Social Robotics

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Felix Lindner, Laura Wächter, Bernhard Nebel

SoSe 2019

UNI FREIBURG

Introduction to Empirical Research Methods

Outline

UNI FREIBURG

- The Empirical Research Method
  - Hypotheses
  - Variables
- Some Basics of Descriptive Statistics
  - Measures of Centrality
  - Measures of Variability

Empirical Questions

UNI FREIBURG

- Empirical questions can be answered by observation:
  - Do people react differently to a human-like robot as compared to a mechanical one?
  - Have people been interacting less often with each other since smartphones exist?
  - Are people more skeptical towards robots the older they are?
- Counterexamples
  - $P \neq NP$?
  - Does God exist?
  - Can a robot be a moral agent?

Research Process

UNI FREIBURG

1. Initial Observation
2. Theory
3. Hypothesis
4. Data Collection
5. Data Analysis (goto 2)

Research Process: Example

UNI FREIBURG

1. Initial Observation
   - A child with autism spectrum disorder (ASD) likes to play with the less expressive robot rather than with the expressive one.
2. Theory (Explanation)
   - Children with ASD cannot deal well with too much stimuli.
3. Hypothesis (Prediction)
   - The less expressive robot R1 yields more therapy success compared to the more expressive one R2.
4. Data Collection (Experiment)
   - Expose 15 children with ASD to R1 (condition 1) and 15 children with ASD to R2 (condition 2). Measure therapy success of the 30 trials.
5. Data Analysis (goto 2)
   - Compare the data of the two conditions using statistics. If hypothesis is supported by data, this adds some evidence to the validity of the theory.

Hypothesis: Types of Hypotheses

UNI FREIBURG

Hypotheses are very central to empirical research. They are the statements that can be tested in an experiment and checked for statistical significance. We distinguish four types of hypotheses:

$$\{\textit{difference}, \textit{relationship}\} \times \{\textit{directional}, \textit{nondirectional}\}$$

Quiz: Classify the following hypotheses

- A treatment with Paro leads to behavior change.
- A treatment with Paro leads to higher approachability.
- A patient's age and its affection towards Paro are related.
- The older a patient the more she benefits from a treatment with Paro.

Data Collection: Dependent and Independent Variables

UNI FREIBURG

- During data collection (experiments) we as researcher manipulate some aspect(s) and measure the effect of the manipulation onto other aspect(s).
  - E.g., we manipulate the expressiveness of the robot and measure the success of the therapy.
- This process is formalized using Variables
  - E.g., one variable for expressivenes that can have one of two values: more expressive, less expressive (V1), and another variable for the success (V2). Because V1 is (presumingly) determined externally (by the researcher), it is called independent. Because V2 is determined by V1, it is called dependent.

Data Collection: Variables, Values, and Scores

UNI FREIBURG

- Variable: any entity that can take on different values
- Value: any number or category
- Score: an individual value

Example

- Age: 44, 22, 18, 19, 27, 22, 18
  - One Variable (Age)
  - Five Values (18, 19, 22, 27, 44)
  - Seven Scores.

Data Collection: Levels of Measurement

- Categorial Variable: Values have no natural order. E.g., cultural background.
- Ordinal: Values can be ordered. E.g., Likert-scale ratings.

	Gar nicht	Wenig	Ziemlich	Absolut
Würdest Du gerne wieder ein Gespräch mit Immanuel führen?	☐	☐	☐	☐

Figure: Likert Scale Example

- Interval: Distances between values are equally spaced. E.g., Age.
- The level of measurement restricts what can be reasonably done mathematically with the data. E.g., it makes no sense to compute arithmetic means in case of categorial data.

Data Collection: Decision Procedure

UNI FREIBURG

1. **If** values can be sorted into categories **then** goto (2) **else** rethink your design
2. **If** values can be put into a meaningful order **then** goto (3) **else** Categorial
3. **If** the distance between consecutive values is exactly the same **then** Interval **else** Ordinal

Quiz

UNI FREIBURG

What type of variable is *grade*?

Data Collection: Validity and Reliability

UNI FREIBURG

During the experiment, variables are measured using appropriate instruments of measurement, e.g., questionnaires. An appropriate instrument fulfills two requirements:

- Validity: The instrument measures what it is intended to measure.
- Reliability: The instrument is robust, i.e., it gets you the same result if you repeat the measurement at different times.

Therefore, make use of validated questionnaires whenever possible, instead of making up your own!

Data, Data, Data

UNI FREIBURG

- After data collection is finished, the researcher faces a huge amount of data.
- The data now has to get analysed. As a first step, we make use of so-called statistics to describe the data.

Data Analysis: Frequency Distributions

UNI FREIBURG

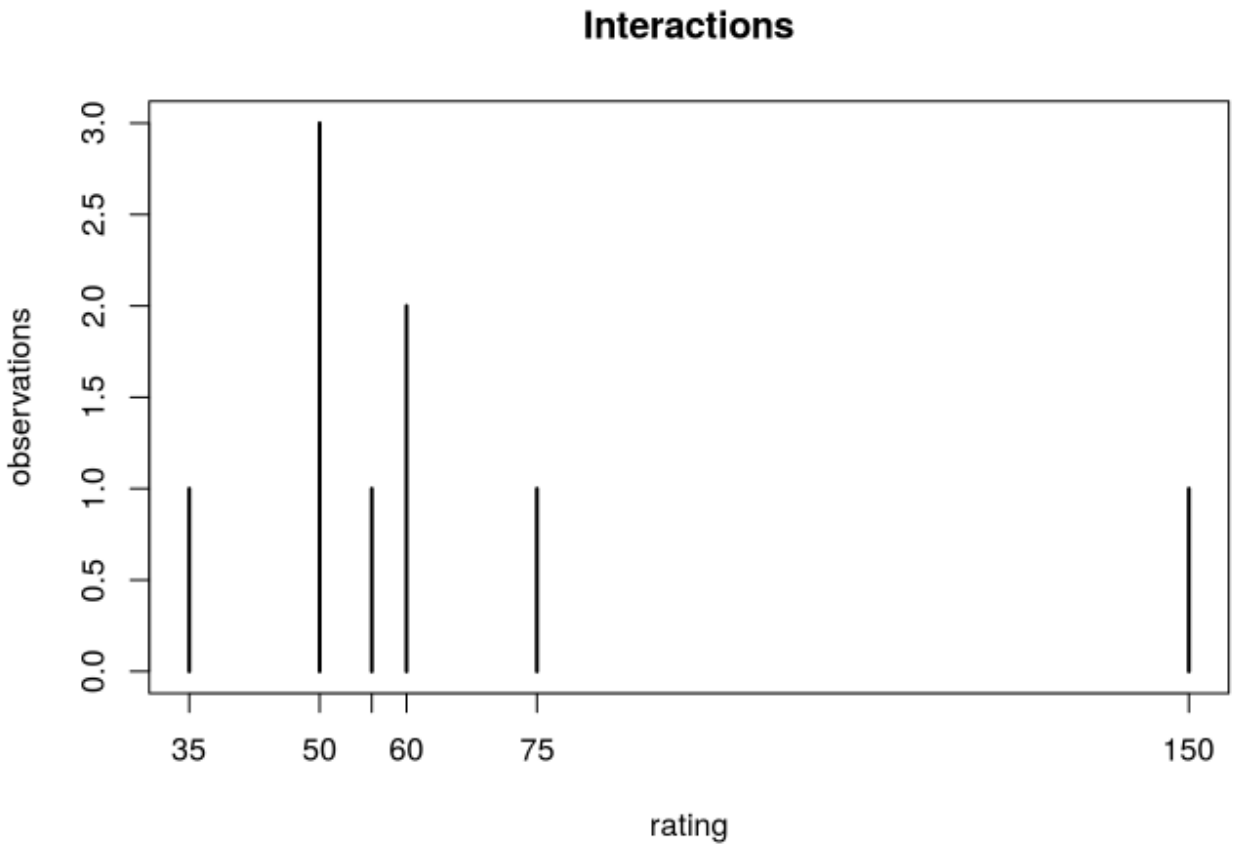
Assume you counted the numbers of interactions with a social robot per day.

- Variable #Interactions of type Interval with values from $\mathbb{N}_0$
- Scores ($N = 9$): 50, 35, 50, 50, 150, 56, 60, 75, 60
- (For today's purpose we will not make the distinction between the sample and the population we took the sample from!)

The first thing you can do to get an overview of the data is to plot a frequency distribution.

Data Analysis: Frequency Distributions

```
data <- c(50, 35, 50, 50, 150, 56, 60, 75, 60)
freq <- table(data)
plot(freq, main="Interactions", xlab="rating", ylab="observations")
```



Data Analysis: Measures of Centrality

UNI FREIBURG

- Mean: The sum of scores X_i divided by the number of scores N in the data.
  - $\overline{X} = \frac{1}{N} \sum_i^N X_i$
  - Only makes sense for interval data.
- Median: The middle score in the data, i.e., 50% of the data is below the median.
  - Order data, pick the middle data point (or the mean between the two middle data points in case $|X|$ is even).
  - Ordinal or Interval Data.
  - Compared to the mean, it is more robust when outliers are present.
    - Example $X = (1, 1, 2, 2, 100)$, $\overline{X} = 21.5$, $Mdn = 2$. The mean actually describes the data not very well, because almost all data points are smaller.
- Mode: The most common value in the data.
  - Categorial scores: Report which score is most typical.

Data Analysis: Measures of Centrality

UNI FREIBURG

```r
# Mean
data.mean <- mean(data)

# Median
data.median <- median(data)

# Mode
max <- which.max(freq)
data.mode <- sort(unique(data))[max]

sprintf("Mean: %f, Median: %f, Mode: %f", data.mean, data.median, data.mode)
```

```
## [1] "Mean: 65.111111, Median: 56.000000, Mode: 50.000000"
```

Data Analysis: Measures of Variation

UNI FREIBURG

- Range: Distance between the minimum and the maximum.
  - $range = X_{max} - X_{min}$
- (Biased) Variance: Average of the squared differences from the mean.
  - $s^2 = \sum_i^N (X_i - \overline{X})^2 / N$, with N being the number of scores in the data.
- Standard Deviation: Square root of the variance.
  - $s = \sqrt{s^2}$

Again, as s^2 and s depend on the mean, these measures only make sense for interval variables.

Data Analysis: Quartiles

UNI FREIBURG

Yet another statistics of variability are quartiles. This also works for ordinal data.

- Quartiles
  - $Q1$ is the value such that 25% of the data is below $Q1$ and 75% is above.
  - $Q2$ is just the median.
  - $Q3$ is the value such that 75% of the data is below $Q3$ and 25% is above.

There is no generally agreed-upon method of how to compute Quartiles (R implements 9 different methods). Simplest method: 1) Order values $x_1 < \ldots < x_n$, 2) compute median, 3) take the values x_l, x_r left and right from the median, 4) compute the median for the values $x_1 < \ldots < x_l$ and $x_r < \ldots < x_n$.

Data Analysis: Quartiles

UNI FREIBURG

Compute quartiles for our data

50, 35, 50, 50, 150, 56, 60, 75, 60

1. 35, 50, 50, 50, **56**, 60, 60, 75, 150
2. $Q2 = Mdn = 56$
3. 35, **50, 50**, 50 | 60, **60, 75**, 150
4. $Q1 = (50 + 50)/2 = 50$ | $Q3 = (60 + 75)/2 = 67.5$

Data Analysis: IQR

UNI FREIBURG

Interquartile Range (IQR) describes the distance between Q3 and Q1: $IQR = Q3 - Q1$. Depending on the data, IQR can be more accurate than standard deviation.

Data Analysis: Measures of Variation

UNI FREIBURG

```r
# Range
data.range <- diff(range(data))

# Variance
data.variance <- var(data)

# Standard deviation
data.sd <- sd(data)

# IQR
q <- quantile(data, type=6) # Type 6 is the SPSS way of doing it
data.q3 <- q["75%"]
data.q1 <- q["25%"]
data.iqr <- q["75%"]-q["25%"]

sprintf("Range: %f, s^2: %f, s: %f, IQR: %f, (Q3: %f, Q1: %f)", data.range, data.variance, data.sd, data.iqr, dat
a.q3, data.q1)
```

```
## [1] "Range: 115.000000, s^2: 1128.861111, s: 33.598528, IQR: 17.500000, (Q3: 67.500000, Q1: 50.000000)"
```

Reporting Descriptive Statistics

UNI FREIBURG

The measures calculated so far can be used to report the statistical results.

Example 1

We recorded the number of interactions with our robot per day for nine days (N = 9). The number of interactions ranged from 35 to 150 ($\overline{X}$ = 65.11, s = 33.59).

Example 2

We recorded the number of interactions with our robot per day for nine days (N = 9). The number of interactions ranged from 35 to 150 (Mdn = 56, IQR = 17.5).

- Remember the data 35, 50, 50, 50, 56, 60, 60, 75, 150.
- Which description seems more appropriate?

Data Analysis: Outliers

UNI FREIBURG

Standard deviation and IQR can be used to judge the “normality” of a given value, i.e., to detect outlier. For IQR, as a matter of convention, a distance of 1.5IQR or more below/above Q1/Q3 is classified as outlier.

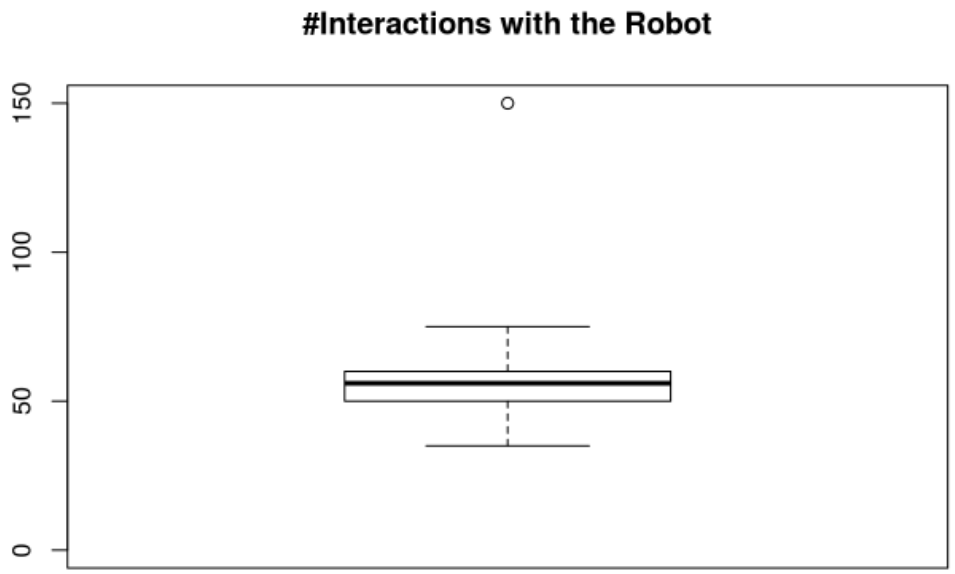
Example

In our data, we have $IQR = Q3 - Q1 = 67.5 - 50 = 17.5$. Hence, values below $50 - 1.5 \times 17.5 = 23.75$ and above $67.5 + 1.5 \times 17.5 = 93.75$ are considered outliers in our data, viz., 150.

- 150 is $(150 - 67.5)/17.5 = 4.71$ IQRs above $Q3$

Data Visualization: Boxplots

```
boxplot(data, main="#Interactions with the Robot", ylim=c(0,150))
```



- Dots signify outliers
- Box bounded by Q1 and Q3
- Thick line signifies Median
- Whiskers connect scores within the 1.5IQR

z-Score

UNI FREIBURG

The z-Score is another measure of distance of some score X in terms of number of deviation, viz., the number of standard deviations from the sample mean $\overline{X}$. The computation of the z-Score is very similar to the computation of the distance in terms of IQR.

Example

We want to measure how many standard deviations 150 is away from the mean. We know from above that $\overline{X} = 65.11$ and $s = 33.59$. Hence, we solve $(150 - 65.11)/33.59 = z = 2.52$. Hence, we know that 150 is 2.52 standard deviations above the mean.

This generalizes to the computation of the z-Score:

$$z = (X - \overline{X})/s$$

z-Score: Yet Another Example

UNI FREIBURG

Let's assume a fleet of robots. We measure number of tasks of each robot over two weeks and get:

- $\overline{X}_{week1} = 120, s_{week1} = 50,$
- $\overline{X}_{week2} = 90, s_{week2} = 5.$

Question: In which week performed our particular robot who completed 130 tasks in the first week, and 100 task in the second week, better relative to the other robots?

- $z_{week1} = (130 - 120)/50 = 0.2$
- $z_{week2} = (100 - 90)/5 = 2.0$

Thus, our robot performed average in week 1 and extremely well in week 2 (relative to the other robots).

z-Score

UNI FREIBURG

If our variable is normally distributed, then also the z-Score is ($z \sim \mathcal{N}(0,1)$). Therefore, if we know the z-Score of some score X, then we can determine the probability of measuring a score at least as extreme as X by computing $1 - P(\leq |z|) + P(\leq -|z|) = 2 \times P(\leq -|z|)$. If we find that the probability is very low, then we can call X extreme.

- Luckily, we can look these probabilities up in tables, e.g., `http://www.z-table.com`, or we ask R using `pnorm`.
- It turns out that a z-Score at least as extreme as 1.96 has a probability of only 0.05 (5%). By convention (most of the time), probabilities of 0.05 or below count as extreme.
- Of course, talking about probabilities of events here already goes beyond our actually observed data. However, we cannot be sure our $\overline{X}$ and s apply beyond our sample!
  $\Rightarrow$Next time: Inferential Statistics

Sketches

Intentionally left blank :-)

UNI FREIBURG