

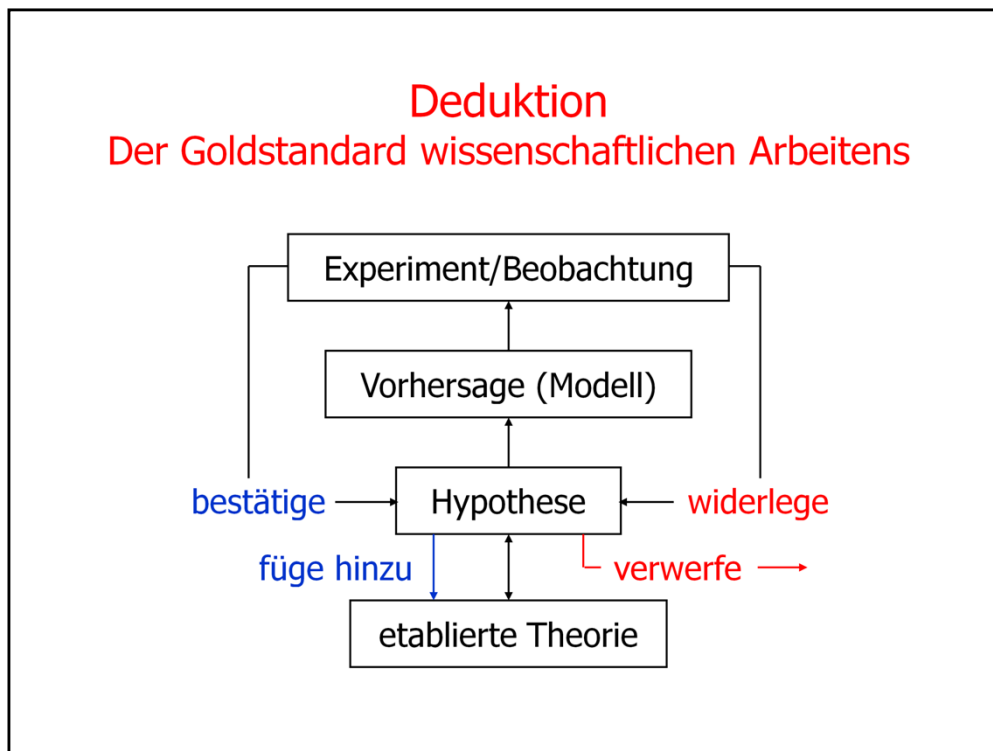
Einführung



A posse ad esse

Das Wissenschaftlerbild vieler Nicht-Wissenschaftler ähnelt dem, das Jerry Lewis als "Verrückter Professor" vermittelt. Es spiegelt die Ansicht einer Vielzahl von Laien wider, dass aufregende wissenschaftliche Entdeckungen durch Geniestreiche zustande kommen, also durch exploratives Arbeiten, das als einziger Motivationsquelle der Neugier des Wissenschaftlers bedarf. Diese stereotype Sicht mag für witzige Hollywoodfilme taugen, mit der Realität hat sie aber wenig zu tun. Wissenschaftliche Forschung ist harte Arbeit, die meistens von Hypothesen getrieben wird, die wiederum der kritischen Bewertung existierenden Wissens entspringen und nicht höherer Eingebung.

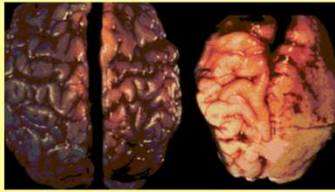
Das Hauptziel medizinisch-wissenschaftlicher Forschung besteht darin, die Wissensbasis für die Diagnose, Prävention und Therapie von Krankheiten zu verbreitern. Die hierbei zur Anwendung kommenden Methoden und Qualitätskriterien sollten daher für jede im medizinischen Bereich tätige Person von höchstem Interesse sein. Dies gilt insbesondere, weil das Wohlbefinden und unter Umständen sogar das Leben jedes einzelnen Patienten von der Urteilsfähigkeit und dem Wissensstand des behandelnden Arztes abhängt.



Im Wesentlichen gibt es zwei Wege, um wissenschaftliche Forschung zu betreiben. Der griechische Philosoph Aristoteles erkannte im 4. Jahrhundert vor Christus wahrscheinlich als einer der Ersten, dass Wissen dadurch zustande kommen kann, dass logische Schlüsse aus empirischen Beobachtungen gezogen werden. Im 19. Jahrhundert führte so z.B. die Beobachtung, dass an Kuhpocken erkrankte Personen sich nie mit schwarzen Pocken infizieren, zu der Hypothese, dass erstere eine Immunisierung gegen letztere bewirkt. Dieser "Induktion" genannte Ansatz war in der Tat die bevorzugte Art und Weise, wie bis zur Mitte des 20. Jahrhunderts wissenschaftliche Forschung betrieben wurde. Induktion ist unmittelbar einleuchtend und reflektiert, wie Menschen alltäglich die sie umgebende Welt kennen lernen. Allerdings sind per Induktion gemachte Verallgemeinerungen nur selten allgemein gültig, da sie stark von der Art und Entstehung der ihnen zugrunde liegenden Daten abhängen.

Der alternative Forschungsansatz, "Deduktion" genannt, wurde ebenfalls zuerst von Aristoteles beschrieben. Die Anwendbarkeit von Deduktion ist jedoch auf solche Disziplinen beschränkt, die einen gewissen (meist durch Induktion erreichten) Entwicklungsstand aufweisen. Das liegt daran, dass Deduktion etabliertes Wissen voraussetzt, aus dem sich neue Hypothesen ableiten lassen, die sich wiederum durch geeignete Experimente bzw. Beobachtungen bestätigen oder widerlegen lassen müssen. Ob Kuhpocken eine Infektion mit schwarzen Pocken verhindern, lässt sich durch epidemiologische Studien überprüfen. Deduktion bietet daher wenig Raum für Subjektivität; zwei Wissenschaftler sollten mit dem gleichen Experiment zum gleichen Ergebnis kommen. Es war genau dieses Fehlen von Subjektivität, das die Deduktion zum "Goldstandard" der modernen medizinischen Forschung werden ließ.

Alkoholkonsum und Schwangerschaft



www.come-over.to

- Theorie:** Alkoholkonsum wirkt teratogen.
- Hypothese:** Alkoholkonsum während der Schwangerschaft beeinflusst die Hirnentwicklung.
- Vorhersage:** Während der Schwangerschaft exponierte Kinder zeigen vermehrt Verhaltensprobleme.
- Beobachtung:** Ermittlung und Vergleich der Häufigkeit von Verhaltensauffälligkeiten unter jeweils 100 exponierten und nicht exponierten Kindern.

Alkohol ist eine teratogen wirkende Substanz, d.h. sie behindert die normale Entwicklung des ungeborenen Kindes. Alkoholkonsum während der Schwangerschaft kann insbesondere das Gehirn des Feten stark schädigen. Welche Areale betroffen sind, hängt davon ab, welche Teile des Gehirns sich während der Zeit des Alkoholkonsums der Mutter gerade entwickelt haben. Da sich das Gehirn und das zentrale Nervensystem während der gesamten Schwangerschaft ausbilden, ist das Gehirn das am meisten durch Alkoholkonsum der Schwangeren gefährdete Organ des Kindes. Allerdings lassen sich nicht alle durch Alkohol verursachten Schäden z.B. mittels Scanner-Untersuchung oder gar im Zuge einer Autopsie entdecken. Viele derartige Läsionen sind für einen direkten Nachweis zu klein, können aber dessen ungeachtet zu schweren neurologischen oder psychiatrischen Problemen führen.

Die Frage, inwieweit die relative Häufung von Verhaltensauffälligkeiten in einer konkreten Stichprobe von jeweils 100 exponierten und nicht exponierten Kindern tatsächlich die Hypothese der Teratogenität des Alkoholkonsums stützt, lässt sich wissenschaftlich valide nur mit statistischen Methoden beantworten. Reicht z.B. ein Verhältnis von 20:10, oder muss der Unterschied deutlich größer ausfallen?

Was ist Statistik ?

Etymologie

lat. *statisticus* ("staatswissenschaftlich")

Definition

"Statistik bezeichnet die Mathematik des Sammelns, Ordnen und Interpretierens numerischer Daten, insbesondere zur Analyse von Populationscharakteristika auf der Grundlage von Stichproben."

American Heritage® Dictionary

Statistik ist ein Teilgebiet der Wahrscheinlichkeitstheorie, die im Kontext der Naturwissenschaften wiederum eine eigenständige mathematische Disziplin bildet, vergleichbar der Algebra oder der Geometrie. Interessante Sichtweisen von Wahrscheinlichkeit wurden aber auch in anderen Wissenschaftsbereichen entwickelt, wie das folgende Zitat des britischen Philosophen John Locke (1632-1704) belegt: "Wahrscheinlichkeit ist die Augenscheinlichkeit der Übereinstimmung oder Nicht-Übereinstimmung zweier Ideen auf der Grundlage von Beweisen, deren Beziehung zur Idee nicht konstant ist, sondern über weite Teile nur so erscheint."

Und Sie dachten, Mathematik bzw. Statistik wäre kompliziert?

Was ist Statistik ?

(fast) ein Kursplan

Zusammenfassung experimenteller und observationaler Daten

Deskriptive Statistik

(Graphiken, Tabellen, Plots etc.)

Schätzung von Behandlungseffekten

Schätztheorie

(Mittelwert, Standardabweichung, Konfidenzintervall etc.)

Entscheidungsfindung auf der Basis empirischer Beobachtungen

Hypothesentests

(Signifikanz, Power, P-Werte etc.)

Modellierung des Zusammenhangs von Faktoren

Assoziations- und Regressionsanalyse

(Korrelationskoeffizient, lineare Regression etc.)

In der empirischen Forschung dient die Statistik meist einem von vier verschiedenen Zwecken. Zunächst sollten Rohdaten grundsätzlich angemessen zusammengefasst und präsentiert werden, was die Auswahl der richtigen Methoden für deren weitere Analyse erleichtert bzw. erst ermöglicht. Meistens sind Forscher dann entweder daran interessiert, das Ausmaß bekannter Effekte oder Parameter (z.B. Remissionsraten) zu schätzen, oder sie möchten nachweisen, dass eine bestimmte Beobachtung "statistisch signifikant" war, d.h. sich nur schlecht durch Zufall erklären lässt. Statistisches Modellieren wiederum stellt eine fortgeschrittene Methode der Datenanalyse dar, die sich sowohl explorativ als auch konfirmatorisch einsetzen lässt. Die Durchführbarkeit und Sinnhaftigkeit statistischer Modellbildung setzt allerdings stets ein gewisses Vorwissen hinsichtlich der relevanten Modellvariablen voraus.

Wozu Statistik ? für Medizinstudenten

kritische **Bewertung** von Beobachtungen

Berücksichtigung des Zufalls

Ergänzung des "gesunden Menschenverstands"

Verständnis der wissenschaftlichen Literatur

Verstehen von Zweck und Design

Beurteilung der wissenschaftlichen Glaubwürdigkeit

Design und **Durchführung** von Experimenten

Wahl der angemessenen statistischen Methode

Wahl des angemessenen Studiendesigns

richtiger Gebrauch von Formeln und Software

richtige Interpretation von Resultaten

Die grundlegenden Prinzipien der Statistik sollten jedem gebildeten Mitglied einer modernen Gesellschaft geläufig sein. Ständig werden um uns herum kritische Entscheidungen getroffen und wichtige Veränderungen in Gang gesetzt, die auf scheinbar "systematischen", "objektiven" oder "wissenschaftlichen" Erkenntnissen fußen. Jede Woche berichten die Massenmedien über hunderte neuer wissenschaftlicher Studien, die alle einen statistischen Hintergrund für sich in Anspruch nehmen. Für Personen, die ein aktives Interesse an der politischen, technologischen und ethischen Entwicklung ihrer Umgebung haben, ist die kritische Auseinandersetzung mit solchen Informationen unerlässlich. Eine Voraussetzung hierfür ist jedoch z.B. die Fähigkeit, statistische Signifikanz von praktischer Bedeutsamkeit zu unterscheiden, mögliche Verfälschungen von Daten zu erkennen, zwischen dem Mangel an statistischer "Power" und dem Fehlen echter Effekte zu unterscheiden, Ursache und Wirkung auseinander zu halten, sowie die wissenschaftliche Sicht von Zufall und Variabilität in Grundsätzen zu verstehen.

Für Medizinstudenten ist ein gutes Verständnis der Statistik Voraussetzung für die kritische Bewertung der Ergebnisse biomedizinischer Forschung. Selbst wenn sie sich später "nur" niederlassen wollen und nie selbst medizinische Experimente durchführen werden, müssen sich Ärzte über aktuelle Fortschritte auf den Gebieten der Epidemiologie, der Therapie, der Pflege und der wirtschaftlichen Folgen wichtiger Erkrankungen auf dem Laufenden halten. Ansonsten riskieren sie das Wohlbefinden ihrer Patienten.

Evidenzbasierte Medizin

"... das systematische Auffinden, Bewerten, und Nutzen aktueller wissenschaftlicher Ergebnisse als Grundlage klinischer Entscheidungen."

Rosenberg W, Donald A (1995). Evidence based medicine: an approach to clinical problem solving. BMJ 310: 1122-1126

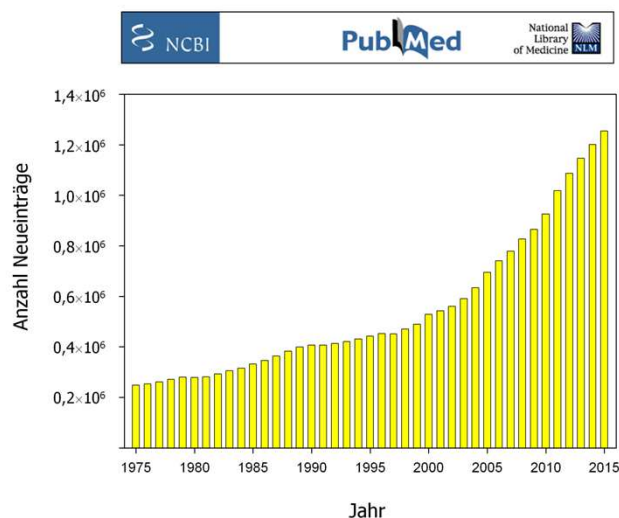
"... verlangt, dass Sie den richtigen Artikel zur richtigen Zeit lesen und anschließend Ihr Verhalten (und, was oft viel schwieriger ist, das Verhalten anderer Leute) dementsprechend ändern"

Greenhalgh T (1997). How to read a paper: the basics of evidence based medicine. BMJ, London

"Evidenzbasierte Medizin" bedeutet, sich bei der individuellen Versorgung von Patienten auf die beste, aktuell verfügbare wissenschaftliche Information zu stützen. In der Praxis setzt dies die Integration individueller klinischer Erfahrung mit externer klinischer Evidenz voraus. Unter "individueller klinischer Erfahrung" ist in diesem Zusammenhang die Leistungs- und Urteilsfähigkeit zu verstehen, die sich ein Mediziner im Laufe seines Berufslebens aneignet. Sie äußert sich in der Verlässlichkeit und Effizienz seiner Diagnosen und in der Art, wie er auf die Nöte, Bedürfnisse und Rechte seiner Patienten eingeht. "Externe klinische Evidenz" ist das Produkt klinisch relevanter Forschung, oft im Bereich der Grundlagenwissenschaften, meist jedoch in Form patientenzentrierter klinischer Studien zur Bewertung der Aussagekraft diagnostischer Tests oder der Wirksamkeit und Sicherheit therapeutischer, rehabilitativer oder präventiver Maßnahmen. Externe klinische Evidenz erlaubt das Verwerfen vormals akzeptierter Test- und Behandlungsverfahren zu Gunsten effektiverer, effizienterer, genauerer oder kostengünstigerer Methoden.

Gute Ärzte verwenden beides, individuelle klinische Erfahrung und externe klinische Evidenz, denn keines von beiden ist für sich genommen für eine optimale Versorgung der Patienten hinreichend. Ohne klinische Erfahrung besteht die Gefahr, dass die klinische Praxis durch die wissenschaftliche Forschung "tyrannisiert" wird, denn auch exzellente externe Evidenz muss nicht immer auf den individuellen Patienten anwendbar sein. Ohne die Nutzung der besten, aktuell verfügbaren externen Evidenz wäre die klinische Praxis aber sehr schnell überholt, was wiederum den Patienten zum Nachteil gereichen würde.

Evidenzbasierte Medizin



Der Umfang der praktisch relevanten Publikationen im Bereich der medizinischen Forschung wächst beständig. Zwischen 1975 und 2015 hat sich das jährliche Publikationsaufkommen in relevanten biomedizinischen Journalen fast **verfünffacht**, und ein Abklingen dieses Trends ist derzeit nicht in Sicht. Das meiste der dabei erzeugten Information ist in elektronischer Form verfügbar und kann von so genannten "Wissensservern" wie z.B. PubMed, einem Service der US-amerikanischen medizinischen Nationalbibliothek, abgerufen werden. Dessen ungeachtet bedeutet die Geschwindigkeit, mit der sich die Generierung neuen Wissens derzeit vollzieht, dass die mit dem Paradigma der Evidenzbasierten Medizin einhergehende Verpflichtung zur Aktualität der genutzten Information für den einzelnen Arzt schon bald nicht mehr zu erfüllen sein wird.

Glücklicherweise mildern systematische Reviews unterschiedlichster Art und Umfangs die geschilderte Problematik etwas ab. So wird beispielsweise die Durchführung und Veröffentlichung klinischer Studien ständig durch die so genannte "Cochrane Collaboration" verfolgt, eine internationale non-profit Organisation mit dem Ziel, aktuelle und verlässliche Informationen über die Rahmenbedingungen der Krankenversorgung zeitnah verfügbar zu machen. Die Cochrane Collaboration wurde 1993 gegründet und hat seitdem ca. **5300** systematische Reviews neuer medizinischer Interventionen erstellt und verbreitet. Sie führt darüber hinaus auch Wirtschaftlichkeits- und Technologiebewertungen im Gesundheitswesen durch.

Variation in Populationen

Gegenstand der Statistik ist die Untersuchung von Art und Variabilität eines gegebenen Merkmals in einer Population.

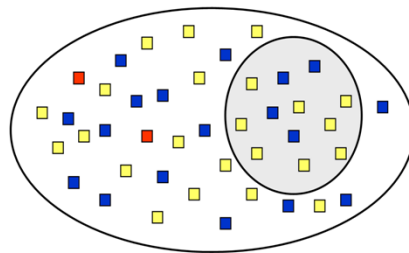
- Welches sind die häufigen oder durchschnittlichen Werte des Merkmals?
- Wie stark variiert das Merkmal?
- Gibt es bezüglich des Merkmals distinkte Teilpopulationen?
- Welches sind die Normalwerte des Merkmals?
- Sind abnormale Werte des Merkmals pathologisch ?
- Ist das Merkmal mit anderen Merkmalen assoziiert ?

In der biomedizinischen Forschung zielen die meisten wissenschaftlichen Fragen in der einen oder anderen Weise auf die Variation eines gegebenen Merkmals innerhalb oder zwischen Individuen ab. Dabei kann es sich um die Reaktion auf ein Medikament oder einen Umweltfaktor handeln, um ein normales oder experimentell verändertes physiologisches Merkmal oder um die genetische Ursache einer Erkrankung. Die Datengrundlage solcher Studien (d.h. die "Stichprobe") enthält üblicherweise Wiederholungen ein und derselben Messung, die zu unterschiedlichen Zeiten, unter geänderten Bedingungen oder an verschiedenen Individuen gemacht wurden. Der wichtigste Aspekt der Messungen ist ihre Variation zwischen den Wiederholungen (natürlich unter gleichen Randbedingungen) und die Tatsache, dass ein besseres Verständnis eben dieser Variation zu einem besseren Verständnis des ursprünglichen wissenschaftlichen Problems führt.

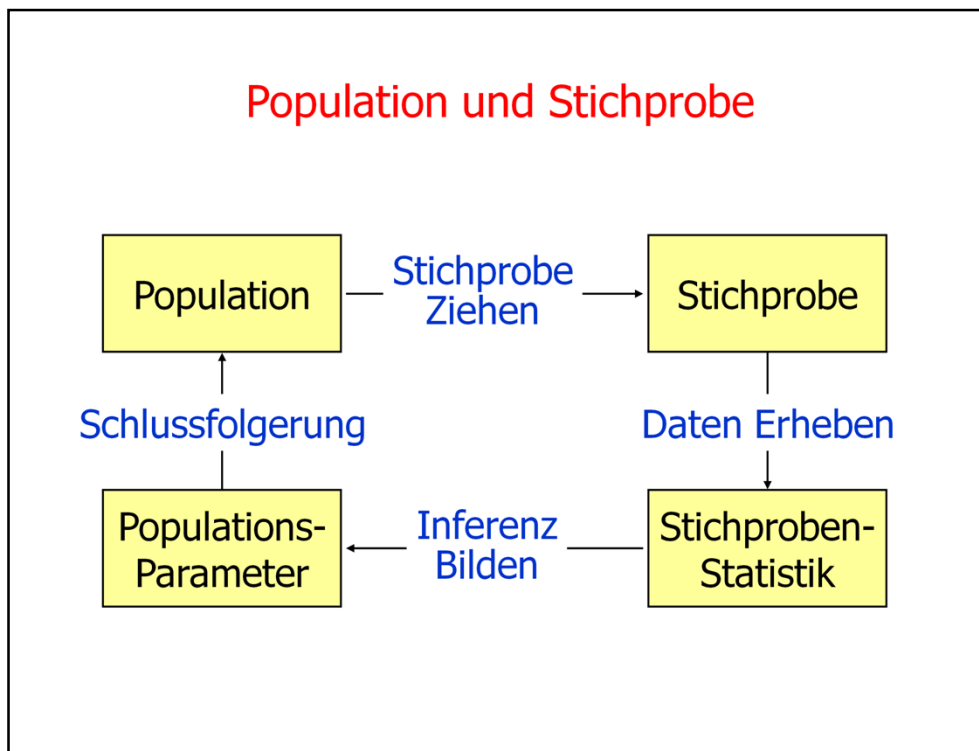
Population und Stichprobe

Eine **Population** ist eine (üblicherweise große) Gruppe von Objekten, über die eine Aussage getroffen werden soll.

Eine **Stichprobe** besteht aus einer (üblicherweise kleinen) Anzahl von Mitgliedern der Population.



Wissenschaftliche Experimente sind üblicherweise so angelegt, dass sie das Ziehen von Schlussfolgerungen über eine große Population auf der Grundlage kleiner Stichproben aus eben dieser Population erlauben. Im biomedizinischen Kontext bezieht sich der Begriff "Population" entweder auf einen tatsächlich existierenden Personenkreis ("alle Knaben der lokalen Schule") oder auf eine hypothetische Gruppe von Menschen ("alle weißen Patienten im gleichen Alter, mit dem gleichen Blutdruck und den gleichen Essgewohnheiten"). Für einen Statistiker hat der Begriff "Population" sogar eine noch viel weiter reichende Bedeutung und kann sich z.B. auf "alle Kohlenstoffatome im Universum", "alle Säugerlungenzellen unter oxidativem Stress" oder "alle am Montagmorgen produzierten Autos" beziehen.



Ziel vieler wissenschaftlicher Studien ist die Charakterisierung eines Populationsparameters. So kann sich ein Forscher z.B. für den physiologischen Zustand von Schulkindern interessieren, die sich nicht regelmäßig sportlich betätigen. Wenn diese Kinder zu Übergewicht neigen, d.h. einen durchschnittlichen Body-Mass-Index (BMI) von über 25 haben, dann könnte der Forscher daraus schließen, dass Bewegungsarmut bei Schulkindern zu Übergewicht führt.

In der Praxis ist es meistens weder möglich noch notwendig, alle Kinder einer Gemeinde oder eines ganzen Landes zu untersuchen. Stattdessen wird aus der fraglichen Population (z.B. den bewegungsarmen Schulkindern) eine Stichprobe gezogen und in dieser die relevanten Daten (z.B. Gewicht und Größe) erhoben. Unter der Bedingung, dass die Stichprobe groß genug und für die fragliche Population "repräsentativ" ist, wird der durchschnittliche BMI der Stichprobe ungefähr dem durchschnittlichen BMI der gesamten Population gleichen. Sind die genannten Bedingungen erfüllt, so kann der Forscher also den Populationsparameter aus der Stichprobenstatistik erschließen.

Weiblicher Body-Mass-Index (BMI)

unterernährt (WHO): BMI<18.5, untergewichtig (NHANES): BMI<19.1
US-Durchschnitt 20-29 Jahre: BMI=22.7

Jahr	Name	BMI
1984	Suzette Charles	17.7
1985	Sharlene Wells	18.2
1986	Susan Akin	16.8
1987	Kellye Cash	17.6
1988	Kaye Lani Rae Rafko	18.8
1989	Gretchen Carlson	19.1
1990	Debbie Turner	17.9
1991	Marjorie Vincent	17.8
1998	Kate Shindle	20.2
1999	Nicole Johnson	19.6
2001	Angela Perez Baraquio	20.3
2002	Katie Harman	19.5

http://www.pbs.org/wgbh/amex/missamerica/sfeature/sf_list.html

Wenn man die vorliegende Liste weiblicher BMIs mit den Grenzwerten für Unterernährung und Untergewicht der WHO bzw. des National Health and Nutrition Examination Survey (NHANES) vergleicht, so zeigt sich schnell, dass die zugehörige Gruppe von Frauen nicht repräsentativ für die US-amerikanische "Durchschnittsfrau" im Alter von 20 bis 29 Jahren sein kann. Als Miss Americas der 1980-iger und 1990-iger Jahre repräsentieren diese Frauen nämlich mit Sicherheit eher die Vorstellung vieler Menschen von Gesundheit und Schönheit als das reale Leben.

Deskriptive Statistik

Die **deskriptive Statistik** dient dazu, Daten in verständlicher Weise zusammenzufassen.

Graphische Methoden sind besser als numerische Methoden dazu geeignet, Muster in Daten zu erkennen. **Numerische Ansätze** sind meistens präziser and objektiver.

Wie der Name schon sagt, geht es bei der "deskriptiven" Statistik darum, Dinge zu beschreiben und nicht darum, Schlüsse über Dinge zu ziehen. Der Mittelwert ist eben zunächst ein gutes Maß dafür, wo die Beobachtungen einer Stichprobe ungefähr angesiedelt sind. Ob und inwieweit das Stichprobenmittel auch eine gute Schätzung des Populationsmittels ist, steht auf einem ganz anderen Blatt. Dieser Frage werden wir uns im Abschnitt zum Thema "Parameterschätzung" zuwenden.

Variablentypen

Qualitative Variable

- **nominal**: Kategorien ohne Anordnung
(z.B. Geschlecht, ABO-Blutgruppe)
- **ordinal**: Kategorien mit Anordnung
(z.B. verbale Examensnoten, Tumor-Stadien)

Quantitative Variable

- **diskret**: ganze Zahlen
(z.B. Zellzahl, Nachkommenzahl)
- **stetig**: reelle Zahlen
(z.B. Körpergewicht, Blutdruck)

Oftmals werden ordinale qualitative Merkmale zusätzlich mit einem „Score“ belegt. Ein klassisches medizinisches Beispiel hierfür ist der Apgar-Score zur Bewertung der Vitalität von Neugeborenen. Definitionsgemäß handelt es sich bei Scores um numerische Werte, so dass Scores auch als quantitative Variable aufgefasst werden können. Entscheidend für die Sinnhaftigkeit einer solchen Transformation ist jedoch die Vergleichbarkeit von einzelnen bzw. paarweisen Merkmalsausprägungen mit gleichen Scores oder Score-Differenzen.

Die Unterscheidung zwischen diskreten und stetigen quantitativen Variablen ist manchmal problematisch. Im Prinzip bedeutet Stetigkeit, dass zwischen zwei beliebige Messwerte theoretisch noch unendlich viele andere Messwerte passen. Diese Annahme ist jedoch im Allgemeinen unrealistisch, da die Auflösung der Messung quantitativer Variabler nun einmal begrenzt ist. So wird z.B. die Körpergröße eines Menschen in der Praxis nur auf den Zentimeter genau gemessen. Streng genommen wäre sie daher eine diskrete Variable. Da sich stetige Variable aber wesentlich einfacher statistisch analysieren lassen als diskrete, werden Wissenschaftler die Körpergröße von Personen immer als stetige Variable behandeln. Die praktischen Konsequenzen dieser Approximation sind meistens ohnehin vernachlässigbar.

Jede stetige Variable ist quantitativ, doch nicht jede diskrete Variable ist automatisch rein qualitativ. Die Anzahl der Kinder eines Ehepaars ist z.B. sowohl diskret als auch quantitativ, wohingegen das Geschlecht des jüngsten Kindes eine qualitative Variable ist. Selbstverständlich sind jedoch alle quantitativen Merkmale auch ordinal skaliert.

Deskriptive Statistik

Stichprobencharakteristika

Stichprobengröße	n	
Beobachtungen	x_i	$i = 1 \dots n$
qualitative Daten		
Anzahl Kategorien	k	
absolute Häufigkeiten	h_i	$i = 1 \dots k$
relative Häufigkeiten	$p_i = \frac{1}{n} \cdot h_i$	

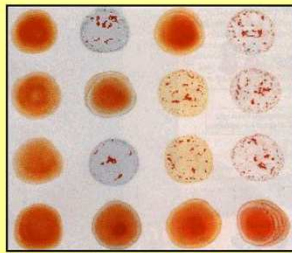
Oh je! Formeln! Aber keine Angst: Mathematische Formeln sind nur ein probates Mittel, um einfache Sachverhalte effizient darzustellen.

Schauen wir uns zunächst die Beobachtungen in unserer Stichprobe an. Wenn diese Beobachtungen bekannt sind, fein! Wenn nicht, und wenn wir trotzdem über sie reden möchten, dann nennen wir sie einfach "x". Also kann x eine Zahl sein, wie z.B. 9, oder eine Farbe, wie grün. Üblicherweise gibt es in einer Stichprobe mehr als eine Beobachtung, so dass wir für jedes x noch klarstellen müssen, welche Beobachtung damit gemeint ist. In der Mathematik verwendet man hierzu Indizes. Wir bezeichnen also mit " x_i " die i-te Beobachtung, wobei i eine ganze Zahl zwischen 1 und der Stichprobengröße ist. Wenn die Stichprobengröße bekannt ist, wieder fein! Wenn nicht, nennen wir sie einfach "n".

Haben wir es mit einer Stichprobe qualitativer Daten zu tun, dann ist die Anzahl der Kategorien von Bedeutung, in die eine einzelne Beobachtung fallen kann. Wenn diese Anzahl bekannt ist, wieder fein! Wenn nicht, nennen wir sie "k". Für den Moment kann k alles zwischen 2 (wie im Fall des Geschlechts) und ... nun ja, irgendetwas sein. Jede Kategorie enthält eine bestimmte Anzahl h_i von Beobachtungen aus unserer Stichprobe. Wir benutzen also " h_i " zur Bezeichnung der Häufigkeit von Beobachtungen in der i-ten Kategorie, wobei i eine ganze Zahl zwischen 1 und k ist.

AB0-Blutgruppen

$k=4$



0 A B AB

	0	A	B	AB
i	1	2	3	4
h_i	28	32	10	5
p_i	0.373	0.427	0.133	0.067
$100 \cdot p_i$	37.3	42.7	13.3	6.7

$n=75$

Datenbeschreibung

Lagemaße

Typ	Lagemaße	Streuungsmaße
qualitativ <i>nominal</i>	Modalwert	Index Qualitativer Variation (IQV)
qualitativ <i>ordinal</i>	Modalwert Quantile (Median)	IQV Interquartilabstand
quantitativ	Quantile (Median) Mittelwert	Interquartilabstand Standardabweichung Mediandeviation

Für einen Forscher sind üblicherweise zwei Charakteristika seiner Daten von näherem wissenschaftlichen Interesse, nämlich die Lage der Daten ("Von welcher Größenordnung sind meine Daten?") und deren Streuung ("Wie stark variieren meine Daten?"). Für nominale Daten lässt sich die Lage einer Stichprobe durch deren häufigste Beobachtung, den so genannten "Modalwert", messen. Sind die Daten auf einer ordinalen Skala angeordnet, so könnte die Anzahl der Beobachtungen ober- oder unterhalb eines bestimmten Werts von Interesse sein, oder man sucht gezielt nach dem Wert ("Median" genannt), der die Stichprobe in zwei Hälften teilt, d.h. der größer bzw. kleiner als jeweils die Hälfte der Beobachtungen ist. Zu guter Letzt ist natürlich auch der Mittelwert ein gutes Maß für die Lage einer Stichprobe.

Lagemaße

qualitative Daten

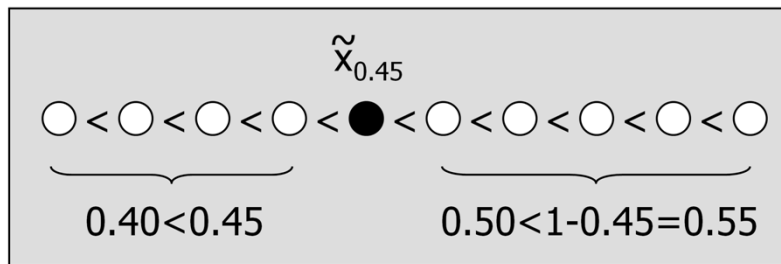
Der **Modalwert** einer Stichprobe ist deren (nicht immer eindeutig bestimmter) häufigster Wert.

	0	A	B	AB
i	1	2	3	4
h_i	28	32	10	5
p_i	0.373	0.427	0.133	0.067
$100 \cdot p_i$	37.3	42.7	13.3	6.7

Lagemaße

ordinale und quantitative Daten

Das α -Quantil \tilde{x}_α einer Stichprobe ist der Wert, für den die relative Häufigkeit kleinerer Werte höchstens α und die relative Häufigkeit größerer Werte höchstens $1-\alpha$ beträgt.



Die hier angegebene Definition der Quantile mag vielleicht ein wenig "sperrig" wirken. Es handelt sich beim Quantil aber um einen der wenigen mathematischen Begriffe, der sich durch eine narrative Darstellung einfacher erläutern lässt, als durch die zugehörige Formel.

Wenn Ihnen also einleuchtet, dass die markierte Beobachtung nicht nur das 0.45-Quantil, sondern auch das 0.41-Quantil (bzw. das 0.49-Quantil) ist, dann haben Sie die Sache im Kern verstanden.

Weiblicher Body-Mass-Index (BMI)

$$\alpha = 0.40$$

BMI
17.7
18.2
16.8
17.6
18.8
19.1
17.9
17.8
20.2
19.6
20.3
19.5

Sortieren
→

[BMI]
16.8
17.6
17.7
17.8
17.9
18.2
18.8
19.1
19.5
19.6
20.2
20.3

$$\left. \begin{array}{l} 17.6 \\ 17.7 \\ 17.8 \end{array} \right\} \frac{4}{12} = 0.33 \leq 0.40$$

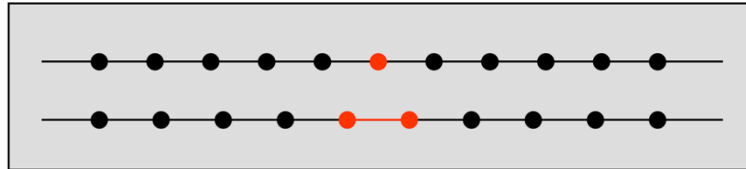
$$\tilde{x}_{0.40} = 17.9$$

$$\left. \begin{array}{l} 18.2 \\ 18.8 \\ 19.1 \\ 19.5 \\ 19.6 \end{array} \right\} \frac{7}{12} = 0.58 \leq 0.60$$

Lagemaße

ordinale und quantitative Daten

$\tilde{x}_{0.50}$ wird als **Median** bezeichnet.



$\tilde{x}_{0.25}(\tilde{x}_{0.50}, \tilde{x}_{0.75})$ heißt 1. (2., 3.) **Quartil**

Das α -Quantil einer Stichprobe ist unter zwei Bedingungen eindeutig definiert:

1. Es gibt eine Beobachtung, für die der Stichprobenanteil kleinerer oder gleich großer Werte strikt kleiner als α ist.
2. Die relative Häufigkeit von Beobachtungen, die kleiner oder gleich der nächst größeren Beobachtung sind, ist strikt größer als α .

Im vorliegenden Beispiel sind diese beiden Bedingungen für den Median (d.h. das 0.50-Quantil) der 11 Beobachtungen in der oberen Messreihe erfüllt. Häufig ist das α -Quantil jedoch uneindeutig, was z.B. für den Median der unteren Messreihe zutrifft. Der 5. und 6. Wert sowie alle dazwischen liegenden Werte erfüllen die Definition des Medians. Leider gibt es keine verbindliche Regel für den Umgang mit solchen Mehrdeutigkeiten. Wenn die Daten quantitativ sind, lässt sich sinnvoller Weise der Mittelwert des größten und des kleinsten in Frage kommenden Wertes als Median verwenden. Für qualitative Daten bietet sich die Beobachtung an, die die Definition des Medians am ehesten erfüllt, d.h. für die die entsprechenden relativen Häufigkeiten am wenigsten von α und $1-\alpha$ abweichen.

Apgar-Score

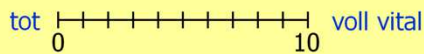
Apgar Scoring			
Apgar Sign	2	1	0
Heart Rate (pulse)	Normal (above 100 beats per minute)	Below 100 beats per minute	Absent (no pulse)
Breathing (rate and effort)	Normal rate and effort	Slow or irregular breathing	Absent (no breathing)
Grimace (Responsiveness or "reflex irritability")	Pulls away, sneezes, or coughs with stimulation	Facial movement only (grimace) with stimulation	Absent (no response to stimulation)
Activity (muscle tone)	Active, spontaneous movement	Arms and legs flexed with little movement	No movement, "floppy" tone
Appearance (skin coloration)	Normal color all over (hands and feet are pink)	Normal color (but hands and feet are bluish)	Bluish-gray or pale all over



benannt nach Dr. Virginia Apgar, der ersten auf die medizinische Versorgung Neugeborener spezialisierten Kinderärztin

erhoben jeweils 1, 5 und 10 Minuten nach der Geburt

Der Apgar-Score ist die Summe der einzelnen Scores zu einem Bewertungszeitpunkt.



Der Apgar-Score hilft dem Arzt bei der Entscheidung, ob ein Neugeborenes unmittelbar medizinischer Hilfe bedarf oder nicht. Wenn ein Baby zwischen 7 und 10 Punkten erreicht, ist es normalerweise in einem guten Zustand und benötigt nicht mehr als die üblichen Routineuntersuchungen. Liegt der Score zwischen 4 und 6 Punkten, braucht das Baby meistens etwas Hilfe beim Atmen, entweder durch das Absaugen der Nasenlöcher, eine leichte Massage oder die Gabe von Sauerstoff. Bei 3 oder weniger Punkten sind unmittelbar lebensrettende Maßnahmen wie z.B. eine kardiopulmonale Reanimation angezeigt. Ein niedriger Score nach 1 Minute bedeutet nicht zwangsläufig, dass sich das Neugeborene im Laufe der Zeit nicht wieder erholt. Gerade Frühgeborene und Babys, die per Kaiserschnitt auf die Welt kommen, haben oftmals unterdurchschnittliche Apgar-Scores, besonders nach 1 Minute. Der 5-Minuten-Test dient dem Arzt zur Überprüfung der Entwicklung des Kindes und der Wirksamkeit eventuell eingeleiteter medizinischer Maßnahmen.

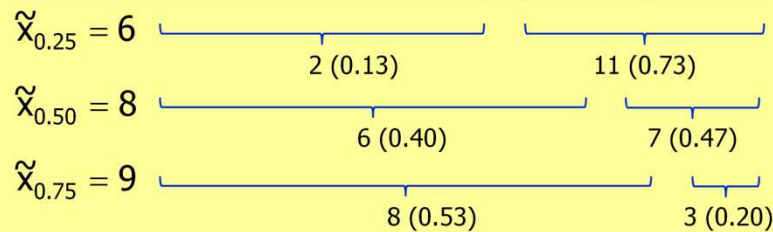
Apgar-Score

n=15

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	7	9	4	6	9	8	10	5	10	9	6	8	9	10	7

k=11

score _i	0	1	2	3	4	5	6	7	8	9	10
h_i	0	0	0	0	1	1	2	2	2	4	3



Der Apgar-Score kann jede ganze Zahl zwischen 0 und 10 annehmen, d.h. er fällt in eine von 11 Kategorien ($k=11$). Von den 15 Neugeborenen ($n=15$) der angegebenen Stichprobe hatten zwei (13%) einen Score kleiner als 6 und 11 einen Score größer als 6 (73%). Da $0.13 \leq 0.25$ und $0.73 \leq 0.75$ gilt, beträgt das 1. Quartil (0.25-Quantil) der Stichprobe also 6. Sechs Babys (40%) wiesen einen Score unter 8 auf, und sieben (47%) hatten einen Score über 8, so dass der Median (0.50-Quantil) gleich 8 ist. Acht Neugeborene (53%) erreichten weniger als 9 Punkte, und drei (20%) hatten einen Score über 9, d.h. das 3. Quartil (0.75-Quantil) beträgt 9.

Lagemaße

ordinale und quantitative Daten

Vorteile der Quantile

robust gegen Ausreißer,
anwendbar auf ordinale Daten,
liefern in der Regel sinnvolle Werte,
intuitiv nahe liegend

Nachteile der Quantile

verwenden nur einen Teil der Information
(nämlich deren Reihenfolge),
sind nicht notwendigerweise eindeutig bestimmt

Lagemaße

quantitative Daten

Der **Mittelwert** \bar{x} einer Stichprobe bezeichnet den Durchschnitt aller Werte, d.h.

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

Vorteil des Mittelwerts

verwendet alle verfügbare Information

Nachteile des Mittelwerts

nur für quantitative Messungen sinnvoll,
empfindlich gegen Ausreißer

Hier ist es, das Ehrfurcht gebietende "Σ". Bei diesem zickzack-artigen Symbol handelt es sich um den großen griechischen Buchstaben "Sigma", der für die Summe von Zahlen steht. Soll eine Reihe von Beobachtungen x_i aufsummiert werden, dann besagt das tief gestellte "i=1", mit welcher Beobachtung man beginnen soll (nämlich der ersten), und das hoch gestellte "n" sagt, mit welcher Beobachtung man aufhören soll (nämlich der letzten, also n-ten). Ist n z.B. gleich 4, so gilt

$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + x_4.$$

Auf der vorliegenden Folie wurde wieder eine mathematische Formel zur Darstellung eines einfachen Sachverhalts benutzt: Für quantitative Daten macht es Sinn, den Mittelwert oder Durchschnitt der Beobachtungen zu bilden, d.h. alle Werte aufzuaddieren und die Summe anschließend durch die Stichprobengröße zu teilen. Der Mittelwert ist zwar ein gutes Lagemaß, er hat aber auch einige Nachteile. So ist er z.B. sehr empfindlich gegen Ausreißer ("Ein faules Ei verderbt den ganzen Brei").

Weiblicher Body-Mass-Index (BMI)

Jahr	Name	BMI
1984	Suzette Charles	17.7
1985	Sharlene Wells	18.2
1986	Susan Akin	16.8
1987	Kellye Cash	17.6
1988	Kaye Lani Rae Rafko	18.8
1989	Gretchen Carlson	19.1
1990	Debbie Turner	17.9
1991	Marjorie Vincent	17.8
1998	Kate Shindle	20.2
1999	Nicole Johnson	19.6
2001	Angela Perez Baraquio	20.3
2002	Katie Harman	19.5

$$\bar{x} = \frac{223.5}{12} = 18.6 \quad \tilde{x}_{0.50} = \frac{18.2 + 18.8}{2} = 18.5$$

Gemäß NHANES Definition (BMI < 19.1) wäre die "durchschnittliche" Miss America untergewichtig. Beachten Sie jedoch, wie gut der Median (18.5) und der Mittelwert (18.6) des BMI in der Stichprobe übereinstimmen.

Weiblicher Body-Mass-Index (BMI)

Jahr	Name	BMI
1984	Suzette Charles	17.7
1985	Sharlene Wells	18.2
1986	Susan Akin	16.8
1987	Kellye Cash	17.6
1988	Kaye Lani Rae Rafko	18.8
1989	Gretchen Carlson	19.1
1990	Debbye Turner	17.9
1991	Marjorie Vincent	17.8
1998	Kate Shindle	20.2
1999	Nicole Johnson	19.6
2001	Angela Perez Baraquio	20.3
2002	Katie Harman	19.5
1630	Venus Rubens	42.6

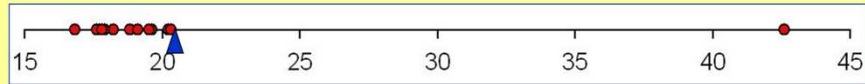
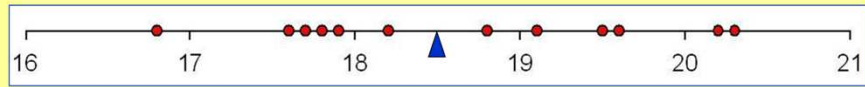


$$\bar{x} = \frac{266.1}{13} = 20.5$$

$$\tilde{x}_{0.50} = 18.8$$

Die (zugegebenermaßen unrealistische) Hinzunahme einer einzelnen barocken Schönheitskönigin würde den mittleren BMI um ungefähr 2 Punkte anheben, während sich der Median nur geringfügig verändern und bei 18.8 liegen würde.

Weiblicher Body-Mass-Index (BMI)



Der Mittelwert markiert den Schwerpunkt der Daten.

Datenbeschreibung

Streuungsmaße

Typ	Lagemaße	Streuungsmaße
qualitativ <i>nominal</i>	Modalwert	Index Qualitativer Variation (IQV)
qualitativ <i>ordinal</i>	Modalwert Quantile (Median)	IQV Interquartilabstand
quantitativ	Quantile (Median) Mittelwert	Interquartilabstand Standardabweichung Mediandeviation

Streuungsmaße dienen der Einschätzung, wie stark Daten in einer Stichprobe variieren. Es hängt wiederum von der Art der Daten ab, welche Streuungsmaße sich sinnvoller Weise aus ihnen berechnen lassen. Für nominale und ordinale Daten gibt der Index Qualitativer Variation (IQV) an, inwieweit die Beobachtungen gleichmäßig über alle Kategorien verteilt sind bzw. in einzelnen Kategorien "klumpen". Die informativsten Streuungsmaße lassen sich natürlich für quantitative Daten ermitteln. Hierzu zählen die empirische Standardabweichung und die Mediandeviation, die jeweils die Streuung der Daten um ein Lagemaß (Mittelwert bzw. Median) messen. Für quantitative Daten ist es auch möglich, den Abstand zwischen bestimmten Beobachtungen anzugeben, so z.B. zwischen dem 1. und 3. Quartil (auch als "Interquartilabstand" bezeichnet).

Streuungsmaße

quantitative Daten

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

Die Spannweite ist empfindlich gegen Ausreißer.

Interquartilabstand $IQR = \tilde{x}_{0.75} - \tilde{x}_{0.25}$

Die Definition von Abständen und Spannweiten setzt voraus, dass man zwischen den Beobachtungen Differenzen bilden kann. Dies bedeutet, dass die entsprechenden Streuungsmaße eigentlich nur für quantitative Daten sinnvoll sind. Oft werden jedoch auch ordinale Daten (trotz ihrer im Grunde qualitativen Natur) auf einer numerischen Skala gemessen. In diesen Fällen kann auch eine Differenzbildung sinnvoll sein, wofür der Apgar-Score ein gutes Beispiel liefert. Ist die Bildung von Differenzen möglich und sinnvoll, so können also Abstände und Spannweiten auch zur Messung der Streuung ordinaler Daten verwendet werden.

Apgar-Score

v_i	0	1	2	3	4	5	6	7	8	9	10
h_i	0	0	0	0	1	1	2	2	2	4	3

$$v_{\min}=4, \quad v_{\max}=10 \quad \longrightarrow \quad R=6$$

$$\tilde{x}_{0.25} = 6, \quad \tilde{x}_{0.75} = 9 \quad \longrightarrow \quad \text{IQR}=3$$

h_i	1	0	0	0	1	1	2	2	2	4	3
-------	---	---	---	---	---	---	---	---	---	---	---

$$v_{\min}=0, \quad v_{\max}=10 \quad \longrightarrow \quad R=10$$

$$\tilde{x}_{0.25} = 6, \quad \tilde{x}_{0.75} = 9 \quad \longrightarrow \quad \text{IQR}=3$$

Ein einziger neonataler Todesfall erhöht die Spannweite des Apgar-Scores in der vorliegenden Stichprobe von 6 auf 10 Punkte, lässt aber den Interquartilabstand unverändert. Beachten Sie, dass die Vergrößerung der Spannweite unabhängig von der tatsächlichen Anzahl der Totgeburten ist. Dies unterstreicht noch einmal die Empfindlichkeit der Spannweite und die Robustheit des Interquartilabstands gegen Ausreißer.

Streuungsmaße

quantitative Daten

Mediandeviation

$$\text{MAD} = \tilde{y}_{0.50}$$

$$\text{mit } y_i = |x_i - \tilde{x}_{0.50}|$$

Standardabweichung

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

Wie der Name schon sagt, handelt es sich bei der Mediandeviation um den Median der Abweichungen der Originaldaten vom Median der Stichprobe. Die Standardabweichung misst demgegenüber die Streuung der Daten um das Stichprobenmittel. Sie errechnet sich als Quadratwurzel aus dem mittleren quadratischen Abstand zwischen Daten und Mittelwert. Die leicht irritierende Verwendung von $n-1$ statt n bei der Mittelung der quadratischen Abstände rührt daher, dass die Standardabweichung der Stichprobe in einem späteren Abschnitt unseres Kurses als "gute" Schätzung der Standardabweichung der gesamten Population dienen soll. Da die Streuung einer vergleichsweise kleinen Stichprobe im Durchschnitt immer etwas kleiner ausfällt als die Streuung in der "großen weiten Welt", muss man die mittlere quadratische Abweichung der Stichprobe systematisch ein wenig erhöhen, und zwar indem man durch $n-1$ statt durch n teilt. Anschließend wird noch die Quadratwurzel gezogen, um die Dimension der Daten zu erhalten. Wenn es sich bei der Stichprobe z.B. um in Zentimeter gemessene Körpergrößen handelt, dann lautet die Dimension der Quadratsumme Quadratzentimeter, was aber nur für die Messung von Flächen sinnvoll wäre. Das Wurzelziehen sorgt also dafür, dass die Standardabweichung wieder die gleiche Dimension hat wie die Originaldaten, in unserem Beispiel also Zentimeter.

Weiblicher Body-Mass-Index (BMI)

$$\bar{x} = 18.6 \quad \tilde{x}_{0.50} = 18.5$$

i	x_i	$(x_i - \bar{x})^2$	$x_{[i]}$	y_i	$y_{[i]}$
1	17.7	0.81	16.8	0.8	0.3
2	18.2	0.16	17.6	0.3	0.3
3	16.8	3.24	17.7	1.7	0.6
4	17.6	1.00	17.8	0.9	0.6
5	18.8	0.04	17.9	0.3	0.7
6	19.1	0.25	<u>18.2</u>	0.6	<u>0.8</u>
7	17.9	0.49	18.8	0.6	0.9
8	17.8	0.64	19.1	0.7	1.0
9	20.2	2.56	19.5	1.7	1.1
10	19.6	1.00	19.6	1.1	1.7
11	20.3	2.89	20.2	1.8	1.7
12	19.5	0.81	20.3	1.0	1.8

$$\text{MAD} = \frac{0.8 + 0.9}{2} = 0.85 \quad s = \sqrt{\frac{13.89}{11}} = 1.12$$

Interessanterweise stimmen Mediandeviation und Standardabweichung der BMIs der Miss America gut überein, genauso wie Median und Mittelwert. Eine solche Übereinstimmung ist ein Hinweis auf eine "schöne" Verteilung der BMI-Werte in der Stichprobe, was bedeutet, dass die Daten symmetrisch verteilt sind (d.h. es gibt annähernd gleich viel große und kleine Werte) und dass es keine Ausreißer gibt.

Weiblicher Body-Mass-Index (BMI)

$$\bar{x} = 20.5 \quad \tilde{x}_{0.50} = 18.8$$

i	x_i	$(x_i - \bar{x})^2$	$y_{[i]}$
1	17.7	7.84	0.0
2	18.2	5.29	0.3
3	16.8	13.69	0.6
4	17.6	8.41	0.7
5	18.8	2.89	0.8
6	19.1	1.96	0.9
7	17.9	6.76	1.0
8	17.8	7.29	1.2
9	20.2	0.09	1.4
10	19.6	0.81	1.5
11	20.3	0.04	1.7
12	19.5	1.00	2.0
13	42.6	488.41	23.8

$$\text{MAD} = 1.00$$

$$s = \sqrt{\frac{544.48}{12}} = 6.74$$

Die Hinzunahme von Venus Rubens verändert die Mediandeviation nur unwesentlich, hat aber einen enormen Einfluss auf die Standardabweichung der Stichprobe. Der Unterschied zwischen beiden Streuungsmaßen lässt eine "ungewöhnliche" Verteilung der Daten vermuten, die im vorliegenden Fall durch eine Ausreißerin verursacht wird, die die Verteilung der BMI-Werte auf der rechten Seite abflacht.

Datenbeschreibung

Histogramm

Stichprobenwerte x_i

Zerlege die Spannweite der x_i in k gleichgroße Intervalle (**Klassen**) der Breite $(x_{\max} - x_{\min})/k$:

$(x_1, a_2]$ $(a_2, a_3]$... $(a_k, x_n]$.

Bestimme die **Klassenhäufigkeiten** h_i .

$$k \approx \sqrt{n} \quad \text{für } n \leq 1000$$
$$k \approx 10 \cdot \log_{10}(n) \quad \text{für } n > 1000$$

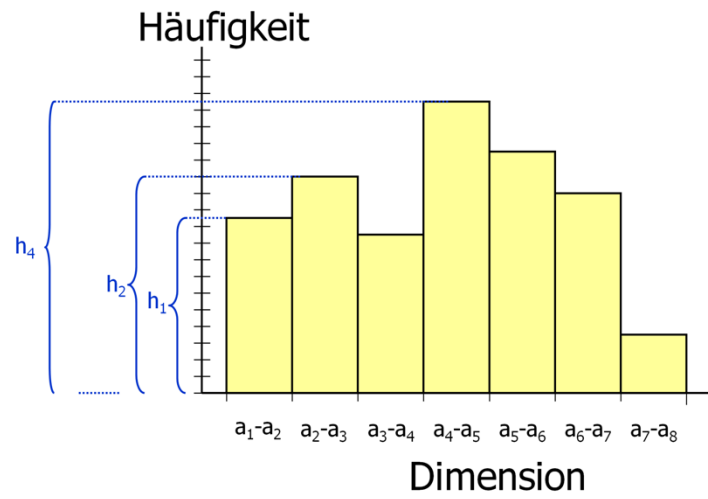
Zur graphischen Darstellung von Stichprobendaten werden oftmals Histogramme verwendet. Histogramme lassen sich für jede Art von Daten zeichnen, egal ob nominal oder quantitativ. Für stetige Daten setzt die Erstellung eines Histogramms jedoch voraus, dass die Daten zuvor in diskrete Klassen transformiert wurden.

Für nominale Daten ergibt sich die Definition der Klassen eines Histogramms von selbst. Allerdings kann es selbst für nominale Daten erforderlich sein, verschiedene Beobachtungswerte in ein und der selben Klasse zusammenzufassen, da ansonsten die Zahl der möglichen Werte zu groß würde. Für quantitative Daten wird der Bereich zwischen dem kleinsten und größten Beobachtungswert in gleich große Intervalle zerlegt, die dann wiederum die Klassen des Histogramms bilden.

Die Anzahl der Klassen eines Histogramms ist weitestgehend beliebig, es gibt jedoch Faustregeln, die einen guten Kompromiss zwischen Auflösungsvermögen (der Stichprobenverteilung) und Robustheit (der Häufigkeiten einzelner Klassen) bilden.

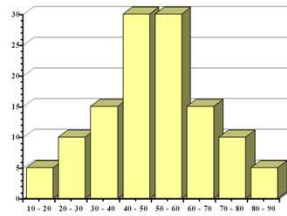
Datenbeschreibung

Histogramm

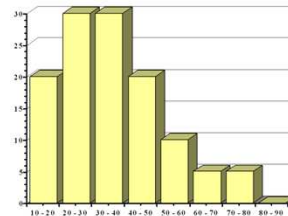


Datenbeschreibung

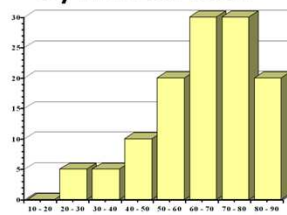
Histogramm



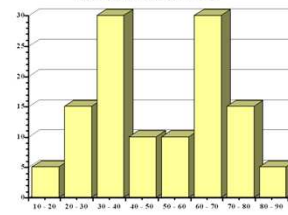
symmetrisch



linkssteil



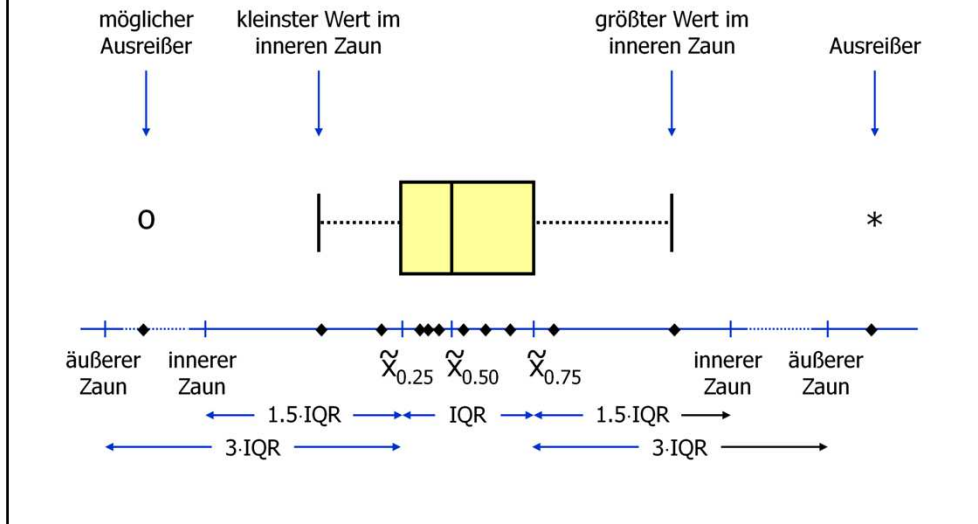
rechtssteil



bimodal

Datenbeschreibung

Box-Whisker-Plot



Ein Box-Whisker-Plot (manchmal auch kurz "Box-Plot" genannt) ist eine graphische Darstellung der wesentlichen Merkmale einer Stichprobe quantitativer Daten. Sie illustriert die Verteilung der Daten zwar nicht so detailliert wie ein Histogramm, ermöglicht aber eine Einschätzung, ob deren Verteilung schief ist oder ob es Ausreißer gibt. Bei einem Box-Whisker-Plot

- markieren die Enden der Box das 1. bzw. 3. Quartil, so dass die Box selbst den Interquartilabstand umfasst.
- markiert die vertikale Linie innerhalb der Box den Median (2. Quartil).
- erstrecken sich die Whisker (also die "Schnurrhaare") bis zum jeweils größten und kleinsten Wert innerhalb des "inneren Zauns", d.h. des Intervalls zwischen 1. Quartil minus 1.5-fachem Interquartilabstand und 3. Quartil plus 1.5-fachem Interquartilabstand.
- werden Daten außerhalb des inneren und innerhalb des "äußeren", Zauns (definiert wie der innere Zaun, aber mit dreifachem IQR) als "mögliche" Ausreißer bezeichnet und durch Kreise markiert.
- werden Daten außerhalb des äußeren Zauns als "echte" Ausreißer bezeichnet und durch Sternchen markiert.

Weiblicher Body-Mass-Index (BMI)

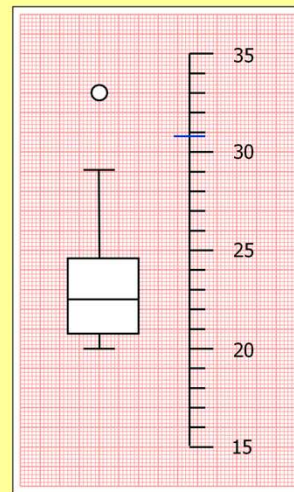
i	$x_{[i]}$
1	20.1
2	20.3
3	20.5
4	20.8
5	21.5
6	21.6
7	22.5
8	22.7
9	23.8
10	24.6
11	24.5
12	24.8
13	25.4
14	29.1
15	33.0

$$\tilde{x}_{0.25} = 20.8$$

$$\tilde{x}_{0.50} = 22.7$$

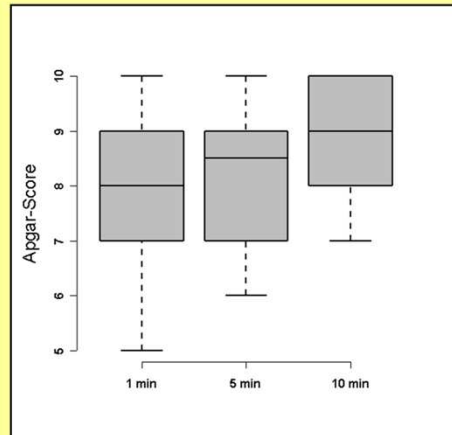
$$\tilde{x}_{0.75} = 24.8$$

$$\text{IQR} = 24.8 - 20.8 = 4.0$$



Bei diesen 15 Body-Mass-Indizes handelt es sich offensichtlich um eine Stichprobe aus der Normalbevölkerung, und nicht um Schönheitsköniginnen.

Apgar-Score



1 Minute: 5, 6, 6, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 10
5 Minuten: 6, 6, 7, 7, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10
10 Minuten: 7, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10, 10, 10, 10

Box-Whisker-Plots sind besonders nützlich zum Vergleich mehrerer Stichproben. Die drei Box-Whisker-Plots im vorliegenden Beispiel illustrieren eindrucksvoll die generell zu beobachtende Verbesserung des Apgar-Scores während der ersten 10 Minuten *post partum*.

Zusammenfassung

- Statistische Aussagen über die **Variation von Merkmalen in Populationen** werden auf der Grundlage **repräsentativer Stichproben** getroffen.
- **Deskriptive Statistik** dient der klaren und verständlichen Zusammenfassung von Daten und erlaubt es so, wichtige Muster oder Regelmäßigkeiten in Daten zu identifizieren.
- Experimentelle Variable lassen sich grob als entweder **qualitativ** oder **quantitativ** klassifizieren, wovon erstgenannte wiederum entweder **nominal** oder **ordinal** skaliert sind.
- Zu den **numerischen** und **graphischen Verfahren** der deskriptiven Datenauswertung gehören Lage- und Streuungsmaße sowie Histogramme und Box-Whisker-Plots, die Besonderheiten der Datenverteilung aufzeigen.

Anhang

klinische Studien



"Kontrollierte klinische Studien, in denen Patienten, die ein Medikament bekommen haben, mit solchen verglichen werden, die einer anderen Behandlung unterzogen wurden, ist der **beste wissenschaftliche Weg** zur Klärung, was ein Medikament tatsächlich bewirkt. Aus diesem Grund sind kontrollierte klinische Studien für die FDA die einzige rechtliche Grundlage zur Feststellung, dass ein neues Medikament seine Wirksamkeit hinreichend unter Beweis gestellt hat."

aus: Flieger, Testing Drugs in People

Anhang

klinische Studien

Phase 0 (Prälinik)

Suche nach neuen therapeutischen Wirkstoffen, Labor- oder Tierexperiment, Pharmakokinetik, Toxizität

Phase I

"first-in-man", freiwillige Probanden, Dosis-Wirkung, Sicherheit, Verträglichkeit

Phase II

"proof of concept", Untersuchung von Effizienz und Wirksamkeit an Patienten (kleine Fallzahlen), Dosis-Findung, eventuell Kontrollgruppe

Klinische Studien am Menschen lassen sich üblicherweise in fünf Phasen einteilen:

Phase 0 dient dem Nachweis und der Charakterisierung neuer Wirkstoffe, was heute zunehmend auch schon *in silico* passiert. Aufwändige bioinformatische Verfahren erlauben es in der Tat, die Interaktion biologischer und chemischer Substanzen hinreichend genau am Computer vorherzusagen. Selbst wenn dabei immer noch erhebliche Fehler auftreten, birgt ein solches Vorgehen angesichts der immensen Entwicklungskosten in der pharmazeutischen Industrie erhebliche Einsparungspotenziale.

In der Phase I wird gesunden Freiwilligen (etwa 20-30) oder Patienten, die nicht auf die Standardtherapie ansprechen, erstmals eine gering oder kurz bemessene Dosis des neuen Präparats verabreicht. Ziel der Phase I ist die Bewertung der Sicherheit und das Gewinnen eines ersten Eindrucks von der notwendigen Dosierung und von möglichen Nebenwirkungen.

In Phase II wird einer größeren Anzahl freiwilliger Patienten (üblicherweise 50-100) das Medikament über einen längeren Zeitraum gegeben. Ziel dieser Phase ist die Bewertung des Nutzens und der Sicherheit des Präparats sowie seiner Dosierung und Darreichungsform.

Anhang

klinische Studien

Phase III

an Patienten (große Fallzahlen), Vergleich mit etablierten Behandlungen oder Placebo (Kontrollgruppe), Nebenwirkungen, Nachweis der Wirksamkeit als Voraussetzung für die Zulassung

Phase IV

Etablierung als Standardbehandlung, Monitoring von Nebeneffekten

In Phase III wird ein neues Medikament an einer großen Anzahl freiwilliger Patienten gegen die Standardtherapie oder (seltener) gegen ein Placebo getestet. Phase III Studien sind sehr umfangreich und sollen objektive Evidenz für die Wirksamkeit des Präparats, die optimale Dosierung und hinsichtlich potenzieller Nebenwirkungen liefern. Fällt die Phase III Studie positiv aus, so bekommt das Medikament seine Zulassung und kann anschließend vermarktet werden.

Phase IV schließt sich an die Zulassung und Vermarktung eines Medikaments an. In Phase IV Studien wird Information über die Verwendung des Medikaments im klinischen Alltag gesammelt. Ziel ist die Identifizierung von Langzeitnebenwirkungen oder von schädlichen Interaktionen mit anderen Medikamenten, die in der Phase III unentdeckt blieben.

Anhang

Good Clinical Practice (GCP)

1. Klinische Studien müssen in Übereinstimmung mit den **ethischen Prinzipien** der **Deklaration von Helsinki** durchgeführt werden, was die Einhaltung der GCP und ihrer regulatorischen Anforderung impliziert.
2. Vor einer Studie müssen die vorhersehbaren **Risiken** gegen den erwarteten **Nutzen** für Studienteilnehmer und Gesellschaft abgewogen werden. Eine Studie darf nur dann durchgeführt werden, wenn der erwartete Nutzen die Risiken rechtfertigt.

© Minister of Public Works and Government Services Canada 1997

Bei der "Guten Klinischen Praxis" (engl. good clinical practice; GCP) handelt es sich um einen internationalen ethischen und wissenschaftlichen Qualitätsstandard für das Design, die Durchführung, die Dokumentation und das Berichtswesen von Studien am Menschen. Die Beachtung dieses Standards sichert die Rechte, die Sicherheit und das Wohlbefinden der Studienteilnehmer und garantiert in gewissem Umfang auch die Vertrauenswürdigkeit klinischer Studien.

Die International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) bringt Experten der pharmazeutischen Industrie und der Regulierungsbehörden aus Europa, Japan und den USA zusammen, um die wissenschaftlichen und technischen Aspekte von Produktzulassungen zu diskutieren. Zweck der ICH ist die Harmonisierung technischer Leitlinien und Anforderungen in Zulassungsverfahren, um dadurch die unnötige mehrfache Testung von Medikamenten zu vermeiden.

Die ICH hat im Januar 1997 GCP Richtlinien erstellt. Basierend auf der Deklaration von Helsinki liefern diese Leitlinien einen detaillierten Rahmen für die ethisch einwandfreie Durchführung klinischer Studien. Klinische Studien zum Zwecke der Marktzulassung innerhalb der EU müssen seither zwingend im Einklang mit den GCP Richtlinien der ICH durchgeführt werden.

Anhang

Good Clinical Practice (GCP)

3. Rechte, Sicherheit und Wohlbefinden der Teilnehmer haben absolute Priorität and stehen über den Interessen von Wissenschaft und Gesellschaft.
4. Die verfügbare klinische und nicht-klinische Information über das zu untersuchende Produkt muss die Notwendigkeit der geplante Studie hinreichend belegen.
5. Klinische Studien müssen wissenschaftlich sinnvoll sein und in einem detaillierten Protokoll klar beschrieben werden.

© Minister of Public Works and Government Services Canada 1997

Anlässlich ihrer 18. Hauptversammlung in Helsinki im Jahre 1974 verabschiedete die World Medical Association Empfehlungen für die medizinische Forschung am Menschen. Diese Empfehlungen bestimmen bis heute die medizinische Wissenschaftsethik und definieren die Regeln für die medizinische Forschung innerhalb und außerhalb des Versorgungszusammenhangs. Die Deklaration von Helsinki wurde mehrfach revidiert und bildet derzeit die Grundlage der Guten Klinischen Praxis (GCP).

Kurz gefasst besagt die Deklaration von Helsinki, dass

- Forschung am Menschen auf vorherigen Labor- oder Tierexperimenten aufgebaut sein sollte,
- Studienprotokolle von unabhängigen Kommissionen zu begutachten sind,
- eine informierte Einwilligung der Patienten zwingend erforderlich ist,
- Forschung nur von qualifizierten Personen durchgeführt werden darf
- und die Risiken von Forschung deren Nutzen nicht übersteigen dürfen.

Anhang

Good Clinical Practice (GCP)

6. Eine Studie muss in Übereinstimmung mit einem Protokoll durchgeführt werden, das zuvor die **Zustimmung einer unabhängigen Ethikkommission** erhalten hat.
7. Medizinische Handlungen und Entscheidungen stehen immer in der **Verantwortung eines qualifizierten Arztes** oder Zahnarztes.
8. Alle an der Durchführung einer Studie Beteiligten müssen durch entsprechende **Ausbildung, Training** und **Erfahrung** für ihre Aufgaben qualifiziert sein.

© Minister of Public Works and Government Services Canada 1997

Anhang

Good Clinical Practice (GCP)

9. Von jedem Teilnehmer ist vor Einschluss in eine Studie die **informierte Zustimmung** zur Teilnahme einzuholen.
10. Informationen zu einer Studie müssen so gespeichert und gehandhabt werden, dass alle diesbezüglichen Berichte und Interpretationen **überprüfbar** bleiben.
11. Die **Vertraulichkeit** von personenbezogenen Daten muss zur Wahrung der Grundrechte der Teilnehmer und unter Beachtung einschlägiger gesetzlicher Regelungen gesichert werden.

© Minister of Public Works and Government Services Canada 1997

Anhang

Good Clinical Practice (GCP)

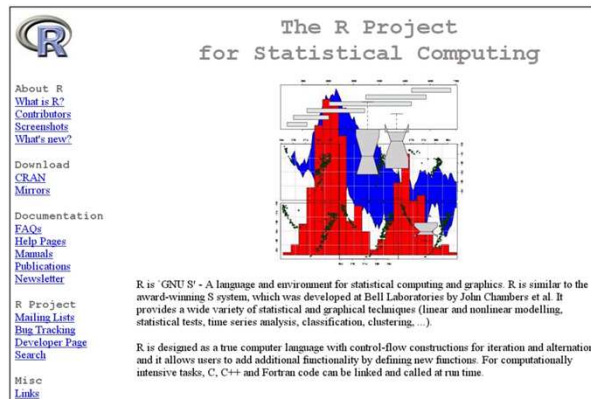
12. Die zum Einsatz kommenden Produkte müssen gemäß der jeweils anwendbaren **Good Manufacturing Practice** (GMP) hergestellt, behandelt und gelagert werden. Sie müssen in Übereinstimmung mit dem Studienprotokoll verwendet werden.

13. Für alle Aspekte der Studie müssen entsprechende Maßnahmen zur **Qualitätssicherung** implementiert werden.

© Minister of Public Works and Government Services Canada 1997

Anhang

Das R-Projekt



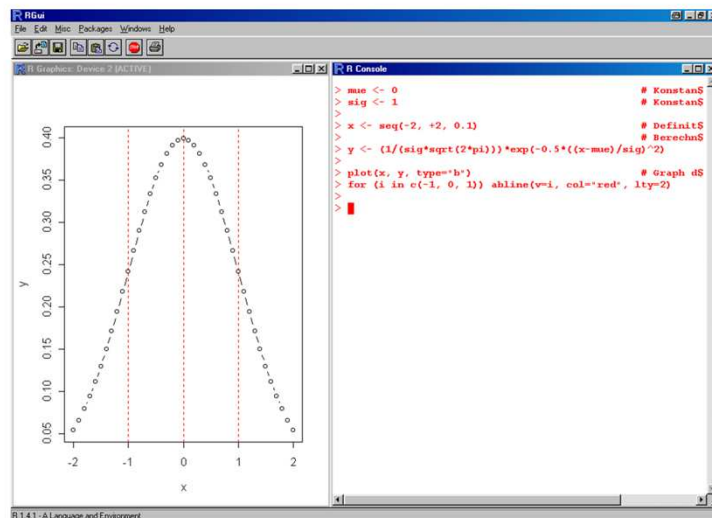
"R ist eine frei verfügbare Programmiersprache und -umgebung für statistische Analysen, die eine große Vielzahl rechnerischer und graphischer Methoden bereithält."

"R" ist eine Programmiersprache und -umgebung für statistische Berechnungen und Graphiken. Sie ähnelt der Sprache "S", die ursprünglich vom Computerwissenschaftler John Chambers und seinen Mitarbeitern für die US-amerikanischen Bell Laboratorien (früher AT&T, heute Lucent Technologies) entwickelt wurde. R gibt es als Free Software. Die Sprache ist leicht zu erlernen, fast so einfach zu benutzen wie ein Taschenrechner und bietet sehr viel mehr Flexibilität als herkömmliche, vorstrukturierte Statistikpakete. R-Programme laufen ohne weiteres auf einer Vielzahl von LINUX-Plattformen und natürlich auch unter Windows und MacOS.

Da R auf der Grundlage einer echten Computersprache konzipiert wurde, erlaubt es Nutzern, neue Funktionalitäten selbst zu entwickeln und hinzuzufügen. Ein großer Teil des Systems ist selbst in der Sprache R programmiert worden, was Nutzern das Verfolgen der Logik einzelner Komponenten deutlich erleichtert. Für rechenintensive Aufgaben können Programme in machtvolleren Sprachen wie z.B. "C", "C++" oder "Fortran" geschrieben und in die R-Umgebung eingebunden werden. Fortgeschrittene Nutzer können sogar Objekte der R-Umgebung selbst manipulieren.

Anhang

Das R-Projekt



Diese Folie illustriert die Nutzung der Graphik-Routinen von R, um z.B. eine Darstellung des typischen glockenförmigen Verlaufs einer Normalverteilung zu erzeugen. Die Normalverteilung, manchmal auch "Gaußverteilung" genannt, ist ein mathematisches Konstrukt, das die zufällige Natur von vielen, wenn nicht gar der allermeisten, natürlichen Prozesse abbildet. Sie liegt im Herzen vieler statistischer Methoden und Konzepte und wird daher eine zentrale Rolle in der sich nun anschließenden Vorlesung zum Thema "Wahrscheinlichkeitstheorie" spielen.