Universität Stuttgart
Institut für Erziehungswissenschaft
und Psychologie

BWT
Abteilung Berufs-, Wirtschafts- und Technikpädagogik

Martin Kenner

Einführung in die Statistik

Literatur:

Benninghaus, Hans: Deskriptive Statistik. Wiesbaden: Verlag für Sozialwissenschaften 2005

Diehl, Joerg M. / Arbinger Roland:: Einführung in die Inferenzstatistik. Eschborn: Klotz 2001

Diehl, Joerg M. / Kohr, Heinz U.: Deskriptive Statistik. Eschborn: Klotz 1999

Sahner, Heinz: Schließende Statistik. Stuttgart: Teubner 1990

1. Vorbemerkungen

1.1 Zur Statistik in der Erziehungswissenschaft

(1) Zum Begriff „Statistik“

> ***„Wissenschaftliche Methode zur zahlenmäßigen Erfassung, Untersuchung und Darstellung von Massenerscheinungen.“[1]***

- **Wissenschaftliche Methode:**
  d.h. ein Verfahren mit definierten und kontrollierbaren Standards
- **zahlenmäßige Erfassung:**
  Aufgegriffen werden quantifizierbare Ausdrucksformen von Phänomenen, z.B. die Zahl der Auszubildenden im IT-Bereich; Die Mathematik fungiert dabei als Vermittlungssprache; Bedingung: Das Phänomen muß sich auch zahlenmäßig ausdrücken lassen
- **Untersuchung/Darstellung:**
  Aufdeckung von Tendenzen, Zusammenhängen etc, mit Hilfe von statistischen Maßzahlen in Orientierung an die forschungsleitenden Fragen/Hypothesen
- **Massenerscheinung:**
  Es interessiert nicht das Einzelphänomen, Statistik sinnvoll ab ≥ 30 Fälle, darunter mit gewissen Einschränkungen[2]

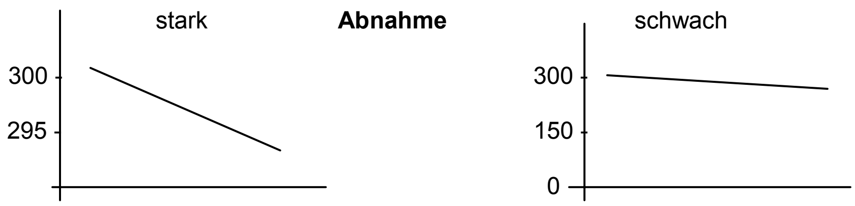
(2) Zur Skepsis gegenüber der Statistik

- **Allgemeine Skepsis gegenüber der Interpretationen statistischer Analysen:**

Statistische Analysen haben nicht den besten Ruf, nicht zuletzt weil in vielen Untersuchungen Daten aus dem Zusammenhang gerissen dargestellt und fehlinterpretiert werden.

<u>Beispiel:</u> Im Januar 1995 waren von 100 Mio. Bäumen 29% geschädigt. Der Holzeinschlag im selben Jahr betrug 7%, davon waren 90% geschädigt. Im Januar 1996 wurde in der Presse bekannt gegeben, dass es dem Wald wieder deutlich besser gehe, denn gegenüber dem letzten Jahr habe der geschädigte Waldbestand um über 5% abgenommen.

Verzerrungen können auch durch unangemessene Visualisierungen hervorgerufen werden.



- **Pädagogisch begründete Kritik (Auszug)**

Die geisteswissenschaftliche Pädagogik kritisiert an statistischen Analysen u.a., dass sich pädagogische Prozesse und Lernerfahrungen nicht adäquat quantifizieren lassen, etwa durch die Zuteilung von Noten. Ferner können die auf eine Gruppe bezogenen Aussagen nicht auf den einzelnen Schüler bezogen werden, individuelle Unterschiede bleiben deshalb unberücksichtigt.
Demgegenüber sind für die empirische Pädagogik, die auf die Gewinnung/Prüfung allgemeiner Erklärungen, Prognosen, etc. ausgerichtet ist, große Fallzahlen eine notwendige Bedingung. Zudem wird mit statistischen Verfahren die Forderung nach intersubjektiver Überprüfbarkeit eingelöst.

---
[1] Duden Band 5: Das Fremdwörterbuch 1974
[2] vgl. Sahner 1997, S.57

(3) Typische statistische Operationen am Beispiel „Schulnoten“

In 2 Schulklassen wurden bei einem Test folgende Ergebnisse erzielt (siehe Datentabelle):

Datentabelle → **Typische Operationen** → **Allg. Umschreibung**

Klasse 1 (K1): Note /Geschlecht		**Klasse 2 (K2): Note/Geschlecht**	
2	w	1	w
1	w	3	m
4	m	2	w
1	w	3	m
4	m	2	w
2	m	3	w
4	w	1	m
3	m	1	w
2	m	2	m
1	w	1	w
4	m	2	w
2	m	3	w
4	w	1	m
3	m	1	w

1. Durchschnittsnote (Mittelwert):
$\varnothing_{K1}$ = ? $\varnothing_{K2}$ = ?

2. Homogenität der Verteilung (Streuung): $S_{K2} > S_{K1}$?

Beschreibung einzelner Variablen

⇒ **Deskriptive Statistik/ Kapitel 2.1**

3. Einfluß des Geschlechts:
Jungs besser als Mädchen?

Beschreibung der Beziehung zwischen zwei Variablen

⇒ **Deskriptive Statistik/ Kapitel 2.2-2.4**

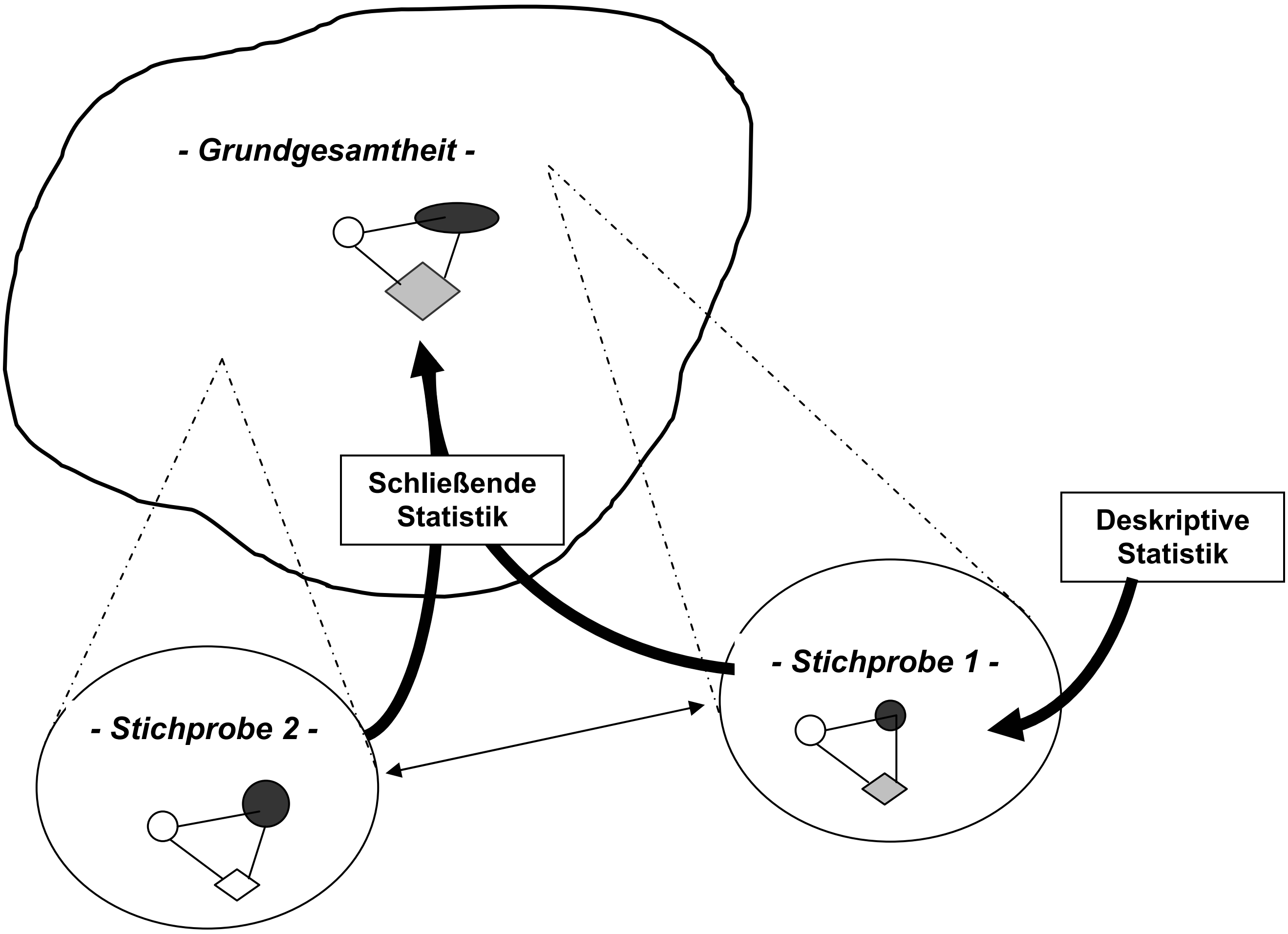
4.Vergleich der Mittelwerte:
$\varnothing_{K2}$ zufällig o. signifikant besser als $\varnothing_{K1}$?

5. Repräsentativität:
Wie wahrscheinlich ist es, dass die Durchschnittsnote **aller** vergleichbaren Klassen bei $\varnothing_{K1K2}$ liegt?

Schlußfolgerung / Generalisierung

⇒ **Schließende Statistik o. Inferenzstatistik/ Kapitel 3**

Entlang dieser Auflistung werden im Folgenden grundsätzliche Aspekte der Deskriptiven Statistik (Kap. 2) und der Schließenden Statistik (Kap. 3) angesprochen.



1.2 Datenmatrix, Skalentypen[3]

(1) Datenmatrix

Die Datenmatrix ist die Grundstruktur, in der die Rohdaten, unabhängig von der Untersuchung dargestellt werden. Die Datenmatrix hat folgenden allgemeinen Aufbau:

<u>Variablen, Merkmale, Stimuli</u>

<u>Untersuchungseinheit / Merkmalsträger</u>

Nr.	SchülerIn	Geschlecht	Übungszeit Englisch (h)	Note Englisch	Note Mathe
1	Gabi	w	3	3	3
2	Holger	m	1	4	4
3	Ingo	m	2	2	2
4	Jasmin	w	3	4	4
5	Klaus	m	1	3	3
6	Ludwig	m	2	1	1
7	Monika	w	1	2	2

<u>Werte, Merkmalsausprägungen, z.B. Antworten</u>

Untersuchungseinheit
- ist das Bezugsobjekt (oder Merkmalsträger)
- häufig Personen aber auch Gruppierungen, Institutionen; *<u>Beispiel:</u> SchülerInnen, z.B. Gabi, Ludwig*

Variable
- das Merkmal ist eine bestimmte Eigenschaft des Merkmalsträgers, die man erforscht
  <u>Beispiel hier</u>: Geschlecht, Übungszeit, Note im Schulfach Englisch,
- Man spricht von Stimuli, wenn die Variable kontrolliert varriert wird, um die Werte zu verändern (Unabhängige Variable wird variiert, die „Reaktion“ der abhängigen Variable wird beobachtet);
  <u>Beispiel</u>: Man varriert in einen Experiment die Übungszeit im Fach Englisch zwischen 1 und 3 h Stunden und beobachtet den Einfluß auf die Schulnote

Werte
- Merkmalsausprägungen, in denen die Variable auftritt. *<u>Beispiel:</u> weiblich, männlich, Note 1, 2, 3;.*

(2) Skalentypen[4] (auch Messniveau)

Die Merkmalsausprägung einer Variable ergibt sich aus einem Meßvorgang. **Messen** im w. S. heißt, die Zuordnung von Zahlen (Zeichen) zu empirisch beobachtbaren Objekten, zum Beispiel die Zuordnung einer Note für die erbrachte Leistung in einem Test. Was die Ausprägungen von Variablen betrifft, lassen sich 4 unterschiedliche Messniveaus unterscheiden.

1. Nominalskala[5]

Das einfachste Messniveau klassifiziert Objekte in einzelnen Klassen, ohne eine Gewichtung zwischen den Klassen vorzunehmen.

<u>Beispiel</u>: Die Variable „Geschlecht“ wird aufgeteilt in die Kategorien „weiblich“ und „männlich“.

Für die Klassenbildung ist grundlegend, dass die Kategorien alle Fälle erfassen (Vollständigkeit), jedoch dabei eine eindeutige Zuordnung erfolgt (Gegenseitiger Ausschluss).

<u>Beispiel:</u> Allen Menschen kann ein Geschlecht zugeordnet werden (Vollständigkeit). Ein Mensch ist entweder „weiblich“ oder „männlich“ (zumindest biologisch).

---

[3] vgl. Benninghaus 2005, S. 16-28
[4] vgl. auch Diehl/Kohr 1999, S. 8-14
[5] Nomen: Substantiv, das eine Eigenschaft bezeichnet

2. Ordinalskala oder auch Rangskala

In vielen Fällen ist es möglich, Variableneigenschaften nach ihrer Intensität oder ihrem Ausprägungsgrad zu ordnen, auch wenn dieser nur unpräzise beziffert werden kann.

<u>Beispiel:</u> Das Anspruchsniveau der Lerninhalte der Sekundarstufe I ist auf dem Gymnasium höher als auf der Realschule, und auf der Hauptschule am geringsten.

Bei der statistischen Verarbeitung ist es unerheblich, ob den drei Kategorien die Zahlenfolge 1 (=Gymnasium), 2 (=Realschule) und 3 (=Hauptschule) oder 30, 21, 15 zugeordnet wird, weil die exakte Differenz zwischen den Kategorien nicht bekannt ist.

3. Intervallskala

Bei dieser Skala existiert eine definierte Maßeinheit, mit der die

> „... Objekte nicht nur geordnet werden können – man kann ihnen auch Zahlen so zuweisen, dass gleiche Differenzen zwischen den den Objekten zugeordneten Zahlen gleiche Differenzen in der Ausprägung des gemessenen Merkmals reflektieren..“[6].

<u>Beispiel</u>: Ein Intelligenztest besteht aus 20 gleichwertigen Aufgaben. Jede erfolgreich bearbeitete Aufgabe wird mit 5 Punkten bewertet. Die erbrachte Leistung spiegelt sich dann exakt in der Punktezahl wieder: Wenn jemand die Hälfte der Punkte erreicht hat, bedeutet es, dass er die Hälfte des im Test abgebildeten Intelligenzbereichs erfüllt.

Damit ist allerdings nicht gesagt, ob ein Testteilnehmer mit 100 Punkten allgemein als hochintelligent bezeichnet werden kann. Möglicherweise ist der Test trivial oder behandelt nur einen spezifischen Bereich von Intelligenz. Ein wesentliches Kennzeichen für Intervallskalen ist also, dass sie nur einen relativen Ausschnitt eines größeren Bereichs abbilden und keinen absoluten Nullpunkt besitzen.

Bemerkung zur Messung von Schulleistungen:
In der Regel werden Schulleistungen auf Intervallskalenniveau angegeben (Noten 1-6, z.B. Abstand 0,25). Das für die Intervallskalierung zugrunde gelegte Kriterium (siehe Zitat oben) kann nur bei vollstandardisierten und geeichten Tests eingelöst werden. Weil solche Tests im Schulalltag eher die Ausnahme bilden, entspricht die alltägliche Notenvergabe eher einer Rangskalierung als einer Intervallskalierung.

4. Ratioskala oder Verhältnisskala

Im Gegensatz zur Intervallskala ist bei der Verhältnisskala ein absoluter Nullpunkt vorhanden.
Die Verhältnisse der zugeordneten Messwerte (Zahlen) reflektieren die Verhältnisse der Ausprägung des gemessenen Merkmals.
<u>Beispiel</u>: Eine Klasse mit 10 Schülern ist halb so groß wie eine Klasse mit 20 Schülern

Tab. 1.2: Skalentypen im Überblick

Skalentyp/ Meßniveau	Eigenschaft	Beispiele
1. Nominalskala	• vollständige, sich gegenseitig ausschließende Kategorien	Geschlecht, Studiengang, Nationalität, etc.
2. Ordinalskala (Rangskala)	• Ordnung nach dem Grad der Ausprägung • Rangfolge sichtbar	Klein - mittel – groß, Schulnoten; Motivationszustand
3. Intervallskala	• Rangordnung mit definierten Abständen • ohne absoluten Nullpunkt	Grad Celcius, Testnoten
4. Ratio- oder Verhältnisskala	• Rangordnung m. definierten Abständen • mit absolutem Nullpunkt	kg, cm, Grad Kelvin

[6] Diehl/Kohr 1999, S.14

2. Deskriptive Statistik

2.1 Univariate Verteilung

Bei der univariaten Verteilung werden einzelne Variablen untersucht. Bedeutsame Informationen liefern die Häufigkeitsanalyse (2.1.1), der arthmetische Mittelwert (2.1.2) und die Standardabweichung (2.1.3).

2.1.1 Häufigkeitsverteilung[7]

Bei der Analyse von einzelnen Variablen beginnt man in der Regel mit der Anfertigung tabellarischer oder graphischer Häufigkeitsverteilungen, weil sie einen ersten charakteristischen (auch visuellen) Überblick über die Daten geben.

Beispiel: Im Eingangstest der Veranstaltung „Forschungsmethoden der BWP 02/03“ wurde den Studierenden folgendes Statement zur Entscheidung vorgelegt: *„Von dieser Veranstaltung erwarte ich, dass sie für die Praxis hilfreich ist“ (trifft voll zu .../ .. trifft überhaupt nicht zu)*

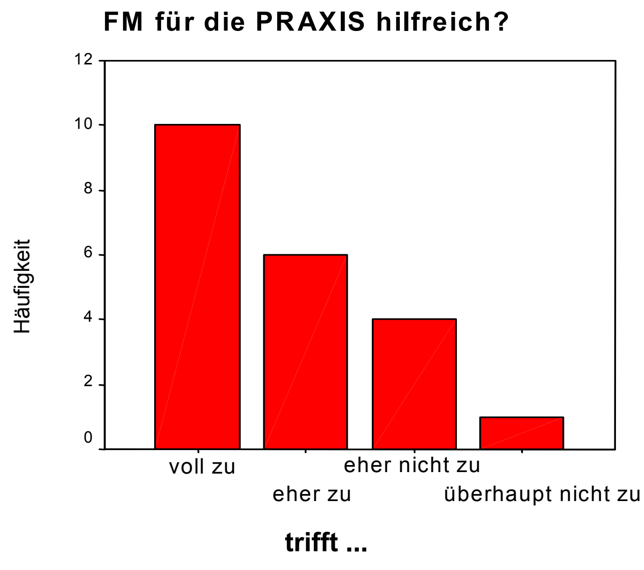
Die Verteilung der rangskalierten Daten läßt sich tabellarisch und graphisch darstellen:

(1) Tabellarische Darstellung

FM für die PRAXIS hilfreich?

trifft ...		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	voll zu	10	45,5	47,6	47,6
	eher zu	6	27,3	28,6	76,2
	eher nicht zu	4	18,2	19,0	95,2
	überhaupt nicht zu	1	4,5	4,8	100,0
	Gesamt	21	95,5	100,0	
Fehlend	weis nicht	1	4,5		
Gesamt		22	100,0		

(2) Graphische Darstellung, z.B. Balkendiagramm



Typische Verteilungsformen:

- unimodal / bimodal
- rechts- /linksschief*
- steil / flach
- j-förmig

* linksschief: rechts ist höchster Wert

Aufgabe: Welche Information ist dieser Häufigkeitsverteilung zu entnehmen?

[7] vgl. Benninghaus 2005, S. 29-35

2.1.2 Arithmetischer Mittelwert[8] „$\bar{x}$“

Mittelwerte - oder auch Maßzahlen der zentralen Tendenz - geben einen verdichteten Eindruck über eine Verteilung einer Variablen wieder. Sie sind deshalb eine gute Schätzung (repräsentativer Wert) für einen beliebigen Wert der Verteilung.

Die folgenden drei Mittelwerte sind bedeutsam:

1. die häufigste vorkommende Variablenausprägung ⇒ **Modus „h“**
2. der Punkt, der exakt zwischen der oberen und der unteren Hälfte der Verteilung liegt ⇒ **Median „$\tilde{x}$“**
3. der Durchschnittswert der Verteilung ⇒ **Arithmetisches Mittel „$\bar{x}$“**

Der **arithmetische Mittelwert „$\bar{x}$“** ist das am häufigsten verwendete Durchschnittsmaß.

<u>Definition:</u> Die Summe aller Messwerte durch die Anzahl der Messwerte.

Messwerte:	⇒	**$x_1, x_2, x_3, x_4, \ldots\ldots x_i$**
Anzahl der Messwerte:	⇒	**N**
Arithmetisches Mittel	⇒	$\bar{x}$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots\ldots x_N}{N} \qquad \bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Kommen Meßwerte mehrmals vor, wird folgende Formel verwendet

$$\bar{x} = \frac{f_1 * x_1 + f_2 * x_2 + f_3 * x_3 \ldots . + f_i * x_i + \ldots\ldots f_k * x_k}{N}$$

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i * x_i}{N} \qquad \text{wobei} \qquad N = \sum_{i=1}^{k} f_i$$

<u>Aufgabe:</u> Wie groß ist der Mittelwert der folgenden Verteilung?

Wert x_i	**Häufigkeit f_i**	**$f_i x_i$**
1	10	
2	6	
3	4	
4	1	
weiß nicht	1	
	N =	**$\sum f_i x_i$ =**

$\bar{x}$ =

[8] vgl. Benninghaus 2005, S. 36-50

Wichtige Eigenschaften des arithmetischen Mittelwertes „$\bar{x}$“:

- Die Summe der Abweichungen aller Werte von ihrem arithmetischen Mittel ist gleich Null

$$\sum (x_i - \bar{x}) f_i = 0$$

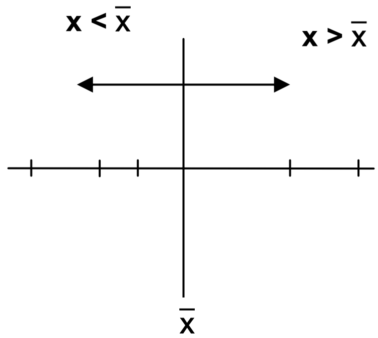
- Die quadrierte Summe der Abweichungen aller Werte von ihrem arithmetischen Mittel ist ein Minimum, d.h. bei jedem anderen Wert wäre sie größer

$$\sum ((x_i - \bar{x}) f_i)^2 = \text{min.}$$

Aufgabe: Prüfung der ersten Eigenschaft - $\sum$ **f_i (x_i-$\bar{x}$) = 0?**

Wert x_i	Häufigkeit f_i	$f_i x_i$	$(x_i - \bar{x})$	$f_i\ (x_i - \bar{x})$
1	1			
2	1			
3	5			
4	1			
5	4			
6	3			
	N =	$\sum f_i x_i$ =		$\sum f_i\ (x_i - \bar{x}) = 0$

$\bar{x}$ =



2.1.3 Standardabweichung[9] „s“

Mittelwerte geben nur eine Information über die zentrale Tendenz einer Datenverteilung wieder; der Grad der Homogenität bzw. Heterogenität der beobachteten Werte ist daraus jedoch nicht ersichtlich. Dieses Merkmal einer Verteilung wird über Streuungswerte ausgedrückt.

Folgende drei Kennwerte kommen in statistischen Analysen vor:

(1) Differenz zwischen dem größten und kleinsten Wert ⇒ **Range „R“**

(2) Die Streuung in der Mitte der Verteilung ⇒ **mittlere Quartilabstand „Q_m“**

(3) Summe der quadrierten Abweichungen vom Mittel ⇒ **Standardabweichung „s“**

Die **Standardabweichung „s“** ist das am häufigsten verwendete Streuungsmaß. Sie ist die Quadratwurzel der **Varianz „s^2“** (Variation), die wiederum aus dem Quotient „Summe der quadrierten Abweichungen vom arithmetischen Mittelwert “ durch die „Fallzahl“ (N) gebildet wird.

Varianz:

$$s^2 = \frac{\sum_{i=1}^{k} f_i * (x_i - \bar{x})^2}{N}$$

Standardabweichung:

$$s = \sqrt{s^2} \qquad s = \sqrt{\frac{\sum_{i=1}^{k} f_i * (x_i - \bar{x})^2}{N}}$$

wobei: $N = \sum_{i=1}^{k} f_i$ k: Anzahl unterscheidbarer Ausprägungen von x

[9] vgl. Benninghaus 1992, S. 51-62

<u>Aufgabe (Fortsetzung):</u> Berechnung der Varianz „s^2" und der Standardabweichung „s"

Wert x_i	Häufigkeit f_i	$f_i x_i$	$(x_i - \bar{x})$	$f_i (x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
1	1	1				
2	1	2				
3	5	15				
4	1	4				
5	4	20				
6	3	18				
	N = 15	$\sum f_i x_i = 60$		$\sum f_i (x_i - \bar{x}) = 0$		$\sum f_i (x_i - \bar{x})^2 =$

$\bar{x}$ **=**

s^2 **=**

s =

<u>Abschließende Übungen</u>

Aufgabe 1: Welche Skalentypen liegen vor?

Körpergröße, Zuordnung der Jahreszeiten in Temperaturklassen, Staatsangehörigkeit, Schularten an beruflichen Schulen, Berufszufriedenheit von Lehrenden an beruflichen Schulen; Kopfnoten in Zeugnissen,

Läßt sich eine Ordinalskala in eine Intervallskala transformieren und umgekehrt?

Aufgabe 2:

Eine Gruppe von Studierenden (N=42) wurde über ihre Mensa-Erfahrungen befragt und dazu folgendes Statement vorgelegt: *„Die Wartezeit an der Essensausgabe ist zu groß"* (stimme voll zu ..... überhaupt nicht zu). Das Ergebnis sieht wie folgt aus:

Antwort: stimme..	**Wert x_i**	**Häufigkeit f_i**	**$(f_i\, x_i)$**	**$(x_i - \bar{x})$**	**$(x_i - \bar{x})^2$**	**$f_i\, (x_i - \bar{x})^2$**
voll zu	1	20				
eher zu	2	12				
eher nicht zu	3	8				
überhaupt nicht zu	4	2				
		N = 42	**$\sum f_i x_i =$**			**$\sum f_i (x_i - \bar{x})^2 =$**

<u>Gesucht sind:</u>

- Modus: **h = ?**
- arithmetisches Mittel: **$\bar{x}$ = ?**
- Standardabweichung, Varianz: **s, s^2 = ?**

2.2 Bivariate Verteilung

Bei bivariaten Verteilungen wird der Zusammenhang/die Beziehung zwischen zwei Variablen untersucht. Bedeutsame Informationen liefern die Kreuztabelle (2.2.1) und Korrelations-, bzw. Assoziationsmaße (2.2.2 bis 2.2.4).

2.2.1 Kreuztabelle (Bivariate Tabelle)[10]

Analog der univariaten Anaylse erscheint es sinnvoll, sich einen Überblick über die Datenlagen zu verschaffen. Dies ist mit der Kreuztabelle möglich, die im Grunde einer (bivariaten) Häufigkeitsverteilung entspricht.

In allgemeiner Darstellung hat die Kreuztabelle folgende Form:

		Variable x				
		x_1	**x_2**	**...**	**x_j**	
Variable y	**y_1**	**f_{11}**	**f_{12}**		**f_{1j}**	$\sum_{Zeile\ 1}$
	y_2	**f_{21}**	**f_{22}**		**f_{2j}**	$\sum_{Zeile\ 2}$
	...					**...**
	y_i	**f_{i1}**	**f_{i2}**		**f_{ij}**	$\sum_{Zeile\ i}$
		$\sum_{Spalte1}$	$\sum_{Spalte2}$	**...**	$\sum_{Spaltej}$	$\sum = N$

Folgende Aspekte sind zu berücksichtigen:

Bezeichnung:

- Allgemein: r x c (r=row =Zeile/c=column=Spalte)
- die Struktur der Tabelle ist abhängig von der Skalierung der Variablen; die kleinste Kreuztabelle ist eine 2x2 Tabelle
- Variablen: x ⇒ unabhängig (Spalte) y ⇒ abhängig (Zeile)

Zuordnung: welche ist die unabhängige, welche die abhängige Variable??

⇒ entlang der festgelegten Hypothesen

Häufigkeiten:

- innerhalb der Tabelle (konditionale/univariate Verteilung): f_{11} bis f_{ij}
- am Tabellenrand (Marginalverteilung):

Zeilensumme 1: $\sum_{Zeile\ 1} = \sum_{j=1}^{c} f_{1j}$ Spaltensumme 1: $\sum_{Spalte1} = \sum_{i=1}^{r} f_{i1}$

⇒ Die Summe aller Zeilen oder Spalten ergibt die Fallzahl N.

[10] vgl. Benninghaus 2005, S. 66-81

Für die Analyse der Kreuztabelle ist bedeutsam:

- Festlegung der gültigen Daten (i. d. R. werden „weiß nicht“- Angaben ausgeschlossen)
- die Intensität/Stärke des Zusammenhangs der beiden Variablen ist an der konditionalen Verteilung sichtbar; die Marginalverteilung ist unabhängig von der Beziehung.
- Je mehr sich die Häufigkeiten in den Zellen **entlang der Diagonalen** der Tabelle bilden, **umso stärker** ist der Zusammenhang

Beispiel: Im Eingangstest der Veranstaltung „Forschungsmethoden der BWP 02/03“ wurden die Studierenden zudem gefragt, wie nützlich die Empirische Sozialforschung ist. Untersucht werden soll nun, welcher Einfluß der Studiengang auf die Einschätzung hat.

Unabhängige Variable: *Zugehörigkeit zum Studiengang*
Abhängige Variable: *Nützlichkeit empirischer Sozialforschung*

Datenmatrix (Rohdaten)

Nr.	Studiengang	Einschätzung Nutzen
1	1	1
2	2	3
3	1	3
4	1	3
5	2	3
6	2	2
7	2	2
8	1	88
9	1	4
10	1	2
11	2	88
12	2	3
13	2	3
14	2	2
15	2	3
16	2	1
17	1	4
18	2	3
19	2	3
20	1	4
21	1	3
22	2	1

Kreuztabelle (bivariate Tabelle)

Einschätzung: Empirische Sozialforschung weniger nützlich.

Dem stimme ich..

	Studiengang: 1 = MA	2 = TP	
1 = voll zu	**1**	**2**	**∑= 3**
2 = eher zu	**1**	**3**	**∑= 4**
3 = eher nicht zu	**3**	**7**	**∑= 10**
4 = überhaupt nicht zu	**3**	**-**	**∑= 3**
88 = weiß nicht	1	1	**∑= 2**
	∑= 9	**∑= 13**	**∑= 22**

Aufgabe/Fragen:

- Wie bildet sich die Kreuztabelle aus der Datenmatrix?
- Ist ein Zusammenhang zwischen den beiden Variablen erkennbar?
- Welche tendenzielle Aussage läßt sich treffen?

2.2.2 Korrelationsmaße im Überblick

Der Darstellung des Zusammenhangs mit Hilfe einer Kreuztabelle schließt sich (anlog der univariaten Verteilung) die Berechnung von Koeffizienten/Kennzahlen an, die eine Aussage über die Stärke/Dynamik der Beziehung/Assoziation geben. Zur Berechnung solcher Kennzahlen gibt es verschiedene Ansätze. Häufig verwendet werden Kennzahlen auf Basis von **"Chi-quadrat[11]",** die vor allem für nominalskalierte Daten in Frage kommen (siehe 2.2.3). Prinzipiell wären sie für alle Skalentypen geeignet, werden aber für anspruchsvollere Skalentypen weniger verwendet, weil man damit ohne Grund Einschränkungen der Aussagekraft in Kauf nehmen würde.
Auf einer anderen Berechnungsüberlegung stehen die sogenannten „**PRE-Maße[12]**", für die es auf allen Skalenebenen angepasste Kennwerte gibt. In diesem Skript werden Pearsons r, r^2 näher vorgestellt (siehe 2.2.4), die für Intervallskalen geeignet sind.

__Hinweis:__ In der Praxis wird häufig Pearsons "r" auch bei rangskalierten Daten verwendet, bzw. rangskalierte Variablen werden wie intervallskalierte Daten verarbeitet, obwohl das streng genommen unzulässig ist.

Abb. 2.2.2: Wichtige Korrelationsmaße im Überblick

Skalentyp	**Berechnungskonzept**	
	Chi-quadrat	**PRE-Maße**
nominal	**Φ (Phi)** **Cramers V**	λ (Lambda)
ordinal/ rangskaliert	-	γ (Gamma)
intervall/ metrisch	-	**Pearsons r, r^2**

2.2.3 „Chi-quadrat"- basierte Koeffizienten[13] „Φ" und „Cramers V"

Im folgenden sollen nun die Berechnung der „Chi-quadrat"- basierten Koeffizienten „∅" und „Cramers V" betrachtet werden.. Der Berechnung der Koeffizienten geht die Berechnung des Faktors „Chi-quadrat" voraus.

„Chi-quadrat" ($\boldsymbol{x^2}$) basiert auf folgender Grundüberlegung:

> **Welcher Unterschied besteht, wenn die existierende Beziehung mit einer theoretischen Nichtbeziehung verglichen wird?**

etwas differenzierter:

1. Wie sähe die Tabelle aus, wenn x und y nicht miteinander assoziiert wären?
2. Welche Differenz stellt sich ein, wenn die aktuelle Tabelle (Kontingenztabelle) mit der Tabelle, die aufgrund der angenommenen Nichtbeziehung gebildet wurde (Indifferenztabelle), verglichen wird?
3. Differieren die Tabellen, folgert man daraus, dass x und y miteinander in Beziehung stehen

---

[11] Chi-Quadrat ist ein Maß für die Abweichung von der statistischen Unabhängigkeit
[12] PRE: Proportional Reduction of Error (Proportionale Fehler Reduktion)
[13] vgl. Benninghaus 1992, S. 104-121

„Chi-quadrat" **(x^2)** berechnet sich nach folgender Formel:

$$\boxed{x^2 = \sum \frac{(f_b - f_e)^2}{f_e}}$$

f_b absolute Häufigkeiten Kontingenztabelle
f_e absolute Häufigkeiten Indifferenztabelle

Beispiel: Eine Untersuchung zur Lohnzufriedenheit ergab folgende Daten. Wie groß ist x^2?

(Kontingenztabelle)

		Beschäfigungsstatus		
		Arbeiter	**Angestellte**	
Lohnzufrieden heit	**gering**	40	10	50
	hoch	10	40	50
		50	50	100

Bei der Berechnung von x^2 geht man folgendermaßen vor:

(1) Bestimmung von f_b
Die Häufigkeiten der Kontingenztabelle entsprechen den Zellhäufigkeiten der beobachteten bzw. vorgefundenen Untersuchungsdaten (s.o.)

(2) Bestimmung von f_e
Nach der erwähnten Vorgehensweise muß eine zweite Kreuztabelle, die sogenannte „Indifferenztabelle" gebildet werden, die die „Nichtbeziehung" beider Variablen repräsentiert. Sie wird aus der Marginalverteilung berechnet, weil diese Verteilung unabhängig vom Zusammenhang beider Variablen ist.

Die Zelle f_{e11} berechnet sich dabei wie folgt:

$$f_{e11} = \frac{\sum \text{Zeile1 x} \sum \text{Spalte1}}{N} = f_{e11} = \frac{50 \text{ x } 50}{100} = 25$$

wenn analog dazu die anderen Zellen berechnet werden, ergibt sich für die Indifferenztabelle folgendes Bild:

(Indifferenztabelle)

		Beschäfigungsstatus		
		Arbeiter	**Angestellte**	
Lohnzufrieden heit	**gering**	25	25	50
	hoch	25	25	50
		50	50	100

(3) Berechnung von x^2 nach obiger Formel:

Zeile i	**Spalte j**	**f_b**	**f_e**	**$(f_b - f_e)$**	**$(f_b - f_e)^2$**	**$\frac{(f_b - f_e)^2}{f_e}$**
1	1					
1	2					
2	1					
2	2					
		N = 100	N = 100			**$\sum = x^2 =$**

Der im Beispiel ermittelte Wert von $\boldsymbol{x}^2$= 36 wird auch von der Fallzahl N bestimmt, d.h. würde man eine Aussage über die Assoziation zwischen den beiden Variablen „Lohnzufriedenheit“ und „Beschäftigungsstatus“ treffen, müßte immer auch die Fallzahl N mit angegeben werden.

Dieser Nachteil wird umgangen, wenn aus dem **$\boldsymbol{x}^2$-Wert** die Koeffizienten „Φ“ und **„CramersV“** berechnet werden

Berechnung von „Φ“ (Phi-Koeffizient) (für 2x2-Tabellen)

$$\boxed{\Phi = \sqrt{\frac{\boldsymbol{x}^2}{N}}}$$

Berechnung von „CramersV“ (für alle rxc-Tabellen)

$$\boxed{\mathbf{V} = \sqrt{\frac{\boldsymbol{x}^2}{\mathrm{N} * \min(\mathrm{r}-1, \mathrm{c}-1)}}}$$

„min“: es wird zunächst geprüft, ob die Anzahl der Spalten oder der Zeilen geringer ist. Der kleinere Wert wird eingesetzt; *<u>Beispiel:</u> 2x3 Tabelle; Anzahl der Zeilen wäre der kleinere Wert*

„Φ“ und **„CramersV“** nehmen Werte an zwischen

0 (≙ kein Zusammenhang) und **1 (≙ max. Zusammenhang)**

<u>Hinweis:</u>

In der Praxis findet man häufig folgende Interpretationen vor:

0,1 - 0,3	schwacher Zusammenhang
0,4 – 0,5	mittlerer Zusammenhang
> 0,5	starker Zusammenhang

In der Literatur wird eine solche Zuschreibung eher vermieden und darauf hingewiesen, dass die Ergebnisse spezifisch, u.a. im Kontext bisheriger Untersuchungswerte, zu interpretieren sind.[14]

---

[14] vgl. z.B. Diehl / Kohr 1999, S.161

Abschließende Übung

Datengrundlage: Eingangstests „Forschungsmethoden WS02/03“ **(siehe Beispiel bei 2.2.1)**

Wie groß ist der Zusammenhang zwischen den Variablen „Studiengang“ und „Einschätzung des Nutzens empirischer Sozialforschung im Vergleich zur Natur-/Ingenieurwissenschaftlichen Forschung“? **Zu berechnen sind „Φ“ und „Cramers V“**

Gegeben: Kontingenztabelle f_b

Einschätzung: Empirische Sozialforschung weniger nützlich. Dem stimme ich..

	Studiengang: 1 = MA	Studiengang: 2 = TP	
1 = voll zu	**1**	**2**	3
2 = eher zu	**1**	**3**	4
3 = eher nicht zu	**3**	**7**	10
4 = überhaupt nicht zu	**3**	**-**	3
	8	12	**N= 20**

Gesucht: Indifferenztabelle f_e

Einschätzung: Empirische Sozialforschung weniger nützlich. Dem stimme ich..

	Studiengang: 1 = MA	Studiengang: 2 = TP	
1 = voll zu			3
2 = eher zu			4
3 = eher nicht zu			10
4 = überhaupt nicht zu			3
	8	12	**N= 20**

Berechnungstabelle

Zeile i	**Spalte j**	f_b	f_e	$(f_b - f_e)$	$(f_b - f_e)^2$	$\frac{(f_b - f_e)^2}{f_e}$
1	1					
1	2					
2	1					
2	2					
3	1					
3	2					
4	1					
4	2					
		N = 20				$\sum = x^2 =$

$\Phi =$

$V = ...$

2.2.4 „PRE"-Koeffizienten „r" und „r²"[9]

(Pearsonscher Produkt-Moment-Korrelations-Koeffizient „r" bzw. „r^2")

1. Vorbemerkungen

- Pearson's „r", bzw. „r^2" basiert auf dem Modell der proportionalen Fehlerreduktion (PRE-Maße): In welchem Maß verringert sich der Vorhersagefehler einer (abhängigen) Variable (AV), wenn zur Vorhersage eine zweite Variable (UV) hinzugezogen wird. Je größer die Reduzierung des Vorhersagefehlers, umso größer ist dann der Zusammenhang zwischen den beiden Variablen.

  Beispiel: Inwiefern läßt sich die Vorhersage von Testleistungen in Mathe (AV) verbessern, wenn bekannt ist, wieviel Zeit zuvor geübt wurde (UV)?

- Bei allen PRE-Koeffizienten (z.B. λ, r) erfolgt die Berechnung in 4 Schritten:
  (1) Vorhersage von AV aufgrund ihrer eigenen Verteilung
  (2) Vorhersage von AV auf der Basis von UV
  (3) Berechnung der jeweiligen Vorhersagefehler von (1) und (2) $\Rightarrow E_1$ und E_2
  (4) Berechnung der proportionalen Fehlerreduktion: $r = (E_1 - E_2) / E_1$

- Vor der Berechnung sollte zunächst aus dem Streudiagramm oder der Kreuztabelle geprüft werden, welcher Zusammenhang (linear, u-fömig, j-fömig, etc.) besteht. Entsprechend gestaltet sich das Berechnungsverfahren. Die folgende Berechnung bezieht sich auf lineare Zusammenhänge.

2. Berechnung von „r^2" (Proportionale Reduktion des Vorhersagefehlers)

Beispiel: Der Zusammenhang zwischen Testleistungen und Übungszeit einer Gruppe von SchülerInnen ergibt folgende Daten:

Tab.1: Ausgangsdaten, Testnote und Übungszeit

SchülerIn	Übungszeit [h] x	Testnote y
1	10	2
2	6	4
3	7	2
4	5	3
5	2	5
6	4	1
7	3	3
8	2	4
9	6	2
10	8	1

Mittelwerte:

$\check{y}$ = 2,7

$\bar{x}$ = 5,3

[9] vgl. Benninghaus 2005, S. 185-227

(1) Vorhersage der Testnote (AV) aufgrund der eigenen Verteilung

Hier wird der Mittelwert herangezogen, weil er die „beste“ Vorhersage liefert (vgl. 2.1.2):

- Die Summe der Abweichungen der einzelnen Werte vom Mittelwert „$\check{y}$“ ist Null.
- Die quadrierten Abweichungen „s^2“ (Gesamtvariation oder Varianz) ist minimal, bzw. bei jedem anderen Punkt der Verteilung wäre diese Summe größer.

Der Vorhersagefehler E_1 kann durch diese Gesamtvarianz „s^2“ ausgedrückt werden. Sie berechnet sich nach der Formel:

Gesamtvarianz: $\boxed{s^2 = \Sigma\, (y - \check{y})^2 / N}$

Die entsprechenden Tabellenwerte:

Tab.2: Werte zur Berechnung der Gesamtvarianz der abhängigen Variable „Testnote“

SchülerIn	**Testnote Y**	**$(y - \check{y})$**	**$(y - \check{y})^2$**
1	2	-0,7	0,490
2	4	1,3	1,690
3	2	-0,7	0,490
4	3	0,3	0,090
5	5	2,3	5,290
6	1	-1,7	2,890
7	3	0,3	0,090
8	4	1,3	1,690
9	2	-0,7	0,490
10	1	-1,7	2,890
N = 10	**$\check{y}$ = 2,7**	**Σ = 0,0**	**Σ = 16,100**

Vorhersagewert: $\check{y}$ = 2, 7

Vorhersagefehler E_1 (Gesamtvarianz): s^2 = 16,1 / 10 = 1,61

(2) Vorhersage der Testnote aufgrund der Information über die Übungszeit

Für diese Vorhersage wird nun die Regressionsgerade[10] herangezogen, weil sie - analog den bereits genannten Kriterien bei der Bestimmung von E_1 - ein „Ort“ ist, für den sich ein Minimum an Abweichungen ergibt, und zwar:

- die Summe der vertikalen Abweichungen (y-Werte) ist gleich Null
- die Summe der quadrierten Abweichungen von y ist minimal

Geradengleichung: $\boxed{y_r = a + bx_r}$

für die Regressionsgerade ergeben sich folgende Werte für b und a:

$\boxed{b = \Sigma(y - \check{y})(x - \bar{x}) \;/\; \Sigma(x - \bar{x})^2}$ $\boxed{a = \check{y} - b\,\bar{x}}$

[10] Der Begriff „Regression“ geht nach Benninghaus auf eine Untersuchung von Galton im 19 Jh. zurück. Galton untersuchte die Körpergröße von Kindern in Abhängigkeit von der Größe ihrer Eltern und stellte dabei zwar einen einigermaßen linearen Zusammenhang fest (große Eltern – eher große Kinder, etc.), gleichzeitig stellte sich jedoch eine Tendenz zur Durchschnittsgröße heraus, die auf die Durchschnittsgröße der Eltern aller Kinder zurückzuführen (regredieren: auf früheres zurückzuführen) ist. Wenn die Eltern größer (kleiner) waren, als der Durchschnitt, tendierten ihre Kinder dahin, kleiner (größer) zu sein als sie (vgl. Benninghaus 2005, S.189).

Auf das Beispiel übertragen bedeutet dies:

Tab. 3: Gesamtvariation und Gesamtvarianz der abhängigen Variable Testnote

$(y - \breve{y})$	$(x - \bar{x})$	$(x - \bar{x})^2$	$(y - \breve{y})\,(x - \bar{x})$
-0,7	4,7	22,09	-3,29
1,3	0,7	0,49	0,91
-0,7	1,7	2,89	-1,19
0,3	-0,3	0,09	-0,09
2,3	-3,3	10,89	-7,59
-1,7	-1,3	1,69	2,21
0,3	-2,3	5,29	-0,69
1,3	-3,3	10,89	-4,29
-0,7	0,7	0,49	-0,49
-1,7	2,7	7,29	-4,59
Σ = 0,0	**Σ = 0,0**	**Σ = 62,1**	**Σ = -19,1**

Berechnung der Regressionsgeraden:

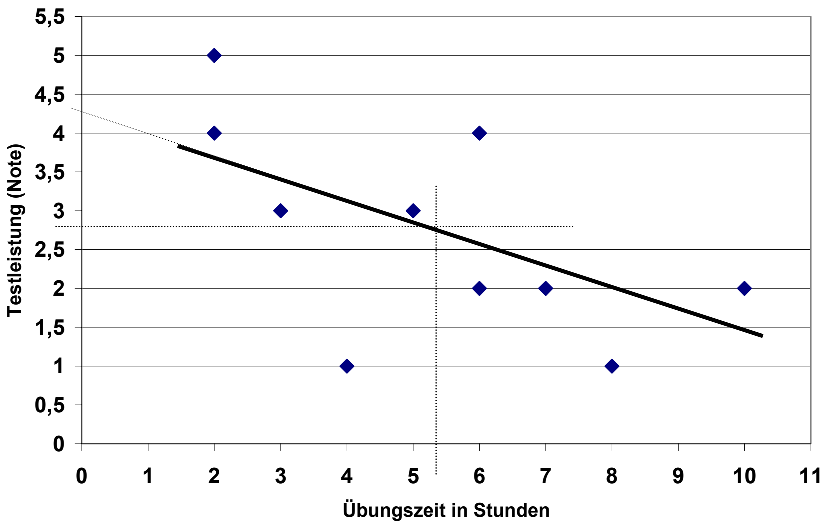
b = - 19,1 / 62,1 = - 0,307

a = 2,7 – (- 0,307 x 5,3) = 4,32

$$\boxed{y_r = 4{,}32 - 0{,}307\, x_r}$$

Im Schaubild sieht die Gerade dann wie folgt aus:

Abb. 1: Testleistung in Abhängigkeit von der Übungszeit



Anmerkungen zur Geraden:

- Die beiden Mittelwerte der Variablen $\bar{x}$ = 5,3 und $\breve{y}$ = 2,7 bilden einen Punkt der Geraden
- Die Gerade gilt nur für den Wertebereich

Mit der Geraden kann nun für jede Übungszeit eine Note „vorhergesagt“ werden. Der Vorhersagefehler E_2 errechnet sich aus der Abweichung zwischen y(x) und $y_r(x)$. Analog dem Vorgehen bei E_1 wird dazu die Gesamtvarianz s_r^2 berechnet:

Gesamtvarianz s_r^2: $$\boxed{s_r^2 = \Sigma\, (y - y_r)^2 / N}$$

Tab. 4: Berechnung der auf y_r bezogenen Variation und Varianz

Übungszeit [h] x	Testnote y	Geradenwert y_r	$(y - y_r)$	$(y - y_r)^2$
10	2	1,250	0,750	0,562
6	4	2,478	1,522	2,316
7	2	2,171	-0,171	0,029
5	3	2,785	0,215	0,046
2	5	3,706	1,294	1,674
4	1	3,092	-2,092	4,376
3	3	3,399	-0,399	0,159
2	4	3,706	0,294	0,086
6	2	2,478	-0,478	0,228
8	1	1,864	-0,864	0,746
	Σ = 27	**Σ = 27**	**Σ = 0**	**Σ = 10,226**

Vorhersagefehler E_2: $s_r^2 = 10{,}226 / 10 = 1{,}02$

Je weiter die Meßpunkte von der Regressionsgeraden entfernt liegen, um so größer wird die Varianz bzw. der Vorhersagefehler.

(3) und (4) Berechnung der proportionalen Fehlerreduktion

Die Vorhersagefehler E_1 und E_2 wurden bereits berechnet. Die proportionale Reduktion des Vorhersagefehlers r^2 berechnet sich wie folgt:

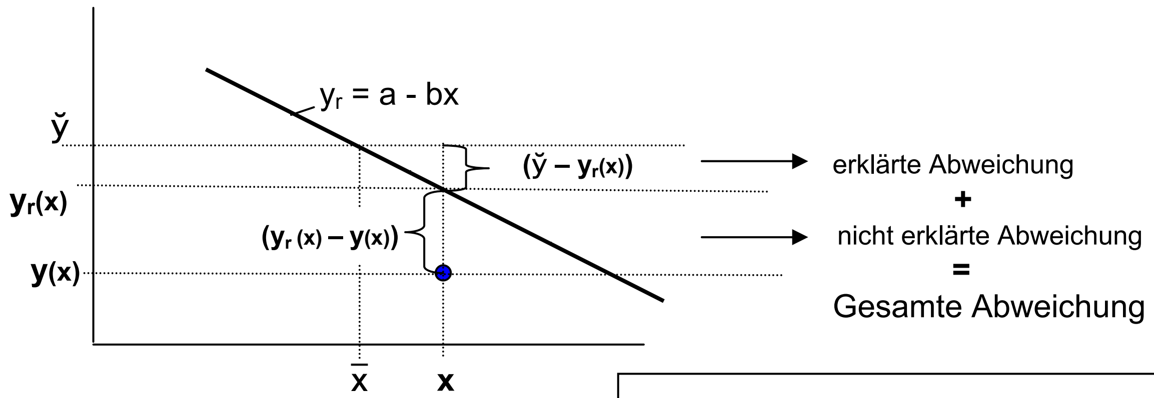
$$r^2 = (E_1 - E_2) / E_1 = (1{,}61 - 1{,}02) / 1{,}61 = 0{,}366$$

Welche Bedeutung dieser Wert hat, kann anhand einer anderen Bezeichnung deutlich gemacht werden:

r^2 = (Gesamtvarianz – Nicht erklärte Varianz) / Gesamtvarianz

oder **r^2 = erklärte Varianz / Gesamtvarianz**

d.h. im vorliegenden Fall können etwa 36% der Variation der y-Variable „Testnote“ durch die x-Variable „Übungszeit“ erklärt werden. Je weiter die y-Werte um die Regressionsgerade streuen, umso kleiner wird die erklärte Varianz. An der Einzelabweichung eines Messpunkts wird der Zusammenhang nochmals deutlich:



$$\boxed{(\breve{y} - y(x)) = (\breve{y} - y_r(x)) + (y_r(x) - y(x))}$$

3. Berechnung von „r“ (Pearsonsche Koeffizient)

Gewöhnlich wird der Koeffizient „r“ angegeben, den man nach dem gleichen Schema berechnen kann. Die Messwerte werden dazu zunächst über die Standardabweichung „s“ standardisiert , d.h. in das Koordinatenkreuz geschoben. Die Werte x und y werden in z_x und z_y transformiert, wobei

$$\boxed{z_x = (x - \bar{x}) / s_x} \qquad \boxed{s_x = \sqrt{\Sigma \; (x - \bar{x})^2 / N}}$$

$$\boxed{z_y = (y - \breve{y}) / s_y} \qquad \boxed{s_y = \sqrt{\Sigma \; (y - \breve{y})^2 / N}}$$

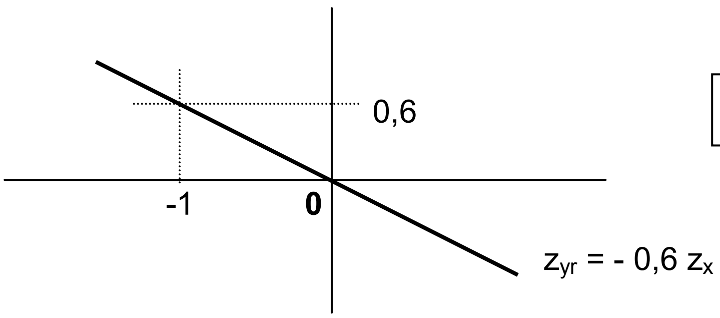
Wenn man nun dazu die Regressionsgerade berechnet, erhält man durch die Standardisierung eine Gerade nach dem Muster

$$\boxed{z_{yr} = b \, z_x}$$

Weil alle z-Werte um den Ursprung herum liegen, verläuft die Gerade durch den Nullpunkt des Koordinatenkreuzes. Die Steigung ergibt sich aus

$$\boxed{b = \Sigma \, (z_x \, z_y) / N}$$

Die Zahlenbeispiele eingesetzt ergibt dies ein Wert von -0,6. Im Schaubild sieht dies wie folgt aus:



$$\boxed{b = r = - 0{,}6}$$

Die Steigung dieser standardisierten Regressionsgeraden (Hinweis: diese Gerade ist mit der oben berechneten Geraden nicht identisch; Ausnahme: das Vorzeichen der Steigung) entspricht dabei dem Koeffizienten **„r“.** Diese Größe gibt einen Anhaltswert darüber, wie sich die abhängige Variable bei Variation der unabhängigen Variablen verändert. Auf das Beispiel bezogen: Welchen Notenunterschied bewirkt eine Stunde mehr Übungszeit.

Der Koeffizient „r“ kann auch direkt aus den Messwerten mit folgender Formel berechnet werden:

$$r = \frac{[N * \sum(x * y) - \sum x * \sum y]}{\sqrt{[(N * \sum x^2) - (\sum x)^2] * [(N * \sum y^2) - (\sum y)^2]}}$$

Auf das Beispiel angewandt ergibt sich folgende Wertetabelle:

Übungszeit [h] **x**	**Testnote** **y**	**x^2**	**y^2**	**x * y**
10	2	100	4	20
6	4	36	16	24
7	2	49	4	14
5	3	25	9	15
2	5	4	25	10
4	1	16	1	4
3	3	9	9	9
2	4	4	16	8
6	2	36	4	12
8	1	64	1	8
Σ = 53	**Σ = 27**	**Σ = 343**	**Σ =89**	**Σ =124**

$$r = \frac{(10 * 124 - 53 * 27)}{\sqrt{[(10 * 343) - 53^2] * [(10 * 89) - 27^2]}} = \frac{(1240 - 1431)}{\sqrt{621 * 161}} = \frac{-191}{\sqrt{99981}} = \frac{-191}{316{,}19}$$

$$r = -0{,}604$$

4. Zusammenhang zwischen „r^2“ und „r“

Die beiden Koeffizienten hängen direkt über die Wurzel miteinander zusammen, d.h.

$$\boxed{r = \sqrt{r^2}}$$

Aus „r“ kann direkt auf „r^2“ geschlossen werden und umgekehrt. Da sich „r“ im Intervall (-1 / 0 / +1) bewegt, ist „r“ immer größer als „r^2“, während r^2 immer positiv ist. Zusammenfassend beschreibt

r^2 die Stärke einer Beziehung

r die Dynamik einer Beziehung

Nach Benninghaus[11] wird häufig der Fehler gemacht, an Hand eines großen „r“-Wertes auf einen starken Zusammenhang zu schließen, obwohl der in Wirklichkeit viel geringer ist. Beispiel: $r = 0{,}6 \Rightarrow r^2 = 0{,}36$ oder $r = 0{,}4 \Rightarrow r^2 = 0{,}16$

[11] Benninghaus 2005, S.221f.

Abschließende Übung

Datengrundlage: Eingangstests „Forschungsmethoden WS02/03“

Wie groß ist der Zusammenhang zwischen den Items:

- *„Empirische Sozialforschung im Vergleich zur Natur-/ingenieurwissenschaftlichen Forschung weniger nützlich“ (Allg. Einstellung: unabhängige Variable)*
  und

- *„Besuch der Veranstaltung vor allem wegen des Scheines..“ (Motivation: abhängige Variable)*

*(*Skala: 1 = stimme voll zu ...... 4 = stimme überhaupt nicht zu)

Motivation: „Besuch vor allem wegen Schein“	*Einschätzung: „Empirische Sozialforschung weniger nützlich“* 1	2	3	4	
1	**1**	**3**	**4**	**-**	**8**
2	**2**	**1**	**2**	**-**	**5**
3	**-**	**-**	**2**	**-**	**2**
4	**-**	**-**	**2**	**3**	**5**
	3	**4**	**10**	**3**	N= 20

Gesucht und zu berechnen ist:

- eine graphische Darstellung der Werte
- die Regressionsgerade aus den nicht standardisierten Werten (nach der Berechnung in die Graphik übernehmen)
- die Varianzaufklärung „r^2“
- der Korrelationskoeffizient „r“

3. Schließende Statistik

Vorbemerkung[12]

Während die *Deskriptive Statistik* über Verteilungen und Kennzahlen die Charakteristik einer Stichprobe zu beschreiben versucht, zielt die *Schließende Statistik* (oder auch Inferenzstatistik) darauf ab, Aussagen zu treffen über den Zusammenhang zwischen einer Stichprobe und der Gesamtheit aller Fälle oder auch zwischen zwei oder mehreren Stichproben. Nachfolgend werden 2 Anwendungen näher betrachtet:

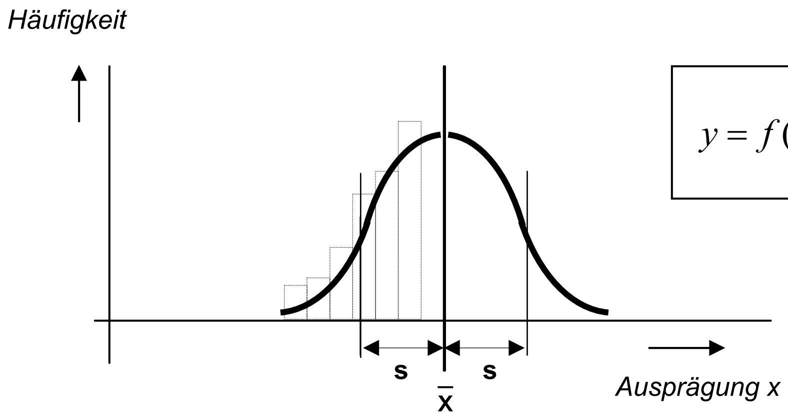
- „Mit welcher Wahrscheinlichkeit kann der in einer Stichprobe gefundene Mittelwert auf die Grundgesamtheit übertragen werden"? ⇒ Repräsentationsschluß (siehe 3.2)
- „Ist der bei einem Stichprobenvergleich sichtbare Unterschied zwischen zwei Mittelwerten zufällig entstanden oder ist er signifikant?" ⇒ t-Test (siehe 3.3)

Voraussetzungen für die Berechnung solcher verallgemeinernder Schlußfolgerungen sind sinnvolle Stichproben. Im Idealfall wären sie ein *zufällig* entstandenes Abbild der Grundgesamtheit. Der Stichprobenauswahl kommt deshalb eine besondere Bedeutung zu.

Innerhalb der *Schließenden Statistik* hat zudem die *Normalverteilung* eine zentrale Bedeutung. Sie dient als Prüfverteilung. d.h. die Messdaten werden bei Berechnungen auf diese „ideale Verteilung" bezogen.

3.1 Normalverteilung[13]

Viele meßbare Größen (z. B. Körpergewicht, Schuhgröße, Schulleistungen, etc.) verteilen sich bei größeren Fallzahlen in ähnlich charakteristischer Weise wie unten in der Kurve dargestellt: Die Häufigkeit ist im mittleren Bereich am größten, während sie an den Rändern gegen Null geht. Bei einer theoretischen, unendlich großen Fallzahl entspricht die Kurve der Normalverteilung (häufig auch „Gauß'sche Kurve" oder „Glockenkurve").



$$y = f(x) = \frac{1}{s\sqrt{2\pi}} * e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{s}\right)^2}$$

Die Normalverteilung hat folgende Eigenschaften:

- Arithmetischer Mittelwert, Modus und Median fallen zusammen
- Kurven-Enden nähern sich asymptotisch der Abszisse an
- Die beiden Kurvenpunkte, bei der die Steigung maximal wird (gleichzeitig auch Wendepunkte), liegen jeweils eine Standardabweichung „s" vom Mittelwert „$\bar{x}$" entfernt.
- Durch die Kenntnis der Kurvengleichung läßt sich über eine Integralrechnung bestimmen, dass im Intervall

	$\bar{x} \pm \mathbf{1s}$	68,3% ,	
zwischen	$\bar{x} \pm \mathbf{2s}$	95,5%	
und zwischen	$\bar{x} \pm \mathbf{3s}$	99,7%	der beobachtbaren Fälle liegen.

[12] Ausführlich dazu Sahner 1997, S. 9-18
[13] vgl. ebd. S. 25-37

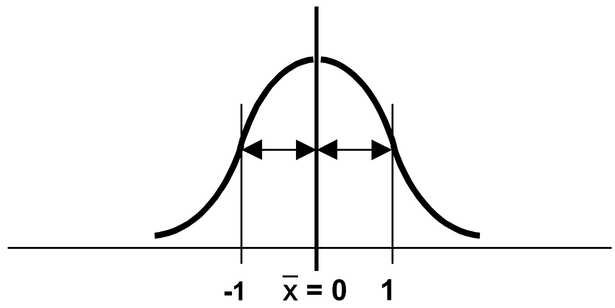
Die letztgenannten Eigenschaften führen also dazu, dass mit der Normalverteilung auch deskriptive Aussagen über die Stichprobe vorgenommen werden können; Voraussetzung: die gewonnene Verteilungscharakteristik entspricht (in etwa) einer Normalverteilung.

Der beschreibende Anwendungsbereich läßt sich auf beliebige Intervallgrenzen ausdehnen, wenn die Normalverteilung standardisiert wird, d.h. wenn die x-Werte mit der folgenden Formel in z_x Werte transferiert werden.

$$\boxed{z_x = \frac{x_i - \bar{x}}{s_x}}$$

$x_i = x_1, x_2, x_3.....$
s_x = Standardanweichung

Die Standardisierung führt dazu, dass **$\bar{x}$ = 0** und **s = ± 1** wird. Die Kurvengleichung vereinfacht sich dadurch entsprechend.



$$\boxed{y = f(x) = 0{,}399 * \frac{1}{\sqrt{e^{x^2}}}}$$

Zur Bestimmung bzw. Abschätzung der Fallzahl innerhalb bestimmter Intervalle wurde für die standardisierte Normalverteilung Tabelle I der nachfolgenden Seite berechnet. Der z-Wert entspricht dem Abstand vom Mittelwert, also von Null, σ_x steht für die Standardabweichung.

<u>Beispiele:</u> Wieviel % der Fälle liegen im Intervall ± 1 (also ± einer Standardabweichung)

z_1-Wert = +1,0 ⇒ Tabellenwert: 0,341 = 34,1 %
z_2-Wert = -1,0* ⇒ Tabellenwert: 0,341 = 34,1 % *Betrag von z_x

Ergebnis: 68,2% der Fälle

Wieviel der Fälle liegen im Intervall

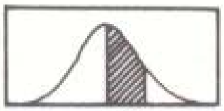
z= ±1,5, z = ± 2 +1,59 ≤ z ≤ 1,73

Wie oben angedeutet, läßt sich dieses Verfahren auf jede beliebige Normalverteilung anwenden, wenn die Intervallgrenzen x_i nach obiger Formel standardisiert werden.

<u>Beispiel:</u> Eine Gewichtsmessung an einer Schule ergab einen Mittelwert **$\bar{x}$ =** 60 kg und eine Standardabweichung $s_x = \pm 10$ kg. Wieviel % der SchülerInnen hat ein Gewicht zwischen 45 und 80 kg?

Intervallgrenzen (z-Werte)	⇒ Tabellenwerte	⇒ **Ergebnis:**
$z_{x1} =$ (45 – 60) kg / 10 kg = -15/10 = -1,5;	0,433	**91 %**
$z_{x2} =$ (80 – 60) kg / 10 kg = 28/10 = +2;	0,477	

Tabelle I: Flächenanteile der Normalverteilung



Die Zahlenwerte entsprechen dem schraffierten Flächenanteil. Die gesamte Fläche unter der Kurve hat den Wert 1.000.

$$z = \frac{x-\bar{x}}{\sigma_x}$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0,0	.000	.004	.008	.012	.016	.020	.024	.028	.032	.036
0,1	.040	.044	.048	.052	.056	.060	.064	.068	.071	.075
0,2	.079	.083	.087	.091	.095	.099	.103	.106	.110	.114
0,3	.118	.122	.126	.192	.133	.137	.141	.144	.148	.152
0,4	.155	.159	.163	.166	.170	.174	.177	.181	.184	.188
0,5	.192	.195	.199	.202	.205	.209	.212	.216	.219	.222
0,6	.226	.229	.232	.236	.239	.242	.245	.249	.252	.255
0,7	.258	.261	.264	.267	.270	.273	.276	.279	.282	.285
0,8	.288	.291	.294	.297	.300	.302	.305	.308	.311	.313
0,9	.316	.319	.321	.342	.326	.329	.322	.334	.337	.339
1,0	.341	.344	.346	.349	.351	.353	.355	.358	.360	.362
1,1	.364	.367	.369	.371	.373	.375	.377	.379	.381	.383
1,2	.385	.387	.389	.391	.393	.394	.396	.398	.400	.402
1,3	.403	.405	.407	.408	.410	.412	.413	.415	.416	.418
1,4	.419	.421	.422	.424	.425	.427	.428	.429	.431	.432
1,5	.433	.435	.436	.437	.438	.439	.441	.442	.443	.444
1,6	.445	.446	.447	.448	.450	.451	.452	.453	.454	.455
1,7	.455	.456	.457	.458	.459	.460	.461	.462	.463	.463
1,8	.464	.465	.466	.466	.467	.468	.469	.469	.470	.471
1,9	.471	.472	.473	.473	.474	.474	.475	.476	.476	.477
2,0	.477	.478	.478	.479	.479	.480	.480	.481	.481	.482
2,1	.482	.483	.483	.483	.484	.484	.485	.485	.485	.486
2,2	.486	.486	.487	.487	.488	.488	.488	.488	.489	.489
2,3	.489	.490	.490	.490	.490	.491	.491	.491	.491	.492
2,4	.492	.492	.492	.493	.493	.493	.493	.493	.493	.494
2,5	.494	.494	.494	.494	.495	.495	.495	.495	.495	.495
2,6	.495	.496	.496	.496	.496	.496	.496	.496	.496	.496
2,7	.497	.497	.497	.497	.497	.497	.497	.497	.497	.497
2,8	.498	.498	.498	.498	.498	.498	.498	.498	.498	.498
2,9	.498	.498	.498	.498	.498	.498	.498	.498	.499	.499

Der Punkt vor jeder Zahl bedeutet, daß dem Wert eine Null vorauszusetzen ist.

vgl. Sahner 1997, S.176

3.2 Repräsentationsschluß[14]

Mit dem Repräsentationsschluß wird aus einem Stichprobenergebnis (z.B. einem Mittelwert) auf das zu erwartende Ergebnis der Gesamtheit aller Fälle (also dem Mittelwert der Gesamtheit) geschlossen. Weil ein exakter Schluß nur dann möglich wäre, wenn Stichprobe und Grundgesamtheit übereinstimmen würden, geht es bei diesem Schließverfahren um die Angabe eines „Toleranzbereiches" (Konfidenzintervall oder auch Standardfehler), innerhalb dessen der Gesamtwert mit definierter Wahrscheinlichkeit liegen müßte.

Zur Unterscheidung der Berechnungsgrößen werden folgende Indizies eingeführt[15]:

Größe	Grundgesamtheit	Stichprobe
Arithmetisches Mittel	μ	$\bar{x}$
Standardabweichung	σ	S
Varianz	σ^2	s^2
Fallzahl	N	N
Proportionen (Prozentwerte)	P	P

<u>Beispiel:</u> Die oben bei 3.1 erwähnten Gewichtswerte $\bar{x}$ **=** 60 kg und die Standardabweichung $s_x = \pm 10$ kg wurden mit 35 zufällig ausgewählten SchülerInnen ermittelt.
<u>Frage:</u> In welchem Bereich wird der Mittelwert **„μ"** aller 1.000 SchulerInnen dieser Schule mit 99%-iger Wahrscheinlichkeit liegen?
Der Mittelwert $\bar{x}$ von 60 kg wird selten mit **μ** exakt übereinstimmen**.** Die Wahrscheinlichkeit, dass beide in der Nähe liegen**,** ist aber bei einer <u>zufälligen</u> Stichprobe relativ groß, denn die Chance, SchülerInnen mit einem mittleren Gewicht auszuwählen ist größer, als mit kleinerem bzw. größerem Gewicht.

Weil die Verteilung der Grundgesamtheit aber nicht bekannt ist, muß der Gesamtmittelwert mit Hilfe der vorhanden Stichprobenkennwerte geschätzt werden. Man geht dabei wie folgt vor:

Vorgehensweise bei der Abschätzung von „μ"

- Angenommen, man zieht sehr viele (unendlich viele) Samples, würden sich jeweils unterschiedliche Mittelwerte ergeben; diese Mittelwerte entsprechen selbst einer Normalverteilung mit einem Mittelwert $E_{\bar{x}}$ und einer Standardabweichung $\sigma_{\bar{x}}$. Bei unendlich vielen Stichproben würde dieser Mittelwert mit dem Mittelwert der Grundgesamtheit übereinstimmen

  $$\mathbf{E_{\bar{x}} = \mu,} \qquad \text{wobei} \qquad \sigma_{\bar{x}} \neq \sigma$$

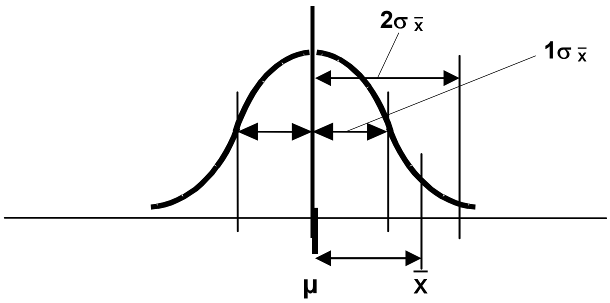
  d.h. für die Standardabweichung dieser Verteilung kann diese Gleichsetzung nicht vorgenommen werden, da es sich um unterschiedliche Verteilungen handelt

- In dem angesprochenen Massenexperiment würden demnach 95,5% der gezogenen Mittelwerte im Bereich **μ,** $\pm 2\sigma_{\bar{x}}$ der Verteilung liegen; (vgl. 3.1)
  oder: wenn nur, wie üblich, <u>eine</u> Stichprobe gezogen wird, besteht eine geringe Wahrscheinlichkeit von nur 5%, dass der Stichprobenmittelwert außerhalb des Intervalls **μ** $\pm 2\sigma_{\bar{x}}$ liegt, bzw. eine Wahrscheinlichkeit von 1%, dass er außerhalb **μ,** $\pm 3\sigma_{\bar{x}}$ liegt.

[14] vgl. Sahner 1997, S. 38-67
[15] vgl. ebd., S.12

Welches Wahrscheinlichkeitsniveau (in anderen Zusammenhängen spricht man auch von Signifikanzniveau oder auch Vertrauensintervall) gewählt wird, 1% oder 5%, ist reine Konvention; verwendet werden üblicherweise beide.

Abb. 3.2: Stichprobenmittelwert $\bar{x}$ in der Verteilung unendlich vieler Mittelwerte



- In der Abb. 3.2 wird beispielsweise (mit etwa 95%-iger Wahrscheinlichkeit) angenommen, dass der Stichprobenmittelwert $\bar{x}$ innerhalb der Intervallgrenzen $\pm 2\sigma_{\bar{x}}$ liegt, und sich dadurch eine Differenz zwischen dem Gesamtmittelwert **μ** und dem Stichprobenmittelwert $\bar{x}$ von $\mathbf{\bar{x} - \mu \leq 2\sigma_{\bar{x}}}$ ergibt.

- In dieser Abbildung ist der Stichprobenmittelwert $\bar{x}$ etwas größer als der Gesamtmittelwert **μ.** Nicht auszuschließen ist aber, dass er auch kleiner wäre, also links der Mitte liegen würde. Die Schätzung ist also prinzipiell von einer Zweiseitigkeit begleitet. Dieser Sachverhalt läßt sich über das Vorzeichen **„±"** berücksichtigen und man gelangt schließlich zu einer ersten Formel für die Abschätzung des Gesamtmittelwertes:

$$\boxed{\mu = \bar{x} \quad \pm z\,\sigma_{\bar{x}}}$$

z 2, 3, .. je nach Wahrscheinlichkeitsintervall

- Mit der angenommen „Hilfsverteilung" wird es also möglich die Differenz zwischen **μ** und $\bar{x}$ abzuschätzen. Die bisherigen Überlegungen beruhten aber nur auf einem hypothetischen Experiment, das in Wirklichkeit nicht vorliegt; die Standardabweichung dieser theoretischen Verteilung $\sigma_{\bar{x}}$ ist natürlich nicht bekannt.
  Es läßt sich jedoch mathematisch nachweisen, dass die Standardabweichung der Verteilung aller Samplemittelwerte, $\sigma_{\bar{x}}$ wie folgt berechnet werden kann, wenn die Stichprobengröße „n" nicht größer als 10% der Grundgesamtheit „N" ist[16]

$$\boxed{\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}}$$

σ_x Standardabweichung der Grundgesamtheit
n Umfang der Stichprobe / des Samples

Falls die Stichprobengröße und Grundgesamtheit bekannt sind, kann mit folgender Formel gearbeitet werden[15]:

$$\sigma_x = \frac{\sigma_x}{\sqrt{n}} * \left[\left(\frac{\mathbf{N-n}}{\mathbf{N-1}}\right)\right]$$

[16] Sonst müsste noch ein weiterer Verzerrungsfaktor berücksichtigt werden; näheres dazu Sahner 1997, S. 47
[15] vgl. Diehl/Arbinger 2001, S.32.ff.

- Weil σ_x ebenfalls unbekannt ist, kann $\sigma_{\bar{x}}$ nicht bestimmt werden. Nun läßt sich aber zeigen, dass für

  Samples von $n \geq 30$ $\boxed{\sigma_x \approx s_x}$ s_x = Standardabweichung des Samples

  ist, d.h. die Standardabweichung des Samples ist ein guter Schätzwert für die Standardabweichung der Grundgesamtheit.

  Die Formel für die geschätzte Standardabweichung von $\sigma_{\bar{x}}$ lautet demnach

$$\boxed{\sigma_{\bar{x}} = \frac{s_x}{\sqrt{n}}}$$

- Damit wären alle „Hürden" aus dem Weg geräumt, der Gesamtmittelwert kann mit Hilfe der Stichprobenkennwerte angegeben werden:

$$\boxed{\mu = \bar{x} \pm z * \frac{s_x}{\sqrt{n}}}$$

μ	Gesamtmittelwert
$\bar{x}$	Mittelwert der Stichprobe
s_x	Standardabweichung der Stichprobe
n	Umfang der Stichprobe / des Samples
z	2, 3, .. je nach Wahrscheinlichkeitsintervall

Auf das Eingangsbeispiel übertragen:

μ = ?;
$\bar{x}$ = 60 kg; $s_x \pm$ 10 kg; n = 35; z = 3 (Wahrscheinlichkeitsbereich ca. 99%)
N = 1.000, d.h. n/10< 10%

In die Formel eingesetzt:

$$\mu = \bar{x} \pm z * \frac{s_x}{\sqrt{n}} = 60\text{kg} \pm 3 * \frac{10\text{kg}}{\sqrt{35}} = 60\text{kg} \pm 3 * 1{,}69\text{kg}$$

$$\boldsymbol{\mu = 60\textbf{kg} \pm 5{,}07\textbf{kg}}$$

Ergebnis: Der Gesamtmittelwert liegt also mit 99%-iger Wahrscheinlichkeit zwischen 55 und 65 kg.

Aus dem Beispiel wird aber auch sichtbar, das die Größe der Stichprobe das Konfidenzintervall (oder auch Standardfehler) beeinflußt; wäre beispielsweise die Stichprobegröße n = 90, würde sich der Standardfehler auf etwa ± 3 kg verringern.

Ergänzung: Repräsentationsschluß mit Prozentwerten

Wenn nur nominalskalierte Daten zur Verfügung stehen, können keine arithmetischen Mittelwerte gebildet werden. Die statistische Analyse beinhaltet dann Angaben über %-Anteile der einzelnen Variablen, die sich ebenfalls „verallgemeinern" lassen, d.h. an Hand eines Stichprobenergebnisses kann auf die Grundgesamtheit geschlossen werden.
Das Paradebeispiel hierfür ist die Wahlforschung.

Beispiel Bundestagswahl: Eine repräsentative Stichprobe (1.100 Personen) im Vorfeld der Wahl 2002 führte zu folgendem Ergebnis CDU 39%, SPD 39%, Grüne 8%, FDP 7%, Sonstige 7%. Wie groß ist der Prognosefehler für das Wahlergebnis der CDU bei 95%-iger Wahrscheinlichkeit?

Die allg. Formel lautet:

$$\boxed{P = p \pm z * \sqrt{\frac{p(1-p)}{n}}}$$

P = Prozentwert der Grundgesamtheit
p = Prozentwert der Stichprobe
z = 2, 3 .. je nach Wahl des Vertrauensintervalls
n = Stichprobengröße

Im Beispiel:

P = ?
p = 39% (0,39); z = 2 (Wahrscheinlichkeit 95,5%); n = 1.100
n/N < 0,1 (10%)

Die Daten eingesetzt:

$$P = 0{,}39 \pm 2 * \sqrt{\frac{0{,}39(1-0{,}39)}{1.100}} = 0{,}39 \pm 2 * \sqrt{\frac{0{,}237}{1.100}} = 0{,}39 \pm 2 * 0{,}0147 = 0{,}39 \pm 0{,}029$$

<u>Ergebnis</u>: Das Wahlergebnis wird mit 95%-iger Sicherheit zwischen 36% und 42% liegen.

Aus der Formel wird sichtbar, dass der Stichprobenumfang und die Größe des Prozentanteils den Prognosefehler bestimmen. Das Ergebnis der kleinen Parteien ist schwieriger zu prognostizieren. Trotzdem ist es erstaunlich, dass man bereits mit etwa 1.000 Personen das Wahlergebnis von über 50 Mio. WählerInnen auf unter 5% genau vorhersagen kann.

Abschließende Hinweise zum Einfluß der Stichprobengröße

Nach Sahner[17] ist bei Schließverfahren hinsichtlich der Samplegrößen
(1) n ≥ 30 und **(2) n < 30** zu unterscheiden.

(1) Stichprobenumfang n ≥ 30

- Die Stichprobengröße von n ≥ 30 (manche Autoren fordern auch 100, doch Sahner verweißt darauf, dass die entstandenen Verzerrungen im Vergleich zu anderen Störgrößen der Messung vernachlässigbar sind) führt dazu, dass (unendlich viel) ermittelte Stichprobenmittelwerte „$\bar{x}$" normalverteilt wären. Bei diesem Sampleumfang ist es deshalb unwesentlich, ob das Merkmal in der Grundgesamtheit normalverteilt ist oder nicht.
- Die Standardabweichung der Grundgesamtheit σ_x kann ab dieser Stichprobengröße durch die Standardabweichung des Samples s_x abgeschätzt und ersetzt werden
$\sigma_x \approx s_x$

(2) Stichprobenumfang n < 30

- Falls das Merkmal in der Grundgesamtheit nicht normalverteilt ist, führen kleine Stichprobengrößen dazu, dass die Verteilung von (unendlich viel) ermittelten Samplemittelwerte „$\bar{x}$" nicht mehr normalverteilt wären. Dies führt zur 1. Bedingung:
Bei Stichproben von n < 30 muß das Merkmal in der Grundgesamtheit normalverteilt sein.
- Bei kleinen Stichproben schwankt die Standardabweichung s_x der Stichprobe sehr stark. **Für die Schätzung der Standardabweichung der Grundgesamtheit muß eine korrigierte Formel angesetzt werden**. Diese ergibt sich:

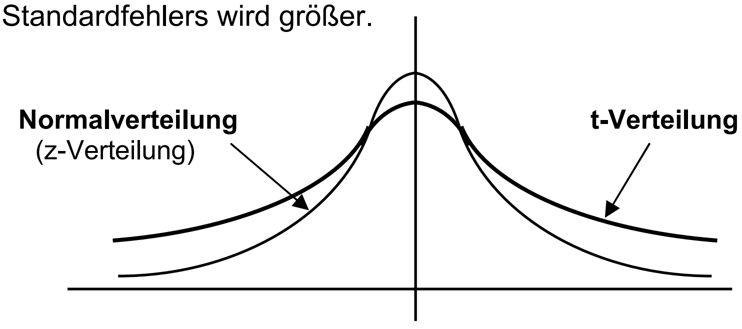
$$\sigma_x = \pm\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

Diese Schätzung reicht aber nicht aus.

- Weitere Konsequenzen werden sichtbar, wenn die angenommene (Normal-) Verteilung der Samplemittelwerte (Mittelwert aller $\bar{x} = E_{\bar{x}} = \mu$) standardisiert wird.

$$\boxed{z_x = \frac{(\bar{x}-E_{\bar{x}})}{\sigma_{\bar{x}}} = \frac{(\bar{x}-E_{\bar{x}})}{s_x} * \sqrt{n-1} = t}$$

- Weil s_x sehr stark schwankt (veränderlich ist), liegt nicht mehr eine lineare Transformation vor (keine Teilung durch eine Konstante sondern durch eine variable Größe). Die transformierten Werte geben keine Normalverteilung, sondern eine t-Verteilung (siehe schematische Abb.). Sie ist „flacher" wodurch der Sicherheitsbereich gestreckt wird, d.h. der Bereich des Standardfehlers wird größer.



[17] Sahner 1997, S.57ff.

3.3 Mittelwertsvergleich zweier Stichproben[18] (t-Test)

Beim Vergleich von Stichproben innerhalb der Schließenden Statistik interessiert, ob die Unterschiede zufällig oder nicht zufällig entstanden sind, und zwar unabhängig davon, wie groß die Unterschiede sind. Man bezeichnet solche Tests auch als Signifikanztests.

<u>Beispiel</u>: Zufällig ausgewählte SchülerInnen aus der 11. Klasse zweier Großstädte bestreiten einen Test, bei dem es um möglichst viele Punkte (x) geht. Folgendes Stichprobenergebnis ergibt sich danach:

SchülerInnen	aus Hambergen:	N_1= 500;	n_1= 45;	$\bar{x}_1 = 13$;	$s^2_{x1} = 3$
	aus Kölnfurt:	N_2= 750;	n_2= 52;	$\bar{x}_2 = 17$;	$s^2_{x2} = 8$

Kam dieser Unterschied zufällig zustande bzw. sind die SchülerInnen aus Kölnfurt signifikant besser als die Schüler aus der 1. Großstadt?

Vor einer solchen Analyse sind folgende grundsätzlichen Fragen zu klären:

1. Abbhängige oder unabhängige Stichproben?

   - abhängig: Beide Stichproben stehen in Beziehung zu einander, d.h. man könnte mit einer Korrelationsanalyse einen Zusammenhang feststellen;
     Beispiel: Der Test wird mit einer Klasse zum Zeitpunkt $t_1$durchgeführt, und mit der gleichen Personengruppe nach einer gewissen Zeit t_2 wiederholt; das 2.Testergebnis wird vom 1. Ergebnis abhängig sein.

   - unabhängig: Die Stichproben sind Teil von unabhängigen Teilpopulationen, die gegenseitig nicht in Beziehung stehen; eine zufällige Auswahl ist möglich
     <u>Beispiel oben:</u> Da die SchülerInnen nur einer Klasse angehören und räumlich von einander getrennt sind kann ein gegenseitiger Einfluß ausgeschlossen werden; d.h. es liegen unabhängige Stichproben vor

     <u>Hinweis:</u> Die Frage ob Abhängigkeit vorliegt oder nicht hat Einfluss auf das nachfolgende Berechnungsverfahren. Hier wird nur das Berechnungsverfahren für unabhängige Stichproben vorgestellt (s.u.); abhängige Stichproben siehe Literatur;

2. Stichprobengröße?

   - Bei n_1 und $n_2 \geq 30$ kann mit der z-Verteilung (Normalverteilung) gerechnet werden; weil die t-Verteilung mit steigender Fallzahl in eine Normalverteilung über geht, wird i. d. R. auch bei großen Stichproben mit der t-Verteilung (t-Test) gearbeitet.[19]

3. Anzahl der Stichproben?
   - Beim Vergleich von 2 Stichproben kommt der t-Test oder z-Test zum Einsatz
   - Werden mehr als 2 Stichproben verglichen, wird eine Varianzanalyse durchgeführt

4. Sind die Varianzen der Teilpopulationen identisch?

   - Diese zusätzliche Prüfung ist bei Stichprobengrößen von n < 30 notwendig; Sie wird mit einem sogenannten F-Test durchgeführt.[20]
     Im obigen Beispiel sind die Stichproben n_1 und n_2 deutlich größer als 30, so dass auf diese Voruntersuchung verzichtet werden kann.

[18] vgl. Sahner 1997, S.104-117; vgl. auch Schnell/Hill/Esser 1999, S. 415-416
[19] vgl. z.B. Schnell/Hill/Esser 1999, S. 416
[20] vgl. Sahner 1997, S. 112f. und S.141ff.

Sind diese Fragen geklärt, kann die Signifikanzprüfung erfolgen. Das Vorgehen ähnelt dem oben beschriebenen Repräsentationsschluß.

Verfahrensschritte bei der Signifikanzprüfung von Mittelwertsunterschieden

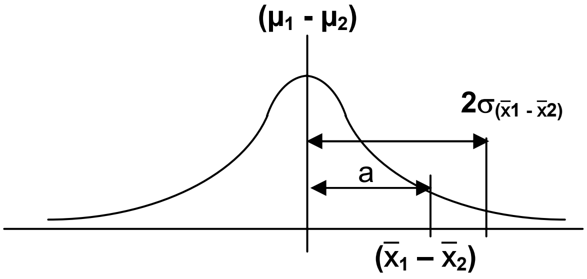
- Angenommen, man zieht in jeder Stadt sehr viele (unendlich viele) Samples, würden sich jeweils unterschiedliche Mittelwertdifferenzen ($\bar{x}_1 - \bar{x}_2$) ergeben; die so gewonnenen Werte würden selbst eine Normalverteilung mit einer mittleren Differenz $E_{(\bar{x}1 - \bar{x}2)}$ bilden. Bei unendlich vielen Stichproben wird diese Differenz mit der tatsächlich existierenden Differenz zwischen den beiden Grundgesamtheiten (oder auch großen Teilpopulationen in einer Grundgesamtheit) übereinstimmen

$$E_{(\bar{x}1 - \bar{x}2)} = (\mu_1 - \mu_2)$$

$\sigma_{(\bar{x}1 - \bar{x}2)}$ Standardabweichung Mittelwertdifferenz

- In Abb. 3.3 liegt die Stichprobendifferenz ($\bar{x}_1 - \bar{x}_2$) innerhalb des Intervalls $\pm 2\sigma_{(\bar{x}1 - \bar{x}2)}$, man könnte auch sagen, sie liegt in einem Streubereich der Verteilung, in dem etwa 95% aller Werte liegen. Dieser Bereich wird als zufälliger Bereich definiert, d.h. der Abstand „a" zwischen den Mittelwertsdifferenzen „Grundgesamtheit-Stichprobe" ist so gering, dass er als zufällig entstanden bezeichnet werden kann. Läge er außerhalb des angegeben Intervalls, würde analog dazu die Bewertung „nicht zufällig" oder „signifikant" zutreffen.

Abb. 3.3: Differenz der Stichprobenmittelwerte $\bar{x}$ in der Verteilung unendlich vieler Stichprobendifferenzen



$$\mathbf{a = (\mu_1 - \mu_2) - (\bar{x}_1 - \bar{x}_2)}$$

<u>**Fragen:**</u> **liegt ($\bar{x}_1 - \bar{x}_2$) noch innerhalb von $\pm 2\sigma_{(\bar{x}1 - \bar{x}2)}$,**
d.h. die Stichprobendifferenz liegt im zufälligen Streubereich der Verteilung
oder liegt ($\bar{x}_1 - \bar{x}_2$) außerhalb von $\pm 2\sigma_{(\bar{x}1 - \bar{x}2)}$
d.h. die Stichprobendifferenz weicht von der als zufällig definierten Streuung der Verteilung ab, bzw. die Stichprobendifferenz liegt außerhalb der zufälligen Mittelwertsdifferenzen innerhalb der Gesamtpopulationen

- Auch hier wird wiederum deutlich, dass die Bewertung „Mittelwertsdifferenz signifikant: ja/nein?" vom zuvor festgelegten Intervall oder Niveau abhängig ist. Die Bewertung „signifikant" wird umso unwahrscheinlicher oder schwieriger, je größer der „zufällige Bereich" definiert wird; Würde z.B. als Zufallsbereich $\pm 3\sigma_{(\bar{x}1 - \bar{x}2)}$ zugrunde gelegt, müßte die Stichprobendifferenz einen derart großen Wert annehmen, der nur auf 1% der Verteilung zutrifft, bzw. in 99% der Fälle ausgeschlossen werden kann.
- Damit einher geht auch die Gefahr einen Fehler zu begehen:

„α - Fehler": Durch ein zu streng angelegtes Niveau (z.B. > $\pm 3\sigma_{(\bar{x}1 - \bar{x}2)}$) werden Mittelwertsdifferenzen als statistisch „nicht signifikant" ausgeschlossen, obgleich sie sich in der Grundgesamtheit bedeutsam erweisen können.

„β-- Fehler“: Wenn der Zufallsbereich zu klein gewählt wird (z.B. < $\pm 2\sigma_{(\bar{x}1 - \bar{x}2)}$) und dadurch die Feststellung von signifikanten Mittelwertsdifferenzen begünstigt wird, obwohl sie in der Grundgesamtheit nicht auftreten.

- In den bisherigen Beschreibungen wurden die grundsätzlichen Überlegungen des Signifikanztestes dargelegt, in deren Zentrum die Verteilung eines hypothetischen Massenexperiments mit unendlich vielen Stichpropbendifferenzen stand. Die Parameter dieser Verteilung, insbesondere die Standardabweichung $\sigma_{(\bar{x}1 - \bar{x}2)}$ sind aber unbekannt und können für die konkrete Berechnung nicht benutzt werden.
  Im folgenden wird nun gezeigt, wie man dieses Manko umgeht

(1) Die in Abb. 3.3 dargestellte Verteilung wird standardisiert, um mit der z-Verteilung arbeiten zu können (siehe auch 3.1). Für einen beliebige Stichprobendifferenz **$(\bar{x}_1 - \bar{x}_2)$** ergibt sich dann

$$\boxed{z = \left| \frac{(\mu_1 - \mu_2) - (\bar{x}_1 - \bar{x}_2)}{\sigma_{(\bar{x}_1 - \bar{x}_2)}} \right|}$$

Problem: (μ_1 - μ_2) ist unbekannt!

(2) Um das Problem zu umgehen, formuliert man nun die Nullhypohtese

H_o: Zwischen μ_1 und μ_2 besteht kein Unterschied; $\mu_1 - \mu_2 = 0$;

d. h., wenn die Hypothese richtig ist, dann darf die Stichprobendifferenz nur zufällig sein, bzw. innerhalb von $\pm 2\sigma_{(\bar{x}1 - \bar{x}2)}$ liegen (alternativ: $\pm 3\sigma_{(\bar{x}1 - \bar{x}2)}$).
Ist dies nicht der Fall, bzw. ist „z“ größer als 2, muß die Nullhypothese verworfen werden, und zwar zu Gunsten der Arbeitshypothese H_1:

H_1: Die Differenz zwischen $\bar{x}_1 - \bar{x}_2$ geht auf Unterschiede zwischen μ_1 und μ_2 zurück

Die veränderte Formel hat nun folgendes Bild:

$$\boxed{z = \left| \frac{(\bar{x}_1 - \bar{x}_2)}{\sigma_{(\bar{x}_1 - \bar{x}_2)}} \right|}$$

Der errechnete Betrag „z“, der darüber entscheidet, ob eine Hypothese angenommen oder verworfen wird, wird als **kritischer Quotient** *(critical ratio)* bezeichnet

(3) Der noch nicht bekannte Standardfehler $\sigma_{(\bar{x}1 - \bar{x}2)}$ wird über die bereits oben beschriebene Grundformel berechnet:

$$\boxed{\sigma_{\bar{x}} = \frac{s_x}{\sqrt{n}}}$$

Da es sich um eine Stichprobendifferenz mit unterschiedlichem „s“ und „n“ handelt, erweitert sich die Formel zu:

$$\boxed{\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{n_1 s_{x_1}^2 + n_2 s_{x_2}^2}{n_1 + n_2 - 2}} * \sqrt{\frac{n_1 + n_2}{n_1 n_2}}}$$

(4) Daraus ergibt sich für den **kritischen Quotient** des **t-Tests** folgende Berechnungsformel:

$$z = \left| \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{n_1 s_{x_1}^2 + n_2 s_{x_2}^2}{n_1 + n_2 - 2}} * \sqrt{\frac{n_1 + n_2}{n_1 n_2}}} \right| = t$$

Auf das Eingangsbeispiel übertragen:

SchülerInnen	aus Hambergen:	N_1= 500;	n_1= 45;	$\bar{x}_1$ = 13;	s_{x1} = 3
	aus Kölnfurt:	N_2= 750;	n_2= 52;	$\bar{x}_2$ = 17;	s_{x2} = 8

$$t = \left| \frac{(13-17)}{\sqrt{\frac{45*3^2+52*8^2}{45+52-2}} * \sqrt{\frac{45+52}{45*52}}} \right| = \left| \frac{4}{\sqrt{1,6}} \right| = \frac{4}{1,27} = 3,13$$

Auf der nachfolgenden Seite sind in Tabelle II Werte für den kritischen Quotienten der „t-Verteilung“ spaltenweise aufgetragen, und zwar abhängig vom Signifikanzniveau. Die Tabelle kann wie folgt interpretiert werden:

- Die Fragestellung von Mittelwertsdifferenzen ist als zweiseitige Fragestellung aufzufassen, denn potentiell kann sowohl H_o oder H_1 zutreffen. Insofern ist die erste Zeile maßgebend und nicht die unterste (Bem: Die Werte der untersten Zeile sind halb so groß)
- Für Signifikanzprüfungen spielen vor allem die letzten drei Spalten eine Rolle, also zwischen 0,05 und 0,01 (Signifikanzniveau 5% bis 1%).
- Der bisher noch nicht angesprochene Freiheitsgrad df (erste Spalte) ist von der Fallzahl n abhängig und berechnet sich mit:

  $\boxed{df = n_1 + n_2 - 2}$ im Beispiel: df = 45 + 52 –2 = 95

- Die Tabellenwerte zeigen, daß ab einem Freiheitsgrad von df ≈ 30 ein kritischer Quotient von 2,0 zu signifikanten Mittelwertsdifferenzen führt (Signifikanzniveau = 0,05 oder 5 %).

  Der im Beispiel oben ermittelte t-Wert von 3,13 übersteigt alle in der Tabelle abgebildeten Werte, d.h. die Mittelwertsdifferenz kann in jedem Fall als hochsignifikant bezeichnet werden (Signifikanzniveau p <1%). Die im Beispiel dargestellten Leistungsunterschiede zwischen den Größstädten sind demnach nicht zufällig entstanden.

Hinweis:

Die Signifikanzprüfung kann in gleicher Weise auch mit Korrelationsmaßen (siehe 2.2) durchgeführt werden. Hier wird dann beispielsweise überprüft, ob die innerhalb einer Stichprobe herausgefundene Beziehung zufällig zustande gekommen ist, oder ob (je nach Signifikanzniveau) anzunehmen ist, dass die Beziehung auch in der Grundgesamtheit vorliegt, d.h. das Stichprobenergebnis signifikant ist.

Tabelle II: t-Verteilung

Signifikanzgrad (Wahrscheinlichkeit) für zweiseitige Fragestellung

df	.90	.80	.70	.60	.50	.40	.30	.20	.10	.05	.02	.01
1	0,16	0,33	0,51	0,73	1,00	1,38	1,96	3,08	6,31	1)	2)	3)
2	0,14	0,29	0,45	0,62	0,82	1,06	1,39	1,89	2,92	4,30	6,97	9,93
3	0,14	0,28	0,42	0,58	0,77	0,98	1,25	1,64	2,35	3,18	4,54	5,84
4	0,13	0,27	0,41	0,57	0,74	0,94	1,19	1,53	2,13	2,78	3,75	4,60
5	0,13	0,27	0,41	0,56	0,73	0,92	1,16	1,48	2,02	2,57	3,37	4,03
6	0,13	0,27	0,40	0,55	0,72	0,91	1,13	1,44	1,94	2,45	3,14	3,71
7	0,13	0,26	0,40	0,55	0,71	0,90	1,12	1,42	1,90	2,37	3,00	3,50
8	0,13	0,26	0,40	0,55	0,71	0,89	1,11	1,40	1,86	2,31	2,90	3,36
9	0,13	0,26	0,40	0,54	0,70	0,88	1,10	1,38	1,83	2,26	2,82	3,25
10	0,13	0,26	0,40	0,54	0,70	0,88	1,09	1,37	1,81	2,23	2,76	3,17
11	0,13	0,26	0,40	0,54	0,70	0,88	1,09	1,36	1,80	2,20	2,72	3,11
12	0,13	0,26	0,40	0,54	0,70	0,87	1,08	1,36	1,78	2,18	2,68	3,06
13	0,13	0,26	0,39	0,54	0,69	0,87	1,08	1,35	1,77	2,16	2,65	3,01
14	0,13	0,26	0,39	0,54	0,69	0,87	1,08	1,35	1,76	2,15	2,62	2,98
15	0,13	0,26	0,39	0,54	0,69	0,87	1,07	1,34	1,75	2,13	2,60	2,95
16	0,13	0,26	0,39	0,54	0,69	0,87	1,07	1,34	1,75	2,12	2,58	2,92
17	0,13	0,26	0,39	0,53	0,69	0,86	1,07	1,33	1,74	2,11	2,57	2,90
18	0,13	0,26	0,39	0,53	0,69	0,86	1,07	1,33	1,73	2,10	2,55	2,88
19	0,13	0,26	0,39	0,53	0,69	0,86	1,07	1,33	1,73	2,09	2,54	2,86
20	0,13	0,26	0,39	0,53	0,69	0,86	1,06	1,33	1,73	2,09	2,53	2,85
21	0,13	0,26	0,39	0,53	0,69	0,86	1,06	1,32	1,72	2,08	2,52	2,83
22	0,13	0,26	0,39	0,53	0,69	0,86	1,06	1,32	1,72	2,07	2,51	2,82
23	0,13	0,26	0,39	0,53	0,69	0,86	1,06	1,32	1,71	2,07	2,50	2,81
24	0,13	0,26	0,39	0,53	0,69	0,86	1,06	1,32	1,71	2,06	2,49	2,80
25	0,13	0,26	0,39	0,53	0,68	0,86	1,06	1,32	1,71	2,06	2,49	2,79
26	0,13	0,26	0,39	0,53	0,68	0,86	1,06	1,32	1,71	2,06	2,48	2,78
27	0,13	0,26	0,39	0,53	0,68	0,86	1,06	1,31	1,70	2,05	2,47	2,77
28	0,13	0,26	0,39	0,53	0,68	0,86	1,06	1,31	1,70	2,05	2,47	2,76
29	0,13	0,26	0,39	0,53	0,68	0,85	1,06	1,31	1,70	2,05	2,46	2,76
30	0,13	0,26	0,39	0,53	0,68	0,85	1,06	1,31	1,70	2,04	2,46	2,75
∞	0,13	0,25	0,39	0,52	0,67	0,84	1,04	1,28	1,65	1,96	2,33	2,58
	.45	.40	.35	.30	.25	.20	.15	.10	.05	.025	.001	.005

Signifikanzgrad (Wahrscheinlichkeit) für einseitige Fragestellung

1) 12,71 2) 31,82 3) 63,66

Anmerkung zu Tabelle II s. S. 184.

vgl. Sahner 1997, S.177

- Schließende Statistik im Überblick -

Operation	**Häufigkeitsanalyse**	**Repräsentationsschluß**	**Mittelwertsvergleich**
	mit der standardisierten Normalverteilung (z-Verteilung) (siehe 3.1)	Stichprobe $\Rightarrow$ Gesamtheit (siehe 3.2)	t-Test, (Varianzanalyse) (siehe 3.3=
Typische Fragestellung	Wieviel der Fälle treten in einem bestimmten Werteinterval auf?	In welchem Bereich liegt bei einer Wahrscheinlichkeit von 95% (z=2) oder 99% (z=3) der Parameter der Grundgesamtheit (z.B. Mittelwert „µ“ oder Prozentwert „P“)?	Sind die Unterschiede zwischen den beiden Stichproben signifikant oder sind sie zufällig entstanden?
Berechnung	$z_x = \frac{x_i - \bar{x}}{s_x}$	$\mu = x \pm z * \frac{s_x}{\sqrt{n}}$	$t = \left\| \frac{(x_1 - x_2)}{\sqrt{\frac{n_1 s_{x_1}^2 + n_2 s_{x_2}^2}{n_1 + n_2 - 2}} * \sqrt{\frac{n_1 + n_2}{n_1 n_2}}} \right\|$
Beispiele	Daten zur Körpergröße einer 1. Klasse: $\bar{x}$ = 1,20 m s = ± 0,1 m. Wieviel % der SchülerInnen haben eine Körpergröße zwischen 1,20 bis 1,35 m ? Voraussetzung: Körpergröße normalverteilt	Daten eines zentralen Statistiktest der Universität Münchgart: Stichprobe: n = 32, Notenschnitt: $\bar{x}$ = 2,8 Standardabw. s = ± 0,7 (n / N < 0,01) Welcher Notendurchschnitt ist mit 95% Wahrscheinlichkeit bei allen Studierenden (N=420) zu erwarten? Die Partei SDU kann in der Stadt Hundsheim (10.000 WählerInnen) 25% der Wählerstimmen verbuchen. Wie groß wird ihr Anteil in der gesamten Bevölkerung mit 99%-iger Sicherheit sein? (n / N = 0,01)	Mit zwei unabhängigen Gruppen von Auszubildenden im Dualen System wurde eine Untersuchung zur Motivation durchgeführt Skala der Motivationsermittlung: M_{max} = 4 …..M_{min} = 0. Folgendes Ergebnis stellte sich ein: Berufsschule 1: M_{BS1}: 2,1, s_{BS1}= 0,4; n_{BS1} = 75; Berufsschule 2: M_{BS2}: 2,9, s_{BS2}= 0,5; n_{BS2} = 89; (n / N < 0,01) Ist die größere Motivation zufällig entstanden oder ist sie signifikant?

vgl. Sahner, Heinz: Schließende Statistik. 4. Aufl. Stuttgart: Teubner 1997

Abschließende Übungen

Aufgabe 1

Eine Gruppe von Studierenden (N=42) wurde über ihre Mensa-Erfahrungen befragt und dazu folgendes Statement vorgelegt: *„Die Wartezeit an der Essensausgabe ist zu groß“* (1= stimme voll zu ..... 4 = überhaupt nicht zu). Die Stichprobe erbrachte einen Mittelwert von 1,8 und eine Standardabweichung von 0,91.
Wie wird mit einer Wahrscheinlichkeit von 95% die Einschätzung aller 7.000 Studierenden ausfallen?

Aufgabe 2

In einer Untersuchung wurde die intrinsische Motivation von gewerblichen Auszubildenden (N=45) untersucht. Die Befragung (Skala: „1 = nicht intrinsisch motiviert .... 6 = voll intrinsisch motiviert“) ergab folgendes Ergebnis:

- Betrieb: $\bar{x}_B = 4{,}2$; $s_B = 0{,}9$
- Schule: $\bar{x}_S = 3{,}9$; $s_s = 1{,}1$

a) Mit welchem Standardfehler kann das Ergebnis auf alle 2.000 Azubis der Region verallgemeinert werden.
b) Unter welchen Bedingungen kann ein t-Test vorgenommen werden?

Aufgabe 3

Studierende im Studiengang Technikpädagogik (N=73) der Universität Stuttgart wurden gefragt, inwiefern sie einen Laissez-faire-Stil als LehrerIn bevorzugen würden, wenn sie merken, dass in einer Klassenarbeit abgeschrieben wird. (Skala: „1 = volle Ablehnung des laissez-faire-Stils .... 6 = volle Zustimmung für laissez-faire-Stils). Die gleiche Frage wurde auch SchülerInnen einer beruflichen Schule (N= 153) vorgelegt. Folgendes Ergebnis wurde dabei ermittelt:

- Studierende TP (Ich als Lehrender würde laissez-faire-Stil..): $\bar{x}_{TB} = 1{,}3$; $s_{TB} = 0{,}5$
- SchülerInnen (LehrerInnen sollten laissez-faire-Stil...): $\bar{x}_S = 3{,}1$; $s_s = 1{,}1$

Ist der Unterschied der Mittelwerte signifikant?

Aufgabe 4

An den Gymnasium des Landkreises Handelsstadt entschieden sich 41% der SchulabgängerInnen eines Jahrgangs (N=434) für die Aufnahme eines wirtschaftswissenschaftlichen Studiums. Wie genau kann damit die Studienwahl des Jahrgangs insgesamt prognostiziert werden?