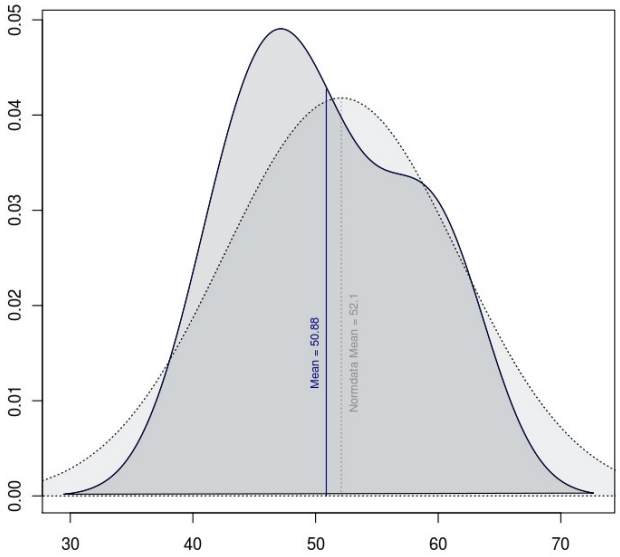
Department Pflegewissenschaft

Skriptum Statistik



Ein Skript zur beschreibenden
und schliessenden Statistik

von:
Cleo Nonn

weitergeführt von:
Jörg grosse Schlarmann

14. März 2014

Cleo Nonn (✝ 2008)

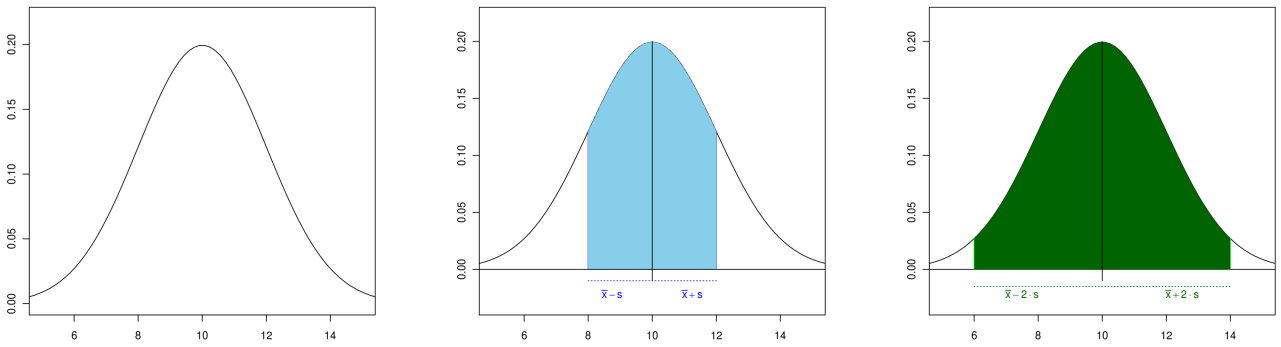
Cleo Nonn war Bachelor- und Masterabsolventin der pflegewissenschaftlichen Studiengänge der Universität Witten/Herdecke und bis zuletzt wissenschaftliche Mitarbeiterin im Institut für Pflegewissenschaft. Sie unterrichtete die Studierenden in Statistik, bot neben den laufenden Veranstaltungen SPSS-Blockseminare an, koordinierte das Studienangebot mit Frau Dr. Zegelin und war in entscheidendem Maße mit der Beratung der Studierenden bei der Erstellung ihrer Bachelor- und Masterarbeiten befasst. Sie erstellte die erste Version des vorliegenden Skriptes als Nachschlagwerk zu ihren Seminaren sowie zur Vorbereitung auf die Pflichtklausur "Statistik" im Bachelorstudiengang.
Cleo Nonn starb 2008 im Alter von 49 Jahren nach kurzer schwerer Krankheit.

Skriptum Statistik.

Ein Skript zur beschreibenden und schliessenden Statistik.

Von Cleo Nonn, weitergeführt von Jörg große Schlarmann.

Dies ist Version 2.5 vom 14. März 2014.



Inhaltsverzeichnis

1. Beschreibende Statistik

1-1. Daten und Merkmale

Daten sind alle Arten von Informationen, die man bei Messungen, Befragungen, Beobachtungen, etc. gewinnt. Man möchte Informationen, also Daten über bestimmte Merkmale erhalten. Ein Merkmal könnte beispielsweise das Geschlecht einer Person sein.
Merkmale, man spricht auch von Untersuchungsvariablen, finden sich bei den Merkmalsträgern (z.B. Patienten) oder den Untersuchungseinheiten (z.B. Stationen). Die verschiedenen Möglichkeiten des Auftretens eines Merkmals nennt man Merkmalsausprägungen. So existieren für das Merkmal ''`Geschlecht`'' die beiden Ausprägungen: ''`männlich` und `weiblich`''.

Merkmal bzw. Untersuchungsvariable	**Merkmalsträger bzw. Untersuchungseinheit**	**Merkmalsausprägungen**
Familienstand	Personen	ledig, verheiratet, geschieden, verwitwet
Kinder in Familien	Familien	vorhanden, nicht vorhanden
Anzahl von Kindern in Familien	Familien	0, 1, 2, 3, 4, ...
Körpergröße von Jugendlichen	Jugendliche	160cm, 161cm, 162cm,
durchschnittliche Bettenzahl der Stationen von Krankenhaus A,B,C,...	Krankenhäuser A, B, C	15, 20, 25, 27, usw.

Tabelle 1.1.: Merkmale und Ausprägungen

Man unterscheidet qualitative und quantitative Daten bzw. Merkmale.

Qualitative Daten unterscheiden sich in ihrer Art: beispielsweise das Merkmal `Geschlecht` mit den Ausprägungen ''`männlich` und `weiblich`'' oder das Merkmal `pflegerische Ausbildung` mit den Ausprägungen ''`Krankenschwester-/Pfleger`, `Altenpfleger`, usw.''.

Quantitative Daten unterscheiden sich in ihrer Größe, die Daten sind quantifizierbar. Sie lassen sich weiter unterscheiden in diskret und stetig.

Diskrete Daten sind nicht beliebig fein zu unterteilen, d.h. zwischen zwei ''benachbarten'' Größen existieren keine Zwischenwerte. Ein Beispiel für ein diskretes Merkmal ist die Anzahl der Kinder in Familien mit den Merkmalsausprägungen

- kein Kind
- ein Kind
- zwei, drei oder x Kinder

aber nicht 1,2548 Kinder!

Stetig: Im Gegensatz hierzu finden sich bei stetigen bzw. kontinuierlichen Daten zumindest theoretisch beliebig viele Zwischenwerte.
Die Körpergröße oder das Gewicht sind typische Beispiele für stetige bzw. kontinuierliche Merkmale. So lassen sich zwischen 173,5cm und 173,6cm noch beliebig viele Zwischenwerte angeben. In der Praxis jedoch stößt jedes Meßverfahren (und ebenso die Dokumentation eines unendlich langen Meßwertes) an seine Grenzen, so dass genau genommen auch stetige Größen nur diskret erfasst und beobachtet werden können.

1-2. Erhebung der Daten

Auf die Erhebung von Daten und was dabei zu beachten ist wird an dieser Stelle nicht eingegangen. Wir verweisen auf Seminare wie ''Quantitative Sozialforschung'' oder ''Forschungsmethodologie'', bzw. auf einschlägige Literatur, in der dann auch auf Begriffe wie Reliabilität, Validität, Stichprobenerhebung / zufällig oder gesteuert etc. ausführlich eingegangen wird.

1-3. Aufbereitung der Daten

Üblicherweise gewinnt man Daten in einer völlig beliebigen, nicht geordneten Reihenfolge. Nachfolgend findet sich ein Beispiel von 100 zufällig ausgewählten Personen, die dazu befragt wurden, wie häufig sie in den vergangenen 10 Jahren stationär in einem Krankenhaus behandelt worden sind. Die sogenannte *Urliste* ergibt sich, indem für jede befragte Person die Anzahl der Krankenhausaufenthalte aufgeschrieben wird:

```
1,0,0,3,1,5,1,2,2,0,1,0,5,2,1,0,1,0,0,4,0,1,1,3,0,
1,1,1,3,1,0,1,4,2,0,3,1,1,7,2,0,2,1,3,0,0,0,0,6,1,
1,2,1,0,1,0,3,0,1,3,0,5,2,1,0,2,4,0,1,1,3,0,1,2,1,
1,1,1,2,2,0,3,0,1,0,1,0,0,0,5,0,4,1,2,2,7,1,3,1,5,
```

Wie man leicht erkennen kann, ist diese Darstellung nicht besonders übersichtlich. Der nächste Schritt besteht nun darin, diese Daten sinnvoll zu ordnen. Es bietet sich hierfür eine Strichliste mit einer Skala von 0 bis 7 für die Anzahl der Krankenhausaufenthalte an.

Anzahl der Krankenhausaufenthalte	**Strichliste**	**absolute Häufigkeit h_i**
0	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~	30
1	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ IIII	34
2	~~IIII~~ ~~IIII~~ IIII	14
3	~~IIII~~ ~~IIII~~	10
4	IIII	4
5	~~IIII~~	5
6	I	1
7	II	2
Gesamt		100

Tabelle 1.2.: Ordnen der Daten

Zählt man ab, wie viele Personen keinmal, einmal, zweimal, etc. im Krankenhaus waren, erhält man die absolute Häufigkeit (siehe Tabelle 1.2), die wir mit $\mathbf{h_i}$ abkürzen werden.

Darüber hinaus interessiert bei vielen Fragestellungen nicht nur die absolute sondern auch die relative Häufigkeit, d.h. der jeweilige Anteil. Dazu setzt man die absoluten Häufigkeiten ins Verhältnis (in Relation) zur Gesamtzahl (**n**). Man dividiert also die jeweiligen absoluten Häufigkeiten durch die Gesamtzahl und erhält so die relativen Häufigkeiten ($\mathbf{f_i}$).

> Wäre beispielsweise die Befragung nach der Anzahl der Krankenhausaufenthalte in einer anderen Stadt mit 250 Personen durchgeführt worden, von denen 60 angaben, bereits einmal in einem Krankenhaus gewesen zu sein, so lässt der Vergleich der absoluten Häufigkeiten (Stadt A = 34, Stadt B = 60) keine Aussagen darüber zu, in welcher Stadt mehr bzw. weniger Leute einmal im Krankenhaus waren. Dies ist nur durch die Angabe der jeweiligen Anteile, also der relativen Häufigkeiten möglich. Für Stadt A ergibt sich eine relative Häufigkeit von 0,34 (34 / 100), für Stadt B wird ein Anteil von 0,24 (60 / 250) errechnet.

Multipliziert man die relativen Häufigkeiten mit 100, erhält man den prozentualen Anteil:

$0,34 * 100 = 34\%$ und $0,24 * 100 = 24\%$

Mit anderen Worten: 34% der Befragten in Stadt A und 24% derjenigen in Stadt B waren bereits einmal im Krankenhaus. Erst durch die Angabe der relativen Häufigkeit oder den prozentualen Anteil wird das Ergebnis der Umfrage vergleichbar. An der folgenden Tabelle 1.3 lassen sich die einzelnen Schritte nachvollziehen:

Anzahl der Kranken-hausaufenthalte	**absolute Häufigkeit $\mathbf{h_i}$**	**relative Häufigkeit $\mathbf{f_i = h_i/n}$**	**relative Häufigkeit in % $\mathbf{f_i}$ P 100**
0	30	0,30	30
1	34	0,34	34
2	14	0,14	14
3	10	0,10	10
4	4	0,04	4
5	5	0,05	5
6	1	0,01	1
7	2	0,02	2
Gesamt	$\sum h_i = n = 100$	$\sum f_i = 1$	$\sum f_i * 100 = 100$

Tabelle 1.3.: Häufigkeitsverteilung der Daten

Addiert man alle relativen Häufigkeiten, erhält man wieder das Ganze, also 1 oder 100%.

1-4. Skalentypen

Nach diesem ersten einführenden Beispiel müssen wir näher auf den Begriff Skala eingehen. Im obigen Beispiel hatten wir für die Anzahl der Krankenhausaufenthalte eine Skala von 0 bis 7 gewählt, um die Daten zu ordnen.
Allgemein dient eine Skala dazu, irgend etwas einzuteilen, zu ordnen oder zu sortieren. In der Statistik gibt es verschieden Arten von Skalen, in die man Daten einsortieren kann (Skalierung der Daten). Die verschiedenen Skalen haben abhängig von ihrem Informationsgehalt und der Möglichkeit der statistischen Auswertung ein sehr unterschiedliches Niveau.

I. Nominalskala

Nomen (latein.) heißt Namen oder Benennung

Fragt man zufällig ausgewählte Personen nach ihrem Familienstand, so wird man die Antworten ledig, verheiratet, geschieden und verwitwet erhalten. Beim sortieren der Urliste spielt es keine Rolle, ob zuerst die ledigen oder zuerst die geschiedenen Personen aufgelistet werden.
Nominaskaliert soll zum Ausdruck bringen, dass die erhobenen Werte nur eine Art von Etiketten sind, ''man gibt dem Kind einen Namen'', ohne dass irgendeine Wertigkeit oder Reihenfolge zugrunde liegt. Man kann auch sagen, Objekte oder Ereignisse werden in Kategorien eingeteilt, die sich gegenseitig ausschließen. Entweder hat ein Objekt bzw. Ereignis ein bestimmtes Merkmal oder nicht.
Weitere Beispiele für nominalskalierte Merkmale sind:

- Geschlecht (männlich, weiblich)
- Augenfarbe (grün, blau, braun,..)
- Körperbau (leptosom, pyknisch, athletisch)
- Krebslokalisation (Lunge, Magen, Darm, etc.)
- Zufriedenheit mit der Pflege (ja, nein)
- Religionszugehörigkeit (evangelisch, katholisch,...)
- Prüfung (bestanden, nicht bestanden)

Die Möglichkeiten der deskriptiven Statistik beschränken sich auf die Ermittlung der Häufigkeiten sowie auf die Bestimmung der Kategorie, in der die meisten Personen (oder Objekte) zu finden sind. Nehmen wir das Beispiel einer Prüfung.
In der Tageszeitung lesen Sie: ''`An der Schule XY haben von 160 Schülern 120 die Abschlussprüfung bestanden. Leider sind 40 Schüler durchgefallen.`''

	absolute Häufigkeit h_i	**relative Häufigkeit in % f_i P 100**
bestanden	120	75
nicht bestanden	40	25
Gesamt	n = 160	100

Tabelle 1.4.: Häufigkeitsverteilung der nominalen Daten

Dieses Beispiel soll verdeutlichen, dass in nominalskalierten Daten wenig Informationsgehalt steckt. Wir können nicht feststellen, mit welchen Noten die Schüler die Prüfung bestanden haben oder wieviel Punkte sie im einzelnen erreicht haben.

II. Ordinalskala

Ordinatus heißt im lateinischen soviel wie geordnet, ordentlich. Eine Ordinalzahl ist eine Ordnungszahl.

Zusätzlich zu den Eigenschaften der Nominalskala zeichnet sich dieses ‘‘nächst höhere’’ Skalenniveau dadurch aus, dass ordinale Merkmale oder Daten einer vorgegebenen Reihenfolge oder Rangfolge unterliegen.

‘‘Gut - Besser - Am Besten’’
‘‘Schlecht - Schlechter - Am Schlechtesten’’

Denkt man an die Stadieneinteilung von Krebserkrankungen, so wird zwar deutlich, dass mit Stadium 1 eine bessere Prognose verbunden ist als mit Stadium 2 oder Stadium 3. Auch wenn Stadium 4 sicherlich prognostisch am ungünstigsten beurteilt werden muss, können wir nicht das ‘‘wieviel besser’’ oder ‘‘wieviel schlechter genauer ausdrücken. Mit anderen Worten, die Abstände zwischen den Werten dieser Skala, also zwischen den verschiedenen Ausprägungen des Merkmals, sind nicht interpretierbar. Die Ziffern im o.g. Beispiel sind lediglich Ordnungs- oder Rangzahlen.
Beim Auflisten und Sortieren ordinaler Daten hat man sich in jedem Fall an die vorgegebene Reihen - oder Rangfolge zu halten. Zusätzlich zu der Information, in welche Kategorie eine Person oder ein Objekt gehört, lässt sich bei ordinalen Merkmalen noch die Zugehörigkeit zu einer bestimmten Rangstufe feststellen. Analog zu den nominalen Merkmalen lassen sich die entsprechenden Häufigkeiten ermitteln. Wie wir später noch sehen werden, existieren darüber hinaus Auswertungsmöglichkeiten, welche die zusätzlich in der Rangfolge liegenden Informationen berücksichtigen.
Ein typisches Beispiel für eine Ordinalskala sind die Schulnoten von sehr gut bis ungenügend, denen die entsprechenden Rangzahlen von eins bis sechs zugeordnet wurden. Kommen wir zurück zu dem Beispiel der Prüfung. Eine andere Tageszeitung veröffentlichte zusätzlich zu der Information ‘‘bestanden - nicht bestanden’’ den Notenspiegel. Der höhere Informationsgehalt der Ordinalskala wird sofort deutlich:

Zensur	**absolute Häufigkeit h_i**	**relative Häufigkeit in % f_i P 100**	
sehr gut	2	1,25	bestanden
gut	21	13,12	
befriedigend	37	23,12	
ausreichend	60	37,50	
mangelhaft	28	17,50	nicht bestanden
ungenügend	12	7,50	
Gesamt	n = 160	100	

Tabelle 1.5.: Häufigkeitsverteilung der ordinalen Daten

III. Metrische Skala

das lateinische Metor bedeutet soviel wie: abstecken, ausmessen oder abgrenzen

Das höchste Niveau besitzt die metrische Skala. Die Werte einer metrischen Skala unterliegen nicht nur einer Reihenfolge, benachbarte Werte weisen auch gleiche Abstände auf. Die Abstände zwischen den Werten sind somit interpretierbar. Man unterscheidet bei der metrischen Skala weiterhin die **Intervallskala** und die **Verhältnisskala**.

- Von einer **Intervallskala** spricht man, wenn kein natürlicher Nullpunkt vorliegt, z.B. °Celsius oder °Fahrenheit. Die Abstände auf einer Intervallskala sind zwar interpretierbar, nicht aber das Verhältnis. Zwischen 10 °C und 20 °C liegen genauso viele Werte, wie zwischen 20 °C und 30 °C, aber 30 °C bedeutet nicht doppelt so warm wie 15 °C, auch wenn dies auf den ersten Blick den Anschein hat. Umgerechnet in °Fahrenheit[1] ergibt sich: 15 °C = 59 °F und 30 °C = 86 °F. Wenn nun 30 °C doppelt so warm wären wie 15 °C, dann müßte es auch für die Werte in °Fahrenheit gelten. Wie man aber leicht erkennen kann, entspricht 86 °F nicht 2 * 59 °F .

- Eine **Verhältnisskala** dagegen besitzt einen natürlichen Nullpunkt (Länge, Gewicht, Lebensalter, usw.). Dies erlaubt Aussagen über das Verhältnis der einzelnen Werte, so ist ein zehnjähriger Junge doppelt so alt wie ein fünfjähriger. Die Unterscheidung von Intervall- und Verhältnisskala ist für die Anwendung von statistischen Verfahren für uns nicht von Bedeutung, sie sollte nur der Vollständigkeit halber erwähnt werden. Wir werden daher im folgenden nur von der metrischen Skala sprechen.

Man könnte jetzt einwerfen, dass die Abstände bei Schulnoten auch immer gleich eins und damit konstant sind. Da z.B. eine 4 doppelt so schlecht erscheint wie eine 2, könnte man auf die Idee kommen, Schulnoten als metrische Größen einzustufen. Hierbei wird jedoch vergessen, dass es sich bei Schulnoten nur um Rangzahlen (bzw. Ordnungszahlen) handelt, die wenig über die Differenz der Leistung aussagen. Darf man sich beispielsweise für die Note 1 nur drei Fehler erlauben, sind es für die Note 2 schon 7 und für die Note 3 schon 20 Fehler. Hieran wird deutlich, dass die Leistungsdifferenz zwischen den Noten nicht konstant ist, obwohl die Differenz der Rangzahlen (der Noten) jeweils den Wert eins besitzt.
Werden in einer Klausur oder Klassenarbeit richtig gelöste Aufgaben mit einer Anzahl von Punkten bewertet, dann handelt es sich wirklich um eine metrische Skala. Erreicht ein Schüler 30 Punkte und ein anderer vielleicht 60 Punkte, so lassen sich klare Aussagen über die Leistungsdifferenz der Schüler machen. Der erste Schüler war nur halb so gut wie der zweite. Die Differenz der beiden beträgt 30 Punkte.

IV. Anmerkungen

Für jeden Skalentyp stehen statistische Verfahren zur Verfügung, die den jeweiligen Informationsgehalt der Skalen optimal nutzen. Hierbei ist zu beachten, dass man Verfahren für nominale Merkmale auch auf ordinale und metrische Merkmale anwenden kann, und Verfahren für ordinale Daten auch bei metrischen Anwendung finden können. Dazu muß man die Daten allerdings auf das niedrigere Skalenniveau ‘‘herunter stufen’’. Wollen oder müssen wir auf metrischskalierte Daten Verfahren für ordinale Merkmale anwenden, sind wir gezwungen, metrische Daten in ordinale umzuwandeln. Dabei nehmen wir allerdings immer einen erheblichen Informationsverlust in Kauf. Ein umgekehrtes Vorgehen ist nicht möglich, d.h. statistische Verfahren dürfen niemals ‘‘unter ihrem Niveau’’ angewendet werden. Vor jeder Datenerhebung ist daher zu überlegen, welches Skalenniveau benötigt wird, um die Fragestellung zu beantworten. Im Zweifelsfall sollte man das höhere Niveau vorziehen. Um noch einmal die verschiedenen Skalen mit ihrem unterschiedlichen Informationsgehalt deutlich zu machen, fügen wir dem Beispiel der Prüfung in Tabelle 1.6 auf der folgenden Seite die Information über die erreichten Punktzahlen hinzu.

[1] die Umrechnungsformel lautet: $F = 1,8 * C + 32$

metrisch			ordinal			nominal		
Punkte	absolute Häufigkeit h_i	relative Häufigkeit in % f_i	**Zensur**	absolute Häufigkeit h_i	relative Häufigkeit in % f_i		absolute Häufigkeit h_i	relative Häufigkeit in % f_i
60	0	0						
59	0	0						
58	1	0,625	**sehr gut**	2	1,25			
57	1	0,625						
56	0	0						
55	0	0						
54	2	1,2						
53	1	0,625						
52	2	1,2						
51	3	1,875	**gut**	21	13,12			
50	2	1,2						
49	6	3,75						
48	5	3,125						
47	3	1,875						
46	3	1,875						
45	3	1,875						
44	5	3,125						
43	3	1,875	**befriedigend**	37	23,13			
42	7	4,375				**bestanden**	120	75
41	5	3,125						
40	5	3,125						
39	3	1,875						
38	3	1,875						
37	6	3,75						
36	6	3,75						
35	9	5,625						
34	4	2,5						
33	4	2,5						
32	9	5,625						
31	2	1,2	**ausreichend**	60	37,5			
30	7	4,375						
29	1	0,625						
28	1	0,625						
27	3	1,875						
26	4	2,5						
25	1	0,625						
24	3	1,875						
23	4	2,5						
22	2	1,2						
21	4	2,5						
20	0	0						
19	3	1,875						
18	1	0,625						
17	3	1,875	**mangelhaft**	28	17,5			
16	1	0,625						
15	0	0,625						
14	2	1,2						
13	3	1,875				**nicht bestanden**	40	25
12	1	0,625						
11	1	0,625						
10	2	1,2						
9	2	1,2						
8	0	0						
7	0	0						
6	1	0,625						
5	4	2,5	**ungenügend**	12	7,5			
4	1	0,625						
3	1	0,625						
2	1	0,625						
1	0	0						
0	0	0						
Gesamt	**n=160**	**100**		**n=160**	**100**		**n=160**	**100**

Tabelle 1.6.: Die Skalenniveaus im Überblick

1-5. Die Darstellung von Daten

Zunächst werden wir uns mit Daten befassen, welche die Ausprägungen eines Merkmals, also einer Variablen beschreiben. In der Literatur findet man Begriffe wie: univariat, monovariabel, eindimensional und ähnliches. Geht es um Beziehungen oder Zusammenhänge zwischen zwei oder mehr Variablen, spricht man von bivariat, bivariabel, zweidimensional oder auch mehrdimensional. Das Anliegen der deskriptiven Statistik ist das Ordnen, Zusammenfassen und Darstellen der Daten, damit sie übersichtlicher und anschaulicher werden. Ziel ist es u.a. zu verdeutlichen, wie häufig einzelne Merkmalsausprägungen vorkommen. Man spricht auch von Häufigkeitsverteilungen. Dazu stehen uns im wesentlichen die folgenden Möglichkeiten zur Verfügung:

- die Darstellung in Tabellenform
- die Darstellung durch Graphiken (Stabdiagramme, Histogramme, etc.)
- Charakterisierung durch sogenannte Kenngrößen

Die zahlenmäßige Darstellung durch sogenannte Kenngrößen wie Mittelwerte, Streuungsmaße und Maße für den Zusammenhang von Merkmalen führt zu einer extremen Reduktion der Daten. Die Gesamtheit der Daten wird hierbei durch typische Vertreter repräsentiert. Diese Kenngrößen werden in einem späteren Kapitel behandelt.

I. Tabellen und Graphiken

Als erster Schritt bietet sich die tabellarische Darstellung der Daten an. Wir hatten dies bereits unter dem Punkt ‘‘Aufbereitung der Daten’’ angesprochen. Ausgehend von den in einer Tabelle geordneten Daten lassen sich die verschiedene Graphiken erstellen. Nehmen wir das Beispiel der Anzahl der Krankenhausaufenthalte von Seite 2 (Tabelle 1.2). Die Daten der Urliste wurden mittels Strichliste sortiert, wir erhielten die absoluten Häufigkeiten und konnten die relativen Häufigkeiten berechnen.
An dieser Stelle führen wir noch den Begriff der Häufigkeitssumme ein. Eine Häufigkeitssumme (absolut oder relativ) erhält man durch schrittweises aufaddieren der einzelnen Häufigkeiten. Dieses schrittweise aufaddieren nennt man kumulieren. Zur Veranschaulichung des Berechnens der Häufigkeitssumme (H_i) wurde in der folgenden Tabelle 1.7 eine Spalte eingefügt, in der die schrittweise Addition der absoluten Häufigkeit nachvollzogen werden kann. Mit dem gleichen Vorgehen wird die relative Häufigkeitssumme (F_i) errechnet.

Anzahl der Krankenhaus-aufenthalte	**absolute Häufigkeit h_i**	**schrittweise Addition**	**absolute Häufigkeit kumuliert H_i**	**relative Häufigkeit in % f_i P 100**	**relative Häufigkeit kumuliert F_i**
0	30	30	30	30	30
1	34	30+34=64	64	34	64
2	14	64+14=78	78	14	78
3	10	78+10=88	88	10	88
4	4	88+4=92	92	4	92
5	5	92+5=97	97	5	97
6	1	97+1=98	98	1	98
7	2	98+2=100	100	2	100
Gesamt	n = 100			100	

Tabelle 1.7.: Schrittweise Addition der Häufigkeiten

Sehen wir uns die relative Häufigkeitssumme in Tabelle 1.7 an. Auf einen Blick lässt sich z.B. sagen, dass mehr als $^3/_4$ der Personen (78%) maximal zweimal im Krankenhaus waren.

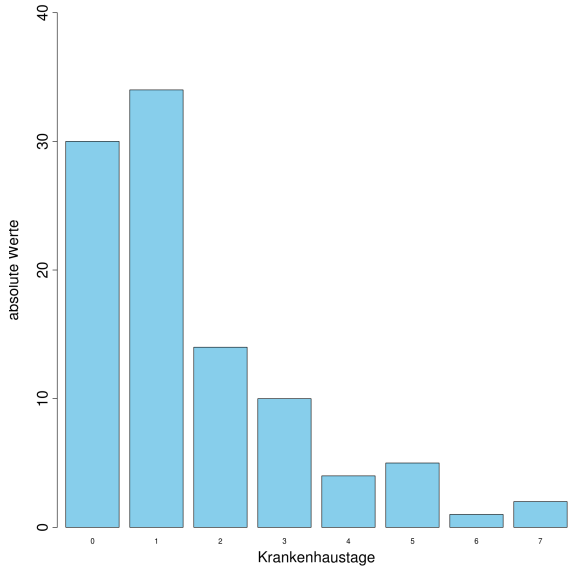


Abbildung 1.1.: Balkendiagramm

Kommen wir zur graphischen Darstellung unserer Daten und stellen die Anzahl der Krankenhausaufenthalte mittels eines Stab- oder Balkendiagramms dar. Die Höhe der Balken entspricht den absoluten Häufigkeiten. Ebenso eignen sich die relativen Häufigkeiten zur Erstellung eines Balkendiagramms. Dabei sollte beachtet werden, dass die Gesamtzahl (n) angegeben wird. Dies gilt insbesondere dann, wenn es sich um eine zahlenmäßig sehr kleine Erhebung handelt. Breite und Abstand der Stäbe spielen keine Rolle und können nach ästhetischen Gesichtspunkten gewählt werden, wobei die Übersichtlichkeit der Darstellung im Vordergrund stehen sollte.
Eine weitere Form der Darstellung ist der Polygonzug (siehe Abbildung 1.2). Trägt man die einzelnen Häufigkeiten nicht als Balken, sondern als Punkte oder Kreuze über dem entsprechenden Wert ein und verbindet diese Punkte, erhält man den Polygonzug der Häufigkeitsverteilung.
Ebenso lassen sich die Zwischenschritte der kumulierten Häufigkeiten darstellen. Man spricht dann vom Summenpolygon (siehe Abbildung 1.3). Analog zur relativen Häufigkeitssumme erlaubt auch das Summenpolygon durch einfaches Ablesen Aussagen vom Typ:

- 78% der Befragten waren maximal zweimal im Krankenhaus oder
- 50% der Schüler erreichten soundsoviel Punkte

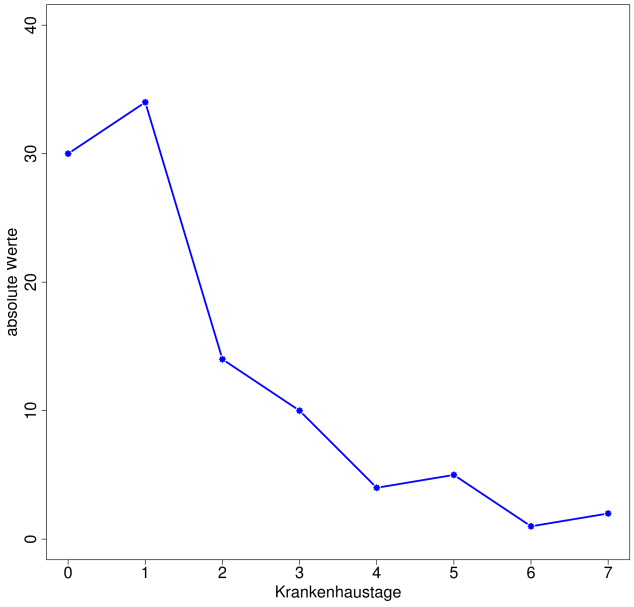


Abbildung 1.2.: Polygonzug

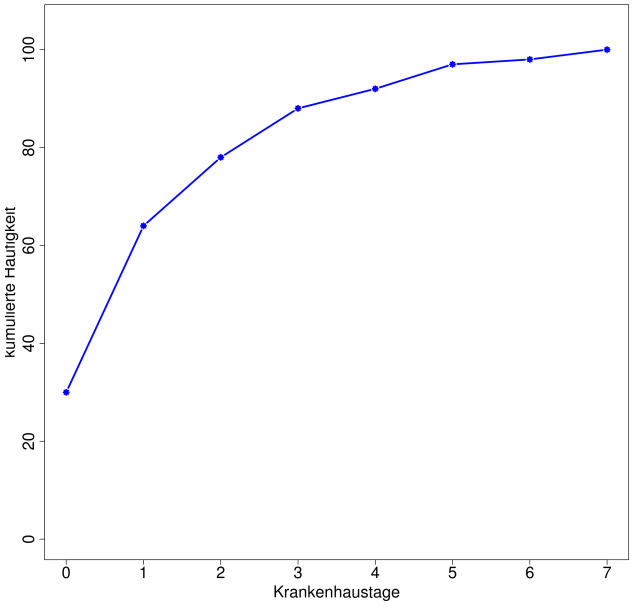


Abbildung 1.3.: Summenpolygon

Für unser Beispiel ‘‘Anzahl der Krankenhausaufenthalte’’ bieten die eben besprochenen Verfahren sicher eine gute Möglichkeit zur übersichtlichen und anschaulichen Darstellung der Daten.

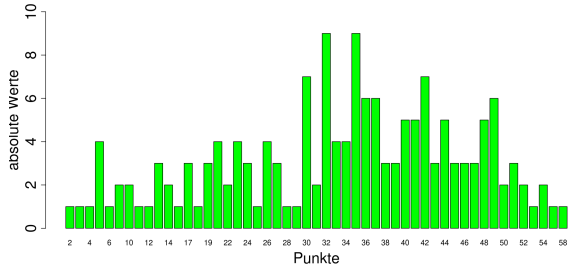


Abbildung 1.4.: Balkendiagramm Schulnoten

Stellen wir allerdings die Auflistung der einzelnen Punkte für die Prüfung (Seite 7) mittels eines Balkendiagramms dar, so kann man sicherlich nicht mehr von Übersichtlichkeit sprechen (Abbildung 1.4).
Wenn also sehr viele Meßwerte vorliegen oder wenn es sich um kontinuierliche Daten (Körpergröße, Gewicht, etc.) handelt, ist es sinnvoller, mehrere benachbarte Werte zu einer Klasse zusammenzufassen. Dadurch lässt sich größere Klarheit und Übersicht erreichen. Man spricht von Gruppierung oder Klassierung der Daten.

Begriffsdefinitionen:

- Eine Klasse ist die Menge sämtlicher Meßwerte, die innerhalb festgelegter Grenzen liegen.
- Diese festgelegten Kassengrenzen werden durch den kleinsten und den größten Meßwert einer Klasse gebildet.
- Die Kassenmitte ist das arithmetische Mittel (wird später eingehend besprochen) aus beiden Klassengrenzen.
- Die Kassenbreite ist bei diskreten Merkmalen die Anzahl der in der Klasse zusammengefaßten Merkmalsausprägungen, bei kontinuierlichen Merkmalen die Differenz der Klassengrenzen.
- Offene Kassen sind solche, in denen Meßwerte zusammengefaßt werden, die über oder unter einem Grenzwert liegen. Bei der Variablen Lebensalter wären z.B. ‘‘unter 18 Jahre’’ oder ‘‘über 70 Jahre’’ offene Klassen.

Es gibt keine verbindlichen Regeln zur Klassierung von Daten. Bei der Wahl einer zweckmäßigen Klassenbreite ist zu bedenken, dass sehr breite Klassen zuviel an Informationen über die Verteilung der Daten verwischen und zu schmale Klassen schnell unübersichtlich erscheinen. Auch bestimmt die Wahl der Klassenbreite unmittelbar die Anzahl der Klassen.
Klassierte oder gruppierte Daten lassen sich mit einem Histogramm graphisch gut darstellen. Ähnlich dem Balkendiagramm, allerdings ohne Abstände zwischen den einzelnen Balken. Bei einem Histogramm geht es um die Darstellung von Werten, die in einem Bereich : ‘‘von hier bis dort’’ liegen, also der Darstellung einer Fläche. Wir werden jetzt anhand des Beispiels ‘‘Punkte in der Prüfung’’ deutlich machen, wie unterschiedlich die Darstellung derselben Daten ausfallen kann, beeinflusst allein von der Art der Klassierung.

> Die Schreibweise **(0-6]** in Tabelle 1.8 bedeutet, ’’*mehr als 0 bis einschließlich 6*’’. Der Wert der Klassengrenze mit der runden Klammer wird nicht berücksichtigt, wohingegen der Wert mit der eckigen Klammer zu dieser Klasse zählt. Gerade bei kontinuierlichen Daten mit theoretisch unendlich vielen Zwischenwerten ist eine klare Abgrenzung wichtig, jeder Wert muß eindeutig einer Klasse zugeordnet werden können. Die Klassengrenzen **(6-12]** der zweiten Variante (Tabelle 1.9) bedeuten demnach: ’’*mehr als 6 bis einschließlich 12*’’. Sollte also ein Schüler 6,3 oder 6,5 Punkte erhalten haben, gehört er eindeutig zur zweiten Klasse.

Klassen-grenzen	Klassen-mitte	absolute Häufigkeit
(0-6]	3	8
(6-12]	9	6
(12-18]	15	10
(18-24]	21	16
(24-30]	27	17
(30-36]	33	34
(36-42]	39	29
(42-48]	45	22
(48-54]	51	16
(54-60]	57	2
	Gesamt	160

Tabelle 1.8.: Klassenbreite 6

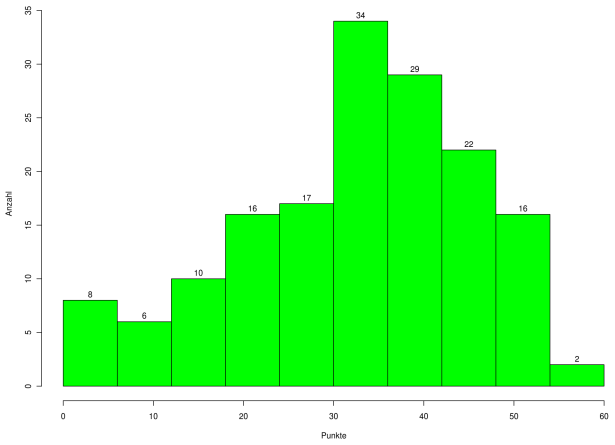


Abbildung 1.5.: Histogramm Klassenbreite 6

Klassen-grenzen	Klassen-mitte	absolute Häufigkeit
(0-12]	6	14
(12-24]	18	26
(24-36]	30	51
(36-48]	42	51
(48-60]	54	18
	Gesamt	160

Tabelle 1.9.: Klassenbreite 12

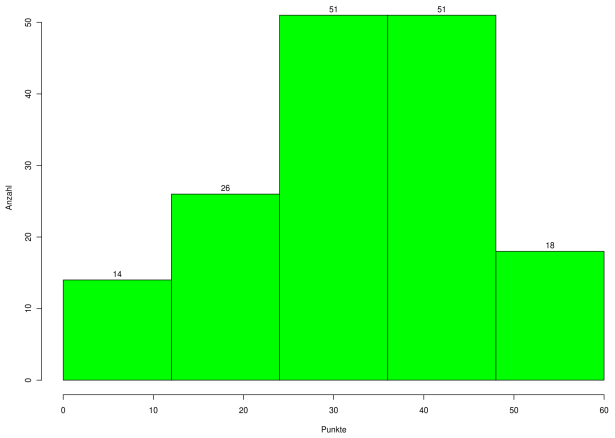


Abbildung 1.6.: Histogramm Klassenbreite 12

Wir könnten an dieser Stelle noch die unterschiedlichsten Histogramme derselben Daten liefern. Eines sollte jetzt klar geworden sein: Gerade weil es keine verbindlichen Richtlinien zur Klassierung von Daten gibt, ist die Gefahr der Manipulation sehr groß.
Ein weiteres Problem bei der Erstellung eines Histogramms tritt dann auf, wenn unterschiedliche Klassenbreiten gewählt werden.

Klassen-grenzen	Klassen-mitte	Klassen-breite	absolute Häufig-keit
(0-18]	9	18	24
(18-30]	24	12	33
(30-36]	33	6	34
(36-42]	39	6	29
(42-48]	45	6	22
(48-60]	54	12	18
		Gesamt	160

Tabelle 1.10.: Unterschiedliche Breite der Klassen

Liegen, wie in Tabelle 1.10, klassierte Daten mit unterschiedlichen Klassenbreiten vor, vermittelt ein **nicht flächenproportionales** Histogramm einen falschen Eindruck von den Daten.

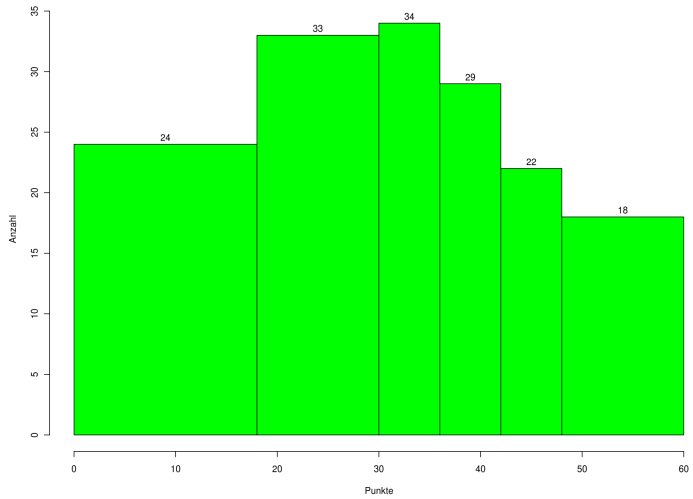


Abbildung 1.7.: nicht flächenproportionales Histogramm

Die Abbildung 1.7 zeigt, dass die erste Klasse mit einer Breite von 18 Punkten größer erscheint als die dritte Klasse mit einer Breit von 6 Punkten, obwohl die absolute Häufigkeit dieser Klasse (=34) die der ersten Klasse (=24) deutlich übertrifft. Auch erscheint die zweite Klasse optisch doppelt so umfangreich wie die dritte klasse, obwohl die absoulte Häufigkeit (=33) sogar einen Wert niedriger ist als in der dritten Klasse (=34).
Nur ein flächenproportionales Histogramm spiegelt den tatsächlichen Sachverhalt wieder. Hierzu setzt man die absolute Häufigkeit ins Verhältnis zur Klassenbreite, d. h. für unser Beispiel:

Klassen-grenzen	**Klassen-mitte**	**Klassen-breite**	**absolute Häufigkeit**	**Quotient aus absoluter Häufigkeit und Klassenbreite**
(0-18]	9	18	24	$24/18 = 1,3$
(18-30]	24	12	33	$33/12 = 2,7$
(30-36]	33	6	34	$34/6 = 5,6$
(36-42]	39	6	29	$29/6 = 4,8$
(42-48]	45	6	22	$22/6 = 3,6$
(48-60]	54	12	18	$18/12 = 1,5$
		Gesamt	160	

Tabelle 1.11.: Unterschiedliche Klassenbreite und Quotient

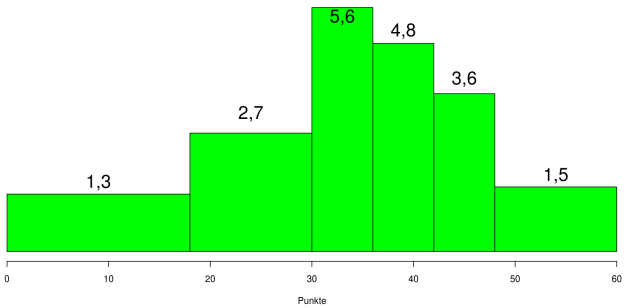


Abbildung 1.8.: flächenproportionales Histogramm

Die Y-Achse entspricht in diesem Histogramm (Abbildung 1.8) nicht mehr er absoluten Häufigkeit, sondern dem errechneten Quotienten.
Ein weiteres Beispiel soll den Unterschied verdeutlichen. Nehmen wir die folgende Altersverteilung von Personen, die in einem Londoner Stadtbezirk einen häuslichen Unfall erlitten haben (Tabelle 1.12, vgl. [1]). Schaut man sich Abbildung 1.9 an, so entsteht der Eindruck, dass Menschen im Alter von 15 bis 45 Jahren das höchste Unfallrisiko aufweisen. Zusätzlich scheint es so, als ob dieses Risiko bei Kleinkindern kleiner sei.

Klassen-grenzen	Klassen-mitte	Klassen-breite	absolute Häufigkeit
(0-5]	2,5	5	206
(5-15]	10	10	154
(15-45]	30	30	247
(45-65]	55	20	111
(65-90]	77,5	25	95
		Gesamt	813

Tabelle 1.12.: Häusliche Unfälle in London

Klassen-grenzen	Klassen-mitte	Klassen-breite	absolute Häufigkeit	Quotient aus absoluter Häufigkeit und Klassenbreite
(0-5]	2,5	5	206	$206/5 = 41,2$
(5-15]	10	10	154	$154/10 = 15,4$
(15-45]	30	30	247	$247/30 = 8,23$
(45-65]	55	20	111	$111/20 = 5,55$
(65-90]	77,5	25	95	$95/25 = 3,8$
		Gesamt	813	

Tabelle 1.13.: Proportionen der häuslichen Unfälle in London

Erst durch ein flächenproportionales Histogramm wird auch in diesem Fall der falsche Eindruck entzerrt. Setzen wir also wieder die absolute Häufigkeit in Bezug zur Klassenbreite (Tabelle 1.13), so erhalten wir das Histogramm aus Abbildung 1.10. Wie man nun erkennen kann, haben in Wirklichkeit die Kleinkinder das höchste Risiko eines häuslichen Unfalls.

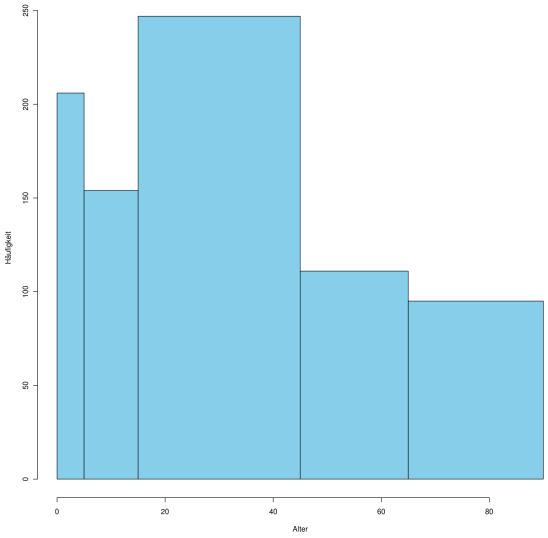


Abbildung 1.9.: Histogramm London

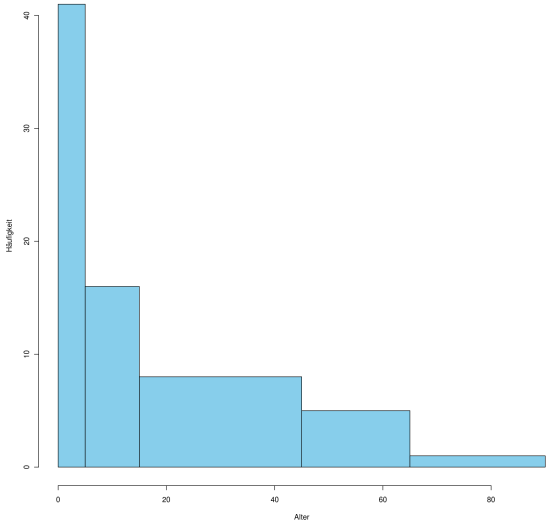


Abbildung 1.10.: Proportionales Histogramm

Es würde den Rahmen dieses Skriptes sprengen, auf alle Möglichkeiten der graphischen Darstellung einzugehen. Neben den besprochenen Diagrammen - Stab- oder Balkendiagramm, Polygonzüge, Histogramme - existieren noch vielfältige Möglichkeiten zur Präsentation der Daten. Spezielle Computerprogramme zur statistischen Datenverarbeitung bieten komfortable und vielfältige Optionen zur graphischen Darstellung des Materials. Hier sei nur noch das Kreisdia-

gramm erwähnt, welches eine sehr anschauliche Form zur Präsentation der Anteile bei nominalen Daten bietet.
Erinnern wir uns an das Beispiel der Prüfung. Auf nominalem Niveau konnten wir nur feststellen, wie viele Schüler die Prüfung bestanden, bzw. nicht bestanden hatten. Die Darstellung dieses Sachverhaltes ist sowohl als Balkendiagramm, als auch mittels eines Kreisdiagramms möglich. Wir überlassen natürlich dem Leser die Entscheidung, welche der beiden Möglichkeiten anprechender ist.

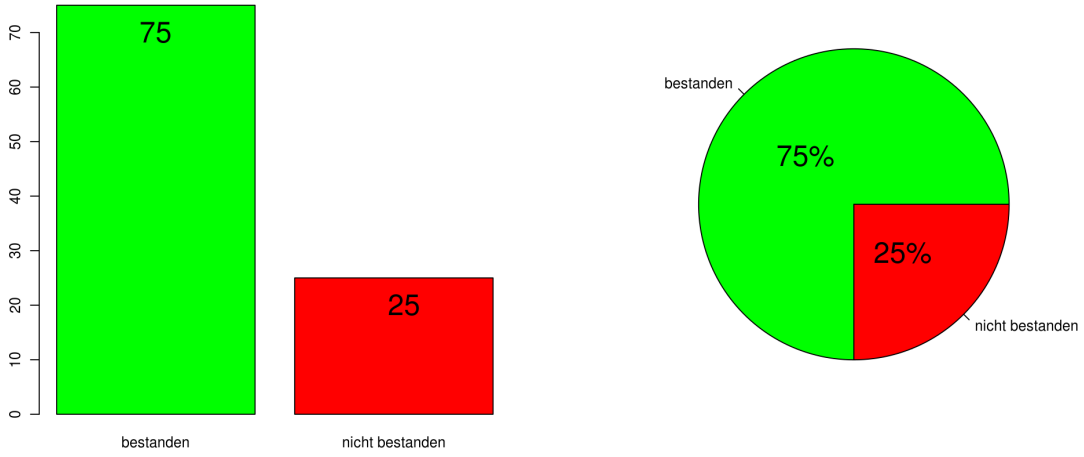


Abbildung 1.11.: Ergebnis der Abiturprüfung

Nach dieser Darstellung der Daten mittels Tabellen und Kenngrößen werden wir uns im nächsten Kapitel mit statistischen Kenngrößen beschäftigen.

1-6. Statistische Kenngrößen

Eine Tabelle oder graphische Darstellung der Daten informiert über die gesamte Verteilung eines Merkmals mit seinen Ausprägungen. Demgegenüber sagen statistische Kennwerte etwas über spezielle Eigenschaften der Verteilung aus. Wie bereits im Kapitel zur Darstellung der Daten behandeln wir zunächst nur Kenngrößen für univariate Verteilungen, also Größen, die sich auf ein Merkmal beziehen.
Die zahlenmäßige Darstellung der Daten durch statistische Kenngrößen beruht auf einer Reduktion der Daten. Die Gesamtheit des Datenmaterials wird durch typische Vertreter repräsentiert, welche die Datenmenge zusammenfassend und gründlich charakterisieren. Dies geht jedoch mit dem Verlust der einzelnen individuellen Daten und deren Informationsgehalt einher.
Sicherlich ist es einleuchtend, dass repräsentative Vertreter einer Häufigkeitsverteilung nicht die selten beobachteten Werte sind. Denken wir an unser Beispiel ''Schulnoten'', so sind die Zensuren ''sehr gut'' und ''ungenügend'' keine typischen Vertreter dieser Verteilung.

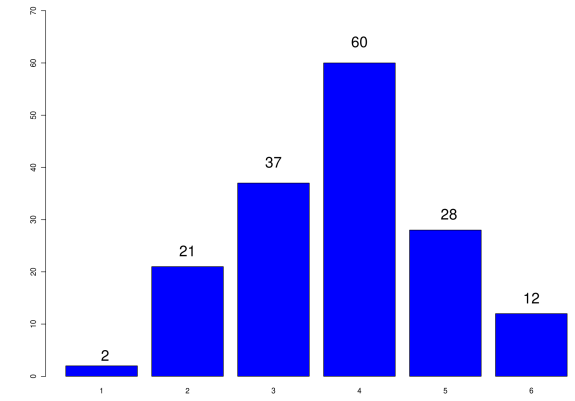


Abbildung 1.12.: Verteilung der Abiturnoten Schule "X"

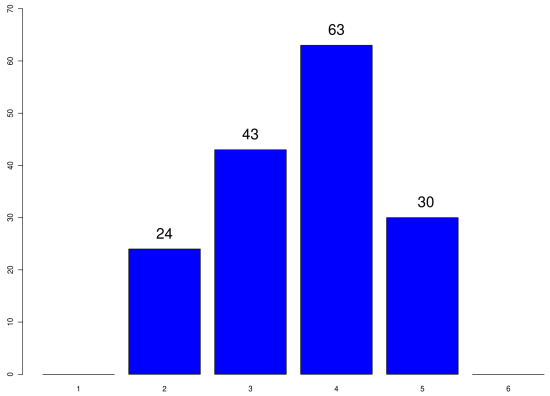


Abbildung 1.13.: Verteilung der Abiturnoten Schule "Y"

Abbildung 1.12 zeigt deutlich, dass der Hauptteil der Abiturienten die Note ausreichend erzielte. Das "Typische" dieser Verteilung lässt sich durch die Angabe, wo sich der "Hauptteil" der Werte befindet, also das Zentrum der Verteilung liegt, beschreiben. Allgemein spricht man von Maßen der zentralen Tendenz oder Lage-Kenngrößen. Maße der zentralen Tendenz, bzw. Lage-Kenngrößen allein genügen oft nicht, eine Häufigkeitsverteilung ausreichend zu beschreiben. Dies soll das folgende Beispiel verdeutlichen:
Betrachtet man die in Abbildung 1.13 dargestellte Verteilung der Noten der Schule "Y", kann festgestellt werden, dass auch hier der Hauptteil der Schüler die Note "ausreichend" erreichte. Auffallend ist aber, dass es an dieser Schule anscheinend keine sehr guten und ebenfalls keine sehr schlechten Schüler gibt, während an Schule "X" durchaus auch die Noten sehr gut und ungenügend vergeben wurden. Gibt man für die Schulen "X" und "Y" lediglich die Lagemaße an (der Hauptteil der Schüler beider Schulen erreichte Noten zwischen befriedigend und ausreichend), kann der Eindruck entstehen, dass beide Schulen bei der Abiturprüfung gleich abgeschnitten haben. Erst die Angabe "*Die Noten der Schule "X" verteilten sich von sehr gut bis ungenügend, die der Schule "Y" von gut bis mangelhaft*" macht die Unterschiedlichkeit der beiden Verteilungen deutlich.
Kenngrößen, die Aussagen über die Unterschiedlichkeit oder auch Variabilität von Meßwerten zulassen, nennt man **Streuungskenngrößen** oder auch **Dispersionsmaße**. Zusammenfassend kann gesagt werden, dass durch die Angabe von Lage- und Streuungskenngrößen das Charakteristische einer Verteilung oft schon recht deutlich wird.
Die Verfahren zur Ermittlung der verschiedenen Lage- und Streuungskenngrößen reichen vom einfachen Ablesen aus der Häufigkeitsverteilung bis hin zur mathematischen Berechnung. Die Frage, welche Kenngrößen verwendet werden und verwendet werden dürfen, hängt direkt mit dem Informationsgehalt, also mit dem Skalenniveau der Daten zusammen. beginnen werden wir mit den Lagekenngrößen.

I. Lage-Kenngrößen

Der Modalwert

Der **Modus** oder **Modalwert** ist der in einer Verteilung am häufigsten auftretende Wert.
Liegen klassierte Daten vor, entspricht der Modalwert der Klassenmitte der Klasse mit der größten Häufigkeit. Die Kategorie (Gruppe, Klasse, etc.) mit den häufigsten Werten nennt man

die Modalklasse. Es ist durchaus denkbar, dass innerhalb einer Verteilung mehrere Modalwerte ermittelt werden können. Sind jedoch die Häufigkeiten der Merkmalsausprägungen annähernd gleich verteilt, macht es keinen Sinn, den Modalwert zu bestimmen. Der Modalwert, bzw. die Modalklasse lässt sich für Daten jeglichen Skalenniveaus angeben. Andererseits bietet er für nominale Merkmale die einzige Möglichkeit, etwas typisches über die Verteilung der Merkmalsausprägungen aufzuzeigen.

> Im eigentlichen Sinne (Lokalisation der Merkmalsausprägungen auf der Meßwertskala) gibt es keine Lage-Kenngrößen für nominale Daten, denn sie verteilen sich weder auf einer Meßwertskala, noch unterliegt die Anordnung der Merkmalsausprägungen einer vorgegebenen Reihenfolge. Lorenz [2] weist darauf hin, dass der Modalwert / die Modalklasse ursprünglich nur auf metrische Merkmale angewendet und erst später auf nominale Daten ausgedehnt wurde.

Der Modalwert lässt sich ohne großen Rechenaufwand durch einfaches Ablesen aus der Häufigkeitsverteilung ermitteln.

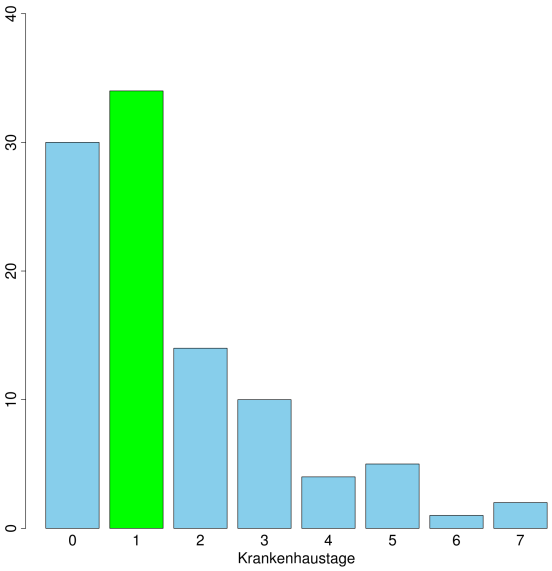


Abbildung 1.14.: Modus liegt bei 1

Der Modalwert, man spricht auch vom Dichtemittel, veranschaulicht eine Häufigkeitsverteilung insofern sehr gut, als es sich ja um den relativ am häufigsten auftretenden Wert handelt.
Er vermittelt eine Vorstellung davon, welcher Wert ''am ehesten auftreten wird und damit ''am wahrscheinlichsten'' ist. In unserm Beispiel zur Befragung nach der Anzahl der Krankenhausaufenthalte (Abbildung 1.14) tritt der Wert ‘‘bereits einmal im Krankenhaus'' am häufigsten auf.
Allerdings ist der Modus relativ unzuverlässig. Betrachtet man im Balkendiagramm die Anzahl derjenigen, die noch nie in einem Krankenhaus waren, wird deutlich, dass eine geringe Änderung der Daten (z.B. durch eine Wiederholungsuntersuchung) eine entscheidende Verschiebung des Modus zur Folge haben kann. Abbildung 1.15 zeigt dies am Beispiel von Zensuren. Durch die Veränderung eines Wertes verschiebt sich die Modalklasse von ‘‘sehr gut'' nach ‘‘befriedigend''.

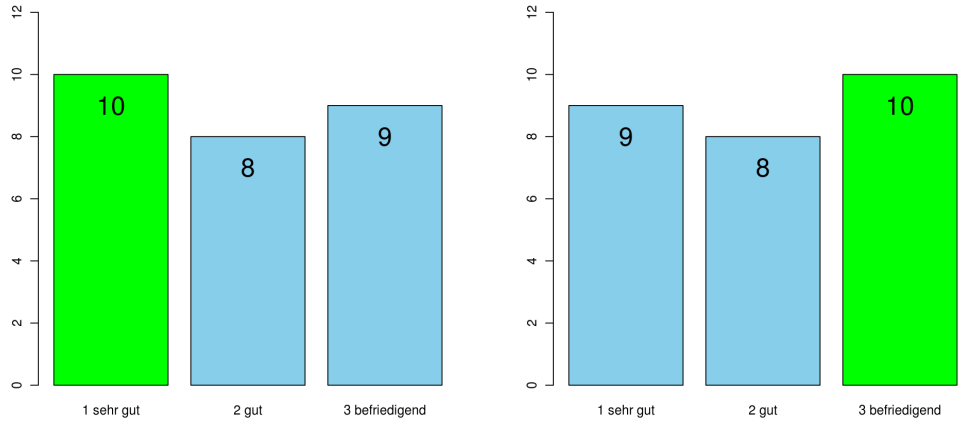


Abbildung 1.15.: Verschiebung des Modus durch Änderung eines Wertes

Der Median ($\tilde{x}$)

Unter Median wird der ''mittelste'' Wert verstanden. Man spricht auch vom Zentralwert.

> Der Median halbiert eine geordnete Meßwertreihe, d.h. je 50% der Werte befinden sich unterhalb und oberhalb des Medians.

Der Begriff ''geordnete Meßwertreihe'' weist auf das Skalenniveau hin, welches als Mindestanforderung für den Median vorliegen muß. Zur Erinnerung: Ordnen lassen sich lediglich ordinale Daten (nach Rangplätzen) und metrische Daten, bei nominalen Daten ist die Anordnung der Merkmalsausprägungen frei wählbar. Die Bestimmung des Medians ist somit nur für ordinale und metrische Merkmale zulässig.
Der Median kann bei geordneten Meßwertreihen mit geringer Anzahl einfach durch Ablesen des mittelsten Wertes ermittelt werden.
Beispiel:

Patient	A	B	C	D	E	F	G
Liegedauer in Tagen	5	7	5	6	6	5	8

Ordnen wir diese Messwerte, erhalten wir die folgende Reihenfolge

5	5	5	6	6	7	8

Aus der nun geordneten Meßwertreihe ergibt sich, dass der vierte Meßwert die Reihe halbiert. Es liegen sowohl drei Werte unterhalb wie auch oberhalb des Medians von 6 Tagen.
Nehmen wir nun an, es wäre nur bei 6 Patienten die Liegedauer in Tagen erfasst worden:

5	5	5	6	6	7

Der Median liegt in diesem Fall genau zwischen dem dritten und vierten Wert. Bei metrischen Daten ergibt der Durchschnitt aus 5 (dritter Wert) und 6 (vierter Wert) den Median, hier

also 5,5 Tage. Beim Vorliegen von ordinalen Daten werden beide Rangplätze benannt, d.h. der Median ergibt sich aus dem dritten und vierten Rangplatz mit den Werten 5 und 6. Da die ''Abzählmethode'' bei umfangreichen Meßwertreihen sehr umständlich ist, ermittelt man den Median durch Anwendung nachfolgender Formeln.
Für Messwertreihen mit **ungerader** Anzahl gilt:

$$\frac{n+1}{2} \tag{1.1}$$

Beispiel für 99 Messwerte:

$$\frac{99+1}{2} = 50$$

Die geordnete Messwertreihe wird durch den fünfzigsten Wert halbiert, sein Wert entspricht dem Median.
Bei **gerader** Anzahl gilt:

$$\frac{n}{2} \quad \textbf{und} \quad \frac{n}{2}+1 \tag{1.2}$$

Beispiel für 100 Messwerte:

$$\frac{100}{2} = 50 \quad \textbf{und} \quad \frac{100}{2}+1 = 51$$

Der Median bestimmt sich aus dem fünfzigsten und einundfünfzigsten Wert der geordneten Messwertreihe.
Etwas schwieriger wird die Bestimmung des Medians bei Daten, die in klassierter Form vorliegen.

Klassengrenzen	absolute Häufigkeit	kumuliert
(0-12]	14	14
(12-24]	26	**40**
(24-36]	**51**	*91*
(36-48]	51	142
(48-60]	18	160
Gesamt	**160**	

Tabelle 1.14.: Median bei klassierten Daten

Greifen wir noch einmal auf unser Beispiel der Anzahl von Punkten von Seite 11 zurück: Sicherlich haben Sie schnell ermittelt, dass der 80. und 81. Wert den Median bezeichnen, und dass diese Werte in der Klasse mit den Grenzen (24-36] liegen.
Möchte man genauere Angaben zum Median haben, benötigt man folgende Größen:

- die Hälfte der Messwerte: $\frac{n}{2} = \frac{\mathbf{160}}{2} = \mathbf{80}$
- die Klassenbreite b: **12**
- die untere Grenze der Medianklasse x_m: **24**
- die absolute Häufigkeit der Medianklasse h_m: **51**
- die kumulierte Häufigkeit der Klasse, die unterhalb des Medians liegt H_{m-1}: **40**

Die Berechnung des Medians erfolgt dann nach der allgemeinen Formel:

$$\tilde{x} = x_m + b \cdot \frac{\frac{n}{2} - H_{m-1}}{h_m} \tag{1.3}$$

Setzen wir unsere gegebenen Werte in diese Formel ein, ergibt sich für den Median:

$$\tilde{x} = 24 + 12 \cdot \frac{\frac{160}{2} - 40}{51} \;=\; \mathbf{33,41} \quad \text{(gerundet)}$$

Eine weiter Möglichkeit zur Ermittlung des Medians bei klassierten Daten ist die graphische Methode (siehe Anhang A-1). Der Median entspricht unserem natürlichen Gefühl von ''Mitte''. Liegen kumulierte Prozentwerte vor, lässt sich der Median leicht aus dem 50% Wert ablesen. Der Median ist nur eine Größe aus einer ganzen Reihe von Kenngrößen, die auf Rangreihen beruhen. Schließlich lässt sich eine Meßwertreihe nicht nur halbieren, man kann sie beispielsweise vierteln oder in 10 Teile teilen. Man spricht dann von den sogenannten **Quantilen**, die je nach Einteilung als **Quartile** (Viertel) oder **Zentile** (Zehntel) bezeichnet werden. Allgemein spricht man von p-Quantilen, die mit dem Symbol x_p abgekürzt werden. Nimmt man beispielsweise das 90%-Quantil oder 0,9-Quantil ($x_{0,9}$) bedeutet dies, dass 90% der Werte unterhalb und 10% der Werte oberhalb von $X_{0,9}$ liegen.

Warum macht es keinen Sinn, bei nominalen Daten den Median zu bestimmen?

Betrachten wir nachfolgend die erhobenen Daten zu einigen Pflegediagnosen.

	Häufigkeit	**Prozenz**	**kumulierte Prozente**
Störung des Körperbildes	10	40	40
soziale Isolation	5	20	60
Inkontinenz, funktional	8	32	92
Körpertemperatur, erhöht	2	8	100
Gesamt	25	100	

Tabelle 1.15.: Häufigkeitsverteilung der Pflegediagnosen

Laut Formel 1.1 würde der 13. Wert den Median darstellen. Da die Pflegediagnose "*Störung des Körperbildes*" 10 Messwerte enthält, liegt der 13. Messwert in der nachfolgenden Diagnose "*soziale Isolation*" (denn diese enthält 5 Messwerte, also die Werte von 11 bis 15).
Aber: die Anordnung, also die Reihenfolge der Kategorien von nominalen Merkmalen (hier die Pflegediagnosen) spielt keine Rolle! Ordnen wir die Pflegediagnosen einmal anders an, ergibt sich folgende Häufigkeitstabelle:

	Häufigkeit	**Prozenz**	**kumulierte Prozente**
Körpertemperatur, erhöht	2	8	8
soziale Isolation	5	20	28
Inkontinenz, funktional	8	32	60
Störung des Körperbildes	10	40	100
Gesamt	25	100	

Tabelle 1.16.: Pflegediagnosen in veränderter Reihenfolge

Jetzt fände sich der Median in der Kategorie "*Inkontinenz, funktional*". Es erscheint uns nicht notwendig, alle Möglichkeiten, die Pflegediagnosen anzuordnen, vorzuführen. Eines sollte klar sein, der "*13. Wert*" liegt jedesmal in einer anderen Kategorie. Wie aber soll ein Wert, der je nach Anordnung in einer anderen Kategorie liegt, ein typischer Repräsentant einer Häufigkeitsverteilung sein?
Fazit: für nominale Merkmale bietet sich lediglich der Modalwert an. Egal, wie die Pflegediagnosen angeordnet werden, die Kategorie "*Störung des Körperbildes*" bleibt die am häufigsten vorkommende Größe!

Der arithmetische Mittelwert ($\bar{x}$)

> Der arithmetische Mittelwert $\bar{\mathbf{x}}$ (sprich: *x quer*) eines Merkmals ist die Summe aller Daten dividiert durch den Stichprobenumfang.

Der arithmetische Mittelwert beruht auf mathematischen Operationen und darf daher ausschließlich bei metrischen Daten angewendet werden. Dieser Mittelwert entspricht eher dem umgangssprachlichen "Durchschnitt".
Die Berechnung von $\bar{\mathbf{x}}$ erfolgt, indem man alle Einzelwerte addiert ($\sum$) und anschließend durch die Anzahl (n) der Werte dividiert. Die allgemeine Formel lautet:

$$\bar{x} = \frac{\sum x_i}{n} \tag{1.4}$$

Beispiel:

Patient	A	B	C	D	E	F	G
Liegedauer in Tagen	5	7	5	6	6	5	8

Eingesetzt in die Formel 1.4 ergibt sich:

$$\begin{aligned} \bar{x} &= \frac{(5+7+5+6+6+5+8)}{7} \\ &= \frac{42}{7} \\ &= \mathbf{6} \end{aligned}$$

Der mittlere stationäre Aufenthalt beträgt für dieses Beispiel sechs Tage.
In der Praxis wird eher selten eine Meßwertreihe mit sieben Einzelwerten wie in diesem Beispiel vorliegen. Oder anders ausgedrückt: In der Regel werden einzelne Meßwerte mehrfach vorkommen. Entsprechend berücksichtigt man bei der Berechnung des arithmetischen Mittelwertes die absoluten Häufigkeiten.

$$\bar{x} = \frac{\sum x_i \cdot h_i}{n} \tag{1.5}$$

Beispiel:

Stationärer Aufenthalt in Tagen (x_i)	1	2	3	4	5	6	7
Anzahl der Patienten (abs. Häufigkeit h_i)	2	4	3	5	6	4	1

Eingesetzt in die Formel 1.5 ergibt sich:

$$\begin{aligned}\bar{x} &= \frac{\sum (1\cdot 2)+(2\cdot 4)+(3\cdot 3)+(4\cdot 5)+(5\cdot 6)+(6\cdot 4)+(7\cdot 1)}{25}\\ &= \frac{\sum (2+8+9+20+30+24+7}{25}\\ &= \frac{100}{25}\\ &= \mathbf{4}\end{aligned}$$

Der mittlere stationäre Aufenthalt beträgt für dieses Beispiel vier Tage.
Eine weitere ‘‘Variante’’ des arithmetischen Mittelwertes kommt zur Anwendung, wenn beispielsweise der Gesamtmittelwert aus beiden Stichproben (beide Beispiele) interessiert. Man spricht dann vom **gewichteten arithmetischen Mittelwert**. Eine Möglichkeit besteht darin, jeden Einzelwert der beiden Beispiele zu summieren und anschließend durch die Anzahl aller Werte zu dividieren.

$$\bar{x} = \frac{Summe_{gesamt}}{Anzahl_{gesamt}} = \frac{Summe_1 + Summe_2}{Anzahl_1 + Anzahl_2}$$

bzw.

$$\bar{x} = \frac{\sum x_{i1} + \sum x_{i2}}{n_1 + n_2} \tag{1.6}$$

Das Addieren aller Einzelwerte dürfte bei umfangreichen Messwertreihen recht mühsam sein. Nutzen wir die folgende Tatsache:

$$\text{Da} \quad \bar{x} = \frac{\sum x_i}{n} \quad \text{gilt dementsprechend} \quad n \cdot \bar{x} = \sum x_i$$

Das mühsame Aufaddieren der Einzelwerte kann demnach entfallen. Wir ersetzen $\sum x_{i1}$ durch $n_1 \cdot \bar{x}_1$ und $\sum x_{i2}$ entsprechend durch $n_2 \cdot \bar{x}_2$
Die Berechnung des Gesamtdurchschnitts erfolgt nach folgender Formel:

$$\bar{x}_{gesamt} = \frac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2}{n_1 + n_2} \tag{1.7}$$

Wir erinnern uns, dass in Beispiel 1 der stationäre Aufenthalt der 7 Patientinnen im Schnitt 6 Tage und in Beispiel 2 der Aufenthalt der 25 Patientinnen durchschnittlich 4 Tage betrug.
Eingesetzt in die Formel 1.7 ergibt sich ein gesamter oder gemeinsamer Durchschnitt von 4,437 Tagen:

$$\begin{aligned}\bar{x}_{gesamt} &= \frac{7\cdot 6 + 25\cdot 4}{7+25}\\ &= \frac{42+100}{32}\\ &= 4,437\end{aligned}$$

II. Streuungskenngrößen

Wie eingangs bereits erwähnt unterscheiden sich Häufigkeitsverteilungen nicht nur durch ihre Lage auf der Meßwertskala, sondern auch durch ihre Streuung. Streuungskenngrößen sind Maßzahlen zur Bewertung oder Quantifizierung der Variabilität der Meßwerte.

Die Spannweite "R" und der Quartilsabstand "Q"

Ein besonders einfaches Streuungsmaß ist die Spannweite, auch Variationsbreite genannt, die mit "R" (engl. range) abgekürzt wird. Sie bestimmt sich aus der Differenz zwischen dem größten und dem kleinsten Meßwert.

$$R = x_{max} - x_{min} \tag{1.8}$$

Der Quartilsabstand "Q", auch Interquartilsabstand genannt, gibt die mittleren 50% der Verteilung an, d.h. sowohl das obere, als auch das untere Viertel entfallen. Mit anderen Worten bezeichnet "Q" den Abstand zwischen dem ersten und dritten Quartil.

$$Q = x_{0,75} - x_{0,25} \tag{1.9}$$

Relativ einfach lassen sich diese beiden Streuungsmaße aus der geordneten Meßwertreihe bestimmen. Tabelle 1.17 zeigt 15 Meßwerte, der Vollständigkeit halber wurde auch der Median gekennzeichnet.

5	5	6	6	6	7	7	7	7	7	8	8	8	9	9
$\mathbf{x_{min}}$			$\mathbf{x_{0,25}}$				Median $\mathbf{x_{0,5}}$				$\mathbf{x_{0,75}}$			$\mathbf{x_{max}}$

Tabelle 1.17.: Stationärer Aufenthalt in Tagen

Aus diesen Werten ergibt sich eine Spannweite von: $R = 9 - 5 = 4$ Tage

Der Quartilsabstand beträgt: $Q = 8 - 6 = 2$ Tage

Andere Streuungsmaße berücksichtigen die Abweichung jedes einzelnen Wertes vom arithmetischen Mittelwert. Diese Abweichungen werden aufsummiert:

$$\sum x_i - \bar{x}$$

Dabei gibt es allerdings (zunächst) ein Problem. Da sich sowohl oberhalb als auch unterhalb des Mittelwertes Einzelwerte befinden, somit also positive und negative Abweichungen vom Mittelwert vorhanden sind, ergibt die Summe dieser Abweichungen Null.

$$\sum x_i - \bar{x} = 0$$

Für unser Beispiel auf dieser Seite beträgt der arithmetische Mittelwert 7 Tage, die nachfolgende Tabelle zeigt die jeweiligen Abweichungen vom arithmetischen Mittelwert:

Messwert:	5	5	6	6	6	7	7	7	7	7	8	8	8	9	9
Abweichung $\mathbf{x_i - \bar{x}}$:	-2	-2	-1	-1	-1	0	0	0	0	0	+1	+1	+1	+2	+2
	Die Summe der negativen Abweichungen beträgt -7										Die Summe der positiven Abweichungen beträgt +7				

Summe der quadrierten Abweichungen "SQ"

Um dennoch etwas über die Streuung der Werte vom Mittelwert aussagen zu können, greift man auf das Quadrieren zurück. Durch dieses Verfahren ergibt die Summe der Abweichungen in jedem Fall eine positive Zahl. Grundlegender Begriff beim Aufbau der nachfolgenden Formeln der Streuungsmaße ist die Summe der quadrierten Abweichungen "SQ".

$$SQ = \sum (x_i - \bar{x})^2 \cdot h_i \tag{1.10}$$

Es versteht sich, dass selbstverständlich bei mehrfachen Auftreten eines Meßwertes die Häufigkeiten (h_i) berücksichtigt werden. Bleiben wir bei unserem Beispiel von Seite 26 und setzen die Werte in die Formel 1.10 ein:

$$\begin{aligned} SQ &= (5-7)^2 \cdot 2 + (6-7)^2 \cdot 3 + (7-7)^2 \cdot 5 + (8-7)^2 \cdot 3 + (9-7)^2 \cdot 2 \\ &= (-2)^2 \cdot 2 + (-1)^2 \cdot 3 + (0)^2 \cdot 5 + 1^2 \cdot 3 + 2^2 \cdot 2 \\ &= 4 \cdot 2 + 1 \cdot 3 + 0 \cdot 5 + 1 \cdot 3 + 4 \cdot 2 \\ &= 8 + 3 + 0 + 3 + 8 \\ &= 22 \end{aligned}$$

Varianz s^2

Zugegebenermaßen ist die Summe der quadrierten Abweichungen nicht besonders anschaulich, wenn wir etwas über die Variabilität der Werte im Beispiel des stationären Aufenthaltes aussagen möchten. Der nächste Schritt ist nun, SQ auf die Anzahl der Meßwerte "n" zu beziehen. Handelt es sich bei den zu berechnenden Daten um eine Stichprobe (was meistens der Fall ist) rechnet man mit "$n-1$". Daraus ergibt sich die **Varianz s^2**

$$s^2 = \frac{SQ}{n-1}$$

bzw.

$$s^2 = \frac{\sum (x_i - \bar{x})^2 \cdot h_i}{n-1} \tag{1.11}$$

In unserem Beispiel errechneten wir $SQ = 22$, welche wir nun durch $n-1$ dividieren.

$$s^2 = \frac{SQ}{n-1} = \frac{22}{15-1} = 1,57$$

Standardabweichung s

Konsequenterweise sollte nun die Quadrierung wieder rückgängig gemacht werden (Lorenz [2] schreibt: "*... den Ausdruck SQ/n-1 auf dieselbe Dimension zu bringen, welche die Meßwerte selbst besitzen*"). Um die Quadrierung rückgängig zu machen, zieht man die Wurzel und erhält damit die **Standardabweichung "s"**:

$$s = \sqrt{s^2} \quad \text{bzw.} \quad s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1} \cdot h_i} \tag{1.12}$$

Für unser Beispiel ergibt sich damit:

$$s = \sqrt{s^2} \quad = \sqrt{1,57} \quad = 1,25$$

In Publikationen findet sich häufig die Abkürzung `sd` (standard deviation) für die Standardabweichung anstelle von `s`.

Die Bedeutung der Standardabweichung

Liegt eine glockenförmige, symmetrische Verteilung vor, eine sogenannte Normalverteilung (siehe Abbildung 1.16), so liegen in dem Bereich von einer Standardabweichung um den Mittelwert, d.h. innerhalb $\bar{x}+s$ und $\bar{x}-s$ ca. $^2/_3$ aller Werte (genau 68,23%, siehe Abbildung 1.17). Innerhalb von 2 Standardabweichungen sind es bereits ca. 95% der Werte (siehe Abbildung 1.18).
Für unser Beispiel bedeutet dies, dass 68,23% der Patientinnen $7-1,25$ bis $7+1,25$ Tage stationär im Krankenhaus lagen, also 68,23% der Patientinnen lagen zwischen 5,75 und 8,25 Tagen stationär.
95% der Patientinnen lagen zwischen 4,5 und 9,5 Tagen im Krankenhaus ($7 \pm 2,5$). Anders ausgedrückt kann man sagen, dass insgesamt nur 5% der Pat. weniger als 4,5 und mehr als 9,5 Tage im Krankenhaus verweilten.

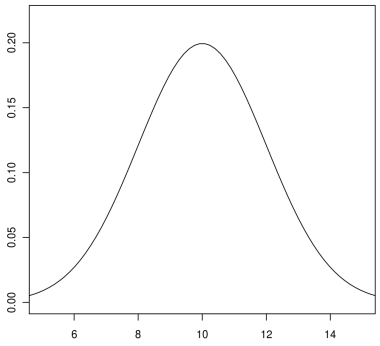


Abbildung 1.16.: Normalverteilung

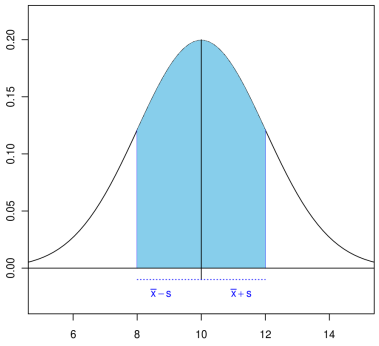


Abbildung 1.17.: Fläche von $\bar{\mathbf{x}} \pm \mathbf{s}$

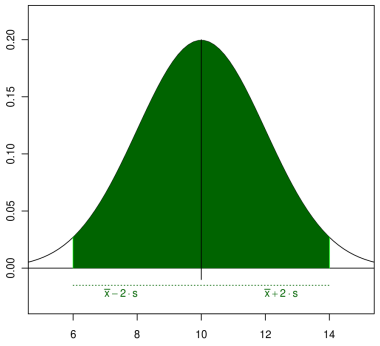


Abbildung 1.18.: Fläche von $\bar{\mathbf{x}} \pm \mathbf{2} \cdot \mathbf{s}$

Der Variationskoeffizient "Vk"

Bei Verteilungen mit unterschiedlichen Mittelwerten könnte die Frage interessieren, in welcher der beiden Verteilungen die Variabilität der Meßwerte größer ist. Um dies zu untersuchen, bietet sich der Variationskoeffizient an. Dieser setzt jeweils die Standardabweichung ins Verhältnis zum Mittelwert. Multipliziert man den Wert mit 100, so erhält man den Variationskoeffizienten in Prozent.

$$VK = \frac{s}{\bar{x}} \quad \text{bzw.} \quad VK' = \frac{s}{\bar{x}} \cdot 100 \tag{1.13}$$

Beispiel:
Auf allgemeinchirurgischen Stationen fallen im Durchschnitt pro Patient/Tag Kosten in Höhe

von 500,- € (s=75) an. Auf den untersuchten Intensivstationen beliefen sich die Kosten auf 1000,- € (s=100) täglich. Es errechnen sich die folgenden Variationskoeffizienten:

$$VK_{Chirurgie} = 550 = 0,15 \quad \text{bzw.} \quad 15\%$$

$$VK_{Intensiv} = 1000 = 0,1 \quad \text{bzw.} \quad 10\%$$

Die Verteilung der Werte auf den Intensivstationen ist demnach homogener, bzw. die Variabilität der Werte ist auf den chirurgischen Stationen ist höher.
Bleiben wir bei dem Beispiel und nehmen wir an, ein Verwaltungsleiter stellt fest, dass in seinem Haus die chirurgische Station im Schnitt täglich 650,- € pro Patient aufwendet, und die Intensivstation sogar 1200,- €. Damit liegt die chirurgische Station 150,- € und die Intensivstation 200.- € über dem Durchschnitt. Auf den ersten Blick hat also die Intensivstation noch ‘‘schlechter’’ gewirtschaftet als die Chirurgie. Bevor man aber ‘‘Äpfel mit Birnen’’ vergleicht, also Werte vergleicht, die aus sehr unterschiedlichen Verteilungen stammen, muß man diese Verteilungen erst einmal vergleichbar machen.

Exkurs: z-Werte und z-Transformation

Sollen also Werte aus unterschiedlichen Verteilungen verglichen werden, so müssen sie laut Bortz[3] S.45:

> ‘‘....zuvor an der Unterschiedlichkeit aller Werte im jeweiligen Kollektiv relativiert werden. Dies geschieht, indem die Abweichungen durch die Standardabweichungen im jeweiligen Kollektiv dividiert werden. Ein solcher Wert wird als z-Wert bezeichnet.’’

Unter den unendlich vielen Normalverteilungen gibt es eine Normalverteilung, die sich dadurch auszeichnet, dass sie einen Erwartungswert von $\mu = 0$ und eine Streuung von $\sigma = 1$ aufweist. Dieser speziellen Normalverteilung wird deshalb eine besondere Bedeutung zugemessen, weil sämtliche übrigen Normalverteilungen durch eine einfache Transformation in sie überführbar sind. Dies erfolgt mit der Formel:

$$z_i = \frac{x_i - \bar{x}}{s} \qquad \text{bzw.} \qquad z_i = \frac{x_i - \mu}{\sigma} \tag{1.14}$$

Transformiert man mit Hilfe der Formel 1.14 alle vorliegenden Werte einer Verteilung, erhält man eine neue Verteilung mit einem Mittelwert von Null und einer Streuung von eins

$$z = 0 \quad \text{und} \quad s_z = 1 \qquad \text{bzw.} \qquad \mu = 0 \quad \text{und} \quad \sigma = 1$$

Durch die z-Transformation können also sämtliche Normalverteilungen standardisiert werden. Deshalb wird die Normalverteilung mit $\mu = 0$ und $\sigma = 1$ auch als *Standardnormalverteilung* bezeichnet.
Setzen wir für unsere Stationen die Werte in die Formel 1.14 ein, ergibt sich:

$$z_{chirurgie} = \frac{650 - 500}{75} = 2 \quad \text{und} \quad z_{intensiv} = \frac{1200 - 1000}{100} = 2$$

Durch die z-Transformation sind beide Stationen vergleichbar geworden. Beide Transformationen ergeben einen z-Wert von ‘‘2’’ d.h., beide Stationen weichen in gleicher Höhe vom Mittelwert ab.
Weitere Übungsaufgaben zur z-Transformation finden sich im Anhang D-1 auf Seite 74.

III. Streuungsmaße und Skalenniveau

Für nominalskalierte Daten gibt es im eigentlichen Sinne keine Streuungskenngrößen. Man kann lediglich die Zahl der Kategorien angeben. Alle besprochenen Streuungskenngrößen beruhen auf Rechenoperationen, selbst bei der Spannweite "R" wurde die Subtraktion des kleinsten vom größten Meßwert durchgeführt. Mathematische Operationen sind aber, - wie bereits mehrfach erwähnt -, lediglich bei metrischem Skalenniveau der Daten zulässig.
Trotzdem ist es möglich, auch bei ordinalen Daten die Spannweite und den Quartilsabstand anzugeben. Allerdings beschränken sich die Angaben dann darauf, über wie viele **Skalenstufen** bzw. **Rangstufen** sich die Spannweite oder der Quartilsabstand erstrecken. Es wird im eigentlichen Sinne nicht mehr gerechnet.
Als Beispiel sind in Tabelle 1.18 sind die Ergebnisse der Zwischenprüfung von 15 Altenpflegerinnen notiert worden. Die Ergebnisse erstrecken sich über vier Noten, es gibt bei diesem Beispiel demnach vier Rangstufen auf der Skala.

1. Stufe			**2. Stufe**						**3. Stufe**				**4. Stufe**	
2	2	2	3	3	3	3	3	3	4	4	4	4	5	5
$\mathbf{x_{min}}$			$\mathbf{x_{0,25}}$				Median $\mathbf{x_{0,5}}$				$\mathbf{x_{0,75}}$			$\mathbf{x_{max}}$

Tabelle 1.18.: Ergebnisse der Zwischenprüfung

Die Spannweite "R" erstreckt sich über 4, der Quartilsabstand "Q" über 2 Ränge.

Nominale Skala	Ordinale Skala	Metrische Skala
• es gibt Unterschiede, aber nur im Sinne von Klassifizierung • die Reihenfolge spiel keine Rolle	• es gibt Unterschiede, größer oder kleiner, besser oder schlechter, die Abstände lassen sich aber nicht interpretieren • die Reihenfolge der Klassen muß beachtet werden	• die Größe des Unterschieds kann gemessen werden • die Abstände sind interpretierbar • gleiche Abstände benachbarter Werte • ist der Ausgangspunkt/ Nullpunkt frei wählbar, handelt es sich um eine **Intervallskala**, z.B. °C oder Fahrenheit, die Abstände sind interpretierbar, nicht aber das Verhältnis • liegt ein natürlicher Nullpunkt vor (Länge, Gewicht, Lebensalter) handelt es sich um eine **Verhältnisskala**. (10 Jahre ist doppelt soviel wie 5 Jahre). Abstände und Verhältnis der Werte sind interpretierbar.
Lagekenngrößen		
• Modalwert	• Modalwert • Median	• Modalwert • Median • arithmetisches Mittel
Streuungskenngrößen		
• keine Streuungskenngrößen, allenfalls Zahl der Kategorien	• Spannweite und Quartilsabstand in Skalenstufen	• Spannweite • Quartilsabstand • Varianz • Standardabweichung • Variationskoeffizient

Tabelle 1.19.: Zusammenfassung

1-7. Bivariate Statistik

In den bisherigen Abschnitten haben wir uns mit der Beschreibung und Darstellung von **einem** Merkmal mit seinen Ausprägungen beschäftigt. In der Praxis liegt das Interesse häufig darin herauszufinden, ob es einen Zusammenhang oder eine Beziehung zwischen zwei (oder mehr) Merkmalen gibt.
Nehmen wir als fiktives Beispiel eine Dekubitusstudie, in der der Zustand der Haut, die Größe des Hautdefektes und die durchgeführte Dekubitusprophylaxe festgehalten wird. In einem solchen Beispiel könnte man den Dekubitus als eine Art von ‘‘Supermerkmal’’ bezeichnen, dass sich aus vielen verschiedenen Einzelkomponenten zusammensetzt. Man spricht dann von einem mehrdimensionalen Merkmal. Im Rahmen dieses Skriptes werden wir uns allerdings auf die Beschreibung von bivariaten Häufigkeitsverteilungen beschränken.
Der einfachste Fall einer solchen Verteilung liegt vor, wenn wir zwei Merkmale betrachten, die jeweils nur zwei Ausprägungen besitzen. Nehmen wir die Merkmale ‘‘Geschlecht’’ mit den Ausprägungen ‘‘weiblich und männlich’’ und ‘‘Rauchen’’ wobei wir nur zwischen Raucher und Nichtraucher unterscheiden. Es gibt dann insgesamt nur vier Kombinationsmöglichkeiten und jeder Proband kann eindeutig einer dieser Gruppen zugeordnet werden:

- Weibliche Raucher
- Weibliche Nichtraucher
- Männliche Raucher
- Männliche Nichtraucher

Für die übersichtliche Darstellung bieten sich sogenannte Kontingenztafeln an, die sich aus Reihen und Spalten zusammensetzen, in denen die Merkmale mit ihren Ausprägungen eingetragen werden können.

Spalten ⇊

Reihen ⇉		Spalte 1 = Raucher	Spalte 2 = Nichtraucher	
	Reihe 1 = Männer	**a = Anzahl männliche Raucher**	**a = Anzahl männliche Nichtraucher**	Summe der Reihe 1 = $a+b$ = Anzahl der Männer
	Reihe 2 = Frauen	**c = Anzahl weibliche Raucher**	**d = Anzahl weibliche Nichtraucher**	Summe der Reihe 2 = $c+d$ = Anzahl der Frauen
	Summe der Spalten	Summe der Spalte 1 = $a+c$ = Anzahl der Raucher	Summe der Spalte 2 = $b+d$ = Anzahl der Nichtraucher	Gesamt n = $a+b+c+d$

Tabelle 1.20.: Beispiel einer Kontingenztafel, hier 2x2-Felder-Tafel

Die schwarz umrandeten Felder in der Mitte der Tabelle nennt man Zellen. Sie ergeben sich aus dem ‘‘Schnittpunkt’’ der Reihen und Spalten. In unserem Beispiel spricht man von einer 2x2

Felder-Tafel (allgemein r x s Tafel, für r = Anzahl der Reihen und s = Anzahl der Spalten). Da es sich in diesem Fall um nominale Merkmale handelt, ist die Anordnung der Merkmalsausprägung frei wählbar. Bei ordinalen oder metrischen Variablen, muß natürlich die Rang- bzw. Reihenfolge eingehalten werden. Nachfolgend findet sich eine Liste von Noten der Vorprüfung und der Abschlußprüfung von 20 Auszubildenden einer Krankenpflegeschule.

Nummer der Auszubildenden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Vornote	2	3	3	4	3	4	3	3	4	1	2	3	2	2	1	2	4	3	1	4
Abschlussnote	1	3	2	4	3	3	3	4	3	1	2	3	2	1	2	2	3	2	1	2

Die Anordnung der Merkmalsausprägung (hier die Noten von 1 bis 5) sind durch die Rangfolge vorgegeben.

		Note der Abschlussprüfung					
		1	2	3	4	5	Summe:
Note der Vor-prüfung	1	2	1	0	0	0	3
	2	2	3	0	0	0	5
	3	0	2	3	0	0	5
	4	0	1	3	1	0	5
	5	0	0	0	1	1	2
	Summe:	4	6	7	2	1	Gesamt: 20

Tabelle 1.21.: Zweidimensionale Darstellung ordinaler Daten

Das Lesen einer solchen Darstellung:

Es gibt z.B. 3 Auszubildende, die sowohl in der Vor- als auch in der Abschlußprüfung die Note ‘‘2’’ erzielten (2. Reihe, 2. Spalte) und 3 Personen, die in der Vorprüfung die Note ‘‘4’’ und in der Abschlußprüfung die Note ‘‘3’’ erhielten (4. Reihe, 3. Spalte). Weiterhin lässt die Darstellung die Vermutung zu, dass es einen Zusammenhang zwischen der Vorprüfung und der Abschlußprüfung gibt. Dafür spricht die stärkere Besetzung der Felder in der Diagonalen oder nahe der Diagonalen.

I. Der Zusammenhang zweier Merkmale

Greifen wir noch einmal das Beispiel ‘‘Rauchen und Geschlecht’’ auf. Nehmen wir an, die Stichprobe besteht aus 300 Probanden (100 Frauen und 200 Männer), von denen die Hälfte Raucher und die andere Hälfte Nichtraucher sind. Wenn es nun keinen Zusammenhang zwischen dem Merkmal Geschlecht und dem Rauchverhalten gibt, würden wir erwarten, dass sowohl die Hälfte der untersuchten Frauen als auch die Hälfte der Männer Raucher sind. In der nachfolgenden Tabelle 1.22 sind die Häufigkeiten, wie man sie bei Unabhängigkeit der Merkmale erwarten würde, dargestellt.
Natürlich sind die Verhältnisse bei tatsächlichen Untersuchungen nicht so ‘‘rund und glatt’’ und auf einen Blick sichtbar. Die erwarteten Häufigkeiten ergeben sich für die einzelnen Zellen, indem man die entsprechende Reihensumme mit der Spaltensumme multipliziert und dann durch die Gesamtzahl dividiert.

Merkmal ‘‘Rauchen’’

Merkmal ‘‘Geschlecht’’		Raucher	Nichtraucher	Summe
	Männer	**männliche Raucher, erwartete Anzahl = 100**	**männliche Nichtraucher, erwartete Anzahl = 100**	200 ($= r_1$)
	Frauen	**weibliche Raucher, erwartete Anzahl = 50**	**weibliche Nichtraucher, erwartete Anzahl = 50**	100 ($= r_2$)
	Summe	150 ($= s_1$)	150 ($= s_2$)	300

Tabelle 1.22.: 2x2-Felder-Tafel für die Merkmale ‘‘Rauchen’’ und ‘‘Geschlecht’’

So errechnet sich zum Beispiel die erwartete Anzahl für die Zelle ‘‘männliche Raucher’’:

$$\frac{Summe_{Reihe1} \cdot Summe_{Spalte1}}{Summe_{Gesamt}} \quad \text{bzw.} \quad \frac{r_1 \cdot s_1}{n}$$

Setzen wir die entsprechenden Zahlen ein, ergibt sich:

$$\frac{200 \cdot 150}{300} = 100$$

Die allgemeine Formel zur Berechnung der erwarteten Häufigkeiten für die Zelle der i-ten Reihe und j-ten Spalte ($Z_{i,j}$) lautet:

$$Z_{i,j} = \frac{r_i \cdot s_j}{n} \tag{1.15}$$

Nachdem die erwarteten Häufigkeiten für alle vier Felder bestimmt sind, vergleichen wir diese jetzt mit den tatsächlich beobachteten Werten (130 männliche Raucher, 70 männliche Nichtraucher, 20 weibliche Raucher, 80 weibliche Nichtraucher):

Merkmal ‘‘Rauchen’’

Merkmal ‘‘Geschlecht’’		Raucher	Nichtraucher	Summe
	Männer	**beobachtete Anzahl = 130 erwartete Anzahl = 100**	**beobachtete Anzahl = 70 erwartete Anzahl = 100**	200 ($= r_1$)
	Frauen	**beobachtete Anzahl = 20 erwartete Anzahl = 50**	**beobachtete Anzahl = 80 erwartete Anzahl = 50**	100 ($= r_2$)
	Summe	150 ($= s_1$)	150 ($= s_2$)	300

Tabelle 1.23.: Vergleich von beobachteten und erwarteten Werten

Die tatsächlich beobachteten weichen von den erwarteten Häufigkeiten ab. Dies spricht für einen Zusammenhang zwischen den Merkmalen ‘‘Geschlecht und Rauchen’’. Der Anteil rauchender

Männer ist größer als erwartet, bzw. der Anteil rauchender Frauen ist geringer als erwartet. Anders ausgedrückt: Würden die beobachteten mit den erwarteten Häufigkeiten übereinstimmen, so spräche dies für die Unabhängigkeit der beiden Merkmale.

Anmerkung

An dieser Stelle wird immer die Frage gestellt:

> "*Wie groß müssen die Abweichungen zwischen beobachteten und erwarteten Häufigkeiten denn sein, um von einem Zusammenhang sprechen zu können?*"

Dies ist Inhalt eines späteren Kapitels sein, welches sich mit Signifikanztesten beschäftigt. Dort werden u.a. Verfahren zur Überprüfung der Unabhängigkeit von Merkmalen behandelt.

II. Die Stärke des Zusammenhanges zweier Merkmale

Bei der Berechnung der Stärke des Zusammenhanges kommt wiederum dem Skalenniveau der Daten entscheidende Bedeutung zu. Die folgende Tabelle gibt einen Überblick über einige Zusammenhangsmaße:

Skalenniveau der Daten	**Bezeichnung des Zusammenhangs**	**Kenngröße des Zusammenhangs**
nominal (2x2-Felder-Tafel)	Assoziation/ Kontingenz	Assoziationskoeffizient/ Kontingenzkoeffizient
ordinal	Rangkorrelation	Rangkorreltaionskoeffizient
metrisch	Maßkorreltation	Maßkorreltationskoeffizient

Tabelle 1.24.: Zusammenhang zwischen Merkmalen

In diesem Skript werden wir uns mit der

- **Rangkorrelation nach Spearman** für ordinales Skalenniveau und der
- **Maßkorrelation nach Bravais-Pearson** für metrische Daten befassen

Zudem beschränken wir uns auf die Beschreibung von linearen Zusammenhängen.
Bevor wir mit der Berechnung von Korrelationskoeffizienten beginnen, müssen wir uns zunächst mittels einer graphischen Darstellung davon überzeugen, dass ein linearer Zusammenhang vorliegt. Nehmen wir als Beispiel die Frage:

> "*Gibt es bei Kindern einen Zusammenhang zwischen Leistungen bei der Rechtschreibung und beim Lesen?*"

Nachfolgend finden sich in der Tabelle die Punkte, die 10 Schüler in einem Rechtschreibtest (x_i) und in einem Lesetest (y_i) erreichten.

Schüler	a	b	c	d	e	f	g	h	i	j
Lesetest x_i	2	12	7	15	10	4	13	9	16	19
Rechtschreibung y_i	3	14	9	17	12	4	16	12	18	20

Tabelle 1.25.: Lese- und Rechtschreibtest

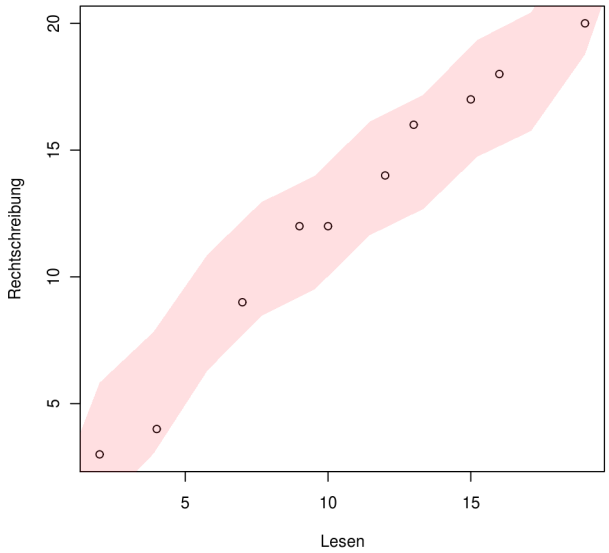


Abbildung 1.19.: positiver Zusammenhang

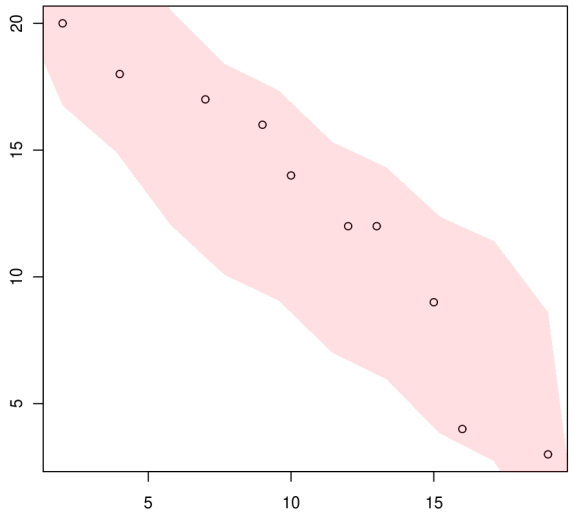


Abbildung 1.20.: negativer Zusammenhang

Zeichnet man die Wertepaare (x,y) in ein Koordinatensystem, so ergibt sich das Streudiagramm aus Abbildung 1.19, auch **Scatterplot** oder **Punktwolke** genannt. Zunächst einmal lässt sich durch ein solches Streudiagramm feststellen, inwieweit von einem linearen Zusammenhang ausgegangen werden kann.
Zur Verdeutlichung der Punktwolke zeichnet man eine Linie um die Punkte. Je ‘‘enger’’ oder auch ausgeprägter die Ellipsenform der Punktwolke, um so enger ist der Zusammenhang. Steigt die Punktwolke an, d.h. niedrige Punkte beim Schreiben gehen mit ebenfalls niedrigen Punkten beim Lesen, bzw. hohe Punktzahl beim Schreiben mit hoher Punktzahl beim Lesen einher, spricht man von einem **positiven** Zusammenhang. Einen sogenannten **negativen** Zusammenhang (Abbildung 1.20) findet man beispielsweise in dem Fall, dass aus der zunehmenden Dosierung eines blutdrucksenkenden Medikaments entsprechend sinkende Blutdruckwerte resultieren.
Erst nach der Feststellung, dass ein Zusammenhang, und zwar ein linearer Zusammenhang vorliegt, ist es sinnvoll, die Stärke des Zusammenhangs zu berechnen. Sowohl für den Maß-, als auch für den Rangkorrelationskoeffizienten gilt, dass sie Werte im Bereich von **-1 bis + 1** annehmen können. Je näher der Korrelationskoeffient bei Eins liegt (sowohl -1 als auch + 1), um so stärker ist der Zusammenhang der beiden Variablen. Von einem schwachen, bzw. keinem Zusammenhang spricht man bei Annäherung des Wertes an Null.
Bleiben wir bei dem Beispiel des Rechtschreib- und Lesetests (Tabelle 1.25). Wie wir später sehen werden, errechnet sich hier ein Korrelationskoeffizient von 0,98. Dies bedeutet, dass es einen besonders starken Zusammenhang zwischen Rechtschreib- und Lesefähigkeiten gibt. Allerdings sagt ein Korrelationskoeffizient nichts über die Ursache des Zusammenhanges aus. Folgende Möglichkeiten müssen in Betracht gezogen werden:

- Beeinflusst die Rechtschreibfähigkeit (R) das Lesen (L)? oder
- Hat die Lesefähigkeit Einfluss auf die Rechtschreibung? oder
- Beeinflussen sich Lesen und Schreiben gegenseitig? oder

- Werden Lese- und Schreibfähigkeiten von einer dritten oder weiteren Variablen wie z.B. Intelligenz oder Alter beeinflusst?

Allgemeiner Ausgedrückt:

R $\rightrightarrows$ L	R $\leftleftarrows$ L	R $\leftrightarrows$ L	dritte Variable $\downdownarrows$ $\downdownarrows$ R L

Achtung: Nicht selten kommt es zu sogenannten ''Scheinkorrelationen'' oder ''Nonsenskorrelationen'', wie bei Lorenz [2] am Beispiel der Korrelation zwischen ''Anzahl nistender Störche'' und ''Anzahl der Neugeborenen'' sehr anschaulich beschrieben.

Der Maßkorrelationskoeffizient nach Pearson (r_p)

Der Maßkorrelationskoeffizient nach Pearson (genau genommen nach Bravais-Pearson) eignet sich prinzipiell nur für metrische Daten und kann für Stichproben ungeachtet ihrer Verteilungseigenschaften berechnet werden. Bei der Berechnung von Pearson's Maßkorrelationskoeffizient bedienen wir uns der Kovarianz $cov(x, y)$ als Hilfsgröße.

$$r_p = \frac{cov(x,y)}{s_x \cdot s_y} \qquad \text{wobei} \qquad cov(x,y) = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n-1} \tag{1.16}$$

Für die Punktwerte aus unserem Beispiel (Schreib- und Lesetest) ergeben sich folgende Größen:

$\mathbf{x_i}$	$\mathbf{y_i}$	$(\mathbf{x_i} - \bar{\mathbf{x}})$	$(\mathbf{x_i} - \bar{\mathbf{x}})^\mathbf{2}$	$(\mathbf{y_i} - \bar{\mathbf{y}})$	$(\mathbf{y_i} - \bar{\mathbf{y}})^\mathbf{2}$	$(\mathbf{x_i} - \bar{\mathbf{x}}) \cdot (\mathbf{y_i} - \bar{\mathbf{y}})$
2	3	-8,7	75,69	-9,5	90,25	82,65
4	4	-6,7	44,89	-8,5	72,25	56,95
7	9	-3,7	13,69	-3,5	12,25	12,95
9	12	-1,7	2,89	-0,5	0,25	0,85
10	12	-0,7	0,49	-0,5	0,25	0,35
12	14	1,3	1,69	1,5	2,25	1,95
13	16	2,3	5,29	3,5	12,25	8,05
15	17	4,3	18,49	4,5	20,25	19,35
16	18	5,3	28,09	5,5	30,25	29,15
19	20	8,3	68,89	7,5	56,25	62,25

Tabelle 1.26.: Lese- und Rechtschreibtest

Die benötigten Werte errechnen sich nun wie folgt (siehe hierzu auch Anhang A-2):

$$\bar{y} = \frac{\sum y_i}{n} = \frac{125}{10} = 12,5 \qquad \bar{x} = \frac{107}{10} = 10,7$$

$$s_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} = \frac{296,5}{9} = 32,94 \qquad \text{und} \qquad s_y = \sqrt{32,94} = 5,74$$

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{260,1}{9} = 28,9 \qquad \text{und} \qquad s_x = \sqrt{28,9} = 5,38$$

$$cov(x,y) = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n-1} = \frac{274,5}{9} = 30,5$$

Setzen wir die errechneten Größen in die Formel für den Maßkorrelationskoeffizienten ein, ergibt sich:

$$r_p = \frac{cov(x,y)}{s_x \cdot s_y} = \frac{30,5}{5,74 \cdot 5,38} = 0,9876$$

Wie bereits besprochen, kann r Werte zwischen -1 und +1 annehmen. Auf unser Beispiel angewendet bedeutet dies, dass es einen sehr starken Zusammenhang zwischen Lese- und Rechtschreibfähigkeiten gibt. Nehmen wir jetzt an, wir hätten Zweifel daran, dass die Punkte der beiden Teste metrisch skaliert sind. In einem solchen Fall weichen wir auf den Rangkorrelationskoeffizienten für ordinales Skalenniveau aus.

Der Rangkorrelationskoeffizient nach Spearman (r_s)

Wie man aus dem Namen Rangkorrelationskoeffizient schon ableiten kann, spielen die Ränge, bzw. die Rangplätze der Daten eine entscheidende Rolle bei der Berechnung von Spearman's Rho. Er bietet sich für ordinale oder aber für metrische, nicht normalverteilte Daten an. Bleiben wir bei unserem Beispiel der Schreib- und Leserfähigkeit der Schüler. Der erste Schritt zur Bestimmung des Rangkorrelationskoeffizienten besteht darin, den einzelnen Punktwerten - für jede Variable getrennt - Rangplätze zu zuordnen. Da bei der Berechnung der Rangkorrelation die Differenzen der Rangplätze, bzw. die quadrierten Differenzen eine große Rolle spielen, haben wir sie in der nachfolgenden Tabelle mit aufgenommen.

Schüler	Punkte Schreiben	Rangplätze Schreiben	Punkte Lesen	Rangplätze Lesen	Differenz der Ränge d	d^2
a	3	1	2	1	0	0
b	14	6	12	6	0	0
c	9	3	7	3	0	0
d	12	4,5*	9	4	0,5	0,25
e	4	2	4	2	0	0
f	16	7	13	7	0	0
g	12	4,5*	10	5	-0,5	0,25
h	17	8	15	8	0	0
i	18	9	16	9	0	0
j	20	10	19	10	0	0
n = 10						$\sum d^2 = 0,5$

Tabelle 1.27.: Rangplätze der Daten

(*) Da der Punktwert ''12'' zweimal vorkommt, wird hier aus den Ranplätzen vier und fünf der Mittelwert gebildet. Man spricht in einem solchen Fall auch von gebundenen Rängen oder Rangbildungen. Läge der Wert ''12'' dreimal vor, würde man z.B. den Mittelwert der Rangplätze 4, 5, und 6 entsprechend vergeben. Es muss hier gleich angemerkt werden, dass die Formel 1.17 beim Vorliegen von mehr als 20% Rangbindungen nicht mehr angewendet werden darf. Liegen Daten ''von Natur'' aus auf ordinalem Niveau vor, wie beispielsweise Schulnoten, erhält man beim Zuordnen der Rangplätze sehr schnell einen großen Anteil gebundener Ränge. In den verschiedenen Lehrbüchern finden sich entsprechende Formeln zur Berechnung des Rangkorrelationskoeffizienten unter Berücksichtigung von gebundenen Rängen.
Die allgemeine Formel zur Berechnung der Rangkorrelationskoeffizienten nach Spearman (auch *Spearman's Rho*) lautet:

$$r_s = 1 - \frac{6 \cdot \sum d_i^2}{n \cdot (n^2 - 1)} \tag{1.17}$$

Setzen wir die aus der Tabelle 1.27 ersichtlichen Werte in diese Formel ein, erhalten wir:

$$\begin{aligned} r_s &= 1 - \frac{6 \cdot 0,5}{10 \cdot (100 - 1)} \\ &= 1 - \frac{3}{990} \\ &= 1 - 0,00303 \\ &= 0,9969 \end{aligned}$$

Auch Spearman's Rho kann Werte zwischen -1 und +1 annehmen, auch hier gilt: je näher an Eins, um so enger/stärker der Zusammenhang. Nähert sich der Wert an Null an, liegt kein bzw. nur ein sehr schwacher Zusammenhang vor. Nachdem wir uns mittels der Korrelationsrechnung mit der Quantifizierung der Stärke des Zusammenhanges zweier Merkmale beschäftigt haben, interessiert im nachfolgenden Kapitel die Art des Zusammenhangs.
Abschließend sind in Abbildung 1.21 verschiedene Punktwolken mit ihrem Korrelationskoeffizienten angebildet.

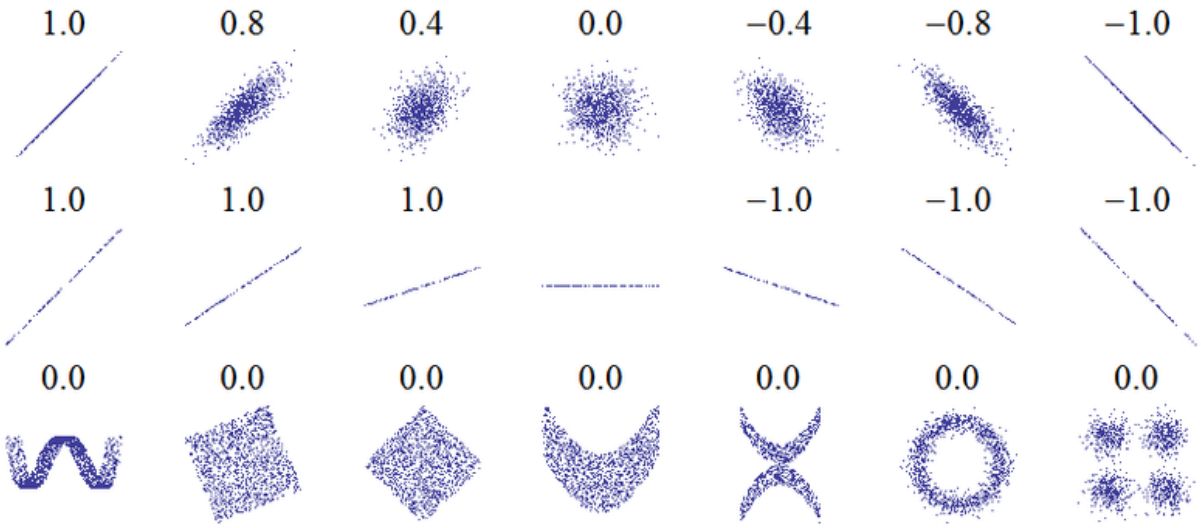


Abbildung 1.21.: Beispiel Korrelationskoeffizienten

1-8. Regressionsanalysen

Im Kaptiel zur Korreltation wurde der Zusammenhang zweier Variablen statistisch untersucht. Als Ergebnis steht der Korrelationskoeffizient `r`, der die Stärke des Zusammenhangs in einer Zahl ausdrückt. Die zentrale Frage wird hierbei aber nicht beantwortet, nämlich: Wie ändert sich die Ausprägung eines Merkmals, wenn die Einflussgröße systematisch verändert wird? Anders ausgedrückt: Wie verhält sich die abhängige Variable (auch Zielgröße), wenn sich die unabhängige Variable verändert (Einflussgröße)? Wissen über die Art des Zusammenhangs zweier Merkmale erlaubt Vorhersagen wie beispielsweise: Eine Erhöhung der Dosis des Schlafmittels um 5 Milligramm verlängert die Schlafdauer um 4 Stunden.

I. Grundgedanken

Um die Art des Zusammenhanges zu berechnen, steht uns die Regressionsanalyse zur Verfügung. Im Gegensatz zur Korrelationsrechnung, die sich auf eine zweidimensionale oder bivariate Häufigkeitsverteilung bezieht, geht es bei der Regressionsrechnung um eine **abhängige** Variable (auch Zielgröße), die von einer **unabhängigen** Variablen (Einflussgröße) beeinflusst wird.
Die *einfache lineare* Regressionsanalyse ist ein Verfahren für metrische Daten. Im Rahmen dieses Skriptes werden wir wiederum nur auf lineare Zusammenhänge eingehen. Dementsprechend ist es auch bei der Regressionsrechnung wichtig, sich zunächst einen optischen Eindruck mittels Streudiagramm zu verschaffen.

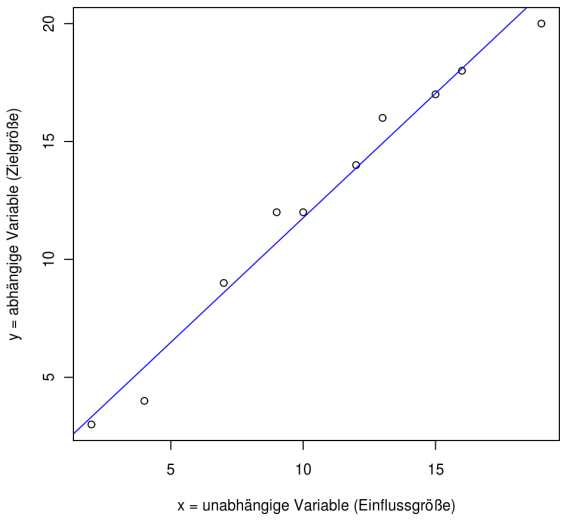


Abbildung 1.22.: Regressionsgeraden

Grundgedanke der Regressionsanalyse ist es, die Beziehung zwischen x und y mittels einer Geraden zu veranschaulichen. Natürlich könnte man diese Linie nach "Augenmaß" durch die Punktwolke ziehen, wobei das Ziel ist, dass alle Punkte möglichst nahe an der Geraden liegen. Das Problem hierbei besteht darin, dass zehn Personen wahrscheinlich zehn verschiedene Linien zeichnen würden. Es bleibt uns also nichts anderes übrig, als die "beste" Gerade rechnerisch zu ermitteln, eine Gerade, von der die Abstände der einzelnen Punkte am geringsten sind.[2]
Dazu benötigen wir die Gleichung einer geraden Linie:

$$y = a + b \cdot x$$

wobei "a" den Schnittpunkt der Geraden durch die y-Achse angibt und "b" als der Steigungskoeffizient, bzw. linearer Regressionskoeffizient bezeichnet wird.
Die allgemeinen Formeln zur Berechnung von "a" und "b" lauten:

$$a = \bar{y} - b \cdot \bar{x} \tag{1.18}$$

$$b = \frac{cov(x, y)}{s_x^2} \tag{1.19}$$

[2] Genaugenommen geht es um die quadrierten Abstände, da die Summe der Abstände ansonsten wieder "Null" ergibt. Die Bestimmung der Regressionsgeraden geschieht nach der Methode der kleinsten Quadrate.

Bleiben wir nochmals bei dem Beispiel der Schreib- und Lesefähigkeit. Nehmen wir an, es sei erwiesen, dass Lesen Einfluss auf die Rechtschreibung hat. Somit stellt sich das Lesen als die Einflussgröße und die Rechtschreibung als die Zielgröße oder abhängige Variable dar.
Es dürfte einleuchten, zunächst "b" zu berechnen, da diese Größe wiederum zur Bestimmung von "a" benötigt wird. Auf Seite 33 wurden bereits Mittelwerte und Varianz, sowie Kovarianz für das Beispiel bestimmt. so dass wir diese jetzt in die Formeln einsetzen können:

$$b = \frac{30,5}{28,9} = 1,055 \qquad \text{und} \qquad a = 12,5 - (1,055 \cdot 10,7) = 1,211$$

Die Regressionsgleichung für das Beispiel lautet demnach (gerundet):

$$y = 1.21 + 1,06 \cdot x$$

Setzt man jetzt einen beliebigen Punktwert für die Lesefähigkeit (x) ein , lässt sich der entsprechende Punktwert für die Rechtschreibung (y) voraussagen.

II. Interpolation und Extrapolation

Ein Vorteil der Regression besteht darin, dass für beliebige Wert x der passende y-Wert vorhergesagt werden kann. In Abbildung 1.23 sehen wir die Punkte einer Erhebung mit der dazugehörigen Regressionsgeraden. Zusätzlich ist jeweils beim Minimum und Maximum von x eine graue vertikale Hilfslinie in die Grafik eingezeichnet worden.

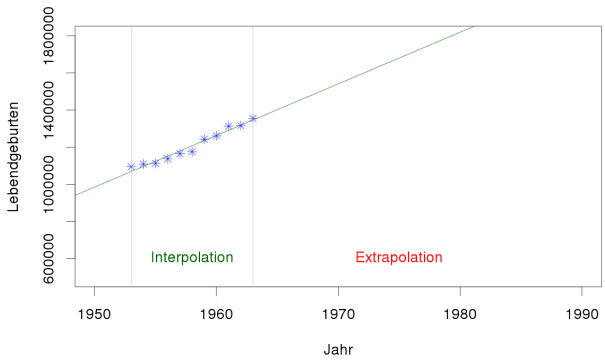


Abbildung 1.23.: Interpolation und Extrapolation

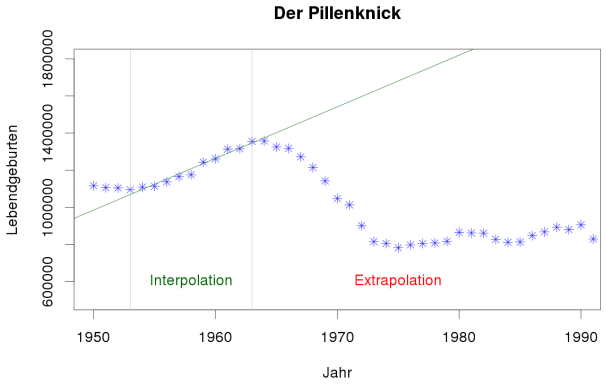


Abbildung 1.24.: Extrapolation: der Pillenknick

Macht man nun Vorhersagen innerhalb dieser Hilfslinien (das x, für welches wir einen y-Wert vorhersagen wollen, liegt also zwischen dem gemessenen Minimum und Maximum von x), so spricht man von `Interpolation`. Möchten man jedoch einen passenden y-Wert für ein x vorhersagen, welches außerhalb des gemessenen Minimum und Maximum von x liegt, so spricht man von `Extrapolation`. Dass Vorhersagen mittels Extrapolation sehr kritisch gesehen werden müssen zeigt Abbildung 1.24. Laut der Geraden müsste die Anzahl der Lebendgeburten (Y-Achse) im Jahr 1980 bei ca. 1.800.000 liegen. Tatsächlich wurden 1980 aber nur 865.789 Kinder geboren. Dieser Abfall der Geburtenrate ist als "Pillenknick" bekannt, da sich die Geburtenrate mit Einführung der Anti-Baby-Pille (1965) deutlich verringerte. Die mittels Extrapolation getroffene Vorhersage für 1980 hat sich in diesem Beispiel um 1.000.000 Babys verschätzt.

III. Die einfache lineare Regression

Soll - wie in unserem Beispiel zur Schreib- und Lesefähigkeit - eine Zielvariabel y durch *eine* Einflussvariabel x erklärt werden, spricht man von der *einfachen* linearen Regression.

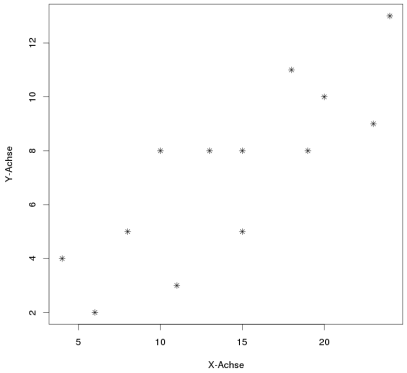


Abbildung 1.25.: Punktwolke

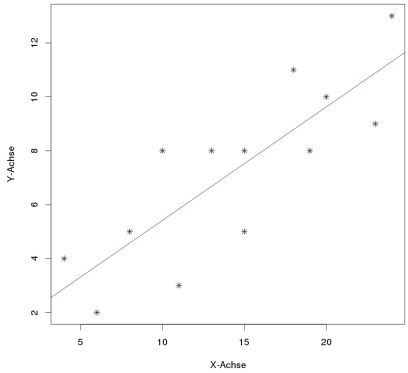


Abbildung 1.26.: Regressionsgerade

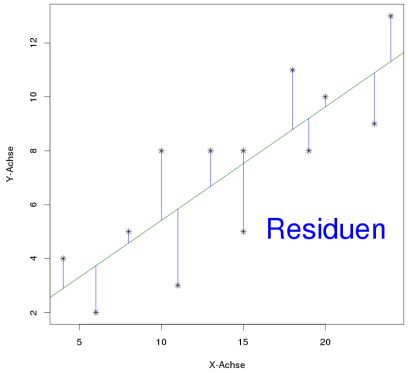


Abbildung 1.27.: Modell-Residuen

Abbildung 1.25 zeigt eine Punktwolke, die einen linearen Trend aufzuweisen scheint. Daher wurde in Abbildung 1.26 eine Regressionsgerade durch diese Punktwolke gezogen. Deutlich zu erkennen ist, dass die gemessenen Punkte nicht direkt auf der Geraden liegen, sondern durchaus nach oben oder unten abweichen können. Diese ‘‘Abweichler’’ nennt man `Residuen`. In Abbildung 1.27 wurden die Abstände der Residuen zur Regressionsgeraden als blaue Linien eingezeichnet. Wenn man die Länge dieser blauen Linien misst, quadriert und dann zusammenzählt, erhält man die so genannte `Summe der quadrierten Abweichungen`. Carl-Friederich Gauß entdeckte im Jahr 1801, dass die ‘‘perfekte’’ Regressionsgerade bestimmt werden kann, indem man die Gerade so anlegt, dass die Summe der quadrierten Abweichungen minimal ist. Man nennt dies auch die ‘‘Methode der kleinsten Quadrate’’ (engl.: least squares). Wie bereits erwähnt lautet der mathematische Ausdruck einer Geraden:

$$y = a + b \cdot x$$

In der Statistik hat sich folgende Schreibweise etabliert:

$$\hat{y} = \beta_0 + \beta_1 \cdot x \tag{1.20}$$

wobei β_0 den Schnittpunkt (engl. *Intercept*) der Geraden durch die y-Achse angibt, und β_1 als der Steigungskoeffizient bzw. linearer Regressionskoeffizient bezeichnet wird. Zur Erinnerung:

$\mathbf{y}$ = gemessene Werte

$\hat{\mathbf{y}}$ = Modellwerte (auch: *gefittete* Werte)

$\bar{\mathbf{y}}$ = arithmetisches Mittel

Während y (ohne Dach und ohne Querstrich) für die tatsächlich gemessenen Werte verwendet wird, bezeichnet $\hat{\mathbf{y}}$ jene Werte, die durch die Regressionsgerade für y vorhergesagt werden. Man spricht hier auch von *gefitteten* Werten oder *Modellwerte*.

1-9. Wahrscheinlichkeitsrechnung

Der Begriff der **Wahrscheinlichkeit** ist insbesondere in den Gesundheitswissenschaften von fundamentaler Bedeutung. Dies liegt z.T. daran, dass Experimente, anders als z.B. in der Physik, aus ethischen und praktikablen Gründen in der Regel nicht durchführbar sind. Aufgrund der Ungleichartigkeit der Versuchsobjekte - damit sind gewöhnlich Probanden oder Patienten gemeint - liegt es auf der Hand, dass die Ergebnisse normalerweise weder beliebig oft noch exakt reproduzierbar sind. Da "jeder Mensch anders ist", sind auch die Bemühungen die Versuchs- und Beobachtungsbedingungen konstant zu halten, zum Scheitern verurteilt.
Um die lebensrettende Wirkung von Antibiotika zu erkennen, brauchte man zu Flemmings Zeiten sicherlich keine Statistik. Bei den heute vorherrschenden Krankheiten oder Pflegeprobleme spielt die Statistik und damit auch der Wahrscheinlichkeitsbegriff aufgrund der oben erwähnten Variabilität eine entscheidende Rolle.
Auch im allgemeinen Sprachgebrauch ist das Wort "wahrscheinlich" gebräuchlich:

1. Morgen wird es wahrscheinlich regnen.
2. Wahrscheinlich wird das nächste Kind ein Mädchen.
3. Er wird höchst wahrscheinlich wieder gesund.
4. Die Wahrscheinlichkeit, dass das Neugeborene ein Jungen wird, beträgt etwa 50:50.
5. Die Wahrscheinlichkeit mit einem Würfel eine sechs zu Würfeln beträgt 1/6

An den Beispielen lassen sich unterschiedliche Wahrscheinlichkeiten unterscheiden: Subjektiv Empirisch Theoretisch

- subjektiv
- empirisch
- theoretisch

I. Subjektive Wahrscheinlichkeit

Wir erkennen, dass unter den in 1. bis 3. genannten Wahrscheinlichkeiten jeder etwas anderes versteht. Wir nennen sie daher subjektive Wahrscheinlichkeiten. Demgegenüber wird in 4. und 5. der Versuch einer Quantifizierung unternommen.

II. Empirische Wahrscheinlichkeit

Stellen wir beispielsweise fest, dass bei zehnmaligem Werfen einer Münze 6 mal Zahl erscheint, (von 1000 Neugeboren 488 Mädchen sind), sprechen wir von einer empirisch festgestellten Wahrscheinlichkeit für Zahl von 6/10 (bzw. 488/1000 für Mädchen). Mit anderen Worten wir verstehen unter dem Begriff der empirischen Wahrscheinlichkeit P für das Ereignis E (abgekürzt $P(E)$) die relative Häufigkeit mit der das Ereignis E auftritt

III. Theoretische Wahrscheinlichkeit

Unter der theoretischen bzw. statistischen Wahrscheinlichkeit $P(E)$ verstehen wir den Quotienten aus der Zahl der interessierenden (günstigen Fälle) und der Zahl aller möglichen Fälle. So bestimmt sich die Wahrscheinlichkeit mit einem idealen nicht gezinkten Würfel eine sechs zu Würfeln $P(6) = 1/6$

Anmerkung:

In der Praxis lässt sich die theoretische Wahrscheinlichkeit oft nicht bestimmen z.B.:

- a) Wahrscheinlichkeit dass ein Neugeborenes weiblich ist.
- b) Wahrscheinlichkeit dass ein Neugeborenes das erste Lebensjahr überlebt.

Übungsaufgaben:

1. Erläutern Sie den Unterschied zwischen der empirischen und theoretischen Wahrscheinlichkeit mit einem idealen Würfel eine drei zu Würfeln.
2. Werfen Sie 30- bis 50-mal eine Münze und bestimmen Sie die empirische Wahrscheinlichkeit für Zahl d.h. $P(Z)$ nach $n = 1; 3; 5; 10; 20; 30$ und gegebenenfalls 50 Würfen.

Definitionen:

$$\begin{aligned} \text{Für ein unmögliches Ereignis gilt:} \quad & P(E) = 0 \\ \text{Für ein sicheres Ereignis gilt:} \quad & P(E) = 1 \end{aligned}$$

Daraus folgt unmittelbar:
Die Wahrscheinlichkeit für das Eintreffen eines zufälligen Ereignis E liegt im Bereich zwischen 0 und 1, d.h.

$$0 \leq P(E) \leq 1$$

Bisher haben wir uns nur mit der Wahrscheinlichkeit **eines** zufälligen Ereignisses E beschäftigt. In der Praxis interessiert man sich aber oft für das Auftreten mehrerer verbundener Ereignisse: Wir unterscheiden hierbei drei Typen zusammengesetzter Ereignisse:

1. Wie groß ist die Wahrscheinlichkeit dass entweder Ereignis A **oder** Ereignis B eintritt, $P(AoderB)$?
2. Wie groß ist die Wahrscheinlichkeit dass Ereignis A **und** Ereignis B eintritt, $P(AundB)$?
3. Wie groß ist die Wahrscheinlichkeit dass Ereignis B eintritt sofern Ereignis A eingetreten ist, $P(B|A)$?

Wir wollen uns dies an einem Beispiel verdeutlichen:
Die (Erkrankungs)wahrscheinlichkeit für A betrage 40%, die für B betrage 30%.

1. Wie groß ist die Wahrscheinlichkeit daß eine Person erkrankt ist (d.h. mindestens an einer Krankheit leidet?
2. Wie groß ist die Wahrscheinlichkeit bei einer Person beide Erkrankungen vorzufinden?
3. Wie groß ist die Wahrscheinlichkeit an B zu erkranken, wenn die Person schon an A erkrankt ist?

Für die Bestimmung der entsprechenden Wahrscheinlichkeiten bedienen wir uns der **Additions-** und der **Multiplikationsregel**.

Additionsregel:

Sind A und B voneinander *unabhängige* Ereignisse mit den Wahrscheinlichkeiten $P(A)$ und $P(B)$ so gilt:

$$P(AoderB) = P(A) + P(B) - P(AundB)$$

wobei $P(AundB)$ die Wahrscheinlichkeit bezeichnet gleichzeitig an A und B erkrankt zu sein.

Multiplikationsregel:

Sind A und B voneinander *abhängige* Ereignisse mit den Wahrscheinlichkeiten $P(A)$ und $P(B)$ so gilt:

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

Anwendung

Setzen wir die Unabhängigkeit der beiden Erkrankungen voraus, so lassen sich die Fragen im obigen Beispiel beantworten:

Zu 3) Da die Ereignisse A und B als **unabhängig** vorausgesetzt werden, hängt die Wahrscheinlichkeit an B zu erkranken nicht davon ab, ob A vorliegt oder nicht, d.h. $P(B|A) = P(B) = 0.3$ **oder** 30%.

Zu 2) Die Wahrscheinlichkeit an A **und** B zu erkranken beträgt gemäß der Multiplikationsregel

$$P(A \text{ und } B) = P(A) \cdot P(B) = 0.4 \cdot 0.3 = 0.12|\text{oder } 2\%$$

Zu 1) Die Wahrscheinlichkeit an A oder B zu erkranken, d.h. das mindestens an einer der beiden Krankheiten zu erkranken beträgt nach der Additionsregel

$$P(A \text{ oder } B) = P(A) + P(B) - P(A \text{ und } B) = 0.4 + 0.3 - 0.12 = 0.58 \text{ oder } 58\%$$

Anmerkung:

Können die Ereignisse A und B niemals gleichzeitig auftreten, d.h. P(A und B) = 0, vereinfacht sich die Additionsregel zu

$$P(A \text{ oder } B) = P(A) + P(B)$$

Beispiel:

Die Wahrscheinlichkeit mit einem Würfel eine 6 oder eine 1 zu Würfel ist demnach
$P(6 \text{ oder } 1) = P(6) + P(1) = 1/6 + 1/6$,
da $P(6 \text{ und } 1)$ bei einem einmaligem Wurf immer Null ist.

1-10. Odds

Als *Odd* wird die Chance bezeichnet, dass ein Ereignis ein- bzw. zutrifft. Man berechnet Odds, indem man die Wahrscheinlichkeit, dass ein Ereignis eintrifft, durch die Wahrscheinlichkeit, dass dieses Ereignis nicht eintrifft, dividiert.

$$Odd = \frac{p}{1-p} \tag{1.21}$$

Nehmen wir einen sechs-seitigen Würfel als Beispiel. Uns interessiert nun, wie hoch die Chance ist eine 6 zu würfeln. Die Wahrscheinlichkeit eine 6 zu würfeln liegt bei $p = \frac{1}{6}$. In die Formel 1.21 eingesetzt rechnen wir:

$$Odd_{W=6} = \frac{p}{1-p} = \frac{\frac{1}{6}}{1-\frac{1}{6}} = \frac{\frac{1}{6}}{\frac{5}{6}} = \frac{1}{6} \cdot \frac{6}{5} = \mathbf{\frac{1}{5}}$$

Die Chance eine 6 zu würfeln liegt also bei $\frac{1}{5}$.
Dieses Beispiel lässt sich noch leicht ohne rechnen lösen, da ein Würfel typischerweise nur auf einer seiner Seiten die Zahl 6 zeigt. Auf den übrigen fünf Seiten sind entsprechend die Zahlen von 1 bis 5 abgebildet. Wenn man Odds so versteht, dass die *Anzahl der Möglichkeiten ein Ereignis zu erreichen* (in diesem Falle = 1, da es nur eine 6 auf dem Würfel gibt) ins Verhältnis gesetzt wird zu der *Anzahl der Möglichkeiten das Ereignis nicht zu erreichen* (in diesem Falle = 5, da es fünf Würfelseiten gibt, die keine 6 zeigen), ergibt sich auch hier die Chance von $\frac{1}{5}$.

I. Odds und Wahrscheinlichkeiten

Odds sind nicht das selbe wie Wahrscheinlichkeiten! Am Beispiel des Würfels lässt sich sagen:

- Die *Wahrscheinlichkeit* eine 6 zu würfeln beträgt $p_{w=6} = \mathbf{\frac{1}{6}}$
- Die *Chance* eine 6 zu würfeln beträgt $Odd_{w=6} = \mathbf{\frac{1}{5}}$

Jedoch lassen sich beide Werte in das entsprechende Gegenstück transformieren:

- Wahrscheinlichkeiten in Odds umrechnen: $Odds = \frac{p}{1-p}$
- Odds in Wahrscheinlichkeiten umrechnen: $p = \frac{Odds}{Odds+1}$

II. Odds Ratio

Die so genannte Odds Ratio (OR) stellt die Odds zweier Gruppen in Relation. Hierbei werden die Odds einer Vergleichsgruppe durch die Odds einer Referenzgruppe dividiert.

$$Odds\ Ratio = \frac{Odds_{Vergleichsgruppe}}{Odds_{Referenzgruppe}} \tag{1.22}$$

Die Odds Ratio kann Werte zwischen 0 und ∞ (unendlich) annehmen.

Die Berechnung der Odds Ratio kann in vielen Fällen auch über eine $2 \cdot 2$-Felder-Tafel erfolgen.

	mit Risiko	**ohne Risiko**
erkrankt	a	b
nicht erkrankt	c	d

Tabelle 1.28.: Odds Ratio $2 \cdot 2$-Felder-Tafel

Die allgemeine Formel zur Berechnung der OR nach Tabelle 1.28 lautet:

$$OR = \frac{a \cdot d}{b \cdot c} \tag{1.23}$$

Tabelle 1.29 zeigt die Ergebnisse einer ausgedachten Untersuchung zur Wirksamkeit von Impfungen.

	nicht geimpft	**geimpft**
erkrankt	231	17
nicht erkrankt	34	534

Tabelle 1.29.: $2 \cdot 2$-Felder-Tafel zur Impfung

Setzt man diese Werte in Formel 1.23 ein, ergibt sich:

$$OR = \frac{a \cdot d}{b \cdot c} = \frac{231 \cdot 534}{17 \cdot 34} = \frac{123.354}{578} = \mathbf{213,41}$$

Das bedeutet, dass `nicht-geimpfte` Personen ein mehr als 213-fach erhöhtes Risiko tragen zu erkranken als `geimpfte` Personen.

Es ist wichtig zu beachten, dass bei der $2 \cdot 2$-Felder-Methode die OR aus Sicht der Gruppe aus Spalte 1 angegeben wird. Im Gegensatz dazu sind ORs aus Regressionsanalysen immer aus Sicht der Vergleichsgruppe (Gruppe + 1) zu interpretieren.

III. Interpretation

Die Interpretation der Odds Ratio folgt grob gesprochen der Logik:

$OR > 1$ $\hat{=}$ höhere Chance in `Spalte 1` bzw. in der `Vergleichsgruppe`

$OR = 1$ $\hat{=}$ kein Unterschied

$OR < 1$ $\hat{=}$ geringere Chance in `Spalte 1` bzw. in der `Vergleichsgruppe`

Eine Odds Ratio von (annähernd) 1 besagt, dass es keinen Unterschied zwischen den untersuchten Gruppen gibt.

Für Werte *größer* als 1 gilt, dass die Vergleichsgruppe eine höhere Chance (bzw. ein höheres Risiko) aufweist als die Ausgangsgruppe. In einer beispielhaften Untersuchung wurde mittels logistischer Regression berechnet, inwieweit das Geschlecht einen Einfluss auf die Berufung in den Aufsichtsrat eines Unternehmens hat. Hierbei wurde das Geschlecht `weiblich` als

Ausgangsgruppe gegen das Geschlecht `männlich` als Vergleichsgruppe getestet. Die resultierende Odds Ratio lag bei $1,58$.

$$OR_{Geschlecht} = \frac{Odds_{maennlich}}{Odds_{weiblich}} = 1,58$$

Das bedeutet, dass Männer eine 58% *größere* Chance haben in den Aufsichtsrat berufen zu werden als Frauen. Es heisst aber *nicht*, dass die Chance bei 58% liegt - sie ist bei Männern um 58% *größer*.

Werte *kleiner* als 1 sind da schon schwieriger zu interpretieren. In einer beispielhaften Untersuchung wurde mittels logistischer Regression berechnet, inwieweit das Rauchverhalten einen Einfluss auf die Bildung eines Bronchialkarzinoms aufweist. Hierbei wurden `Raucher` als Ausgangsgruppe gegen `Nicht-Raucher` als Vergleichsgruppe getestet. Die resultierende Odds Ratio lag bei 0.232.

$$OR_{Karzinom} = \frac{Odds_{Nicht-Raucher}}{Odds_{Raucher}} = 0,232$$

Hier kann man lediglich sagen, dass die Chance (bzw. das Risiko) ein Bronchialkarzinom zu entwickeln bei `Nicht-Rauchern` *kleiner* ist als bei `Rauchern`. Um welchen Faktor dieses Risiko nun wirklich kleiner ist, lässt sich aus der OR nicht direkt ablesen. Der *Trick* besteht darin, die Daten so umzuschreiben, dass Referenzgruppe und Vergleichsgruppe vertauscht werden:

$$OR_{Karzinom} = \frac{Odds_{Raucher}}{Odds_{Nicht-Raucher}}$$

In unserem ausgedachten Beispiel ergibt sich so eine Odds Ratio von:

$$OR_{Karzinom} = \frac{Odds_{Raucher}}{Odds_{Nicht-Raucher}} = 4,31$$

Jetzt kann man ablesen, dass `Raucher` ein mehr als 4-fach erhöhtes Risiko haben ein Bronchialkarzinom auszubilden als `Nicht-Raucher`.

2. Schließende Statistik

2-1. Einführung

In Kapitel 1 haben wir uns mit der *Beschreibenden Statistik* befaßt. Unsere Stichprobe wurde mittels Häufigkeitstabellen, Graphiken, Lagekenngrößen und Streuungsmaßen beschrieben und - je nach Fragestellung und Datenniveau - Korrelationskoeffizienten oder Regressionsberechnungen durchgeführt. Es dürfte jedem klar sein, dass unsere berechneten Werte zunächst einmal nur für die untersuchte Stichprobe gelten.
In der Wahrscheinlichkeitsrechnung wurde verdeutlicht, dass es möglich ist, bei bekannter Grundgesamtheit, bzw. bei bekannter Verteilungsform eines Merkmals (ob es sich z.B. um eine Normal-, Binomial-, Hypergeometrische- oder Poissonverteilung handelt) etwas darüber auszusagen, wie eine Zufallsstichprobe aus dieser Grundgesamtheit aussehen wird. Umgekehrt ist es möglich, bei bekannter Zusammensetzung der Stichprobe Rückschlüsse auf die zugrundeliegende Grundgesamtheit zu ziehen. Genau dies ist Inhalt der *Schließenden Statistik*.
Zwar liefert die schließende Statistik keine Verfahren, die es ermöglichen, mit 100%iger Sicherheit von der Stichprobe auf die Grundgesamtheit zu schließen, aber ihre Verfahren ermöglichen es, den *Grad der Unsicherheit* zu bestimmen. Wir unterscheiden dabei zwei Schwerpunkte:

- Methoden zum Schätzen unbekannter Größen
- Statistische Tests zum Prüfen von Hypothesen

I. Das Schätzen unbekannter Größen: Punkt- und Intervallschätzungen

Nehmen wir an, die durchschnittliche Verweildauer in einer Stichprobe von 25 per Zufall ausgewählter Patientinnen des Krankenhauses A beträgt 9 Tage. Jedem wird es einleuchten, dass man jetzt nicht behaupten kann, dies sei die durchschnittliche Liegezeit für das gesamte Krankenhaus. In einer nächsten Stichprobe sind es vielleicht 10 Tage, in einer weiteren eventuell 11 Tage. Hiermit soll ausgedrückt werden, dass ein einzelner Befund einer Stichprobe nur ein ungefährer Wert, also ein **Punktschätzwert** für jenen unbekannten Wert sein kann, den man erhalten würden wenn man alle Patienten des Krankenhauses erfassen würde. Wo liegt nun dieser unbekannte **wahre Wert**? Es ist also unser Anliegen, den Punktschätzwert näher zu präzisieren.
Unter der Voraussetzung, dass es sich tatsächlich um Zufallsstichproben (des gleichen Umfanges) handelt, kann davon ausgegangen werden, dass einige Stichprobenmittelwerte etwas unterhalb und andere wiederum etwas oberhalb des wahren Wertes liegen. Die Wahrscheinlichkeit für extreme Abweichungen ist relativ gering. Theoretisch (siehe Wahrscheinlichkeitsrechnung) ist genau bekannt, wie sich die Werte einer Zufallsstichprobe um den wahren Wert herum verteilen. Demzufolge kann man fragen:

> *Wie weit kann der unbekannte wahre Wert von dem Schätzwert der Stichprobe mit einer großen Wahrscheinlichkeit höchstens entfernt sein?*

Wir wollen die Überlegungen dazu an einem Beispiel mit normalverteilten Werten erläutern.
Wiederholen wir kurz unser Vorgehen aus der Wahrscheinlichkeitsrechnung, als es um die Bestimmung von Flächen (Wahrscheinlichkeiten) innerhalb einer Normalverteilung ging. Als

fiktives Beispiel wollen wir davon ausgehen, dass von unserem Beispiel-Krankenhaus die durchschnittliche Liegedauer ermittelt worden ist und uns somit der wahre Wert der Grundgesamtheit bekannt sei. Nehmen wir an, die Liegedauer sei normalverteilt mit einem Mittelwert μ von 10 Tagen.
Ausgehend von den Überlegungen zur Mittelwerte-Verteilung können wir dann behaupten, dass sich der Mittelwert einer beliebigen Zufallsstichprobe des Umfangs n mit einer Wahrscheinlichkeit von 95,44% im Bereich $\mu \pm 2 \cdot \sigma_{\bar{x}}$ befindet.
Angenommen, in unserem Beispiel sei $\sigma = 2$ Tage, dann lautet dieser Bereich 6 bis 14 Tage. Dies bedeutet anders ausgedrückt, dass sich für 95,44% der Stichproben eine mittlere Verweildauerdauer ergibt, die zwischen 6 und 14 Tagen liegt. Umgekehrt heißt dies aber auch, es gibt lediglich bei 2,28% der Stichproben Verweildauern von weniger als 6 Tagen und ebenfalls bei 2,28% die länger als 14 Tage angaben. Drückt man dies als Wahrscheinlichkeitsaussage aus, so beträgt die Wahrscheinlichkeit bei einer zufällig gezogenen Stichprobe eine Liegedauer von weniger als 8 Tage zu finden, ganze 2,28%. Dieser Sachverhalt wird in Abbildung 2.1 nochmals verdeutlicht.
Wenn wir also per Zufall eine Stichprobe aus der Grundgesamtheit (hier das Krankenhaus) auswählen, ist die Wahrscheinlich sehr hoch (95,44%), eine mittlere Liegezeit zwischen 6 und 14 Tagen zu finden.

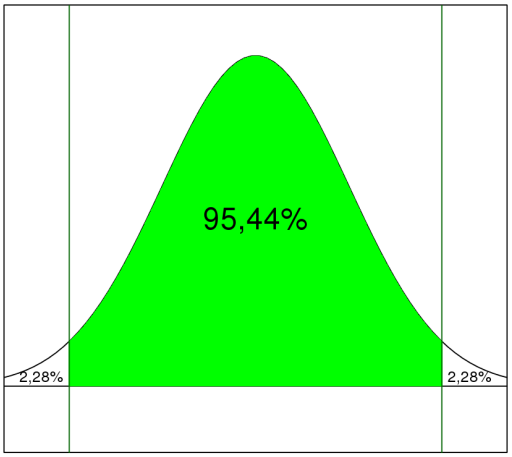


Abbildung 2.1.: Bereich von 2 Standardabweichungen

Anders herum könnte man auch sagen, dass sich aus den erhobenen Daten unseres Krankenhauses ein Wertebereich ergibt, in dem mit großer Wahrscheinlichkeit bestimmte Stichprobenmittelwerte zu finden sind. Nehmen wir jetzt an, die durchschnittliche Liegedauer betrüge 12 Tage mit einem Standardfehler von 2 Tagen. Daraus ergibt sich ein Bereich von 8 bis 16 Tagen.
Auch hier findet sich unser Stichprobenmittelwert von 9 Tagen innerhalb dieser Grenzen. D.h., eine mittlere Liegedauer von 12 Tagen ist mit unserem Stichprobenmittelwert vereinbar. Bei einer durchschnittlichen Liegedauer von 14 Tagen lautet der 95,44%-Bereich 10 bis 18 Tage. Jetzt liegt unser Stichprobenmittelwert von 9 Tagen jetzt nicht mehr innerhalb dieser Grenzen. Dies bedeutet, betrüge der tatsächliche Mittelwert (den wir normalerweise ja nicht kennen) 14 Tage, wäre es sehr unwahrscheinlich (weniger als 2,28%), einen Stichprobenmittelwert von 9 Tagen zu ermitteln.
Anders ausgedrückt heißt dies, dass bestimmte Mittelwerte der Grundgesamtheit als „Erzeuger" des Stichprobenmittelwertes von 9 Tagen als sehr unwahrscheinlich ausscheiden. Verdeutlichen wir dies mit einigen Mittelwerten der Grundgesamtheit bei gleichbleibendem Standardfehler von 2 Tagen und deren möglichen Wertebereichen (siehe Tabelle 2.1).
In der Tabelle 2.1 sind jene Populationsmittelwerte grau hinterlegt, bei denen die Wahrscheinlichkeit sehr gering ist, einen Stichprobenmittelwert von 9 Tagen zu „produzieren".
Setzen wir in die Formel, mit der wir den 95,44%-Werte-Bereich für den Mittelwert der Grundgesamtheit bestimmt haben jetzt einfach einmal unseren Stichprobenmittelwert von 9 ein:

$$\bar{x} \pm 2 \cdot \sigma_{\bar{x}} = 9 \pm 2 \cdot 2$$

μ der Grundgesamtheit	Bereich $\mu \pm 2 \cdot \sigma_{\bar{x}}$
4	0 bis 8
5	1 bis 9
6	2 bis 10
7	3 bis 11
8	4 bis 12
9	5 bis 13
10	6 bis 14
11	7 bis 15
12	8 bis 16
13	9 bis 17
14	10 bis 18

Tabelle 2.1.: Beispiel von möglichen 95,44% Wertebereichen

dann ergibt sich ein 95,44%-Werte-Bereich von 5 bis 13 Tagen. Also genau jener Bereich, den wir unter der Annahme, dass die Verhältnisse in der Grundgesamtheit und damit der wahre Wert μ bekannt sei, ermittelten. Die 95,44% bedeuten nicht, dass mit dieser Wahrscheinlichkeit der wahre Wert der Grundgesamtheit innerhalb dieses Bereichs liegt. Denn schließlich liegt der wahre Wert entweder innerhalb oder außerhalb. Es bedeutet lediglich, dass theoretisch in 95,44% der Stichproben Bereiche ermittelt werden können, die tatsächlich den wahren Wert umschließen.
Tatsache ist, dass unsere konkrete Stichprobe allerdings auch zu den 4,46% gehören könnte, die „daneben liegen".
Ganz vereinfacht ist dies der eingangs erwähnte Sachverhalt: Bei bekannter Grundgesamtheit sind Aussagen über die Stichprobe möglich und im Umkehrschluß lässt eine Stichprobe Aussagen über die zugrundeliegende Population zu.
Nachfolgend sollen einige Begriffe verdeutlicht werden. Im Beispiel der Liegedauer war der Mittelwert $\bar{x}$ der Zufallsstichprobe ein **Punktschätzwert** für den wahren Wert μ der Grundgesamtheit (das gesamte Krankenhaus). Weitere Punktschätzwerte, die wir aus Zufallsstichproben ermitteln können, sind z.B. Anteile ($\hat{p}$), Differenzen von Anteilen ($\hat{d}$), Korrelationskoeffizienten ($\hat{r}$), etc. Um diese Punktschätzwerte vom wahren Wert zu unterscheiden, versehen wir sie mit einem „Dach": $\hat{}$
Es ist also möglich, ausgehend von Punktschätzwerten einer Zufallsstichprobe mit einer hohen Wahrscheinlichkeit (im folgenden sofern nichts anderes gesagt 95%) einen Bereich abzuschätzen, in dem der wahre Wert liegt. Bereiche werden üblicherweise mit Grenzen abgesteckt. Wie in Tabelle 2.1 dargestellt, benötigen wir die Angabe einer unteren und einer oberen Grenze, da der wahre Wert sowohl unterhalb als auch oberhalb unseres einzelnen Punktschätzwertes liegen kann.
Diese Grenzen nennt man **Konfidenzgrenzen** oder Vertrauensgrenzen (im englischen confidence limits), die mit einer bestimmten statistischen Wahrscheinlichkeit den wahren Wert einschließen. Üblicherweise geht man von 95% oder 99% Wahrscheinlichkeit aus und bezeichnet dies als **Konfidenzniveau**. Das Intervall zwischen den Konfidenzgrenzen bezeichnet man als **Konfidenzintervall**.
Die Wahrscheinlichkeit, dass der ermittelte Bereich den wahren Wert **nicht** einschließt, wird als **Irrtumswahrscheinlichkeit** α (alpha) bezeichnet. Geläufig sind hierbei Werte von 5% oder 1%. Die Irrtumswahrscheinlichkeit liegt üblicherweise je zur Hälfte oberhalb und unterhalb der Konfidenzgrenzen, was mit dem Begriff „zweiseitig" ausgedrückt wird. Das Konfidenzniveau

ergibt sich, indem man von der Gesamtwahrscheinlichkeit die Irrtumswahrscheinlichkeit abzieht $(1-\alpha)$. Abbildung 2.2 soll die Begriffe verdeutlichen.

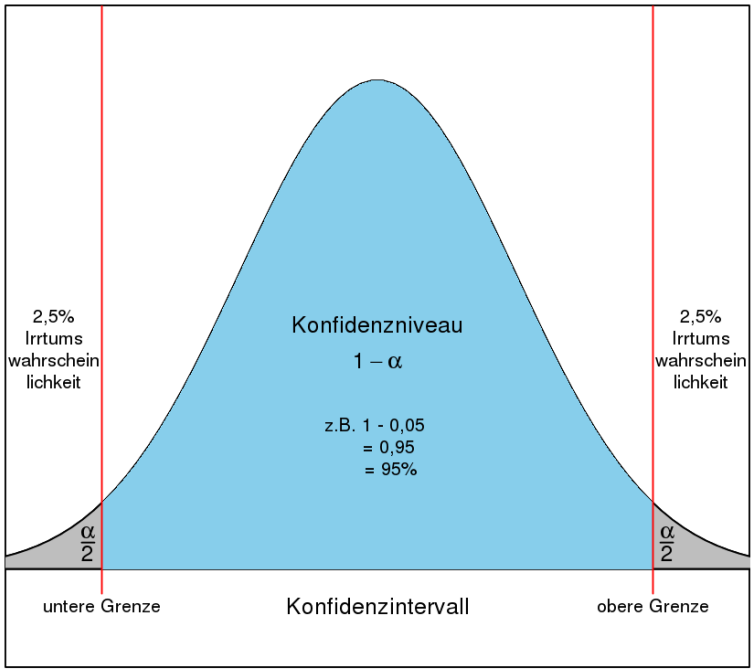


Abbildung 2.2.: Einige Begriffe zur Intervallschätzung

Im Rahmen dieses Skripts beschränken wir uns auf die Bestimmung von Konfidenzgrenzen für Mittelwerte (Normalverteilungen), Anteile, sowie Differenzen von Anteilen (Binomialverteilungen).

2-2. Konfidenzgrenzen bei Normalverteilungen

Die wesentlichen Überlegungen zur Bestimmung von Konfidenzgrenzen bei Normalverteilung sind am Beispiel der Liegedauer in unserem Beispiel-Krankenhaus bereits besprochen worden. Zieht man sehr viele Zufallsstichproben aus einer Grundgesamtheit, so ist die Wahrscheinlichkeit für das Auftreten von Stichprobenmittelwerten, die extrem weit vom wahren Wert entfernt liegen, gering. Viel wahrscheinlicher dagegen ist es, einen Wert zu ermitteln, der innerhalb des Bereichs liegt, in dem 95% oder 99% aller Stichprobenmittelwerte liegen.
Bestimmt man für die Stichproben-Mittelwerte wiederum 95% oder 99% Wertebereiche, kann man im Umkehrschluss davon ausgehen, dass diese Bereiche (Intervalle) mit hoher Wahrscheinlichkeit den wahren beinhalten (einschließen).
Zur Bestimmung der Konfidenzintervalle legt man zunächst die Irrtumswahrscheinlichkeit α fest. Geläufig sind Angaben von $\alpha = 0,05$ und $\alpha = 0,01$. Daraus ergibt sich das Konfidenzniveau $(1-\alpha)$, zu 0,95 und 0,99 bzw. zu 95% und 99%. Wir erinnern uns, dass in der Mittelwertverteilung im Bereich von 2 Standardfehlern um μ 95,44% der Werte liegen. Der 95%-Bereich einer Mittelwertverteilung (symmetrisch um den Mittelwert) wird von den z-Werten +1,96 und -1,96 begrenzt.

Dies bedeutet, dass im Bereich von 1,96 Standardfehlern um μ 95% der Stichprobenmittelwerte liegen.
Der Vollständigkeit halber soll noch angemerkt werden, dass für den 99%-Bereich ein z-Wert von 2,576 gilt. Mit anderen Worten liegen im Bereich von 2,576 Standardfehlern rechts und links vom Mittelwert 99% der Werte. Natürlich lässt sich jeder gewünschte Wertebereich (90%, 80%, etc.) ermitteln. Es sei daran erinnert, dass zur Ermittlung des entsprechenden z-Wertes die gesamte Fläche links von dem Wert berücksichtigt werden muß. Dies wird in der Abbildung 2.3.

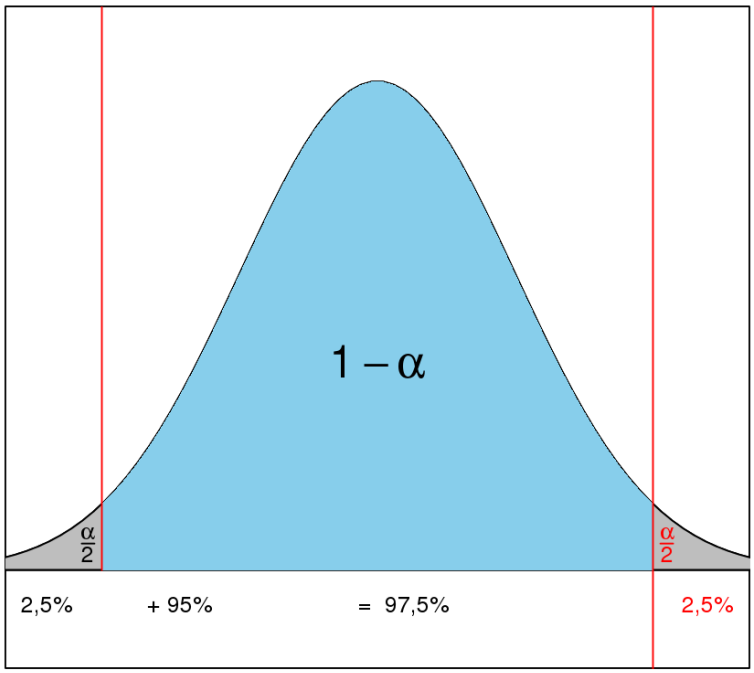


Abbildung 2.3.: Konfidenzintervall und z-Wert

Suchen wir beispielsweise einen z- Wert zur Ermittlung der oberen Grenze des 95%-Konfidenzintervalls (d.h. α entspricht 5%), so ist dies der z-Wert, der 97,5% der Verteilung „abschneidet". Er ergibt sich aus den Flächen

$$\frac{\alpha}{2} + 1 - \alpha \qquad \text{oder} \qquad 1 - \frac{\alpha}{2}$$

$$\text{(d.h. hier} \quad 1 - \frac{0,05}{2} = 0,975)$$

Aus Referenztabellen (z.B. Tabelle B.2) lässt sich dann ersehen, dass $z_{0,975} = 1,96$ beträgt.
Kommen wir jetzt zur Ermittlung der Konfidenzintervalle, die nach der folgenden Formel berechnet werden:

$$\left.\begin{matrix}\mu_o \\ \mu_u\end{matrix}\right\} \bar{x} \pm z \cdot \sigma_{\bar{x}} \tag{2.1}$$

μ_o, μ_u = die obere und untere Grenze des Intervalls
$\bar{x}$ = Mittelwert der Stichprobe
z = Wert, der die entsprechende Fläche einer Normalverteilung ''abschneidet''
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ = Standardfehler des Mittelwerts

Wir verdeutlichen dies nochmals an einer Beispielaufgabe: Eine Zufalls-Erhebung über die jährlichen Ausgaben für Pflegehilfsmittel ergab bei 100 Befragten ($n = 100$) einen Mittelwert von 400,-€. Bekannt sei die Standardabweichung der Grundgesamtheit $\sigma = 20$,-€. Bestimme das 95%- Konfidenzintervall für den wahren Wert μ. Der entsprechende z-Wert für den 95%-Bereich lautet 1,96.

$$95\%\text{-Konfidenzintervall} = 400 \pm (1,96 \cdot \frac{20}{\sqrt{100}}) = 400 \pm 3,92$$

Das 95%-Konfidenzinterval liegt zwischen 396,08€ und 403,92€. Eigentlich gestaltet sich die Bestimmung von Konfidenzintervallen rein rechnerisch nicht sehr schwierig. Allerdings gibt es das Problem, dass in den meisten Fällen nichts über die Standardabweichung der Grundgesamtheit bekannt ist. Dann wird mittels der Standardabweichung s der Stichprobe eine Punktschätzung für die Standardabweichung σ der Grundgesamtheit vorgenommen. Man geht dann allerdings davon aus, dass der geschätzte Standardfehler $s_{\bar{x}}$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad \text{häufig findet sich die Schreibweise} \qquad \hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}}$$

nicht einer Normalverteilung folgt, sondern einer t-Verteilung. Für die Bestimmung der Konfidenzintervalle bedeutet dies, das statt des z-Wertes der entsprechende t-Wert mit $n - 1$ Freiheitsgraden (z.B. Lorenz[2]) eingesetzt werden muß.

I. Exkurs: Freiheitsgrade (nach Bortz)

Freiheitsgrade bezeichnen die Anzahl der ''frei wählbaren'' Beobachtungen (Messungen). Beispiel: Eine Schulklasse mit 30 SchülerInnen hat einen Notenschnitt von 2, 9. Nun kann man bei 29 SchülerInnen die Noten frei variieren – doch beim jeweils Letzten ist die Note vorgegeben, wenn der Mittelwert bei 2, 9 liegen soll. Das arithmetische Mittel hat also $n - 1$ Freiheitsgrade.
Für die Berechnung des Standardfehlers benötigen wir die Varianz s^2 in der die Summe der Abweichungsquadrate zunächst einmal - nämlich vor der Quadrierung - Null ergeben hätte

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

Ergeben sich beispielsweise bei einer Stichprobe von $n = 5$ die ersten vier Abweichungen mit $x_1 - \bar{x} = -5, x_2 - \bar{x} = +3, x_3 - \bar{x} = +1$ und $x_4 - \bar{x} = 2$, dann muss $x_5 - \bar{x} = -1$ sein, damit die Summe der Abweichungen wieder Null ergibt:
$-5 + 3 + 1 + 2 - 1 = 0$
Bei einer Stichprobe von $n = 5$ ist eine der Abweichungen festgelegt. Die Varianz hat vier $(n - 1)$ Freiheitsgrade. Freiheitsgrade kürzt man häufig mit v (gesprochen "nü") oder mit *df* (englisch: degrees of freedom) ab.
Kehren wir zu unserem Beispiel der durchschnittlichen Ausgaben für Pflegehilfsmittel zurück. Aus der Stichrobe von $n = 100$ errechnete sich ein Stichprobenmittelwert von $\bar{x} = 400$ und eine Standardabweichung $s = 20$. Da jetzt σ unbekannt ist, schätzen wir den Standardfehler mit der Standardabweichung aus der Stichprobe.

Die Formel zur Bestimmung des Konfidenzintervalls ändert sich dann wie folgt:

$$\left.\begin{matrix}\mu_o \\ \mu_u\end{matrix}\right\} = \bar{x} \pm t \cdot s_{\bar{x}} \tag{2.2}$$

μ_o, μ_u = die obere und untere Grenze des Intervalls
$\bar{x}$ = Mittelwert der Stichprobe
t = Wert, der die entsprechende Fläche einer t-Verteilung ‘‘abschneidet’’
$s_{\bar{x}} = \frac{s}{\sqrt{n}}$ = Standardfehler des Mittelwerts

Setzen wir unsere Werte jetzt in die Formel ein: $400 \pm t \cdot \frac{20}{\sqrt{100}}$

und ermitteln wir laut Tabelle (z.B. Tabelle B.3) einen Wert für $t_{\substack{\alpha=0,05 \\ v=99}} = 1,985$ (gemittelt),

dann lautet das 95%-Konfidenzintervall $= 400 \pm 1,985 \cdot 2 = 400 \pm 3,97$

Sicherlich fällt auf, dass der z-Wert (1,96) und der t-Wert (1,985) sich in unserem Beispiel nur unwesentlich unterscheiden. Wie in der Wahrscheinlichkeitsrechnung angesprochen, nähern sich einige Verteilungen (z.B. die Binomialverteilung) bei großen Stichproben einer Normalverteilung an. Dies gilt auch für die t-Verteilung. Hätte der Umfang der Stichprobe beispielsweise $n = 5$ betragen, dann wäre der t-Wert mit 2,776 deutlich größer als der entsprechende z-Wert in der Normalverteilung und damit auch das Intervall breiter.

2-3. Schärfe und Sicherheit von Konfidenzaussagen

Die Länge eines Konfidenzintervalls - also der Abstand zwischen der oberen und unteren Grenze des Intervalls, ist ein Maß für die Präzision oder die **Schärfe** einer Konfidenzaussage. Das Konfidenzniveau, z.B. 95%, gibt die **Sicherheit** der Aussage an. Im Beispiel der durchschnittlichen Aufwendungen für Pflegehilfsmittel (Seite 50) lag das 95%-Konfidenzintervall zwischen 396,08,-€ und 403,92,-€. Es hatte demnach eine Länge von 7,84,-€, was sicherlich als eine recht präzise Aussage gewertet werden kann (zumindest wenn es - wie in unserem Bespiel um mehrere 100€ geht). Nehmen wir an, uns reicht die Sicherheit von 95% nicht aus und wir hätten gerne eine Aussage mit 99%iger Sicherheit.
Dann ergibt sich laut der Formel 2.2 von Seite 51 für den Mittelwert der Stichprobe von 400,-€ mit bekannter Standardabweichung der Grundgesamtheit von 20,-€ bei $n = 100$ das folgende Intervall:

$$400 \pm (2,576 \cdot \frac{20}{\sqrt{100}}) = 400 \pm 5,15$$

Die untere Konfidenzgrenze lautet jetzt 394,85€, die obere 405,15€. Das heisst die Intervalllänge hat sich auf 10,30€ vergrößert. Anders ausgedrückt ist das Intervall hier bei höherer Sicherheit, nämlich 99%, breiter und damit unschärfer geworden. Es sieht also so aus, dass eine schärfere Konfidenzaussage zu Lasten der Sicherheit geht und umgekehrt eine höhere Sicherheit der Aussage diese wiederum unschärfer macht. Dies gilt zumindest solange die Stichprobengröße n konstant bleibt.
Die Breite und damit die Schärfe einer Konfidenzaussage wird von verschiedenen Größen beeinflusst:

- die Irrtumswahrscheinlichkeit α, die direkt den z-Wert (bzw. t-Wert) beeinflusst. Ein großes α bedeutet niedriges Konfidenzniveau (niedrige Sicherheit) und bewirkt ein schmaleres (schärferes) Intervall und umgekehrt bedeutet ein kleines α ein höheres Konfidenzniveau (höhere Sicherheit) mit einem breiterem Intervall.

- die Stichprobengröße n, die (neben der Bestimmung des z- bzw. des t-Wertes) zur Berechnung des Standardfehlers der Mittelwertverteilung benötigt wird. Betrachtet man nochmals die Formel, müßte deutlich sein, dass ein „großes n" den Wert des Standardfehlers (und damit auch die Intervallbreite) verkleinert und umgekehrt (und zwar quadratisch, siehe unten).

- die Variabilität des Merkmals in der Stichprobe, d.h. die Standardabweichung der Grundgesamtheit, bzw. der Stichprobe. Ein hoher Wert oberhalb des Bruchstriches wird den Standardfehler (und damit auch das Intervall) vergrößern und umgekehrt.

Möchten wir eine sichere (hohes Konfidenzniveau) und gleichzeitig präzise (schmales Intervall) Konfidenzaussage, so geht dies **nur** über den Stichprobenumfang.

> **Die Halbierung eines Konfidenzintervalls macht einen vierfachen Stichprobenumfang erforderlich!**

2-4. Konfidenzgrenzen für Anteile (bei Binomialverteilungen)

Die Befragung über den Aufwand für Pflegehilfsmittel hatte neben den durchschnittlichen Ausgaben noch ergeben, dass insgesamt 25% der Befragten Zuzahlungen leisten müssen. Dieser Anteil von 25% ist wiederum der Punktschätzwert ($\hat{p}$) für den wahren Anteil (p) in der Grundgesamtheit, für den wir ein Konfidenzintervall mit p_u und p_o schätzen möchten. Für binomialverteilte Anteile gibt es verschiedene Methoden zur Bestimmung von Konfidenzintervallen. Wir werden uns hier auf die Beschreibung der graphischen Methode beschränken. Für die quantitative Bestimmung sei auf Lorenz[2] verwiesen.

I. Die graphische Bestimmung von Konfidenzgrenzen für binomialverteilte Anteile

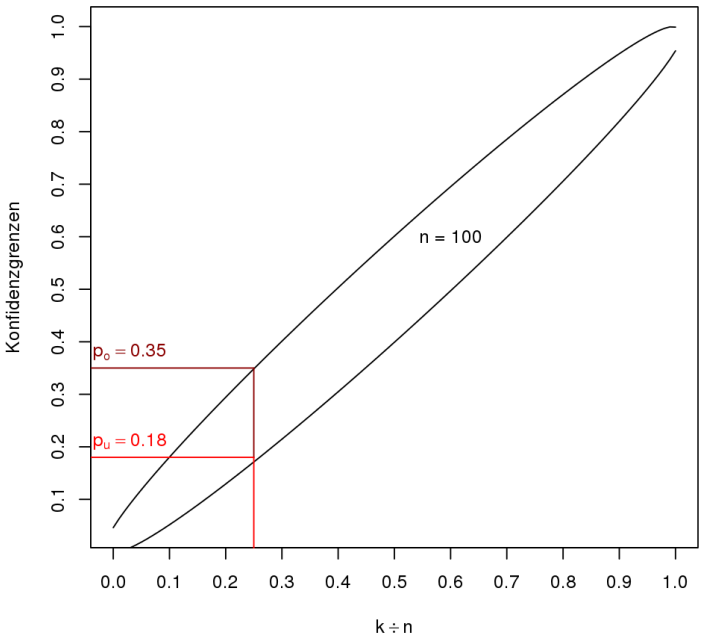


Abbildung 2.4.: Konfidenzgrenzen mittels Nomogramm

Die graphische Methode, die lediglich eine näherungsweise Bestimmung darstellt, erfolgt mittels Nomogrammen (wie z.B. im Lorenz[2] für 95% und für 99%). Auf der unteren waagerechten Achse liest man den Anteil ab, den man über die Stichprobe geschätzt hat. Zieht man von diesem Anteil aus eine Linie nach oben, trifft man auf Kurven, die für verschieden Stichprobengrößen eingezeichnet sind, wobei sich jedesmal eine untere und eine obere Kurve (für die untere und obere Konfidenzgrenze) finden läßt. Geht man jetzt

von den Kurven zur linken senkrechten Achse, lassen sich dort die Konfidenzgrenzen ablesen.

In der Abbildung 2.4 wurde das Vorgehen sehr vereinfacht dargestellt. Auch finden sich dort aus Gründen der Übersicht nur die Werte für unser Beispiel wieder. In unserem Beispiel liegt demnach der wahre Anteil derjenigen, die Zuzahlungen bei den Pflegehilfsmitteln leisten müssen, zwischen 18% und 35%.

Konfidenzgrenzen für binomialverteilte Anteile sind zunächst einmal nicht symmetrisch zu dem Punktschätzwert. Dies begründet sich durch die Schiefe der Binomialverteilung. Bei Anteilen, die nahe bei 0,5 liegen, sind auch die Grenzen annähernd symmetrisch. Konfidenzgrenzen für binomialverteilte Anteile lassen sich auch näherungsweise berechnen. Allerdings sind an die Berechnung entweder Bedingungen geknüpft (z.B. große Stichprobe) oder die Berechnung ist sehr aufwendig.

Bei vielen praktischen Fragestellungen interessieren wir uns allerdings für Unterschiede in Anteilen, so daß wir im Rahmen dieses Skripts darauf kurz eingehen werden.

II. Konfidenzgrenzen für den Unterschied von Anteilen (bei Binomialverteilungen)

Nehmen wir wieder unser Beispiel der Zuzahlungen bei Pflegehilfsmitteln. In Bundesland A gaben 25 von 100 Befragten an, Zuzahlungen leisten zu müssen. In Bundesland B wurden 150 Personen befragt, von denen 60 von Zuzahlungen betroffen waren. Bundesland A hat dementsprechend einen Anteil von 0,25 (25%) und Bundesland B einen Anteil von 0,40 (40%) Zuzahlern. Der Punktschätzwert für die Differenz $\hat{d}$ beträgt folglich $0,40 - 0,25 = 0,15$. Gesucht ist jetzt das 95% Konfidenzintervall für die wahre Differenz d, welches wir mit der folgenden Formel bestimmen können:

$$\left.\begin{matrix} d_o \\ d_u \end{matrix}\right\} = \hat{d} \pm \frac{n_1 + n_2}{2 \cdot n_1 \cdot n_2} \pm c \cdot \sqrt{\frac{\hat{p}_1 \cdot (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 \cdot (1 - \hat{p}_2)}{n_2}} \tag{2.3}$$

n_1, n_2 = die Umfänge der beiden Stichproben
$\hat{p}_1, \hat{p}_2$ = die geschätzten Anteile in beiden Stichproben
c = eine Konstante

So wie wir in den voran gegangenen Formeln z-Werte oder t-Werte benötigten, ist es bei der Ermittlung von Konfidenzintervallen für Unterschiede in Anteilen die Konstante c, deren Werte der folgenden Tabelle zu entnehmen sind:

α	0,01	0,05	0,10	0,20
c	2,58	1,96	1,64	1,28

Tabelle 2.2.: c-Werte für ausgewählte α

Setzen wir jetzt die Zahlen in Formel 2.3 ein:

$$
\begin{aligned}
d_u, d_o &= 0,15 \pm \frac{100+150}{2 \cdot 100 \cdot 150} \pm 1,96 \cdot \sqrt{\frac{0,25 \cdot (1-0,25)}{100} + \frac{0,40 \cdot (1-0,40)}{150}} \\
&= 0,15 \pm \frac{250}{30.000} \pm 1,96 \cdot \sqrt{\frac{0,1875}{100} + \frac{0,24}{150}} \\
&= 0,15 \pm 0,00833 \pm 1,96 \cdot \sqrt{0,003475} \\
&= 0,15 \pm 0,00833 \pm 1,96 \cdot 0,05895 \\
&= 0,15 \pm 0,00833 \pm 0,11554
\end{aligned}
$$

$$
\begin{aligned}
\text{Es ergibt sich für} \quad d_u &= 0,15 - 0,00833 - 0,11554 = 0,02613 \\
\text{und für} \quad d_o &= 0,15 + 0,00833 + 0,11554 = 0,27387
\end{aligned}
$$

Das 95%-Konfidenzintervall für die geschätzte Differenz von 15% zwischen den Bundesländern A und B wird begrenzt von den Werten 2,6% und 27,4%. Dies ist ein recht „breites" Intervall, d.h. die Aussage ist ziemlich unscharf (Wir erinnern uns, dass wir um die Intervalllänge zu verringern die Stichproben vergrößern müssen!). Nehmen wir an, die Differenz zwischen den Bundesländern hätte lediglich 7% betragen, da in Land B 48 von 150 Befragten Zuzahlungen leisten müssen (dies sollte zu Übungszwecken einmal nach gerechnet werden!), dann hätte sich das folgende Intervall ergeben: für die untere Grenze: **-0,05** und für die obere Grenze **+0,19** Errechnet sich ein Intervall, welches die „0" einschließt, bedeutet dies, dass der wahre Unterschied auch „0" sein könnte. Anders ausgedrückt, kann bei einer gefundenen Differenz von 7% zwischen den beiden Bundesländer nicht sicher genug ausgeschlossen werden, dass der Unterschied nicht zufällig entstanden ist. Der Unterschied ist nicht bedeutsam (**signifikant**). Würde man jetzt einen entsprechenden Sifnifikanztest durchführen, wie wir sie im nächsten Kapitel besprechen werden, müßte dieser ebenfalls ein nicht signifikantes Ergebnis liefern.

Bemerkung:

Konfidenzgrenzen für den Unterschied zweier Mittelwerte werden hier nicht besprochen. Wichtig hierbei ist, dass es zwei unterschiedliche Formeln zur Bestimmung gibt (siehe Lorenz[2]), deren Anwendung sich danach richtet, ob die Varianzen der beiden Stichproben gleich (homogen) oder unterschiedlich (heterogen) sind. Kommen wir jetzt zu den bereits erwähnten Signifikanztests.

2-5. Signifikanztests

I. Einführung

Die Fragestellung, um die es beim Testen auf Signifikanz geht, ist die, inwieweit ein gefundenes Ergebnis, z.B. ein Unterschied zufällig oder nicht zufällig zufällig zustande gekommen sein kann. Im Rahmen dieses Skripts werden wir uns in der Hauptsache mit Unterschieden zwischen zwei Stichproben befassen. Der Grundgedanke, der sich dahinter verbirgt, ähnelt ein wenig dem Prinzip bei der Bestimmung von Konfidenzgrenzen. Greifen wir nochmals auf das Beispiel der durchschnittlichen Liegedauer zurück, die 9 Tage betrug. Unsere Stichprobe entstammte einer Grundgesamtheit mit einen Mittelwert $\mu = 10$ Tage und einen Standardfehler von $\sigma_{\bar{x}} = 2$ Tage. Im Bereich von 1,96 Standardfehlern rechts und links vom Mittelwert, also zwischen 6,08 und 13,92 Tagen, liegen 95% aller Werte. Die Wahrscheinlichkeit dafür, einen Stichprobenmittelwert von weniger als 6,08 Tagen oder einen von mehr als 13,92 Tage zu finden, beträgt jeweils nur 2,5%.
Wenn in einer weiteren Stichprobe eine durchschnittliche Liegedauer von 45 Tagen ermittelt wird, müssen wir fragen, ob diese Stichprobe wohl eher einer anderen Grundgesamtheit entstammt. Denn es ist sehr unwahrscheinlich, dass unsere Grundgesamtheit ($\mu = 10, \sigma_{\bar{x}} = 2$) einen Mittelwert von 45 Tagen und mehr „produziert" dass man sicherlich spontan sagen würde, der Unterschied zwischen Stichprobe A mit 9 Tagen und Stichprobe B mit 45 Tagen ist signifikant. Die beiden Stichproben entstammen zwei völlig verschiedenen Grundgesamtheiten (vielleicht Liegedauer in Allgemeinen Krankenhäusern und Liegedauer in Reha-Einrichtungen). Hätte Stichprobe B eine durchschnittliche Liegedauer von 7 oder 11 oder 12 Tagen angegeben, hätten wir mit unserem Vorwissen sofort gesagt, es ist sehr wahrscheinlich, dass Stichprobe B zu der gleichen Grundgesamtheit gehört wie Stichprobe A.
Wie schon bei der Bestimmung von Konfidenzgrenzen ist es auch bei Signifikanzaussagen nicht möglich, Aussagen mit 100%iger Sicherheit zu erhalten. So ist in unserem Beispiel eine durchschnittliche Liegedauer von 16 Tagen oder mehr Tagen sehr unwahrscheinlich (0,13%), es ist aber nicht unmöglich. Das heißt, wenn wir nach der Durchführung eines Signifikanztestes behaupten, das Krankenhaus gehöre zu einer anderen Grundgesamtheit, kann es sein, dass wir uns irren, dass wir also einen **Fehler** gemacht haben. Auf die Möglichkeit, Fehlentscheidungen zu treffen, gehen wir an späterer Stelle ausführlich ein. Vor der Durchführung eines Signifikanztestes legen wir (wie auch bei den Konfidenzgrenzen) fest, wie hoch denn das Irrtums-Risiko höchstens sein darf. Auch hierbei sind Irrtumswahrscheinlichkeiten α von 5% oder 1% geläufig, so daß das Signifikanzniveau bei 95% oder 99% liegt.
Dadurch ergeben sich Schwellenwerte, die im Beispiel der Normalverteilung die z-Werte sind. Mittels z-Transformation unseres Stichprobenmittelwertes ist es uns dann möglich zu entscheiden, wo genau unser Stichprobenwert liegt. Findet er sich innerhalb des 95% (oder 99%) -Bereiches, so ist davon auszugehen, dass die Stichprobe zu dieser Grundgesamtheit gehört. Die vorher formulierte statistische **Null-Hyothese** (es gibt keinen Unterschied) kann somit nicht verworfen werden. Liegt er links vom unteren Wert bzw. rechts vom oberen Wert, so ist es sehr wahrscheinlich, dass unsere Stichprobe aus einer anderen Verteilung stammt. Dies würde die statistische **Alternativ-Hypothese** (es gibt einen Unterschied) bestätigen.
Zusammenfassend kann man sagen, dass es um die Frage geht, inwieweit die aufgetretenen Unterschiede „echt" sind (d.h. es existiert ein Unterschied) oder zufällig zustande kamen (es existiert kein Unterschied). Dazu formuliert man statistische Hypothesen. Gibt es tatsächlich Unterschiede, so stammen die beiden Stichproben aus zwei unterschiedlichen Grundgesamtheiten mit den Mittelwerten μ_1 und μ_2. Sind die Unterschiede in den Stichproben zufällig, so entstammen sie derselben Grundgesamtheit mit dem Mittelwert μ. Durch festlegen der Irrtumswahrschein-

lichkeit α lassen sich Schwellenwerte bestimmen, mit denen man den Stichprobenwert nach Transformation (Teststatistik) vergleicht. In unserem Beispiel ging es um eine Normalverteilung, so daß wir mit z-Werten und z-Transformation arbeiten konnten. Je nach Art der Daten (Skalenniveau, Verteilung, Variabilität) gibt es unterschiedliche Verfahrem zur Ermittlung der Teststatistik. Bevor wir auf die verschiedenen Signifikanzteste eingehen, müssen wir noch die angesprochenen statistischen Hypothesen und mögliche Fehlerarten besprechen.

II. Statistische Hypothesen

Grundsätzlich formulieren wir in der Statistik vor den Untersuchungen zwei Hypothesen. Zum ersten sprechen wir von der **Null-Hypothese ($\mathbf{H_0}$)**, die vereinfacht ausgedrückt besagt, dass es keinen Effekt / Unterschied gibt, die Stichproben also aus der selben Grundgesamtheit stammen. Der transformierte Stichproben-Wert, die Teststatistik also, liegt innerhalb eines möglichen Bereiches, der vom Schwellenwert abgegrenzt wird.
Die **Alternativ-Hypothese ($\mathbf{H_1}$ oder $\mathbf{H_A}$)** ist das (logische) Gegenteil der Nullhypothese, also es gibt einen Effekt / Unterschied. Somit gehören die Stichproben zu unterschiedlichen Grundgesamtheiten. Die Teststatistik hat den Schwellenwert überschritten (bzw. unterschritten). Alle Testverfahren überprüfen die Nullhypothese. Wird diese nach Durchführung des Signifikanztestes abgelehnt, wird dadurch die Alternativhypothese bestätigt.
Beim Formulieren von Hypothesen unterscheidet man weiterhin, ob es sich um eine **einseitige** oder **zweiseitige** Fragestellung handelt. Man spricht dann von **gerichteten** oder **ungerichteten** Hypothesen. Formuliert man zur Liegedauer im Krankenhaus beispielsweise die folgenden Hypothesen:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Die Nullhypothese lautet: es gibt keinen Unterschied in der durchschnittlichen Liegedauer zwischen den beiden Stichproben und entsprechend sagt die Alternativhypothese aus, dass es einen Unterschied gibt. Dies ist eine ungerichtete Hypothese bei zweiseitiger Fragestellung, da wir nicht konkretisieren, ob die Verweildauer länger oder kürzer ist. Wir sagen lediglich, dass sie anders ist.
Sprechen wir unsere Vermutung deutlich aus, handelt es sich um eine gerichtete Hypothese bei einseitiger Fragestellung. Für unser Beispiel lautete die Nullhypothese, dass die durchschnittliche Verweildauer in Stichprobe 1 höchsten genauso lang ist wie in Stichprobe 2. Die Alternativhypothese wäre, dass die durchschnittliche Verweildauer in Stichprobe 1 länger ist als in Stichprobe 2.

$$H_0 : \mu_1 \leq \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

Die Verwendung der Relationssymbole $>$, $\geq$, $\leq$, $<$ richtet sich selbstverständlich nach der jeweiligen Fragestellung. Möchte man beispielsweise mit einem Signifikanztest herausfinden, dass ein neues Medikament (N) weniger Nebenwirkungen mit sich bringt als ein altes, bisher verwendetes Medikament (A), so könnten die Hypothesen lauten:

> Nullhypothese H_0: $\mu_N \geq \mu_A$, das neue Medikament hat genauso viele oder mehr Nebenwirkungen wie das alte.

Alternativhypothese H_1: $\mu_N < \mu_A$, das neue Medikament hat weniger Nebenwirkungen als das alte.

Jetzt muß noch geklärt werden, welche Konsequenzen sich aus ein- bzw. zweiseitiger Fragestellung ergeben. Wir verdeutlichen dies am Beispiel der durchschnittlichen Kosten für Intensivbetten pro Tag.
Angenommen wir wüßten, dass sich diese Kosten in Deutschland auf durchschnittlich 1000 € pro Tag belaufen. Bekannt sei ebenfalls der Standardfehler von 100 €. Finden wir jetzt in einer Stichprobe durchschnittliche Kosten von 1180 € und fragen, ob sich diese Klinik signifikant von der Grundgesamtheit unterscheidet. Bei zweiseitiger Fragestellung mit einem Signifikanzniveau von 95% liegt die Irrtumswahrscheinlichkeit gleichmäßig an beiden Seiten der Verteilung. Der Wert von 1180 € überschreitet nicht den oberen Schwellenwert, unserer Schlußfolgerung lautet, dass der Unterschied nicht signifikant ist. Man sagt auch, das Ergebnis ist mit der Nullhypothese (es gibt keinen Unterschied) verträglich.
Nehmen wir jetzt die einseitige Fragestellung und formulieren die gerichtete Hypothese, dass der Wert von 1180 größer ist als der Mittelwert der Grundgesamtheit von 1000 €, dann bietet sich das Bild aus Abbildung 2.6. Die Irrtumswahrscheinlichkeit von 5% findet sich hier an einer Seite der Verteilung. Der Wert von 1180 überschreitet den Schwellenwert und wir kämen zu dem Ergebnis, dass sich die untersuchte Klinik signifikant von der Grundgesamtheit unterscheidet. Man sagt auch, das Ergebnis ist mit der Nullhypothese nicht verträglich, die Nullhypothese wird abgelehnt.
Fassen wir zusammen: in unserem Beispiel wurde ein Ergebnis, welches bei zweiseitiger Fragestellung nicht signifikant war, bei einseitiger Fragestellung plötzlich signifikant. Man kann auch sagen, ein zweiseitiger Test entscheidet eher für die Nullhypothese, man bezeichnet dies als konservativ.
Das gleiche „Phänomen" tritt übrigens auch bei unterschiedlichem Signifikanzniveau auf. Ein Test auf dem Signifikanzniveau von 99% kann unter Umständen nicht signifikant ausfallen, während er bei einem Niveau von 90% signifikant ist (Allerdings liegt hier die Irrtumswahrscheinlichkeit bei 10%).

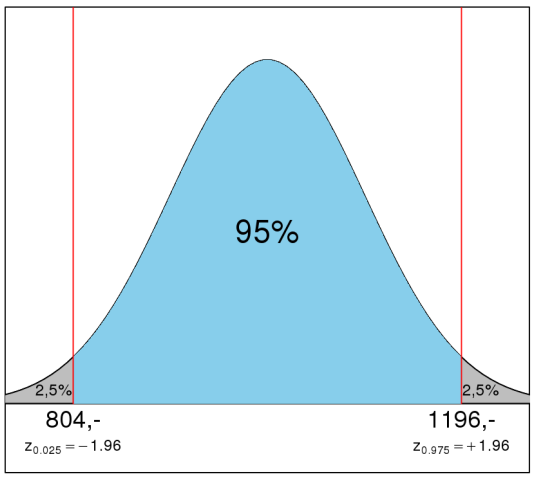


Abbildung 2.5.: zweiseitige Fragestellung

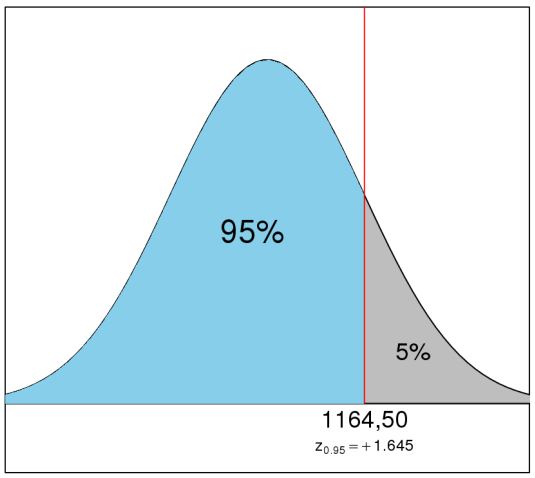


Abbildung 2.6.: einseitige Fragestellung

Vielleicht kommt an dieser Stelle wieder einmal der Gedanke auf, dass statistische Ergebnis willkürlich und manipulierbar sind. Bis zu einem gewissen Grad sind sie es sicherlich. Daher ist es ungeheuer wichtig, in Forschungsberichten und Veröffentlichungen genau zu begründen, warum man welche Methode angewendet oder aber die Fragestellung so und nicht anders formuliert oder auch, warum man welches Signifikanzniveau gewählt hat. Nur wenn ein Leser oder Anwender genau nachvollziehen kann, wie ein Ergebnis zustande kam und wie hoch das Risiko einer Fehlentscheidung ist, kann er Vertrauen in die Forschung haben.
Gerade bei der Überlegung, inwiefern man eine ein- oder zweiseitige Fragestellung wählt, weisen die meisten Autoren darauf hin, dass einseitige Fragestellungen nur sparsam zu verwenden sind. Guggenmoos-Holzmann & Wernecke[4] gehen soweit zu sagen, dass eine einseitige Fragestellung nur dann berechtigt ist, wenn sich die postulierte Richtung der Änderung als absolut sicher (im deterministischen Sinne) erwiesen hat. Bei Polit&Hungler[5] ist zu finden, dass die Frage ein- oder zweiseitig kontrovers diskutiert wird und die meisten Forscher der Konvention folgen, ihre Hypothese zwar gerichtet zu formulieren, den Test dann aber zweiseitig durchzuführen. Bei der Nutzung von Statistikprogrammen wie z.B. SPSS erübrigt sich die Überlegung, da dort meisten zweiseitig getestet wird.
Es ist in den vergangenen Abschnitten immer wieder von der Möglichkeit gesprochen worden, Fehlentscheidungen zu treffen. Dies wollen wir im folgenden näher ausführen.

III. Fehlerarten bei statistischen Entscheidungen

Jede Entscheidung in der schließenden Statistik ist mit einer gewissen Wahrscheinlichkeit behaftet, falsch zu sein. Denken wir an unser Beispiel. Eine durchschnittliche Liegedauer von mindestens 16 Tagen in einer Grundgesamtheit zu finden, deren Mittelwert 10 Tage (mit einem Standardfehler von 2 Tagen) betrug, war mit 0,13% sehr unwahrscheinlich. Unsere Testentscheidung würde bei solch einem Ergebnis die Nullhypothese (es gibt keinen Unterschied) verwerfen. Wir kämen zu einem „positiven Befund" nämlich den, dass es einen Unterschied gibt, der nicht zufällig ist. Anders als in unserem fiktiven Beispiel kennen wir üblicherweise die Verhältnisse in der Grundgesamtheit nicht. Genau genommen gibt es allerdings genau zwei Möglichkeiten, wie die tatsächlichen Verhältnisse sein können und es gibt genau zwei Testentscheidungen, so daß sich insgesamt vier Kombinationen ergeben.

1. Die untersuchte Klinik stammt tatsächlich aus einer anderen Grundgesamtheit und wir haben durch den Test eine **richtige positive Entscheidung** getroffen.

2. Die untersuchte Klinik gehört in Wirklichkeit doch noch zu der Grundgesamtheit der Allgemeinen Krankenhäuser. Der Test hat allerdings ein positives Ergebnis geliefert (es gibt einen Unterschied, die Nullhypothese wird verworfen), so daß wir zu einer **falsch positiven Entscheidung** kommen. Die Wahrscheinlichkeit dafür, eine falsch-positive Entscheidung zu treffen, betrug in unserem Beispiel 0,13%. Es wurde schon erwähnt, dass wir durch das Festlegen der Irrtumswahrscheinlichkeit α von üblicherweise 5% oder 1% bestimmen, welche Wahrscheinlichkeit wir höchstens akzeptieren wollen, einen Irrtum zu begehen. Da die Möglichkeit einer falsch-positiven Entscheidung durch das Festlegen der Irrtumswahrscheinlichkeit a beeinflusst werden kann, spricht man von einem **α-Fehler**, der oftmals auch **Fehler 1. Art** genannt wird. Die ausgerechnete Wahrscheinlichkeit, einen solchen Fehler zu begehen, nennt man α-Fehlerwahrscheinlichkeit (nicht zu verwechseln mit der o.g. vorher festzulegenden Irrtumswahrscheinlichkeit) und sie betrug in unserem Beispiel 0,13%.

3. Der Signifikanztest findet keinen Unterschied, die Nullhypothese kann nicht verworfen werden, das Testergebnis ist negativ. Im günstigen Fall gibt es in der Realität auch tatsächlich keinen Unterschied. Dann handelt es sich um eine **richtig-negative Entscheidung**.

4. Verbleibt noch die letzte Möglichkeit. Der Signifikanztest ist **negativ**, in der Realität gibt es aber sehr wohl einen „Befund". Dann würden wir eine **falsch-negative Entscheidung** treffen. Dieser Fehler wird als β-Feher oder auch **Fehler 2. Art bezeichnet**. Das Problem bei diesem Fehler liegt darin, dass er nicht so leicht - zumindest nicht mit unserem jetzigen Wissen - kontrolliert werden kann.

Exkurs: p-Werte

Bisher war das allgemeine Vorgehen bei einem Signifikanztest so, dass man eine Teststatistik (z.B. der z-transformierter Mittelwert der Stichprobe) mit einem Schwellenwert (z-Wert, der 97,5% der Verteilung abschneidet) vergleicht und so zu einer Signifikanzaussage kommt. Genauso ist es möglich, die ausgerechnete a-Fehler-Wahrscheinlichkeit (im Beispiel 0,13%) mit der vorher festgelegten Irrtumswahrscheinlichkeit α (z.B. 5%) zu vergleichen. Ist die ausgerechnete Wahrscheinlichkeit kleiner als α, so verwerfen wir die Null-Hypothese und sprechen von einem **signifikanten Ergebnis**. Ist die ausgerechnete Wahrscheinlichkeit größer als α, so behalten wir die Null-Hypothese bei, der Unterschied ist **nicht** signifikant, er ist zufällig. In Statistikprogrammen für Computer (z.B. SPSS oder R[1]) wird die entsprechende Teststatistik ausgerechnet und man kann diese dann anhand von tabellierten Schwellenwerten vergleichen. Einfacher ist allerdings, sich den „p-Wert" anzuschauen. Dieser gibt die ausgerechnete α-Fehler-Wahrscheinlichkeit wieder. Für unser Beispiel hätte SPSS einen p-Wert von 0,0013 angegeben, der deutlich kleiner ist als $\alpha = 0,05$.

Wir wollen die vier möglichen Kombinationen zwischen Testentscheidung und tatsächlichen Verhältnissen in der Realität in einer Tabelle zusammen fassen:

		in der Realität gilt:	
		H_1, es gibt tatsächlich einen Unterschied	H_0, es gibt wirklich keinen Unterschied
Testergebnis:	H_0 wird verworfen, es gibt einen Unterschied	**richtig positive Entscheidung**	Fehler 1. Art **falsch positive Entscheidung**
	H_0 wird **nicht** verworfen, es gibt keinen Unterschied	Fehler 2. Art **falsch negative Entscheidung**	**richtig negative Entscheidung**

Tabelle 2.3.: Fehlerarten bei statistischen Entscheidungen

Wie eben erwähnt, ist es nicht so ohne weiteres möglich, den ß-Fehler zu kontrollieren, bzw. die ß-Fehler-Wahrscheinlichkeit zu berechnen. Da allerdings α- und β-Fehler sich gegenseitig beeinflussen, können wir indirekt Einfluss auf den β-Fehler nehmen. Bleiben wir bei unserem Beispiel mit der durchschnittlichen Verweildauer (Grundgesamt ist die Liegedauer in Allgemeinen

[1] http://www.r-project.org

Krankenhäusern). Die Wahrscheinlichkeit, einen Mittelwert zu finden, der größer als 13,92 Tage ist, war sehr gering, allerdings nicht unmöglich. Wenn ein Signifikanztest entscheidet, ein solches Krankenhaus gehört zu einer anderen Grundgesamtheit, ist nicht auszuschließen, dass die H_0 fälschlicherweise verworfen wurde, also ein α-Fehler vorliegt.
Von den Langzeit- und Reha-Einrichtungen liegen uns keine Daten vor. Allerdings ist es denkbar, dass es auch dort angesichts der Sparmaßnahmen im Gesundheitswesen Kliniken gibt, die sich auf kurze Rehabilitationen eingerichtet haben und somit eventuell auch Stichproben mit Mittelwerten von 12 oder 13 Tagen dabei sein können. Ein Signifikanztest würde allerdings bei einem solchen Mittelwert entscheiden, die Nullhypothese (fälschlicherweise) nicht zu verwerfen. Das ist dann die Situation, die als β-Fehler bezeichnet wird.
Der Idealfall ist, wenn zwei völlig verschiedene (getrennt voneinander liegende) Verteilungen vorliegen, was aber in der Realität meist nicht der Fall ist (Abbildung 2.7)

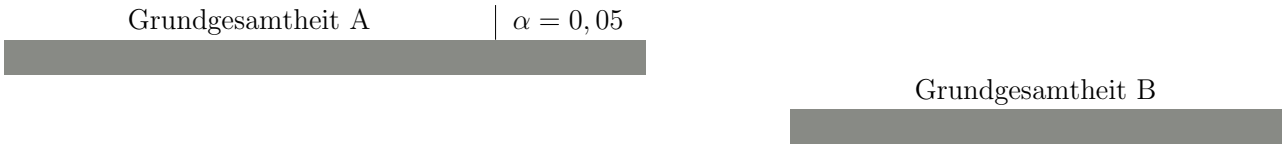


Abbildung 2.7.: verschiedene Verteilungen

Bei einer Irrtumswahrscheinlichkeit von 5% erhöht sich zwar das Risiko, fälschlicherweise die Nullhypothese zu verwerfen, gleichzeitig verringert sich die Wahrscheinlichkeit dafür, fälschlicherweise die Nullhypothese beizubehalten (Abbildung 2.8, roter Bereich).

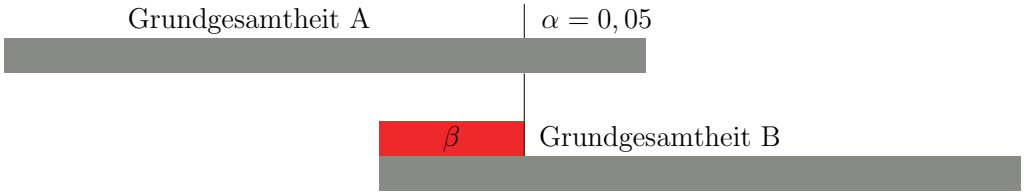


Abbildung 2.8.: α und β-Fehler

Bei einer geringeren Irrtumswahrscheinlichkeit (z.B. 0,01) verhält es sich genau umgekehrt, der α-Fehler verringert sich, der β-Fehler wird größer (Abbildung 2.9)

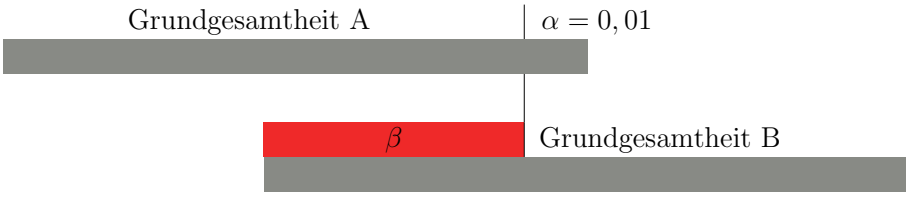


Abbildung 2.9.: α und β-Fehler

Wie man mit einem solchen Problem in der Praxis umgehen kann, kann an einem Beispiel (Lorenz[2]) verdeutlicht werden:
Es sollen alle Fälle von Lungentuberkulose in der Bevölkerung erfasst werden. Dazu veranlasst man eine Röntgen-Reihenuntersuchung. Um möglichst niemanden mit einer TB zu übersehen, nimmt man erst einmal in Kauf, auch gesunde Personen fälschlicherweise als krank zu identifizieren (falsch-positiv, α-Fehler) und wählt daher eine möglichst große Irrtumswahrscheinlichkeit (α=10%,20% oder sogar mehr). Dadurch verringert sich das Risiko, kranke Personen

irrtümlicherweise als gesund zu betrachten (falsch-negativ, β-Fehler). In einem zweiten Schritt werden dann alle erfassten Personen, die wirklich Erkrankten und fälschlicherweise als krank identifizierten, gründlich untersucht. Jetzt wählt man eine geringe Irrtumswahrscheinlichkeit (z.B. α= 1% oder sogar 0,1%), damit wird das Risiko von falsch-positiven Diagnosen sehr klein und die Gesunden können heraus gefunden werden.
Die Entscheidung, welche Irrtumswahrscheinlichkeit a man für eine statistische Analyse wählt, hängt also unmittelbar damit zusammen, welche Konsequenzen jeweils der α-Fehler bzw. der β-Fehler nach sich zieht. Auf die Frage, welcher Fehler denn „schlimmer" sei, wird häufig geantwortet, dies sei der β-Fehler. Allerdings lässt sich dies nur durch inhaltliche Überlegungen klären. Nachfolgend ein paar Beispiele.

Beispiel 1: Untersuchung einer neuen Unterrichtsmethode

Nullhypothese: es gibt keinen Unterschied zwischen neuer und alter Methode.

Alternativhypothese: die neue Methode ist besser

α-Fehler: die Nullhypothese wird fälschlicherweise verworfen, man geht davon aus, die neue Methode sei besser. Dies kann neben immensen Kosten für neue Schulmaterialen noch zur Folge haben, dass Kinder umgeschult werden etc.

β-Fehler: die Nullhypothese wird fälschlicherweise nicht verworfen, die neue und wirklich bessere Methode wird nicht eingeführt.

Beispiel 2: Untersuchung eines neuen Medikaments

Nullhypothese: es gibt keinen Unterschied zwischen dem neuem und dem altem Medikament

Alternativhypothese: das neue Medikament ist besser

α-Fehler: das neue Medikament wird eingeführt, da es angeblich besser wirkt. Man nimmt ernsthafte Nebenwirkungen in Kauf, außerdem sind die Kosten sehr hoch.

β-Fehler: es wird nicht erkannt, dass das neue Medikament besser ist, Es verschwindet in der „Schublade" obwohl vielen Menschen damit hätte geholfen werden können.

Beispiel 3: Untersuchung, inwieweit Fernsehen die Konzentrationsfähigkeit von Kindern senkt

Nullhypothese: Fernsehen hat keinen Einfluss auf die Konzentration

Alternativhypothese: Fernsehen senkt die Konzentrationsfähigkeit von Kindern

α-Fehler: Irrtümlicherweise wird angenommen, Fernsehen schadet der Konzentration und man empfiehlt den Eltern, die tägliche Fernsehzeit einzuschränken.

β-Fehler: es wird fälschlicherweise angenommen, Fernsehen schadet nicht. Die Konsequenz daraus könnte sein, dass Kinder weiterhin uneingeschränkt fernsehen dürfen.

Einige letzte Anmerkungen bevor wir kurz auf die Schritte eines Signifikanztestes und auf einzelne Testverfahren eingehen: Eine Irrtumswahrscheinlichkeit α von z.B. 5% mit einem Signifikanzniveau von 95% bedeutet nicht, dass mit 95%iger Wahrscheinlichkeit der Unterschied in der Realität vorhanden ist. Wie schon bei den Konfidenzgrenzen gibt es entweder einen Unterschied oder es gibt ihn nicht! Die Irrtumswahrscheinlichkeit von 5% sagt viel mehr aus, dass man bei 100 durchgeführten Untersuchungen im Schnitt 5 mal ein falsch-positives Ergebnis erhalten kann! (Natürlich besteht theoretisch auch hier wiederum die Gefahr, dass je nach

Interessenlage nur diese falsch-positiven Ergebnisse veröffentlicht werden). Sowohl den α-Fehler wie auch den β-Fehler möglichst klein zu halten gelingt nur über eine entsprechende große Stichprobe. Allerdings sprengt es den Rahmen eines Skriptes, das Thema Fallzahlschätzungen zu behandeln.

IV. Die Struktur statistischer Tests

1. Formulierung von Hypothesen
   Prinzipiell werden Hypothesen vor der Durchführung der Untersuchung formuliert. Wird Datenmaterial im Sinne von „Exploration" und „hypothesenerkundend" analysiert, muß dies kenntlich gemacht werden, indem die Untersuchung ausdrücklich als Erkundungsexperiment oder explorative Studie gekennzeichnet wird. Es ist unwissenschaftlich, aus dem selben Datenmaterial Hypothesen abzuleiten und gleichzeitig zu überprüfen.

2. Auswahl eines Signifikanztests
   Dabei spielt das Skalenniveau und die Verteilung der Daten eine wichtige Rolle. Hinzu kommt die Überlegung, ob es sich um verbundene oder unverbundene Daten (Stichproben) handelt. Von verbundenen oder auch abhängigen Stichproben geht man aus, wenn es sich um mehrere Meßwerte von ein und den selben Personen / Objekten handelt.

3. Festlegen der Irrtumswahrscheinlichkeit α und des notwendigen Stichprobenumfangs

4. Durchführung der Untersuchung / Datenerhebung

5. Berechnung der Teststatistik

6. Vergleich der Teststatistik mit einem (tabellierten) Schwellenwert

7. Testentscheidung

V. Überblick Signifikanztests

				METRISCH			
NOMINAL		ORDINAL		nicht normalverteilt		normalverteilt	
unabhängig	abhängig	unabhängig	abhängig	unabhängig	abhängig	unabhängig	abhängig
χ^2 $k \cdot l$ -Felder $2 \cdot 2$ -Felder	χ^2 **McNemar-Test** $2 \cdot 2$ -Felder	**Mann-Whitney**	**Wilcoxon**	**Mann-Whitney**	**Wilcoxon**	**F-Test** ist das Ergebnis homogen: **t-Test** heterogen: **Welch Test**	**t-Test** für verbundene Stichproben
nicht-parametrische Testverfahren						parametrische Tests	

Tabelle 2.4.: Übersicht ausgewählter Signifikanztests

Literaturverzeichnis

[1] Lange S, Bender R: **Histogramm**. *DMW - Deutsche Medizinische Wochenschrift* 2001, **126**(15):31--32.

[2] Lorenz RJ: *Grundbegriffe der Biometrie*. Stuttgart: Fischer 1996.

[3] Bortz J: *Statistik für Sozialwissenschaftler*. Berlin: Springer 1999.

[4] Guggenmoos-Holzmann I, Wernecke KD: *Medizinische Biostatistik*. Blackwell Science Ltd. 1995.

[5] Polit DF, Hungler BP: *Nursing Research*. Philadelphia: Lippincott 1999.

[6] Bortz J, Lienert GA: *Kurzgefasste Statistik für die Klinische Forschung*. Heidelberg: Springer-Verlag, 3rd edition 2008.

Abbildungsverzeichnis

Tabellenverzeichnis

Quellcodeverzeichnis

A. ANHANG A

A-1. Die graphische Bestimmung des Medians bei klassierten Daten

Gegeben sind die Daten aus Tabelle 1.14 auf Seite 18.

	Häufigkeit	**Prozent**	**kumulierte Prozente**
bis 12	14	8,8	8,8
bis 24	26	16,3	25
bis 36	51	31,9	56,9
bis 48	51	31,9	88,8
bis 60	18	11,3	100,0
Gesamt	160	100	

Für die graphische Ermittlung des Medians benötigen wir die kumulierten Prozentwerte, aus denen ein Summenpolygon (Abbildung A.1) angefertigt wird.

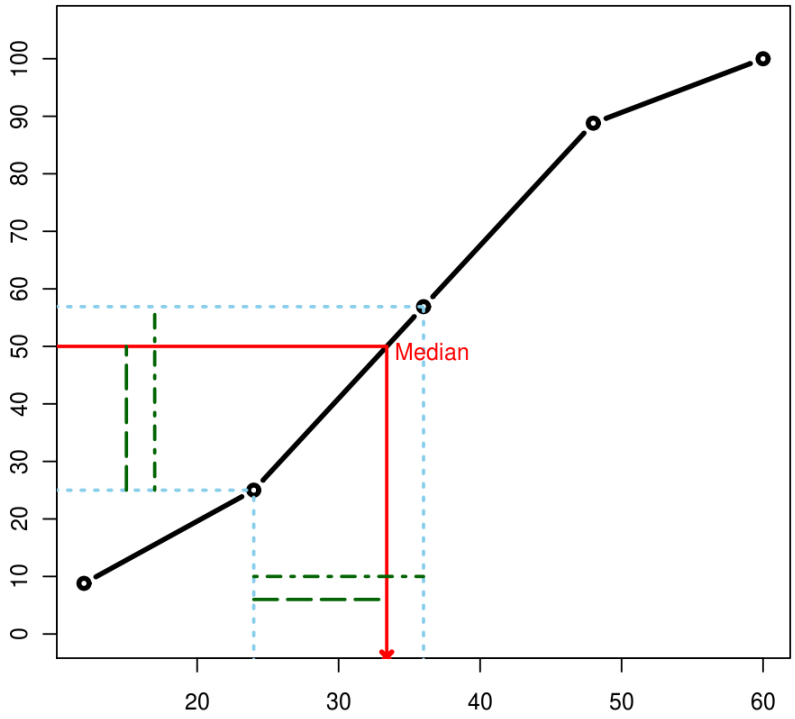


Abbildung A.1.: graphische Bestimmung des Median

Aus der Mathematik (Strahlensätze) wissen wir, dass das Verhältnis der beiden Strecken auf der x-Achse dem Verhältnis der Strecken auf der y-Achse entspricht.

Setzen wir also die Strecke ins Verhältnis, ergibt sich:

$$\frac{x-24}{36-24} = \frac{50-25}{56,9-25} \qquad \cdot 12$$

$$x - 24 = \frac{25}{31,9} \cdot 12 \qquad +24$$

$$x = 9,4 + 24$$

$$x = 33,4$$

A-2. Rechenweg der Beispielaufgabe

Hier nochmal der Rechenweg zum Beispiel aus Tabelle 1.26 auf Seite 33:

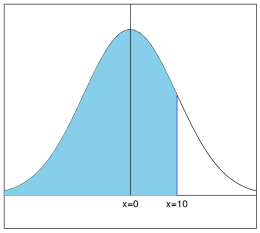
$\mathbf{x_i}$	$\mathbf{y_i}$	$(\mathbf{x_i} - \bar{\mathbf{x}})$	$(\mathbf{x_i} - \bar{\mathbf{x}})^\mathbf{2}$	$(\mathbf{y_i} - \bar{\mathbf{y}})$	$(\mathbf{y_i} - \bar{\mathbf{y}})^\mathbf{2}$	$(\mathbf{x_i} - \bar{\mathbf{x}}) \cdot (\mathbf{y_i} - \bar{\mathbf{y}})$
2	3	-8,7	75,69	-9,5	90,25	82,65
4	4	-6,7	44,89	-8,5	72,25	56,95
7	9	-3,7	13,69	-3,5	12,25	12,95
9	12	-1,7	2,89	-0,5	0,25	0,85
10	12	-0,7	0,49	-0,5	0,25	0,35
12	14	1,3	1,69	1,5	2,25	1,95
13	16	2,3	5,29	3,5	12,25	8,05
15	17	4,3	18,49	4,5	20,25	19,35
16	18	5,3	28,09	5,5	30,25	29,15
19	20	8,3	68,89	7,5	56,25	62,25
$\sum x_i = 107$	$\sum y_i = 125$		$\sum (x_i - \bar{x})^2 = 260,1$		$\sum (y_i - \bar{y})^2 = 296,5$	$\sum (x_i - \bar{x}) \cdot (y_i - \bar{y}) = 274,5$
$\bar{x} = \frac{107}{10} = 10,7$	$\bar{y} = \frac{125}{10} = 12,5$		$s_x^2 = \frac{260,1}{10-1} = 28,9$		$s_y^2 = \frac{296,5}{10-1} = 32,94$	$cov(x,y) = \frac{274,5}{10-1} = 30,5$
			$s_x = \sqrt{28,9} = 5,38$		$s_y = \sqrt{32,94} = 5,74$	

B. ANHANG B - Referenztabellen

B-1. Verteilungsfunktion und Quantile der Standardnormalverteilung

ϕ **z**	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
p =	0.5000	0.7500	0.8000	0.9000	0.9500	0.9750	0.9900	0.9950	0.9975	0.9990
z_p =	0.0000	0.6745	0.8416	1.2816	1.6449	1.9600	2.3263	2.5758	2.8070	3.0902

Tabelle B.1.: Verteilungsfunktion und Quantile der Standardnormalverteilung

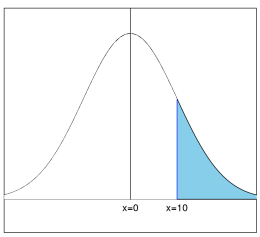


Bei der z-Transformation (vgl. Seite 25) benötigt man die Werte dieser Tabelle beispielsweise für Fragen wie "*Wie hoch ist der prozentuale Anteil für Werte, die* ***höchstens*** *X sind*"

Im Gegensatz dazu zeigt Tabelle B.2 die Überschreitungswahrscheinlichkeiten P für Abszissenwerte z der Standardnormalverteilung ($\mu = 0$ und $\sigma = 1$) [6, S. 380]. Die p-Werte entsprechen der Fläche unter der Standardnormalverteilung zwischen z und $+\infty$ und damit einer *einseitigen* Fragestellung. Sie müssen bei *zweiseitigen* Fragestellungen verdoppelt werden!

ϕ **z**	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2887	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1464	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1291	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
p =	0.5000	0.7500	0.8000	0.9000	0.9500	0.9750	0.9900	0.9950	0.9975	0.9990
z_p =	0.0000	0.6745	0.8416	1.2816	1.6449	1.9600	2.3263	2.5758	2.8070	3.0902

Tabelle B.2.: Überschreitungswahrscheinlichkeiten P für Abszissenwerte z der Standardnormalverteilung ($\mu = 0$ und $\sigma = 1$)



Bei der z-Transformation (vgl. Seite 25) benötigt man die Werte dieser Tabelle beispielsweise für Fragen wie "*Wie hoch ist der prozentuale Anteil für Werte, die* ***mindestens*** *X sind*"

B-2. Kritische Werte t* der t-Verteilung

	Irrtumswahrscheinlichkeit $\alpha/2$ für einseitige Fragestellung					
	0.10	**0.05**	**0.025**	**0.01**	**0.005**	**0.0005**
v	Irrtumswahrscheinlichkeit α für zweiseitige Fragestellung					
	0.20	**0.10**	**0.050**	**0.020**	**0.010**	**0.0010**
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.611
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
70	1.294	1.667	1.994	2.381	2.648	3.211
80	1.292	1.664	1.990	2.374	2.639	3.195
90	1.291	1.662	1.987	2.368	2.632	3.183
100	1.290	1.660	1.984	2.364	2.626	3.174
200	1.286	1.653	1.972	2.345	2.601	3.131
300	1.284	1.650	1.968	2.339	2.592	3.118
400	1.284	1.649	1.966	2.336	2.588	3.111
500	1.283	1.648	1.965	2.334	2.586	3.107
∞	1.282	1.645	1.960	2.326	2.576	3.090

Tabelle B.3.: Kritische Werte t^* der t-Verteilung

B-3. Effektstärke und Power

Die folgenden Tabellen stammen aus Polit&Hungler [5, S.492].

$\alpha = \mathbf{0.05}$										
	Effektstärke									
Power	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**
.60	977	434	244	156	109	61	39	27	20	15
.70	1230	547	308	197	137	77	49	34	25	19
.80	1568	697	392	251	174	98	63	44	32	25
.90	2100	933	525	336	233	131	84	58	43	33
.95	2592	1152	648	415	288	162	104	72	53	41
.99	3680	1636	920	589	409	230	147	102	75	58

Tabelle B.4.: Anzahl Probanden *pro Gruppe* nach Power und Effektstärke bei $\alpha = 0,05$

$\alpha = \mathbf{0.01}$										
	Effektstärke									
Power	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**
.60	1602	712	400	256	178	100	64	44	33	25
.70	1922	854	481	308	214	120	77	53	39	30
.80	2339	1040	585	374	260	146	94	65	48	37
.90	2957	1324	745	477	331	186	119	83	61	47
.95	3562	1583	890	570	396	223	142	99	73	56
.99	4802	2137	1201	769	534	300	192	133	98	

Tabelle B.5.: Anzahl Probanden *pro Gruppe* nach Power und Effektstärke bei $\alpha = 0,01$

C. ANHANG C - Formelsammlung

C-1. Konfidenzgrenzen

I. Konfidenzgrenzen bei Binomialverteilungen

Diese Formel ist kompliziert, kann aber dafür auch bei kleineren Stichproben verwendet werden.

$$\left.\begin{matrix} p_u \\ p_o \end{matrix}\right\} = \frac{1}{n+c^2} \cdot \left[k \pm \frac{1}{2} + \frac{c^2}{2} \pm c \cdot \sqrt{(k \pm \frac{1}{2}) \cdot (1 - \frac{k \pm 0,5}{n}) + \frac{c^2}{4}} \right]$$

Diese etwas einfachere Formel sollte nur bei großen Stichproben verwendet werden.

$$\left.\begin{matrix} p_u \\ p_o \end{matrix}\right\} = \frac{k}{n} \pm \frac{c}{n} \cdot \sqrt{\frac{k \cdot (n-k)}{n}}$$

Für beide Formeln sind hier die Faktoren c und c^2 in Abhängigkeit von der Irrtumswahrscheinlichkeit α aufgeführt.

α	0,01	0,05	0,10	0,20
c	2,58	1,96	1,64	1,28
c^2	6,66	3,84	2,69	1,64

Beispiel

> In einer klinischen Studie wurde an 150 Patienten die Dekubitusprävalenz ermittelt. Dabei wurde eine Prävalenz von 35% ermittelt. Berechnen Sie das 95% Konfidenzintervall für die Dekubitusprävalenz.

Zunächst muss der Wert k für 35% von 150 ermittelt werden. Dies geht am einfachsten, indem man $150 \cdot 0,35$ rechnet. So erhalten wir $k = 52,5$
Jetzt werden die Werte entsprechend in die Formel eingesetzt:

$$\begin{aligned}
p_o &= \frac{52,5}{150} + \frac{1,96}{150} \cdot \sqrt{\frac{52,5 \cdot (150 - 52,5)}{150}} \\
&= 0,35 + 0,0131 \cdot \sqrt{\frac{52,5 \cdot 97,5}{150}} \\
&= 0,35 + 0,0131 \cdot \sqrt{34,125} \\
&= 0,35 + 0,0131 \cdot 5,8417 = 0,35 + 0,0765 \\
&= 0,4265 = \mathbf{42,65}\% \\
p_u &= 0,35 - 0,0765 \\
&= 0,2735 = \mathbf{27,35}\%
\end{aligned}$$

Das Konfidenzintervall erstreckt sich demnach von 27,35% bis 42,65%

D. ANHANG D - Übungsaufgaben

D-1. z-Transformation

Zur z-Transformation siehe auch Kapitel II auf Seite 25. Die Lösungen finden sich auf Seite 75.

I. Aufgabe 1

Man nimmt an, dass die Ergebnisse eines Tests sich um einen Mittelwert von $\mu = 60$ verteilen und eine Streuung von $\sigma = 20$ aufweisen. Welcher prozentuale Anteil der möglichen Testergebnisse lieg bei...

1. über 85?
2. unter 50?

II. Aufgabe 2

Die Verteilung der Ergebnisse des letzten Statistiktests ist normalverteilt mit einem Mittelwert von $7,2$ und einer Standardabweichung von $3,5$. Wie groß ist die Wahrscheinlichkeit,

1. mit mindestens 10 Punkten zu bestehen,
2. mit der Punktzahl zwischen 5 und 8 zu liegen,
3. höchstens 4 Punkte zu erreichen?

E. ANHANG E - Lösungen

E-1. z-Transformation

Die Übungsaufgaben zur z-Transformation befinden sich auf Seite 74.

I. Aufgabe 1

Teil 1

Mittels der Formel 1.14 berechnen wir den z-Wert für den Messwert 85:

$$z_i = \frac{x_i - \mu}{\sigma}$$

$$z_i = \frac{85 - 60}{20} = 1,25$$

Da wir Anteile errechnen sollen, die *mindestens* ein Testergebnis von 85 aufweisen, muss in Referenztabelle B.2 der Wert für $z = 1,25$ abgelesen werden; in diesem Fall liest man den Wert $0,1056$ ab. Das bedeutet, dass ein Anteil von 10,56% ein Testergebnis von mindestens 85 erzielt.

Teil 2

Mittels der Formel 1.14 berechnen wir den z-Wert für den Messwert 50:

$$z_i = \frac{x_i - \mu}{\sigma}$$

$$z_i = \frac{50 - 60}{20} = -0,5$$

In diesem Beispiel erhalten wir einen negativen z-Wert. Das Problem besteht darin, dass Tabellen (wie z.B. Tabelle B.2) nur positive z-Werte auflisten. Hier macht man sich die Symmetrie der Normalverteilung zu eigen. Durch die Parabelform ergibt sich der Zusammenhang:

$$P(Z < -0,5) = P(Z > +0,5)$$

Man sucht also ‘‘einfach’’ die Wahrscheinlichkeit für den z-Wert $+0,5$, und zieht diesen Wert von 1 ab. Da wir Anteile errechnen sollen, die *höchstens* ein Ergebnis von 50 erzielen, muss in Tabelle B.1 der Wert für $z = 0,5$ abgelesen werden; in diesem Falle liest man dort 0.6915 ab. Dieser Wert wird nun von 1 abgezogen:

$$1 - 0,6915 = 0,3085$$

Somit liegt der Anteil, der maximal ein Testergebnis von 50 erreicht, bei 30.85%.

II. Aufgabe 2

Teil 1

Mittels der Formel 1.14 berechnen wir den z-Wert für den Messwert 10 Punkte:

$$z_i = \frac{x_i - \mu}{\sigma}$$

$$z_i = \frac{10 - 7,2}{3,5} = 0,8$$

Da wir Anteile errechnen sollen, die *mindestens* ein Testergebnis von 10 Punkten aufweisen, muss in Referenztabelle B.2 der Wert für $z = 0,8$ abgelesen werden; in diesem Fall liest man den Wert $0,2119$ ab. Das bedeutet, dass ein Anteil von 21,19% ein Testergebnis von mindestens 10 Punkten erzielt.

Teil 2

Zunächst berechnet man die z-Werte für die Punktzahlen 5 und 8:

$$z_i = \frac{5 - 7,2}{3,5} = -0,63$$

$$z_i = \frac{8 - 7,2}{3,5} = 0,23$$

Für den z-Wert von 5 Punkten ($= -0,63$) muss der entsprechenden Wert aus Tabelle B.2 abgelesen werden, da wir die Anteile jener Schüler benötigen, die *mindestens* 5 Punkte erhalten. Da es sich um einen *negativen* z-Wert handelt, lesen wir den Wert bei $+0,63$ ab, und subtrahieren ihn von 1:

$$1 - 0,2643 = 0,7357$$

Für den z-Wert von 8 Punkten ($= 0,23$) muss der entsprechenden Wert aus Tabelle B.1 abgelesen werden, da wir die Anteile jener Schüler benötigen, die *maximal* 8 Punkte erhalten. So erhalten wir den Wert $0,5910$.

Zu guter letzt wird von den Werten der Obergrenze (8 Punkte) die Werte der Untergrenze (5 Punkte) abgezogen:

$$p = P(z < 0,23) - P(z < -0,63)$$

$$p = 0,5910 - 0,7357 = -0,1447$$

Der Prozentanteil liegt somit bei 14,47%.

Teil 3

Mittels der Formel 1.14 berechnen wir den z-Wert für den Messwert 4 Punkte:

$$z_i = \frac{x_i - \mu}{\sigma}$$

$$z_i = \frac{4 - 7,2}{3,5} = -0,91$$

Da wir Anteile errechnen sollen, die *höchstens* ein Testergebnis von 4 Punkten aufweisen, muss in Referenztabelle B.1 der Wert für $z = -0,91$ abgelesen werden. Da die Referenztabelle B.1 keine negativen z-Werte auflistet, macht man sich die Symmetrie der Parabelform zu nutze:

$$P(Z < -0,5) = P(Z > +0,5)$$

Man sucht also ‘‘einfach’’ die Wahrscheinlichkeit für den z-Wert $+0,91$ (in diesem Falle $0,8186$), und zieht diesen Wert von 1 ab.

$$1 - 0,8186 = 0,1814$$

Das bedeutet, dass ein Anteil von 18,14% ein Testergebnis von höchstens 4 Punkten erzielt.

Index