf die Stichprobe oder kann damit auch eine Aussage zur Angst für alle Personen getroffen werden? Alle Personen, oder generell alle möglichen Beobachtungen eines Merkmals, werden als **Population** oder **Grundgesamtheit** bezeichnet. Eine Stichprobe ist für viele Analyseverfahren repräsentativ für eine Population, wenn sie zufällig aus dieser Population gezogen. Ist dies gegeben, wird die Stichprobe auch als **Zufallsstichprobe** bezeichnet.

*Hinweis.* Viele Studien basieren auf Testresultaten von Studierenden, weil diese nahe am Forschungsbetrieb sind und damit über Studien informiert sind oder für wenig Geld oder Bildungsanerkennung an Studien teilnehmen. Einige dieser Studien generalisieren ihre Forschungsergebnisse nachher auf alle Personen. Dies ist in der Regel falsch, da Studierende nicht repräsentativ für die Gesamtbevölkerung sind (Altersstruktur, Geschlechtsverteilung, Vermögen, usw.). Die Frage, wie eine repräsentative Stichprobe würde den Rahmen dieses Buches sprengen.

### 3.1 Was ist das Problem der Stichprobenziehung?

Es wird angenommen, dass sich alle Personen der Population in einem Zimmer befinden. In Abbildung 3.1 ist dieses Zimmer aus der Vogelperspektive dargestellt, wobei jeder Punkt im schwarzen Kasten einer Person der Population. Die Personen im Zimmer, respektive die Beobachtungen in der Population sind normalerweise nicht sichtbar. Aus diesem Zimmer wurden also 30 Personen geholt und befragt also sichtbar gemacht, was der Zufallsstichprobe entspricht. Die Zufallsstichprobe ist gekennzeichnet durch die Punkte über dem Zimmer, oberhalb des Pfeils. Die Farben der Punkte sind jetzt bekannt und entsprechen der jeweiligen Zustandesangst der beobachteten Personen.

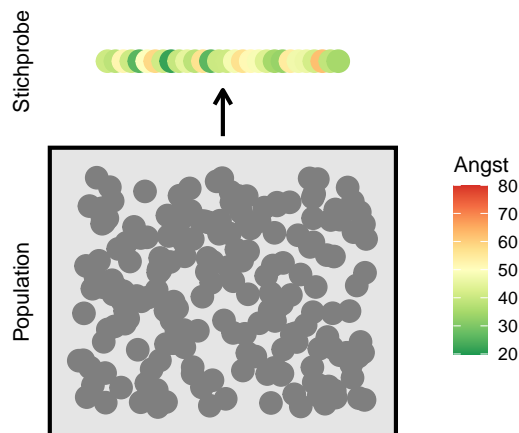


Abbildung 3.1: Population mit unbekannter Angst.

Da die Stichprobe nun eben zufällig gezogen wurde, das heisst zufällig Personen aus dem Zimmer geholt wurden, kann es nun sein, dass die Stichprobe einer Population wie in Abbildung 3.2 entstammt.

Es könnte aber auch sein, dass die Stichprobe einer Population mit viel höherer Zustandsangst, wie in Abbildung 3.3 dargestellt, entstammt. Dies wird zwar weniger häufig vorkommen als der Fall oben, aber ist trotzdem möglich.

Das Problem der zufälligen Stichprobenziehung ist also, dass nie ganz klar ist, wie die darunterliegende Population aussieht. Sind die Werte der Stichprobe

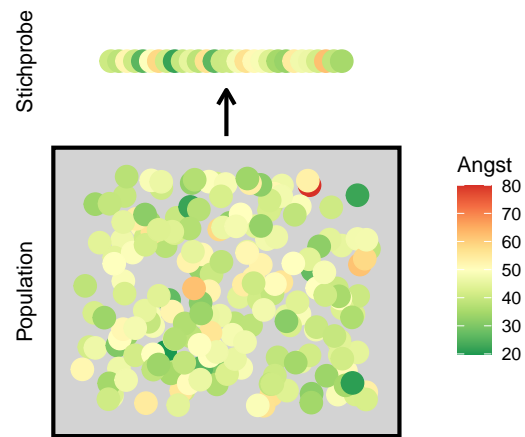


Abbildung 3.2: Population mit ähnlichen Angst-Werten wie in der Stichprobe.

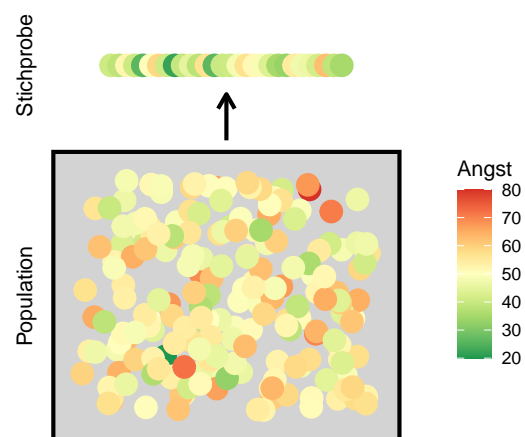


Abbildung 3.3: Population mit höheren Angst-Werten als in der Stichprobe.

tief, weil zufällig gerade Studierende mit tiefer Angst beobachtet wurden, oder haben tatsächlich die meisten Studierenden eine tiefe Angst?

### 3.2 Wie kann man Aussagen über die Grundgesamtheit machen?

Die Lösung dieses Problems funktioniert intuitiv wie folgt: Man stellt sich vor, die Stichprobenziehung würde erneut gemacht, und dann nochmal und dann nochmal. So oft, bis man einen guten Eindruck davon hat, wie häufig eine Stichprobe mit eher tiefen Angst-Werten wie bei der Stichprobe im Beispiel vorkommt. Im Szenario, in welchem in der Population tatsächlich tiefe Werte häufig vorkommen, kann dies aussehen wie in Abbildung 3.4. Stichproben mit eher tiefen Angst-Werten kommen hier häufig vor.

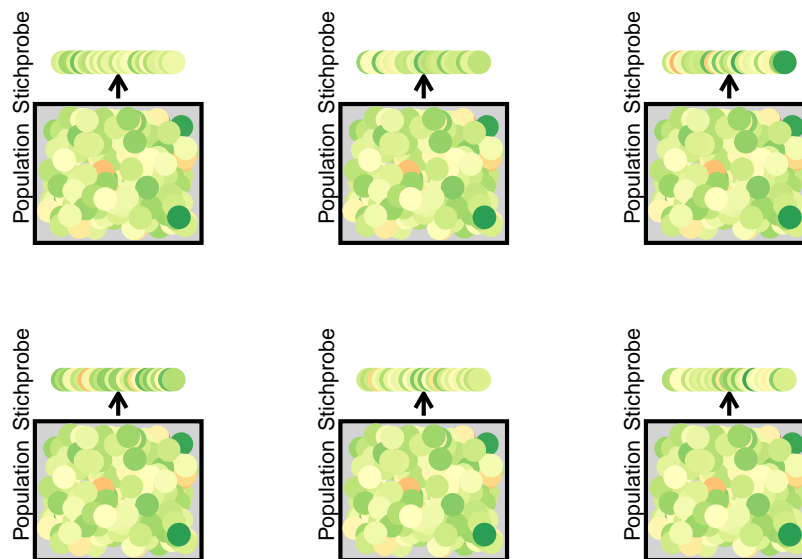


Abbildung 3.4: TODO.

Im Szenario, in welchem in der Population tatsächlich höhere Werte häufig vorkommen, kann dies aussehen wie in Abbildung 3.5. Stichproben mit eher tiefen Angst-Werten kommen hier selten oder gar nicht vor.

Es kann also zusammenfassend gesagt werden, dass die gezogene Stichprobe wohl eher aus einer Population mit tiefen Angst-Werten gezogen wurde als aus einer Population mit eher höheren Angst-Werten. Ganz sicher kann man jedoch

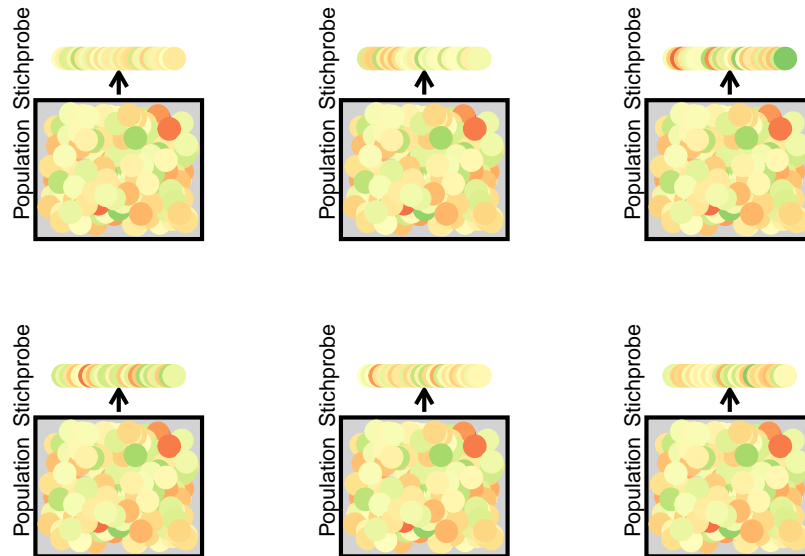


Abbildung 3.5: TODO.

nie sein, da die Werte in der Population eigentlich unbekannt sind. Eine genaue Quantifizierung dieser Unsicherheit kann mit Hilfe der Statistik erreicht werden und wird in den folgenden Kapiteln dieses Buches erläutert.

### 3.3 Übungen





## Kapitel 4

# Durchschnitt und Standardabweichung schätzen

Wie die in Abschnitt 3.2 skizzierte Lösung für das Problem der zufälligen Stichprobe konkret umgesetzt wird, hängt von der Problemstellung ab. Im folgenden wird ein Verfahren zur Generalisierung der Schätzung der zentralen Tendenz und eines für die Schätzung der Variabilität basierend auf einer Stichprobe präsentiert.

### 4.1 Wo liegt der Durchschnitt der Grundgesamtheit?

Ein Parameter über welchen wir gerne eine Aussage treffen würden ist die zentrale Tendenz in der Grundgesamtheit. Diese wird **Erwartungswert** (Symbol  $\mu$  [gr.: mü]) genannt. Wenn das arithmetische Mittel der Stichprobe berechnet wird, ergibt dies auch ein Schätzwert für besagten Erwartungswert. Aufgrund der zufälligen Stichprobenziehung ist jedoch auch klar, dass dieser Schätzwert nie genau dem wahren Erwartungswert entspricht.

In Beispiel 3.1 liegt das arithmetische Mittel in der Stichprobe der Studierenden bei  $M = 43.2$ . Dieser Wert entspricht nun auch der Schätzung des Erwartungswertes, also der geschätzten durchschnittlichen Angst aller Menschen. Die Folgefrage ist also wie genau unsere Schätzung ist. Um dies zu quantifizieren, wiederholen wir die Stichprobenziehung und berechnen das arithmetische Mittel dieser zweiten Stichprobe. Dann wiederholen wir diesen Prozess, zum Beispiel 1000 mal.

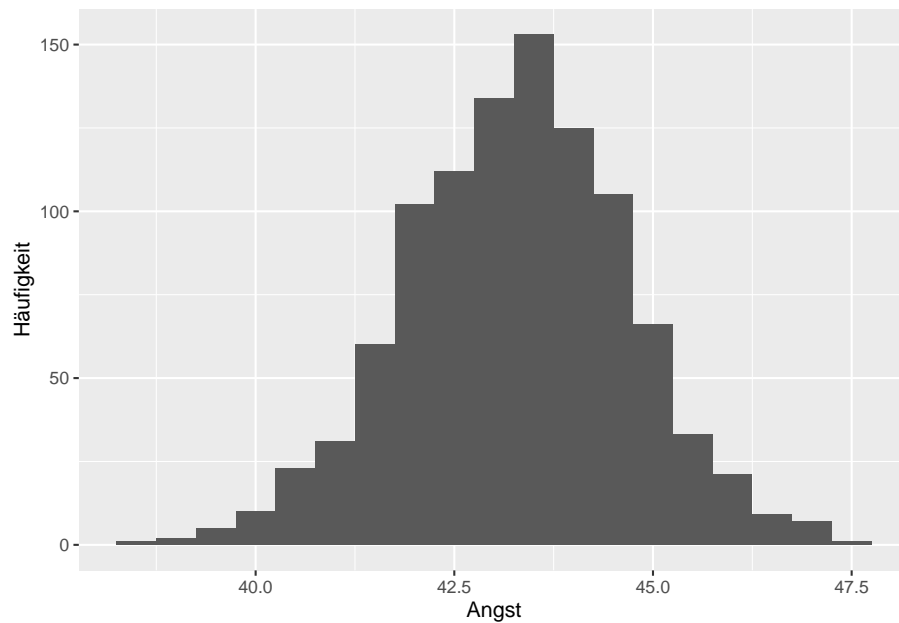


Abbildung 4.1: Verteilung der arithmetischen Mittel von 1000 zufällig gezogenen Stichproben der Angst.

Die Häufigkeitsverteilung der berechneten arithmetischen Mittel in Abbildung 4.1 lässt nun Aussage über die Häufigkeit und damit über die Wahrscheinlichkeit von gewissen Werten als Erwartungswert zu. Ein Durchschnittswert der Zustandesangst um die 30 ist hier am wahrscheinlichsten und ein Wert tiefer als 27 oder höher 33 eher selten. Um diese Aussage präziser zu gestalten, werden konventionell die 95% häufigsten Werte (die höchsten Balken im Histogramm) als wahrscheinlich betrachtet. Die 5% verbleibenden Werte, verteilt auf das untere und obere Extrem, werden als unwahrscheinlich betrachtet. Das 2.5% Perzentil trennt die 2.5% tiefsten arithmetischen Mittel ab und liegt im Beispiel bei 40.4. Das 97.5%-Perzentil trennt die höchsten 2.5% (oder eben die tiefsten 97.5%) arithmetischen Mittel ab und liegt bei 46. Dies ist in Abbildung 4.2 ersichtlich.

**Beispiel 4.1** (Verträglichkeit). Einer der Big-5 Persönlichkeitszüge ist die Verträglichkeit. Eine einfache Art die Big-5 zu messen ist mit den 10 Fragen aus dem ten-item personality inventory (TIPI) (Gosling et al., 2003). Für die Verträglichkeit müssen zwei Items (Item 1: Critical, quarrelsome; Item 2: Sympathetic, warm) auf einer Likert-Skala von 1 bis 7 eingeordnet werden. Anschliessend werden die Antworten gemittelt. Ein Student möchte herausfinden, ob mit diesem Messinstrument die durchschnittliche Verträglichkeit aller Menschen mittig also bei 4 liegt. Dafür befragt er  $n = 100$  Personen und findet die Werte  $M = 4.14$ ,  $s = 1.69$ .

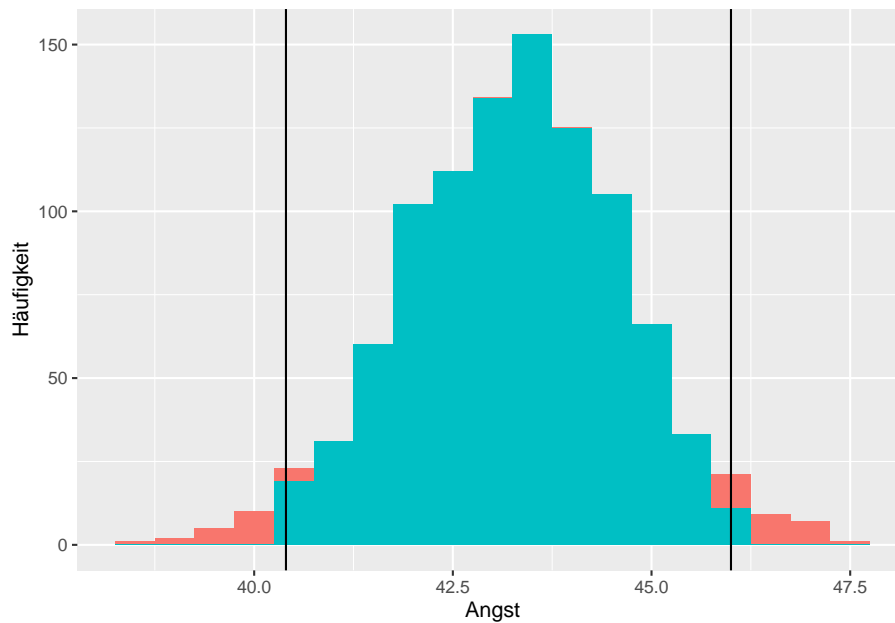


Abbildung 4.2: Verteilung der arithmetischen Mittel von 1000 zufällig gezogenen Stichproben der Angst.

Die Verteilung der Beobachtungen, siehe Abbildung 4.3, zeigt, dass alle Werte zwischen 1 und 7 vorkommen, aber keine zentrale Tendenz greifbar ist. Um herauszufinden wie zutreffend die Schätzung des Erwartungswertes der Verträglichkeit von  $M = 4.14$  ist, stelle man sich wieder vor, dass der Student 1000-mal die Stichprobenziehung wiederholt und jedes mal das arithmetische Mittel  $M$  von neuem berechnet. Die Verteilung der arithmetischen Mittel dieser Stichproben ist in Abbildung 4.4 dargestellt. Bei dieser Verteilung kann erneut links und rechts 2.5% der Werte abgeschnitten werden, um zum Schluss zu gelangen, dass das arithmetische Mittel in 95% der Fälle zwischen 3.7 und 4.3 zu liegen kommt.

Das Problem mit diesem Vorgehen ist, dass es aus finanziellen oder technischen Gründen selten möglich ist mehrere Stichproben aus derselben Population zu ziehen. Glücklicherweise haben Statistiker:innen herausgefunden, dass die Häufigkeitsverteilungen wie in Abbildungen 4.2 und 4.4 immer dieselbe Verteilung haben und dies unabhängig davon wie die ursprüngliche Verteilung des Merkmals aussah. Diese Verteilung ist eine sogenannte **Normalverteilung**.

Die Normalverteilung sieht eine Glocke ähnlich. Deshalb wird sie auch Gaußsche Glockenkurve nach Carl F. Gauss (1777-1855) benannt. Die Normalverteilung kann mit nur zwei Parametern beschrieben werden.

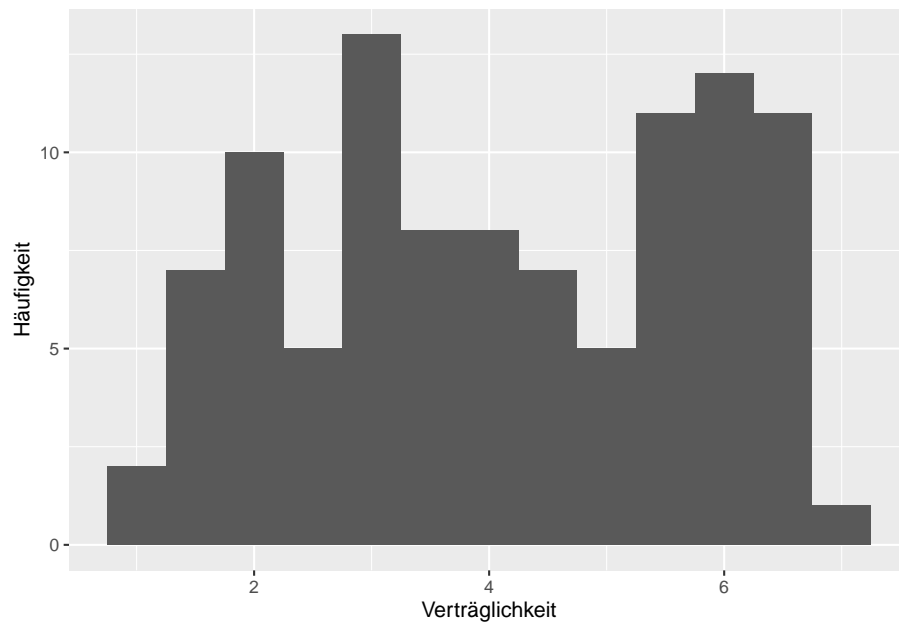


Abbildung 4.3: Verteilung der 100 beobachteten Verträglichkeitswerte einer zufällig gezogenen Stichprobe.

- $\mu_g$  gibt an, wo auf der x-Achse der höchste Punkt der Glocke liegt
- $\sigma_g$  gibt an, wie flach die Glockenform ist (ein grosser Wert entspricht einer flachen Glockenform, ein tiefer Wert einer steilen Glockenform).

Auf [seeing-theory.brown.edu](http://seeing-theory.brown.edu) > Continuous > Normal kann der Einfluss von  $\mu$  und  $\sigma$  auf die Normalverteilung erfahren werden.

Diese Tatsache, dass die Durchschnitte aller Merkmale normalverteilt sind, ist so zentral für die Statistik, dass sie **Zentraler Grenzwertsatz** genannt wurde. Der zentrale Grenzwertsatz besagt geneauer, dass bei einem Merkmal mit Erwartungswert  $\mu$  und Standardabweichung  $\sigma$ , der Durchschnitt aller Stichprobenwerte einer Normalverteilung mit  $\mu_g = \mu$  und  $\sigma_g = \frac{\sigma}{\sqrt{n}}$  entspricht, wobei  $n$  die Stichprobengrösse bezeichnet.

*Hinweis.*

- $\mu_g = \mu$  bedeutet, dass der Wert, welcher unter der normalverteilung am wahrscheinlichsten ist, genau dem Erwartungswert des untersuchten Merkmales entspricht.
- $\sigma_g = \frac{\sigma}{\sqrt{n}}$  hat zwei Implikationen:
  - je grösser die Streuung des Merkmals (grosses  $\sigma$ ) desto breiter ist auch die Streuung der arithmetischen Mittel (grosses  $\sigma_g$ ). Dies be-

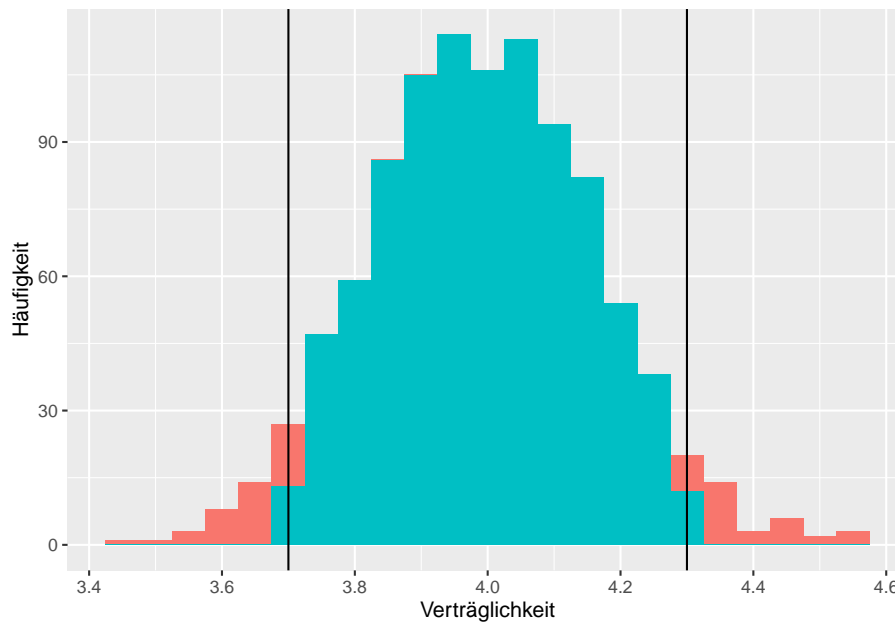


Abbildung 4.4: Verteilung der arithmetischen Mittel von 1000 zufällig gezogenen Stichproben der Verträglichkeit.

deutet, je weniger Streuung das Merkmal aufweist, desto genauer ist die Bestimmung des Erwartungswertes des Merkmales.

- je grösser die Anzahl Beobachtungen  $n$ , desto kleiner die Streuung der arithmetischen Mittel (kleines  $\sigma_g$ ). Dies bedeutet, je grösser die Stichprobe ist, desto genauer ist die Bestimmung des Erwartungswertes des Merkmales.

Die Abbildungen 4.5 und 4.6 illustrieren den zentralen Grenzwertsatz für Beispiel 3.1 und 4.1 respektive, wobei die Normalverteilung der roten Linie entspricht. Dabei wird einstweilen angenommen, dass  $\mu$  und  $\sigma$  bekannt sind. Diese Annahme wird später aufgelöst und dient hier lediglich der Illustration.

Die Erkenntnis des zentralen Grenzwertsatz macht also das wiederholte ziehen von Stichproben unnötig. Die Normalverteilung ist theoretisch konstruiert und ihr 2.5%- und 97.5%-Perzentil können theoretisch hergeleitet werden. Tabelle 4.1 wird kann beobachtet werden, dass für unsere zwei Beispiele die Perzentile der Stichprobe und der Normalverteilung sehr ähnlich, wenn auch nicht exakt gleich sind. Die Ungenauigkeit rührt daher, dass der zentrale Grenzwertsatz nur dann exakt funktioniert, wenn die Anzahl Beobachtungen (unendlich) gross ist.

Einstweilen wurde hier angenommen, dass die Streuung des Merkmals  $\sigma$  bekannt ist. Dies ist in der Realität nie der Fall. Wenn  $\sigma$  also auch aus der Stichprobe

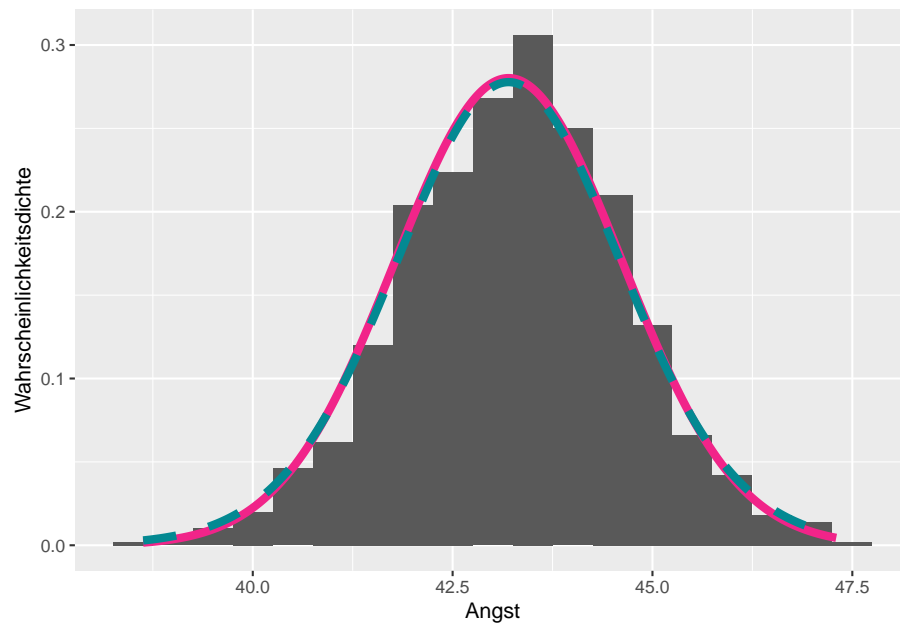


Abbildung 4.5: Die arithmetischen Mittel sind Normalverteilt mit Parametern  $\mu_g = 43.34$  und  $\sigma_g = 9.72/\sqrt{30}$ .

Tabelle 4.1: Vergleich Perzentile der Stichprobe und der theoretischen Verteilung.

Beispiel	Stichprobe		Normalverteilung		t-Verteilung	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
Angst	40.4	46.0	42.23	44.46	42.18	44.50
Vertraeglichkeit	3.7	4.3	3.66	4.34	3.66	4.34

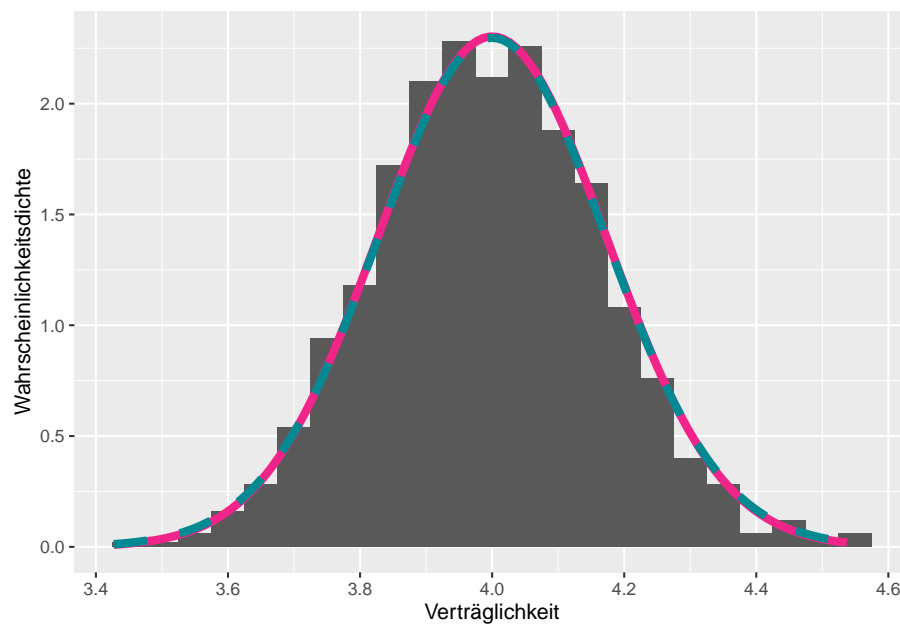


Abbildung 4.6: Die arithmetischen Mittel sind Normalverteilt mit Parametern  $\mu_g = 4.14$  und  $\sigma_g = 1.69/\sqrt{100}$ .

geschätzt werden muss, ist die Annäherung der Verteilung der arithmetischen Mittel besser gegeben mit einer **Student- $t$ -Verteilung** oder kurz  $t$ -Verteilung. Die grüne Linie in Abbildungen 4.5 und 4.6 entspricht der  $t$ -Verteilung im jeweiligen Beispiel.

Der Unterschied zwischen der Normalverteilung und der  $t$ -Verteilung ist nur sichtbar, wenn  $n$  klein ist. In Beispiel 3.1 mit  $n = 30$  ist ein kleiner Unterschied, in Beispiel 4.1 mit  $n = 100$  ist kein Unterschied zwischen der Normalverteilung und der  $t$ -Verteilung sichtbar. Tatsächlich wird die  $t$ -Verteilung mit einem Parameter charakterisiert, welcher Freiheitsgrade [eng. degrees of freedom,  $df$ ] genannt wird. In Abbildung 4.7 wird die  $t$ -Verteilung mit verschiedenen Freiheitsgraden mit der Normalverteilung verglichen. Bei der  $t$ -Verteilung mit den kleinsten Freiheitsgraden sind extremere Werte wahrscheinlicher als  $t$ -Verteilungen mit grösseren Freiheitsgraden.

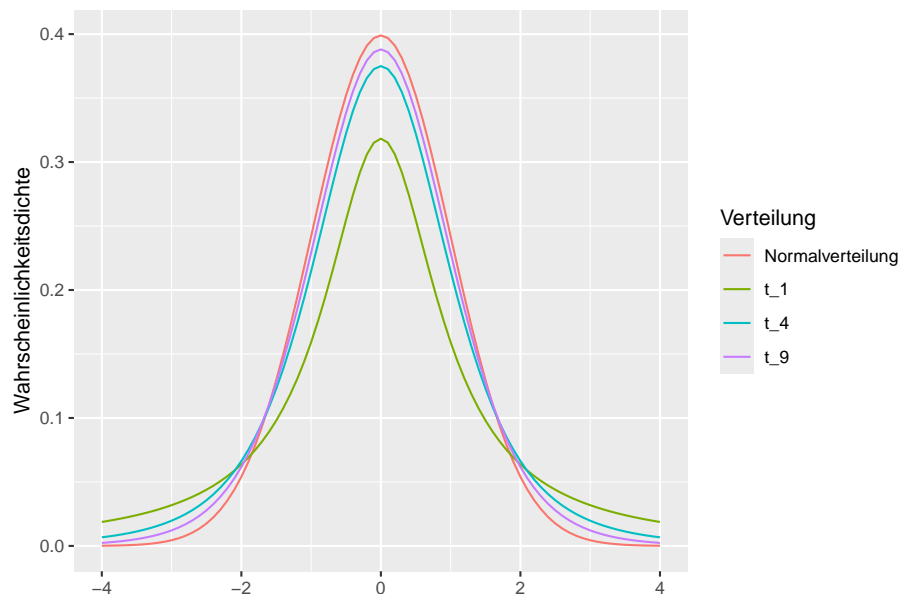


Abbildung 4.7: Student- $t$ -Verteilungen mit 1, 4 und 9 Freiheitsgraden im Vergleich zu der Normalverteilung.

Die Freiheitsgrade der  $t$ -Verteilung in der Annäherung oben entsprechen der Anzahl Beobachtungen minus 1, also  $df = n - 1$ . Die höhere Wahrscheinlichkeit von extremen Werten bei kleinen Freiheitsgraden spiegelt die grössere Unsicherheit der Schätzung des Erwartungswertes wieder, wenn die Standardabweichung unbekannt und damit auch geschätzt werden muss. Je kleiner  $n$  ist, desto stärker fällt diese Unsicherheit aus.

Die arithmetischen Mittel bei unbekannter Standardabweichung sind bei wieder-



holter Stichprobenziehung genau  $t$ -verteilt. Um die Genauigkeit der Schätzung des Erwartungswertes zu bestimmen genügt es folglich, das 2.5% und das 97.5% Perzentil der  $t$ -Verteilung mit  $n-1$  Freiheitsgraden zu bestimmen. Diese Perzentile können mit

$$\bar{x} - \frac{s}{\sqrt{n}} \cdot t_{97.5\%, n-1} < \mu < \bar{x} + \frac{s}{\sqrt{n}} \cdot t_{97.5\%, n-1}$$

berechnet werden, wobei  $\bar{x}$  das arithmetische Mittel,  $s$  die Standardabweichung und  $t_{97.5\%, n-1}$  dem Wert des 97.5%-Perzentil einer auf 0 zentrierten  $t$ -Verteilung mit  $n-1$  Freiheitsgraden entspricht. Letzere Perzentile der  $t$ -Verteilung können bei Bedarf in entsprechenden Tabellen nachgeschlagen werden. Als Gedankenstütze kann für  $t_{97.5\%, n-1}$  immer 2 gedacht werden, da dies ungefähr dem wahren Wert entspricht, wenn  $n$  grösser als 50 ist.

Das 2.5% und das 97.5% Perzentil der Verteilung der arithmetischen Mittel ergeben nun die untere respektive obere Schranke eines **Intervalles**. Ein Intervall bezeichnet durch die Symbolik [untere Schranke, obere Schranke] beinhaltet alle Zahlen zwischen der unteren und der oberen Schranke. Ein Intervall mit den oben beschriebenen Perzentilen als Schranken wurde so berechnet, dass bei wiederholter Stichprobenziehung der wahre Erwartungswert in 95% der Fälle umschlossen wird. Grob übersetzt bedeutet dies, dass wir zu 95% sicher oder *konfident* sind, dass der Erwartungswert in diesem Intervall liegt. Dieses Intervall wird deshalb als 95%-**Konfidenzintervall** (symbol KI) bezeichnet.

In Beispiel 3.1, kann aus der Tabelle 4.1 entnommen werden, dass die Angst in der Population bei  $M = 43.34$  95%-KI [42.18, 44.5] liegt. In Beispiel 4.1, kann aus der Tabelle 4.1 entnommen werden, dass die Verträglichkeit in der Population bei  $M = 4.14$  95%-KI [3.66, 4.34] liegt.

Es ist nun spannend zu explorieren, wie sich die Stichprobengrösse  $n$  oder die geschätzte Standardabweichung  $s$  auf die Länge des Konfidenzintervalls auswirkt. Dies kann in den Übungen 4.4 und 4.5 selbst erforscht werden.

## 4.2 Übungen

Übung 4.1. TODO

```
## [1] "hi"
```

Lösung. TODO

Übung 4.2. TODO

```
## [1] "hi"
```

*Lösung.* TODO

**Übung 4.3.**

## [1] 0.5342439

## [1] 0.2388751

## [1] 0.1677798

*Lösung.* TODO

**Übung 4.4.** Eine Mensa will herausfinden, wie lange die Leute um 12h durchschnittlich anstehen müssen. Dazu befragt sie 5 Kund:innen. Das Resultat der Untersuchung ist, dass die Kund:innen im Durchschnitt 0.4 Stunden anstehen müssen. Leider ist das Konfidenzintervall sehr gross. Da die Mensa nicht weiss, wie viele Leute befragt werden müssen, um ein kleineres Konfidenzintervall zu erhalten befragt sie in 4 weiteren Runden jeweils 20, 50, 100 und 1000 Kund:innen. Die Daten aller 5 Untersuchungen sind unter `04-exr-stichprobengroesse.sav` abgelegt. Für jede der 5 Stichproben:

- Was ist die Schätzung des Erwartungswertes der Wartezeit?
- Wie gross ist die Standardabweichung der Wartezeit?
- Wie gross ist die Standardabweichung der arithmetischen Mittel?
- Bestimmen Sie die 95%-Konfidenzintervalle.
- Berechnen Sie die Länge jedes Konfidenzintervalls.

Vergleichen Sie die Resultate der Berechnungen für jede Stichprobe:

- Weshalb ist die Schätzung für den Erwartungswert für jede Stichprobe unterschiedlich?
- Was lässt sich über den Zusammenhang zwischen Stichprobengrösse und der Länge des Konfidenzintervalls sagen?

*Lösung.* Abbildung 4.8 zeigt die Berechnungsanweisungen für Jamovi und die resultierende Tabelle daraus.

- Der Erwartungswert der Wartezeiten (das heisst der Populationmittelwerte der Wartezeiten) wird mit dem arithmetischen Mittel der Stichprobe geschätzt und kann in der Tabelle bei **Mittelwert** abgelesen werden. Der Erwartungswert der Wartezeiten beträgt bei allen Stichproben ausser bei der ersten ungefähr 0.22 Stunden, also ein bisschen weniger als eine Viertelstunde.
- Der Standardabweichung der Wartezeiten der Stichprobe sind in der Tabelle bei **Std.-abw.** abzulesen. Die Standardabweichungen sind für alle Stichproben ausser der ersten ungefähr bei 0.23.

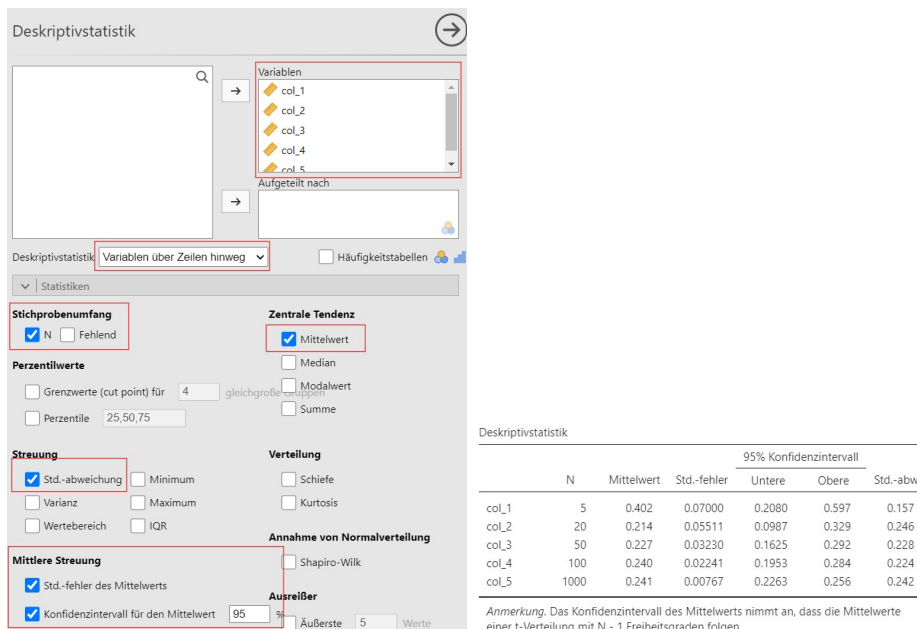


Abbildung 4.8: Links: Jamovi-Anleitung zur Erstellung der Tabelle mit den relevanten Kenngrößen; rechts: Tabelle mit relevanten Kenngrößen.

- c. Die Standardabweichung der arithmetischen Mittel liegt bei  $s/\sqrt{n}$ . Für die erste Stichprobe ist dies  $0.157/\sqrt{5} = 0.0702125$ . Diese Werte werden auch als Standardfehler bezeichnet und sind in der Tabelle bei **Std.-fehler** ablesbar.
- d. Die untere und obere Schranke der 95%-Konfidenzintervalle sind bei **Untere** und **Obere** respektive abzulesen.
- e. Die Länge des Konfidenzintervalls entspricht jeweils dem höheren Wert minus dem tieferen Wert. Für die erste Stichprobe ist dies  $0.597 - 0.208 = 0.389$ , für die anderen 0.23, 0.13, 0.09 und 0.03.
- f. Die Schätzung des Erwartungswertes ist das arithmetische Mittel der Stichprobe. Da jedesmal eine neue Zufallsstichprobe gezogen wurde und diese nicht dieselben Beobachtungen enthalten, ergeben sich auch jedesmal andere Stichprobenmittelewerte.
- g. Je grösser  $n$ , desto kleiner ist das Konfidenzintervall. Wenn man also ein kleines Konfidenzintervall erreichen will, braucht man eine grössere Stichprobe.

#### Übung 4.5. TODO

```
## [1] "hi"
```

*Lösung.* TODO

## Kapitel 5

# Durchschnitt testen

- 5.1 Entspricht der Durchschnitt der Grundgesamtheit einem gewissen Wert?
- 5.2 Weicht der gefundene Durchschnitt stark vom hypothetischen Wert ab?
- 5.3 Übungen



## Teil II

# Gruppenvergleich einer intervallskalierten Variable





## Kapitel 6

# Gruppenvergleich einer intervallskalierten Variable

6.1 Zwei Gruppen vergleichen

6.2 Was ist das Problem der Stichprobenziehung?

6.3 Wie kann man Aussagen über die Grundgesamtheit machen?

6.4 Übungen



## Kapitel 7

# Welch-Test

7.1 Zwei Gruppen vergleichen

7.2 Sind die Durchschnitte der beiden Gruppen in der Grundgesamtheit gleich?

7.3 Wie stark unterscheiden sich die Durchschnitte?

7.4 Übungen



# Begriffsverzeichnis

- Erwartungswert
- Grundgesamtheit
- Intervalles
- Konfidenzintervall
- Median
- Modus
- Normalverteilung
- Population
- Stichprobe
- Stichprobenziehung
- Student- $t$ -Verteilung
- Zentraler Grenzwertsatz
- Zufallsstichprobe
- arithmetische Mittel
- intervallskaliert



# Literaturverzeichnis

- Gosling, S. D., Rentfrow, P. J., and Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, 37(6):504–528.
- R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Selker, R., Love, J., and Dropmann, D. (2024). *jmv: The jamovi Analyses*. R package version 2.5.6.
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., and Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press, Palo Alto, CA.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.
- Xie, Y. (2016). *bookdown: Authoring Books and Technical Documents with R Markdown*. Chapman and Hall/CRC, Boca Raton, Florida.