

# Der Zufallsvorgang und der Inferenzschluss

Dr. Mariana Nold

Institut für Soziologie,  
Fakultät für Sozial- und Verhaltenswissenschaften,  
Lehrstuhl für empirische Sozialforschung und Sozialstrukturanalyse

16. Oktober 2017



# Übersicht

- 1 Themen und Ziele
- 2 Rückblick: Deskriptive Statistik
- 3 Statistische Inferenz
  - Der klassische Inferenzschluss
  - Die Intention der empirischen Forschung
- 4 Zufallsvorgang
- 5 Plan für die nächsten Wochen

# Kernkompetenzen

Folgende Kernkompetenzen sollen Sie im Laufe des Semesters aufbauen:

## Statistische Methoden

- 1 Daten selbst analysieren: Thüringen-Monitor 2015 und Pisa-Daten mit STATA
- 2 Statistische Informationen nutzen und sachadäquat interpretieren: STATA-Output und Information aus Presse oder Fachliteratur
- 3 Statistische Ergebnisse verständlich kommunizieren, wesentliche Aussagen in eigenen Worten ausdrücken können

### **Voraussetzung dafür:**

- ▶ Zentrale Definition kennen und verstehen
- ▶ Den Umgang mit STATA üben

# Ziel der heutigen Veranstaltung ...

ist es die folgenden Fragen beantworten zu können:

## Zielfragen für heute

- 1 Welche Intention hat die deskriptive Statistik?
- 2 Welche Intention hat die induktive Statistik?
- 3 Was ist ein Zufallsvorgang?
- 4 Was ist eine einfache Zufallsstichprobe?
- 5 Was ist die Normalverteilung und wie kann man sie nutzen, um einen Datensatz zu beschreiben?
- 6 Was versteht man unter einem Inferenzschluss?

# Thematische Schwerpunkte in diesem Semester

Es gibt in diesem Semester drei thematische Schwerpunkte:

## ① Probabilistische Gesetze:

- ▶ Was ist ein Zufallsvorgang bzw. eine Zufallsvariable?
- ▶ Was ist eine parametrische Wahrscheinlichkeitsverteilung?

## ② Testen und Schätzen:

- ▶ Wie kommt man zu einer Punktschätzung?
- ▶ Was ist ein Konfidenzintervall?
- ▶ Auf welchen Prinzipien beruht das Konzept des Hypothesentests?

## ③ Regression und Varianzanalyse

- ▶ Wie interpretiert man die Regressionskoeffizienten im linearen Regressionsmodell?
- ▶ Wie nutzt man die Varianzanalyse zum Mittelwertvergleich bei mehr als zwei Gruppen?

# Übersicht über die Vorlesungen (jeweils Montag 10-12 Uhr)

Tabelle: Es gibt ein Aufgabenblatt zur Abgabe

Nr.	Datum	Thema
0	16.10.2017	Der Zufallsvorgang und der Inferenzschluss
1	23.10.2017	Die parametrische Wahrscheinlichkeitsverteilung
2	13.11.2017	Grundlagen des statistischen Testens
3	27.11.2017	Punktschätzer und Konfidenzintervalle
4	11.12.2017	Einfaktorielle Varianzanalyse und lineare Regression (F-Test)
5	08.01.2018	Multiple lineare Regression: Modellannahmen und Regressionsdiagnostik
6	22.01.2018	Mehrfaktorielle Varianzanalyse

# Übersicht über die Übungsgruppen

- Die Übung findet 14-tägig statt (1. Woche Gruppen 1 und 3, 2. Woche Gruppen 2 und 4 )
- Die Übung findet dienstags statt, jeweils von 10-12 Uhr (1 und 3) und 14-16 Uhr (2 und 4)

Übung Nr.	Datum Gruppe 1 u. 3	Datum Gruppe 2 u. 4
1	24.10.2017	07.11.2017
2	14.11.2017	21.11.2017
3	28.11.2017	05.12.2017
4	12.12.2017	19.12.2017
5	09.01.2018	16.01.2017
6	23.01.2018	30.01.2017

# Übersicht über die Tutorien

- Das Tutorium findet 14-tägig statt (1. Woche Gruppen 1 und 3, 2. Woche Gruppen 2 und 4 )
- Das Tutorium findet dienstags statt, jeweils von 16-18 Uhr und 18-20 Uhr
- Der Besuch ist freiwillig und sehr zu empfehlen, sie dürfen mehrere Tutorien besuchen.

Übung Nr.	Datum Gruppe 1 u. 3	Datum Gruppe 2 u. 4
1	24.10.2017	07.11.2017
2	14.11.2017	21.11.2017
3	28.11.2017	05.12.2017
4	12.12.2017	19.12.2017
5	09.01.2018	16.01.2017
6	23.01.2018	30.01.2017



# Die Abgabe der Aufgabenblätter

- Sie geben wieder in Gruppen (3-4 Personen) ab, in diesem Semester gibt es ein standardisiertes Deckblatt, dieses finden Sie unter ...
- Falls Sie das Aufgabenblatt aus gesundheitlichen Gründen nicht abgeben können, wird das auf dem Deckblatt vermerkt.
- Bitte Gruppe anmelden bei ...
- Die Lösungen werden in den Übungen zurückgegeben. Bei der Anmeldung wird festgelegt, in welcher Gruppe ihr Blatt zurück gegeben werden soll.
- Das letzte Aufgabenblatt ist das Blatt 5, es erscheint am Freitag den 5. Januar. Die Abgabe des letzten Aufgabenblattes ist am Freitag den 19. Januar

# Die Klausur

- Sie müssen bei vier von fünf Aufgabenblättern mindestens 50% richtig gelöst haben, um an der Klausur teilnehmen zu dürfen.
- Die Klausur findet am Montag den 19. Februar um 10-12 Uhr statt.
- Die Nachholklausur findet am Montag den 19. März um 10-12 Uhr statt.
- Das Aufgabenblatt 6 wird durch eine Probeklausur ersetzt.
  - ▶ Es wird dann auch vom Umfang einer erwarteten Bearbeitungszeit von 90 Minuten entsprechen.
  - ▶ Zeilenabstände werden größer sein.
  - ▶ Sie brauchen keine STATA-Befehle zu kennen, sollen aber STATA-Output interpretieren können.

# Mein Kontaktdaten

## Kontakt

- ① Sprechstunde: Freitag von 10 – 11 Uhr
- ② Mein Büro: Bachstraße 18k, 1. OG, R132
- ③ Telefonnummer: 03641 9 45013
- ④ E-Mail-Adresse: [mariana.nold@uni-jena.de](mailto:mariana.nold@uni-jena.de)

## Material zur Vorlesung und Übung

wo ?

# Univariate deskriptive Statistik: Den Datensatz verstehen

- Man braucht deskriptive Methoden um den Datensatz kennen zu lernen:
  - ▶ Was bedeuten die Merkmale im Datensatz inhaltlich?
  - ▶ Wie wurden latente Merkmale operationalisiert?
  - ▶ Auf welchem Messniveau wurden sie erhoben?
  - ▶ Gibt es unrealistische Messwerte? (Beispiel: Person gibt an 100 Kinder zu haben)
- Wir haben im letzten Semester Lage- und Streuungsmaße für die univariate Deskription kennen gelernt.
- Die univariate Deskription wird mit Hilfe von Grafiken, wie dem Boxplot, der empr. Verteilungsfunktion, dem Balkendiagramm oder Histogramm illustriert.

# Bivariate deskriptive Statistik: Erste Eindrücke von Zusammenhängen

- Kreuztabellen und insbesondere die Spaltenprozentuierung sind wichtige Werkzeuge der bivariaten Deskription für zwei qualitative Merkmale.
- Gruppierte Balkendiagramme dienen der Veranschaulichen der Spaltenprozentuierung.
- Der Vergleich von Boxplots von verschiedenen Gruppen dient der bivariaten Deskription von einem qualitativen und einem metrischen Merkmal.
- Zwei metrische Merkmale können mit Hilfe der Korrelation und der Regression beschrieben werden.
- Das Streudiagramm (mit Regressionsgerade) dient der grafischen Analyse.

# Die beschreibende Statistik und ihre Intention

Die deskriptive Statistik hat die Intention einen gegebenen Datensatz zu beschreiben, dazu nutzt sie

- Lagemaße wie Mittelwert, Modus, Median und andere empirische Quantile
- Streuungsmaße wie empirische Varianz oder Interquartilsabstand
- Kreuztabellen und bedingte relative Häufigkeiten, insbesondere die Spaltenprozentuierung
- Grafiken wie die empirische Verteilungsfunktion, Boxplots, Histogramme oder (gruppierte) Balkendiagramme
- Eine **gute** Deskription ist eine **notwendige** Voraussetzung für die sinnvolle Analyse des Datensatzes.

## Sie möchten Begriffe nachschlagen oder wiederholen?

- ① (ThULB Jena) Buch:  
**Statistik: Eine Einführung für Sozialwissenschaftler (Grundlagentexte Soziologie),**  
Wolfgang Ludwig-Mayerhofer , Uta Liebeskind , Ferdinand Geißler
- ② ThULB Jena E-Book:  
**Statistik- Eine Einführung mit interaktiven Elementen,**  
Hans-Joachim Mittag
- ③ ThULB Jena E-Book:  
**Statistik- Der Weg zur Datenanalyse** (8. Auflage), Ludwig  
Fahrmeir und andere

### Die Formelsammlung:

[wiwi-online.net/brochures/wiwi\\_formelsammlung\\_statistik\\_b28.pdf](http://wiwi-online.net/brochures/wiwi_formelsammlung_statistik_b28.pdf)

# Lexikon und Online-Karteikarten

- ① Online Lexikon von Prof. Ludwig-Mayerhofer:  
<http://wlm.userweb.mwn.de/Ilmes/>
- ② Karteikarten online:  
<https://www.cobocards.com/pool/de/cardset/7yua50512/online-karteikarten-33209-statistik-i/>



# Was ist eigentlich Statistik?

Zwei Wiener Statistiker haben sich mit dieser Frage beschäftigt

Während die meisten Wissenschaften eine zumindest **formal klare Definition besitzen** und sich somit von anderen Wissenschaften abgrenzen können, gelingt dies bei der Statistik nicht so einfach. Ein Anhaltspunkt dafür sind schon die vielen sehr unterschiedlichen Definitionen von "Statistik", die in den verschiedenen Lehrbüchern zu finden sind.

(Quelle: Marcus Hudec, Christian Neumann, Institut für Statistik der Universität Wien)

Fazit von Hudec und Neumann ...

Statistik wird als Hilfswissenschaft aufgefasst. Sie ist eine der Methoden, mit der die Theorie und Erfahrung (Empirie) systematisch reflektiert wird.

# Die schließende Statistik und ihre Bedeutung

Welche Bedeutung bzw. Aufgabe hat die Statistik als Wissenschaft?

## Antwort von Yudi Pawitan

Uncertainty is pervasive in problems that deal with the real world, but statistics is the only branch of science that puts systematic effort into dealing with uncertainty.

Statistics is suited to problems with **inherent uncertainty due to limited information**,

- 1 it does not aim to remove uncertainty,
- 2 but in many cases it merely quantifies it,
- 3 uncertainty can remain even after analysis is finished.

# Was wird als gültiger Zusammenhang anerkannt?

...im Laufe der Zeit immer wichtiger wird

Die Bedeutung der Statistik ist zur Zeit einem großen Wachstum unterworfen.

Ein Grund für die zunehmende Bedeutung

Statistische Methoden sind oft entscheidend dafür, welche Zusammenhänge in der Gesellschaft als gültig anerkannt werden.

Beispiel:

- ❶ Pisa-Studie: Sollte man Geld investieren, um die Jungs bei Lesen lernen zu unterstützen?
- ❷ Unfallstatistik: Sollte man Geld investieren, um für mehr Sicherheit auf deutschen Autobahnen zu sorgen?

# Beispiel: PISA-Studie

Warum sind viele wissenschaftliche Fragestellungen von Unsicherheit durchdrungen?

- Wir wollen basierend auf den PISA-Daten eine Antwort auf die Frage finden, ob die Leseleistung von 15-jährigen Mädchen in Deutschland besser ist als die von 15-jährigen Jungen.
- Unsere Stichprobe enthält nicht die Information über die Leseleistung aller 15-Jährigen in Deutschland.  
(inherent uncertainty due to limited information)
- Wenn wir eine Aussage über die Grundgesamtheit machen, kann diese Aussage falsch sein. Die Wissenschaftlichkeit der Aussage entsteht dadurch, dass wir ihre Unsicherheit quantifizieren.

## Beispiel: Unfalltod auf der Autobahn

- Wir wollen eine Aussage darüber machen, ob die Wahrscheinlichkeit für den Unfalltod auf deutschen Autobahnen gestiegen ist.
- Die Zahl der Unfalldoten ist von 358 (2012) auf 387 (2013) gestiegen. Ist das dann eine “wirkliche” Veränderung?
- Die Statistik fragt: Hat sich etwas an dem stochastischen Mechanismus geändert der diese Unfalldoten hervorbringt?
- Um wie viel muss die Zahl steigen bzw. sinken, damit man davon ausgeht, dass sich dieser Mechanismus verändert hat?
- Diese Frage kann nicht willkürlich beantwortet werden. Es muss klar sein, wann man von einer signifikanten Veränderung spricht und was damit gemeint ist.

# Inferenzkonzepte

- Durch welche Legitimation ist es erlaubt davon zu sprechen, dass Mädchen besser lesen als Jungen oder dass die Anzahl der Unfalltoden signifikant gestiegen ist?
- Wenn ich die entsprechenden statistischen Tests rechne, dann ist der Unterschied in der Leseleistung signifikant, der Unterschied in der Zahl der Unfalltoden nicht.
- Was bedeuten solche Aussagen bzw. Behauptungen und wodurch werden sie legitimiert?
- Unter einem Inferenzkonzept versteht man ein statistisches Konzept, das die Schlussfolgerung von Beobachtungen (Daten) auf inhaltliche Aussagen rechtfertigt.

# Klassische Inferenz

- Wie werden innerhalb der Statistik Schlüsse von Beobachtungen auf Hypothesen gerechtfertigt?
- Angesichts der Vielfalt wissenschaftstheoretischer und philosophischer Ansätze zur Erklärung des empirischen Forschungsansatzes ist es nicht verwunderlich, dass es auf diese Frage innerhalb der Statistik mehrere Antworten gibt und **nicht eine einzige, in sich geschlossene statistische Inferenztheorie existiert.**
- Die drei bekanntesten Inferenztheorien sind die Likelihood-Inferenz, die Bayes-Inferenz und die Klassische Inferenz.
- Wir beschäftigen uns in diesem Semester mit Ansätzen aus der Klassischen Inferenz.

# Der Inferenzschluss der Klassischen Inferenz

- Um ein valides Abbild der Grundgesamtheit zu bekommen, zieht man eine Zufallsstichprobe.
- Nur bei Realisierung einer Zufallsauswahl kann von einer Stichprobe **mit kontrollierter Irrtumswahrscheinlichkeit** auf die zugrundeliegende Grundgesamtheit zurückgeschlossen werden.
- Dieser auch **Inferenzschluss** bezeichnete Rückschluss von Eigenschaften der Stichprobe auf Eigenschaften der Grundgesamtheit ist Gegenstand der induktiven Statistik.
- Ein Inferenzschluss ist stets mit Unsicherheit behaftet. Dies ergibt sich zwingend daraus, dass nur eine Teilinformation zu Verfügung steht.

Leseempfehlung: ([http://www.fernuni-hagen.de/ksw/neuestatistik/content/MOD\\_27531/html/comp\\_27634.html](http://www.fernuni-hagen.de/ksw/neuestatistik/content/MOD_27531/html/comp_27634.html))



# Empirische Forschung: Muster in Daten erkennen und bewerten

- 1 Die Statistik stellt Methoden zur Verfügung um Information aus Beobachtungen gut zusammen zu fassen und diese Information zielgerichtet zu verarbeiten.
- 2 Die Statistik gibt uns Werkzeuge um empirische Information übersichtlich darzustellen. (Deskriptive Statistik)
- 3 Beruhend auf einer übersichtlichen Darstellung können wir z. B. Vermutungen über Zusammenhänge ableiten. (Explorative Statistik)
- 4 Die Statistik erlaubt u. U. von der beobachteten empirischen Information auf eine über die Stichprobe hinaus gültige Aussagen zu schließen. (Induktive Statistik)

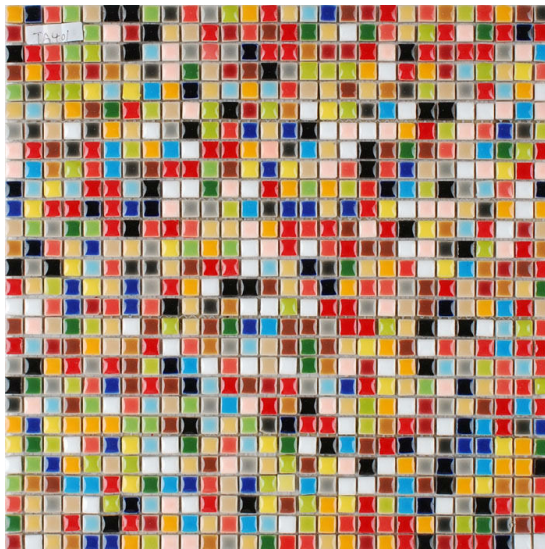
Empirische Forschung ist vergleichbar damit ein Mosaik zu restaurieren



# Restauration: Was es zu beachten gilt

- 1 Transparenz ist wichtig. Man muss sagen, welche Teile man zu Beginn hatte und durch welche Annahmen und Methoden man zu dem Gesamtbild kommt.
- 2 Das Bild das, nach der Restauration entsteht, kann (in Teilen) fehlerhaft sein. Es ist mit Unsicherheit behaftet.
- 3 Je weniger Teile man zu Beginn hat bzw. je komplizierter das Bild ist, desto größer ist die Unsicherheit über das Ergebnis.
- 4 Es gibt viele Möglichkeiten zu dem Ergebnis zu kommen, es gibt nicht den einen richtigen Weg, auch wenn es nur ein richtiges Bild gibt.

# Gibt es überhaupt ein Muster?



Können wir aus den gegebenen Informationen etwas schließen?



Wie viel darf hier fehlen, damit man eine Chance hat das Muster zu finden?



Welche Teile dürfen fehlen, damit man eine Chance hat das Muster zu finden?



# Werkzeugkiste für die Restauration

Wir wollen einen Ausschnitt der sozialen Wirklichkeit erforschen und erheben Daten, um aus dieser “limited information” eine Aussage über die dahinter stehenden Mechanismen machen.

Was gehört in die Werkzeugkiste?

- 1 Kreativität und Fingerspitzengefühl im Umgang mit den Daten
- 2 Sachwissen über den Hintergrund der Daten (inhaltliches Wissen)
- 3 Grundlagen der Methoden der empirischen Sozialforschung
- 4 Grundlagen der deskriptiven Statistik
- 5 Grundlagen der Inferenzstatistik, insbesondere der Test- und Schätztheorie
- 6 Die Fähigkeit Statistik-Software zu nutzen



## Beispiel: Befragung nach Höhe des Taschengeldes

- Fiktiv: In einem Artikel einer regionalen Zeitung aus dem Jahr 2015 steht, dass Grundschulkinder in Jena im Durchschnitt elf Euro Taschengeld im Monat bekommen.
- Wir möchten empirische Information zu diesem Thema sammeln und befragen  $n = 100$  Grundschulkinder in Jena nach ihrem monatlichen Taschengeld (gerundet auf ganze Euro).
- Zwei Eigenschaften kennzeichnen den Vorgang:
  - ① Man kennt im Vorfeld bereits die möglichen Ausgänge, wobei unbekannt ist, welcher Eintritt.
  - ② Es hängt "vom Zufall" ab, welchen Ausgang man beobachtet.

# Zufallsvorgang

Gerade diese beiden Eigenschaften kennzeichnen den Zufallsvorgang. Das Zufallsexperiment ist ein Spezialfall des Zufallsvorgangs.

## Definition: Zufallsvorgang

Ein Zufallsvorgang führt zu einem von mehreren, sich gegenseitig ausschließenden Ergebnissen. Es ist vor der Durchführung ungewiss, welches Ergebnis tatsächlich eintreten wird.

- Die Ziehung einer Stichprobe ist ein Beispiel für einen Zufallsvorgang.
  - ▶ Befragung von Grundschulkindern nach der Höhe ihres Taschengeldes
  - ▶ Ergebnisse des Kompetenztest Lesen in der PISA-Studie
- Viele Prozesse die wir beobachten können sind Zufallsvorgänge.
  - ▶ Beispiel: Wie viel Personen nehmen heute an der Vorlesung teil?
  - ▶ Wie viele Unfalltode gab es im Jahr 2016?

# Zufallsvariable

Ein ganz zentraler Begriff in diesem Semester ist der der Zufallsvariable:

## Definition: Zufallsvariable

Eine Variable oder ein Merkmal  $X$ , dessen Werte oder Ausprägungen die Ergebnisse eines Zufallsvorgangs sind, heißt Zufallsvariable  $X$ . Die Zahl  $x$ , die  $X$  bei der Durchführung eines Zufallsvorgangs annimmt, heißt Realisierung oder Wert von  $X$

Für unser Beispiel definiere ich die Zufallsvariable  $X$  :

$X :=$  Höhe des Taschengeld von Grundschulkind aus Jena in Euro (1)

# Die Notation der Zufallsvariable und ihrer Realisation

- Großbuchstaben bezeichnen den Zufallsvorgang selbst. Sie also eine Abkürzung für den Zufallsvorgang.
- Die entsprechenden Kleinbuchstaben werden mit einem Index versehen und sind die Werte, die konkret beobachtet wurden.
- So bedeutet z. B.  $x_1 = 14$ , dass das erste Kind in der Befragung 14 Euro Taschengeld im Monat bekommt.
- Entsprechend bedeutet  $x_2 = 9$ , dass das zweite Kind 9 Euro Taschengeld bekommt.

# Die Intention der deskriptive Modellierung

- Man beschreibt die beobachteten Daten mit Hilfe einer Modellverteilung.
- Ziel ist es, die in den Daten bestehenden Sachverhalte zu kommunizieren.
- Ziel ist es **nicht**, eine Aussage zu machen, die über das Beschreiben der Daten hinausgeht.
- Insbesondere bezieht sich ein deskriptives Modell nur auf die vorliegende Stichprobe.

# Modellebene und Datenebene

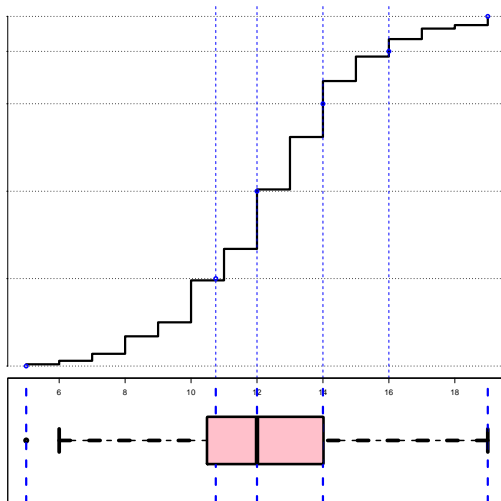
- Die Normalverteilung ist ein Modell, dass sich für symmetrische Verteilungen eignet.
- Die Parameter  $\mu$  und  $\sigma$  sind Platzhalter für den Ort der Symmetrieachse und die Breite der Dichte.
- Wenn man diese Parameter aus Daten schätzt, werden die Schätzer üblicherweise mit  $\hat{\mu}$  und  $\hat{\sigma}$  bezeichnet.
- Durch die Spezifikation der Parameter mit Hilfe der Daten wendet man das Modell auf der Datenebene an.
- Datenebene bedeutet: Man nimmt Bezug zu den vorliegenden Daten
- Modellebene bedeutet: Man spricht über Modellannahmen

## Ergebnis der fiktiven Erhebung in Zahlen

Wir haben die fiktive Erhebung zum Thema Taschengeld abgeschlossen und wollen herausfinden, ob sich die Normalverteilung eignet um die Daten zu beschreiben.

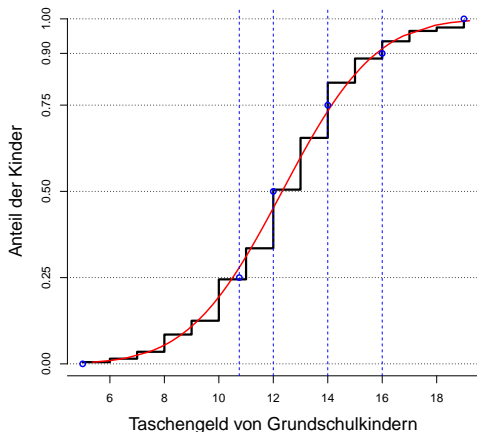
- Ist die beobachtete Verteilung der Daten symmetrisch?
  - ▶ Mittelwert:  $\bar{x} = 12.33$ , Median (= 50% Quantil):  $\tilde{x} = 12,34$
  - ▶ 1. Quartil (= 25% Quantil):  $x_{0.25} = 10.52$
  - ▶ 3. Quartil (= 75% Quantil):  $x_{0.75} = 14.07$
  - ▶ Minimum: 5.36 und Maximum: 19.2
- Sprechen die Lagemaße für eine symmetrische Verteilung?
- Grafische Verfahren werden genutzt um einen Eindruck von den Daten zu bekommen. Wir setzen  $\mu \equiv \hat{\mu} = \bar{x} = 12.33$  und  $\sigma^2 \equiv \hat{\sigma}^2 = 7.26$ .
- Das Zeichen  $\equiv$  bedeutet, dass wir für den Modellparameter einen bestimmten Wert einsetzen.

# Grafische Darstellung: Empirische Verteilungsfunktion und Boxplot



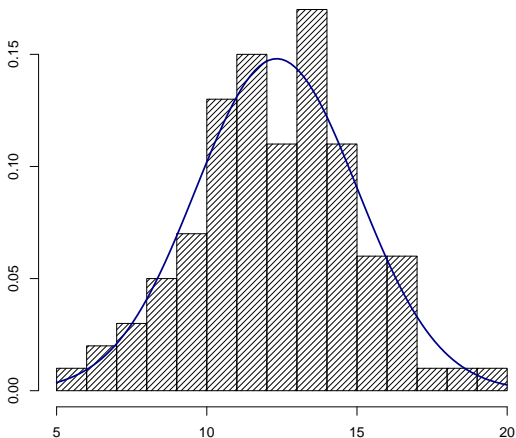


# Grafische Darstellung: Empirische Verteilungsfunktion und theoretische Verteilungsfunktion



# Grafische Darstellung: Histogramm und deskriptive Normalverteilung

## Dichte und Histogramm



# Was haben wir bisher gemacht?

- Wir haben  $n = 100$  Realisationen der Zufallsvariable Taschengeld beobachtet und beschreiben.
- Bisher haben wir nur Methoden der deskriptiven Statistik genutzt. Wir machen also noch keine Aussage die über die Stichprobe hinausgeht.
- Wir nutzen die Normalverteilung als Modell um unsere Daten zu beschreiben und gewinnen den Eindruck, dass diese Modell gut zu den Daten passt
- Wie können Sie die 68-95-99.7-Regel nutzen um die Daten zu beschreiben?

# Darf ich aus der Stichprobe eine allgemeinere Aussage machen?

- Können wir mit Bezug auf unsere Stichprobe behaupten, dass sich das Taschengeld der Grundschulkinder signifikant erhöht hat?
- Es waren im Durchschnitt elf Euro im Jahr 2015.
- Eine Zufallsstichprobe erlaubt genau das zu tun.
- Bei der Art und Weise zu schließen hat man verschiedene Inferenzkonzepte zu Auswahl. Oft führen sie zu ähnlichen Schlüssen.

# Die Zufallsstichprobe

## Bedingungen für eine Zufallsauswahl (aus Methoden I)

- 1 Regelgeleitete Auswahl und
- 2 für jedes Element der Grundgesamtheit ist die Wahrscheinlichkeit in die Stichprobe aufgenommen zu werden bekannt und
- 3 größer Null

## Definition: Einfache Zufallsstichprobe

Eine einfache Zufallsstichprobe liegt dann vor, wenn jede der möglichen auf diese Weise gezogenen Stichproben vom Umfang  $n$  aus einer Grundgesamtheit vom Umfang  $N$  die gleiche Chance hat aufzutreten.

# Was wir in den nächsten drei Wochen machen (bis 10.11.2017)

- Wir wiederholen den zentralen Zusammenhang: Theoretische Dichte und Verteilungsfunktion
- Wir nutzen STATA um den fiktiven Taschengelddatensatz zu beschreiben
- Es wird auch das Geschlecht erhoben, so dass wir die Verteilung des Taschengelds in den beiden Gruppen vergleichen können.
- In der nächsten Vorlesung lernen Sie das Modell der Binomialverteilung kennen.
- In der Übung werden wir auch dieses Modell mit STATA erkunden.
- Inferenzschlüsse werden noch nicht gezogen, aber wir sprechen darüber, was es bedeutet das zu tun und unter welchen Voraussetzungen wir es tun dürfen.

# Bekommen Jungen mehr Taschengeld als Mädchen?

Bitte lesen sie die beiden Artikel

① Statistik und Wahrheit - Hauptsache spektakulär

<http://www.spiegel.de/wissenschaft/mensch/taschengeld-gibt-es-wirklich-einen-gender-pay-gap-a-1100475.html>  
Erscheinungsdatum: 27.7.2016

② Umfrage: Jungen bekommen mehr Taschengeld als Mädchen

<http://www.spiegel.de/lebenundlernen/schule/jungen-bekommen-mehr-taschengeld-als-maedchen-a-1161915.html>  
Erscheinungsdatum: 8.8.2017

Machen Sie sich ein paar Stichpunkte zu dem Ergebnis,  
dass der zweite Artikel Ihnen Nahe legt und kommentieren Sie es kritisch.