# Graphical Representation
# Introduction to Probability Theory

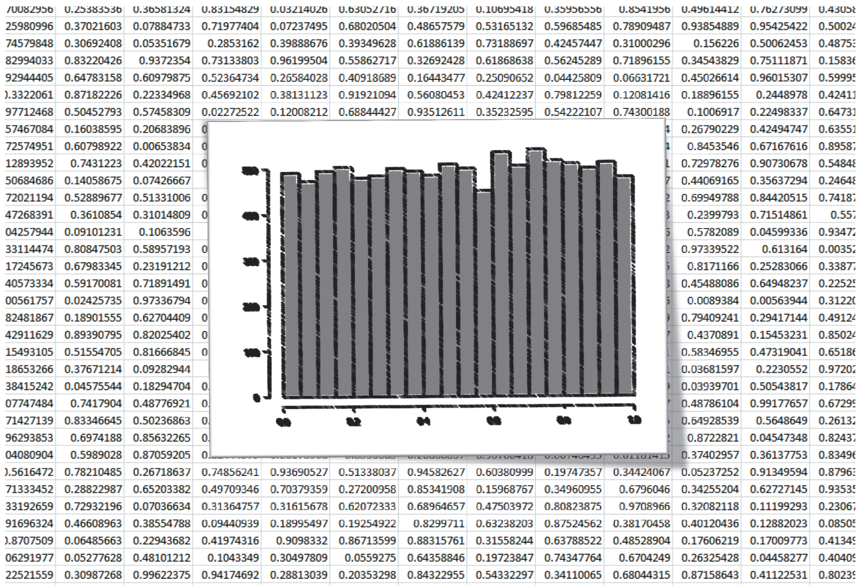
Peter Büchel

HSLU W

SA: Week 2



## Graphical Representation of Data





- Plotting data: Very important aspect of statistical data analysis
- Structure of data is often visible at a glance
- Often sees patterns that are not recognisable from key figures
- Two methods to graphically represent one-dimensional data:
  - Box plot
  - Histogram

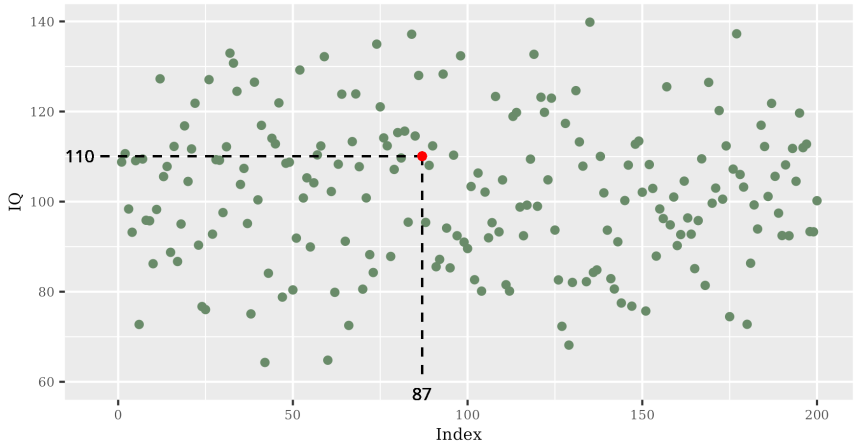


## Bad Graphical Representation of Data

- Choosing “wrong” graphical representation is *not* useful
- Important to choose “correct” graphical representation

## Example: Not Useful Graphical Representation

- Simulate the IQ test of 200 persons
- Plot data in coordinate system:



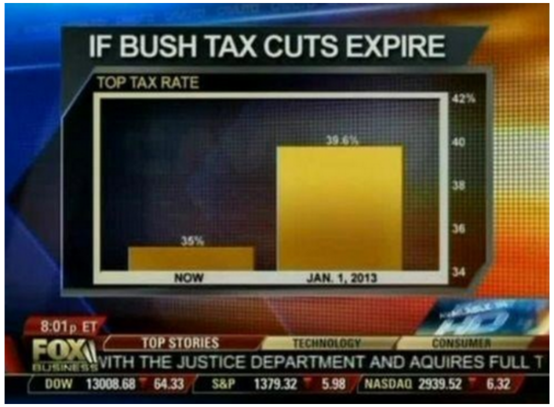



- Horizontal axis (`index`): 200 peoble
- Vertical axis (`iq`) corresponding IQ
- Red dot: Corresponding to IQ (about) 110 of 87th person
- Obviously no clear pattern can be seen
- Reason: IQ of people are not ordered by ascending IQ
- This type of graph is therefore not useful
- Not every graphical representation is simply helpful in itself
- Graphical representations can be more confusing than helpful if they have been created inappropriately



## Example

- Fox News showed following chart of what would happen if Bush tax cuts expired

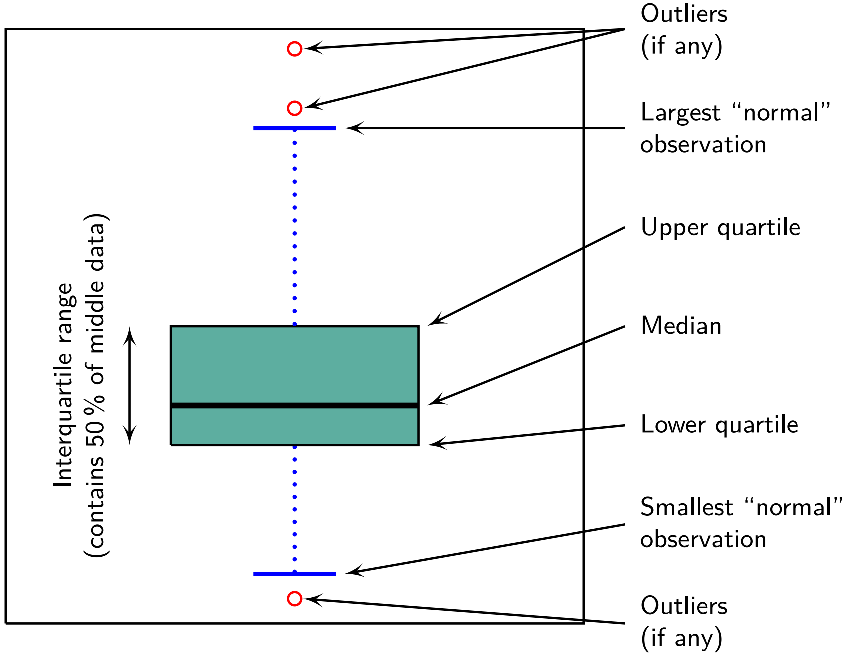



- Disastrous, right? Seems about a fivefold increase



- However, check scale: Starts at 34 and ends at 42
- All is not what it seems: Increase is only about 5 %

## Boxplot: Schematically



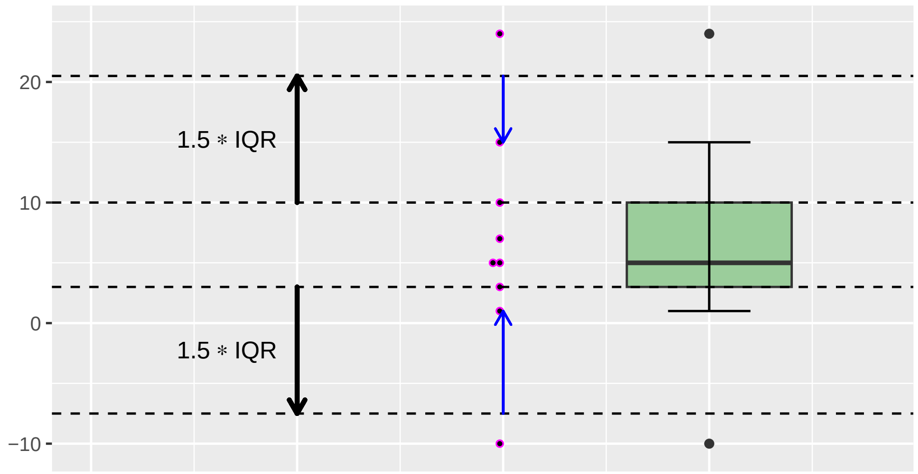



- Box: Height is bounded by lower and upper quartile
- Height of box is interquartile range: Range of 50 % of middle observations
  - If height is small: Small spread
  - If height is large: Large spread
- Horizontal line in box: Median (black)
- Lines, which lead from box to the smallest or largest "normal" observation (blue)
  - Definition: "Normal" observation no more than 1.5 times the interquartile range from one of the two quartiles
  - Why 1.5? Introduced by inventor John Tukey (around 1970)
- Outlier: Small circles (red)



## Remark

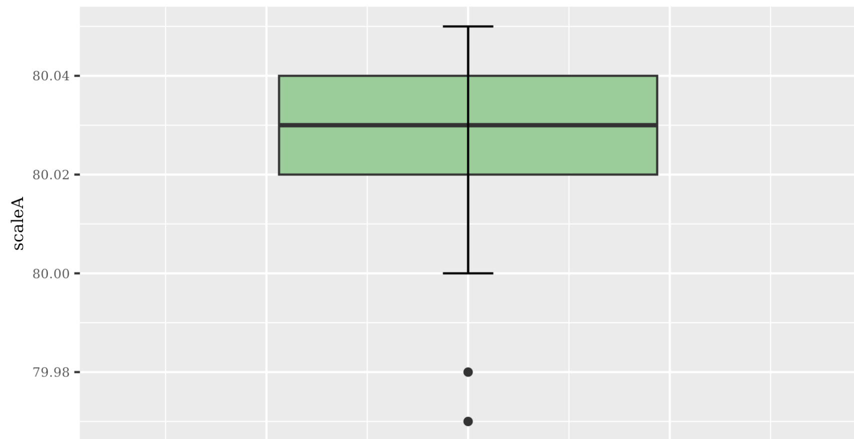
- Upper and lower whisker do *not* have to have the length:






## R: Example Scale *A*

```
boxplot(scaleA,
        col = "darkseagreen3"
)
```



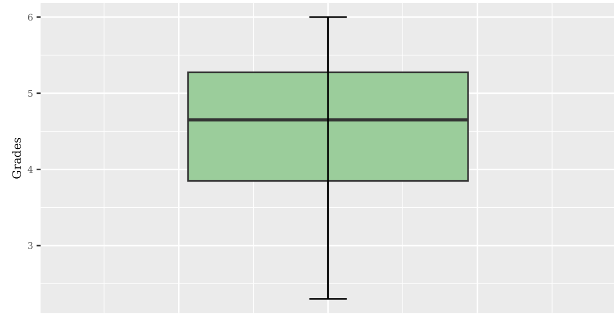
- Half of observations: Between upper quartile 80.04 and lower quartile 80.02, interquartile range 0.02
- Median: 80.03
- "Normal" range of values: Between 80.00 and 80.05
- Two outliers: 79.97 and 79.98
- First two points: Previously calculated
- Boxplot: Median and quartiles graphically displayed



## Examples: Grades

- Box plot of grade:

```
boxplot(grades,
        col = "darkseagreen3"
)
```



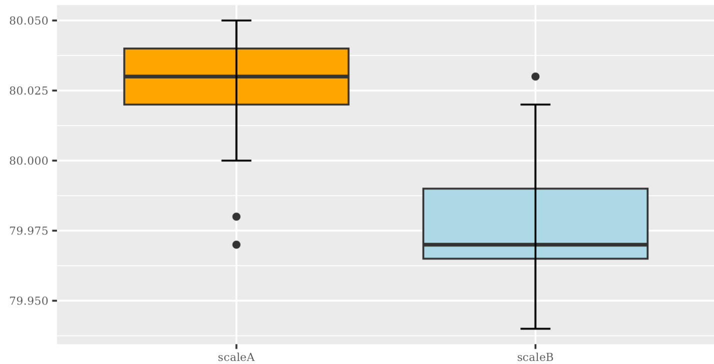

- Again: Values for median and quartiles correspond to values that already calculated



## Comparison of Datasets

- Boxplot: Display of different groups

```
boxplot(scaleA, scaleB,
        xlab = "scale",
        col = c("orange", "light blue")

axis(side = 1, at = c(1, 2), labels = c("scaleA", "scaleB"))
```






- Scale $A$ larger values than scale $B$: Median of $A$ is larger
- Boxes do not overlap
- Data from scale $A$ have less spread than data from scale $B$
  $\rightarrow$ Box less high (interquartile range!)
- R code `axis(...)`: See lecture notes

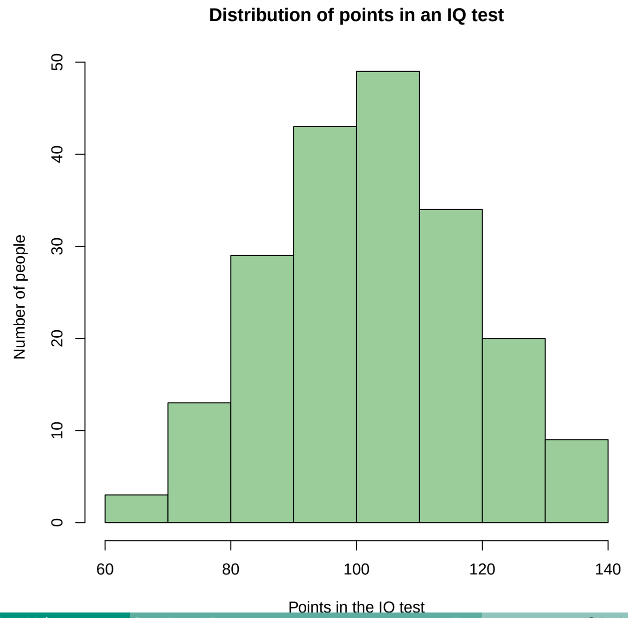
# Histogram

- *Histogram*: Graphical overview of occurring values
- Draw a *bar* for each class: Height proportional to number of observations in that class



# Example: IQ test

- Figure: Histogram of result of an IQ test of 200 people






- Data simulated
- Width of classes: 10 IQ points, same for each class
- Height of bars: Number of people falling into that class
- Example: About 20 people fall into class between 120 and 130 points
- Shape of this histogram typical for many histograms: Normal distribution (later)



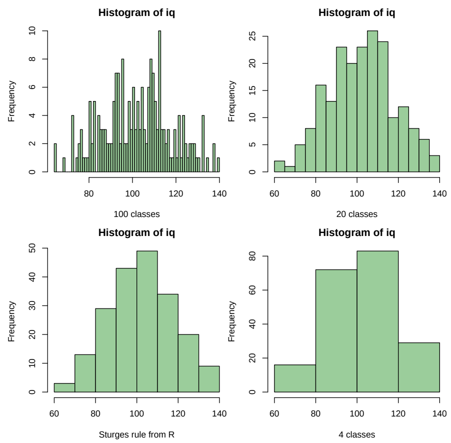
- Code: Histogram above

```
iq <- rnorm(n = 200, mean = 100, sd = 15)

hist(iq,
     col = "darkseagreen3",
     xlab = "Points in the IQ test",
     ylab = "Number of people",
     main = "Distribution of points in an IQ test"

)
```

- `rnorm(n = 200, mean = 100, sd = 15)`: Selects randomly 200 normally distributed data (see later) with mean 100 and standard deviation 15
- Command `hist(iq, …)`: Histogram for data `iq`
- Further options should be clear:
  - `xlab`: $x$-label, label of $x$-axis
  - `ylab`: $y$-label, label of $y$-axis
  - `col`: Colour
  - `main`: Main title

## Choice of Classes

- Selection of number of classes relevant for interpretation of histogram
- There is no general rule how to choose number of classes
- Figure: IQ data of example with different number of classes






- Code: See lecture notes
- Histogram top left: Much too detailed to recognise a pattern
- Histogram bottom right too rough
- Default number of classes for R: *Sturges* rule (see lecture notes)
- Produces generally good results
- Change number of classes: Use option `breaks =`



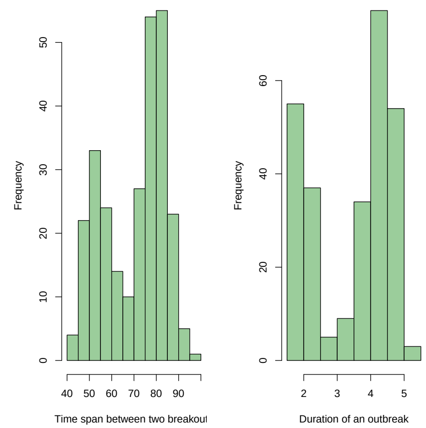
## Old Faithful Geyser (Yellowstone NP)

- Geyser Old Faithful (Yellowstone National Park): Known hot spring
- Time between two eruptions and duration of eruptions of great interest to spectators and National Park Service
- 299 measurements of successive eruptions
- Dataset included in R:

```
geyser <- faithful
head(geyser)

  eruptions waiting
1     3.600      79
2     1.800      54
3     3.333      74
4     2.283      62
5     4.533      85
6     2.883      55
```



- Illustration Histograms:




- Duration of an outbreak (right)
- Time span between two outbreaks (left)

- For both histograms: *Bimodal* behaviour visible
- There are two "humps" in the histogram:
  - Time span between two outbreaks: Duration relatively short (around 50 minutes) or rather long (around 80 minutes)
  - Duration between two outbreaks not "evenly" distributed
  - Same behavior for the duration of an outbreak: Either outbreak is relatively short (about 1.5–2 minutes) or long (about 4–4.5 minutes)

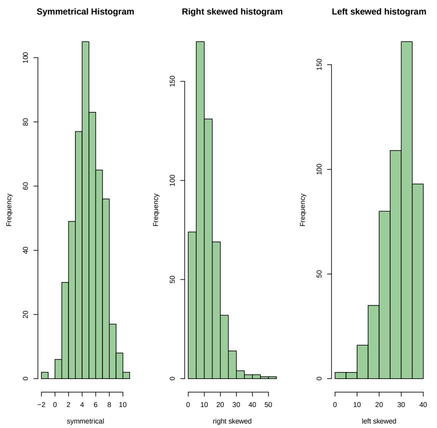


- Question: Is there a correlation between eruption duration and time span between two eruptions?
- Or in other words:
  - Does it take long after a long outbreak until there is another outbreak?
  - Or does an outbreak return very quickly?
  - Or is there no connection at all?
- Questions answered later



## Skewness of Histograms

- Illustration:






- Left Histogram: Symmetrical with respect to approximately 5
- Data is similarly distributed by 5 on both sides
- Middle histogram: Data concentrated left and flatten out towards right: Called a *right skewed* histogram
- Right histogram: Data concentrated right and flatten out towards left: *Left skewed* histogram
- Term "right" and "left": Always refers to direction where it has *less* data (tail of distribution)

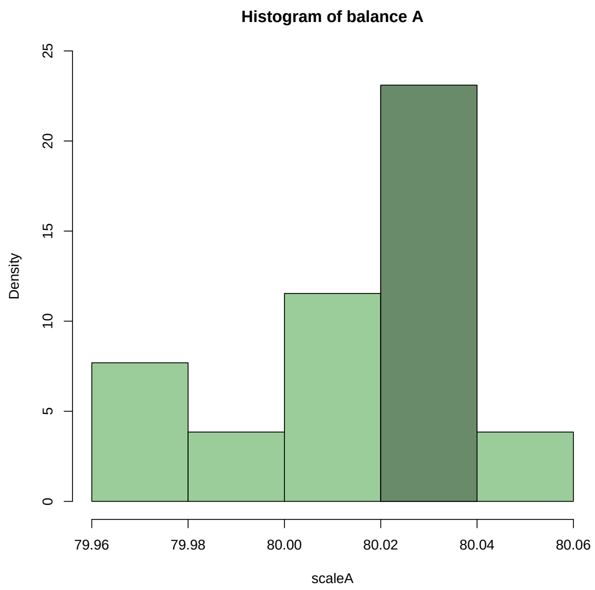
## Normalised Histogram

- In histograms so far: Height of bars corresponds to number of observations in a class
- Often better interpretable: Select bar height such that *bar area* corresponds to percentage/proportion of respective observations in total number of observations
- Total area of all bars must be equal 1
- *Density*: Indicated on the vertical axis
- Important: Density values are *not* percentages



## Example: Scale $A$

- Normalised histogram:






- Density of class of $80.02 - 80.04$ is about 23
- *Area* of this bar (dark green area in figure):

$$(80.04 - 80.02) \cdot 23 = 0.46$$

- Area multiplied by 100: Percentage of data within this bar
- About 46 % of data lies between 80.02 and 80.04



## R-Code

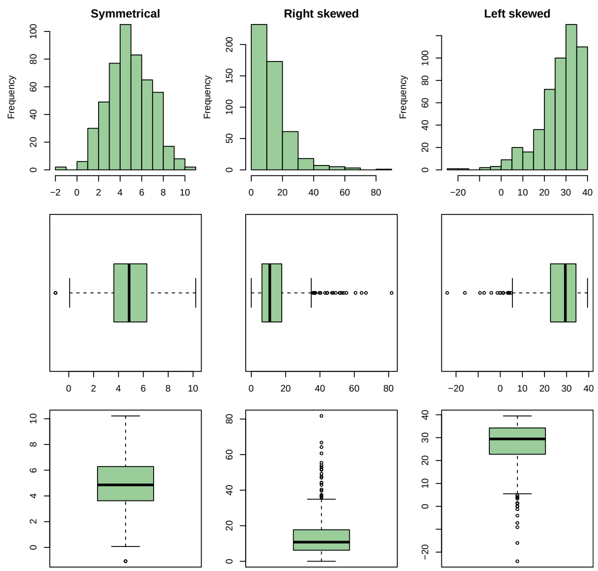
- Code:

```
par(mar=c(4,4,2,0))
hist(scaleA,
     freq = F,
     main = "Histogram of balance A",
     col = "darkseagreen3",
     ylim = c(0, 25)

)
rect(80.02, 0, 80.04, 23.1, col="darkseagreen4")
```

- Option `freq = F` (*frequency false*): Histogram is drawn normalised
- Option `ylim = c(0, 25)`: See lecture notes
- `rect(80.02, 0, 80.04, 23.1, col = "darkseagreen4")`: See lecture notes

## Skewness in Boxplot

- Illustration:



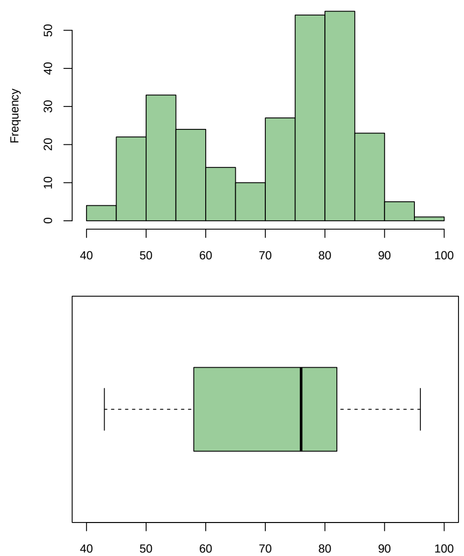



- Symmetrical diagram left: Median in center of box
- Right skewed histogram (center): Median no longer in center of the box, but shifted to left
- Distance from lower quartile to median is smaller than distance from median to upper quartile
- From lower quartile to median: A lot of data lie within a small range
- From median to upper quartile: Much larger range is needed until 25 % of data lie within this interval
- Left skewed histogram: Interpretation other way round



## Example: Old Faithful

- Figure: Histogram and box plot of time between two eruptions of Old Faithful:






- Data are left skewed
- Box plot: 50 % of “middle” time spans between 60 and 80 minutes
- Median is about 75 minutes
- Data between median and upper quartile are in a range of 5 minutes (from 75-80 minutes)
- There are a relatively large number of time spans in this area compared to 15-minute interval from lower quartile to median

## Preliminaries: Reminder Set Theory

- Probability models: Using set theory as language
- Don't despair: Little more than notation is needed
- *Set*: Collection of "things"
- These "things" can be very general, but "our" sets are quite simple



## Example

- Rolling a die
- Possible outcomes: Numbers 1 to 6
- Can regard these numbers as a collection of numbers, i.e. as a set
- Sets generally denoted by a capital letter
- Let's denote set above $A$:

$$A = \{1, 2, 3, 4, 5, 6\}$$

- Members or *elements* of a set: Written within curly brackets
- Say that 2 is an element of $A$ and write

$$2 \in A$$

- Symbol $\in$: "... is element of ...".



- Number 7 is not an element of $A$:

$$7 \notin A$$

- Symbol $\notin$: "... is not element of ..."
- Often interested in sets which are part of larger set
- Let:

$$B = \{2, 5\}$$

- $B$ part of $A$: *All* elements of $B$ are also elements of $A$
- $B$ is a *subset* of $A$:

$$B \subset A$$

- Symbol $\subset$ stands for "... is subset of ..."



- A set is subset of itself:

$$A \subset A$$

- Set, which has no elements (unassuming but important): *Empty set*

$$\{\} \quad \text{or} \quad \emptyset$$

- By *definition*: Empty set is subset of any set

$$\{\} \subset A \quad \text{or} \quad \{\} \subset \{\}$$

## Probability

- Everybody has an intuitive feeling what probability is
  - Probability to roll a 4 with a fair die is one sixth
- But: Exact interpretation of probability surprisingly difficult
- Statement "It rains tomorrow with a probability of 80 %"
  $\rightarrow$ Anything but obvious what is meant
- See remarks end of Chapter 3 in lecture notes



## Probability Model

- *Random experiments*: Outcome is not predictable:
  - Rolling a die
  - Tossing a coin
  - Number of calls to a call center in one hour
- *Probability model* consists of:
  - *Events* that are possible in such an experiment
  - *Probabilities* for different results occurring
- Example: Rolling a die
  - Possible results: $1, 2, 3, 4, 5, 6$
  - Probability to roll one of these numbers: $\frac{1}{6}$ (if die is fair)



- Probability models have following components:
  - *Sample space* $\Omega$: Contains *all* possible elementary events $\omega$
  - *Events* $A$, $B$, $C$: Subsets of sample space
  - *Probabilities* $P$ associated with events $A$, $B$, $C$
- *Elementary event*: *Possible* result (outcome) of random experiment
- All elementary events form sample space:

$$\Omega = \{\underbrace{\text{all possible elementary events } \omega}_{\text{all possible outcomes/results}}\}$$



## Example: Rolling a Die

- Sample space (all possible results):

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- Element $\omega = 2$ is an elementary event
- Interpretation: Number 2 rolled
- Number 7: *Not* an elementary event, because not in sample space $\Omega$

# Example: Incoming Calls to Call Center

- Number of calls in one hour to call center
- Sample space (at least theoretically any number of calls possible):

$$\Omega = \{0, 1, 2, 3, 4, \dots\}$$

- Elementary event $\omega = 6$: Six incoming calls in one hour



# Example: Tossing Coin Twice

- Tossing coin twice
- Notation $H$: "head", $T$: "tail"
- All possible results of experiment (sample space)

$$\Omega = \{HH, HT, TH, TT\}$$

- Elementary event: $\omega = HT$ for example
- Tossing $H$ first, then $T$
- Note: $HT$ and $TH$ are different elementary events



# Event

- *Event*: More general and more important than elementary events, but consist of these
- *Event A*: Subset of $\Omega$:

$$A \subset \Omega$$

- "Event $A$ occurs": Result $\omega$ of experiment belongs to $A$



# Example: Tossing Coin Twice

- Event $A$: Exactly one $H$ is tossed
- Event $A$: Consists of elementary events $HT$ and $TH$
- Event $A$ is set

$$A = \{HT, TH\}$$

- Tossing $TT$: Event $A$ does *not* occur
- Probability that $A$ occurs (if coin is fair):

$$P(A) = \frac{2}{4} = \frac{1}{2}$$

- Statistics: Probabilities often denoted by $P$ or $p$

# Example: Rolling a Die

- Event $A$: "Number rolled is odd"
  - Then
    $$A = \{1, 3, 5\}$$
  - Event $A$ occurs, e.g. if number 5 is rolled
  - Probability that $A$ occurs (fair die):
    $$P(A) = \frac{3}{6} = \frac{1}{2}$$
- Event $B$: "Toss a number smaller than 7"
  - Of course, this is always the case and therefore
    $$B = \Omega$$
  - $B$: *Certain* event
  - Probability that $B$ occurs (fair die):
    $$P(B) = \frac{6}{6} = 1$$



- Event $C$: "Rolling a 7"
  - This is impossible:
    $$C = \{\}$$
  - Empty set $\{\}$ (or $\emptyset$) contains no element
  - Event $C$: *Impossible* event
  - Probability that $C$ occurs (fair die):
    $$P(C) = \frac{0}{6} = 0$$

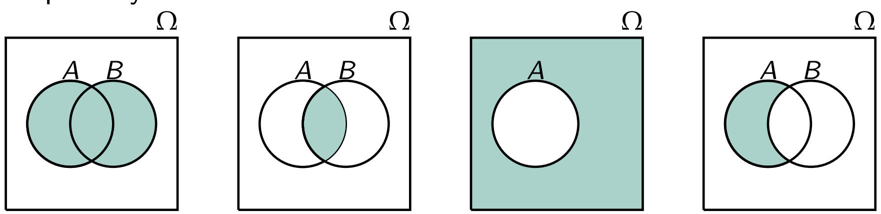


# New Sets from Known Ones

- Operations of set theory for events:

| Name | symbol | meaning |
|---|---|---|
| Union | $A \cup B$ | A *or* B, non-exclusive "or" |
| Intersection | $A \cap B$ | A *and* B |
| Complement | $\overline{A}$ | *not* A |
| Difference | $A \backslash B = A \cap \overline{B}$ | A *without* B |

- Graphically:

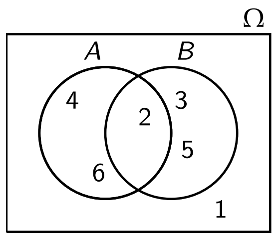



$A \cup B$ $\quad$ $A \cap B$ $\quad$ $\overline{A}$ $\quad$ $A \backslash B$



# Example: Rolling a Die

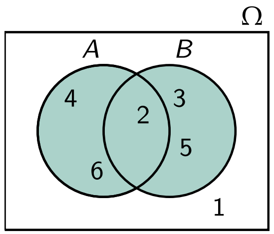
- Event $A$: Number rolled is even:
  $$A = \{2, 4, 6\}$$
- Event $B$: Number rolled is prime:
  $$B = \{2, 3, 5\}$$
- $\Omega$ as usual:
  $$\Omega = \{1, 2, 3, 4, 5, 6\}$$
- Figure:



- *Union*: All elements that are either in $A$ *or* in $B$ *or* in both sets:

$$A \cup B = \{2, 3, 4, 5, 6\}$$

  - Figure:



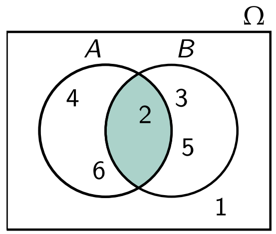

  - Element 2 is in set $A$ *and* in set $B$



- *Intersection*: All elements that are in $A$ *and* in $B$:

$$A \cap B = \{2\}$$

  - Figure:



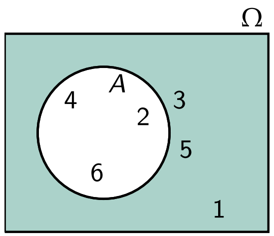

  - Element 2 only element that is both in set $A$ and in $B$



- *Complement*: All elements of $\Omega$ that are *not* in corresponding set:

$$\overline{A} = \{1, 3, 5\}, \qquad \overline{B} = \{1, 4, 6\}$$

  - Figure:



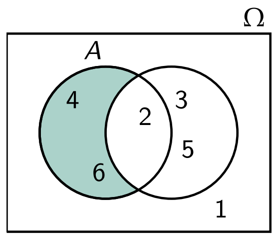

  - Set $\overline{A}$: Odd numbers



- *Difference*: All elements of set $A$, but which are not in set $B$:

$$A \backslash B = \{4, 6\}$$

  - Figure:




  - 2 is in $A$ *and* in $B$ and therefore does *not* belong to difference

## Axioms of Probability

- Properties of probabilities

**Kolmogorov Axioms of Probability**

Each event $A$a *probability* $P(A)$ is assigned, with properties:

A1: $P(A) \geq 0$

A2: $P(\Omega) = 1$

A3: $P(A \cup B) = P(A) + P(B) \quad \text{if } A \cap B = \{\}$

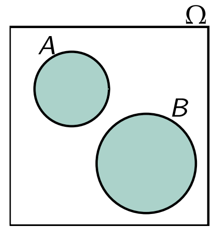
- Notation $P(A)$: Probability that event $A$ occurs
- Event $A$: Rolling an odd number (with fair die)

$$P(A) = \frac{1}{2}$$

- Letter $P$ stands for *probability*



- A1: Probability cannot be negative
- A2: With $P(\Omega) = 1$: Probability of an event between 0 and 1
- Mathematics (Statistics): Probabilities almost never in percent
- A3: For two *disjoint* events:




  - Probability that one of the two occurs, equals to add probabilities of the two events
  - A3 does *not* apply, if events are *not* disjoint



## Example for Not Disjoint Sets

- Example fair die: $A = \{2, 4, 6\}$, $B = \{2, 3, 5\}$
- Then $P(A) = P(B) = \frac{1}{2}$
- $A \cup B = \{2, 3, 4, 5, 6\}$
- Apply A3:

$$P(A \cup B) = P(A) + P(B) = \frac{1}{2} + \frac{1}{2} = 1$$

- Can't be, because $P(A \cup B) = \frac{5}{6}$
- Reason why A3 fails: $A \cap B = \{2\} \neq \{\}$



## Example

- Tossing different two coins:

$$\Omega = \{HH, HT, TH, TT\}$$

- Plausible that all 4 elements of are equally probable (if coin fair)
- Because $P(\Omega) = 1$: Probabilities must add up to one:

$$P(KK) + P(KZ) + P(ZK) + P(ZZ) = 1$$

- Because all elementary events equally probable:

$$P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$$

## Deriving Rules from Axioms

**Laws for Calculating Probabilities**

If $A$, $B$ and $A_1, \dots A_n$ events, then

$$\begin{aligned}
P(\overline{A}) &= 1 - P(A) && \text{for all } A \\
P(A \cup B) &= P(A) + P(B) - P(A \cap B) && \text{for all } A \text{ and } B \\
P(A_1 \cup \ldots \cup A_n) &\leq P(A_1) + \ldots + P(A_n) && \text{for all } A_1, \ldots, A_n \\
P(B) &\leq P(A) && \text{for all } A \text{ and } B \text{ with } B \subseteq A \\
P(A \backslash B) &= P(A) - P(B) && \text{for all } A \text{ and } B \text{ with } B \subseteq A
\end{aligned}$$

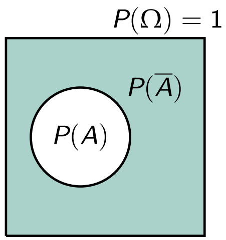


- Probabilities as areas in Venn diagrams
- Total area of $\Omega$ equal to 1 or $P(\Omega) = 1$
- Laws obvious



## 1st Rule

- Illustration:




- $P(A)$: Area of $A$
- $P(\overline{A})$: Remaining area in $\Omega$
- Obviously, following applies:

$$P(A) + P(\overline{A}) = P(\Omega) = 1$$

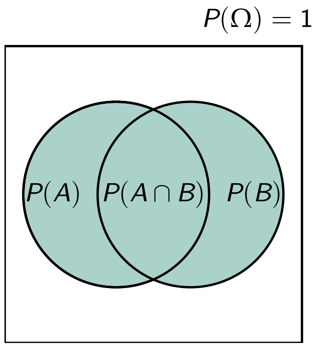
- And so:

$$P(\overline{A}) = 1 - P(A)$$



## 2nd Rule

- Illustration:

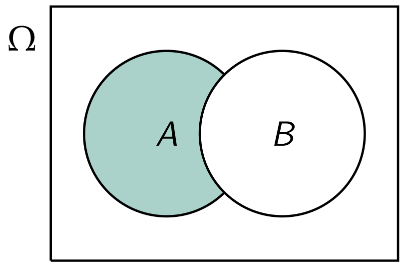



- $P(A \cap B)$ with $P(A) + P(B)$: Counted twice, subtract once:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

## Example

$P(A \setminus B) = \ldots$




"*A* without *B*"

1. $P(A) - P(B)$
2. $P(A) + P(B)$
3. $P(A) - P(A \cap B)$
4. $P(A) + P(B) - P(A \cap B)$



## Discrete Probability Models

- Now: *Discrete* probability models
- Means: Sample space finite or infinite and discrete
- Term "discrete": *finite* set, like:

$$\Omega = \{0, 1, \ldots, 10\}$$

- *Infinite*, but still *discrete* set, like

$$\Omega = \mathbb{N}_0 = \{0, 1, 2, \ldots\}$$

- Set $\Omega = \mathbb{R}$ (set of all decimal fractions): *Not* discrete
- Will play a very important role for measurement data later



## Probabilities for Discrete Models

- Calculation of probabilities for discrete models

> Probability of event
>
> $$A = \{\omega_1, \omega_2, \ldots, \omega_n\}$$
>
> is *determined* by sum of probabilities $P(\omega)$ of corresponding elementary events:
>
> $$P(A) = P(\omega_1) + P(\omega_2) + \ldots + P(\omega_n) = \sum_{\omega_i \in A} P(\omega_i)$$

- All probabilities of elementary events from event $A$ are added up
- Follows from axioms 1–3



## Example: Tossing Coin Twice

- Event $A$: "Tossing $H$ exactly once":

$$A = \{HT, TH\}$$

- Probability $P(A)$:

$$P(A) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

- Event $B$: "At least one head tossed"
- Probability of $B = \{HT, TH, HH\}$ occurring:

$$P(B) = P(HT) + P(TH) + P(HH) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

- Easier to calculate with so-called *complementary probability*
- The complement $\overline{B}$ of $B$ is

$$\overline{B} = \{TT\}$$

- From first calculation law (see above):

$$P(B) = 1 - P(\overline{B}) = 1 - \frac{1}{4} = \frac{3}{4}$$



## Example: Unfair (Biased) Die

- Probabilities to roll different numbers are not equal:

| $\omega$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $P(\omega)$ | $\frac{1}{3}$ | $\frac{1}{6}$ | $\frac{1}{12}$ | $\frac{1}{4}$ | $\frac{1}{12}$ | $\frac{1}{12}$ |

- From A1:

$$\begin{aligned} P(\Omega) &= P(1) + P(2) + P(3) + P(4) + P(5) + P(6) \\ &= \frac{1}{3} + \frac{1}{6} + \frac{1}{12} + \frac{1}{4} + \frac{1}{12} + \frac{1}{12} \\ &= 1 \end{aligned}$$

- Probability of $A = \{1, 2, 4\}$ occurring:

$$\begin{aligned} P(A) &= P(1) + P(2) + P(4) \\ &= \frac{1}{3} + \frac{1}{6} + \frac{1}{4} \\ &= \frac{3}{4} \end{aligned}$$



- Note: Result not the same if die would be fair: $\frac{1}{2}$
- Example: Calculate probability to roll a number smaller than 6
- Event $B$:

$$B = \{1, 2, 3, 4, 5\}$$

- Probability for $B$ occurring:

$$\begin{aligned} P(B) &= P(1) + P(2) + P(3) + P(4) + P(5) \\ &= \frac{1}{3} + \frac{1}{6} + \frac{1}{12} + \frac{1}{4} + \frac{1}{12} \\ &= \frac{11}{12} \end{aligned}$$



- Simpler with *complementary probability*: $P(\overline{B})$
- 1. calculation rule: Complement $\overline{B}$ from $B$:

$$\overline{B} = \{6\}$$

- Then it follows:

$$P(B) = 1 - P(\overline{B}) = 1 - P(6) = 1 - \frac{1}{12} = \frac{11}{12}$$

## Laplace Model

- Assume: Every elementary event has same probability
- Event $E = \{\omega_1, \omega_2, ..., \omega_f\}$
- Sample space $p$ Elements
- Probabilities of all elementary elements add up to 1:

$$P(\omega_k) = \frac{1}{|\Omega|} = \frac{1}{p}$$

> Event $E$: Laplace model:
>
> $$P(E) = \frac{f}{p} = \sum_{k:\omega_k \in E} P(\omega_k)$$

- Divides number of *"favorable"* elementary events by number of *"possible"* elementary events



## Example: Laplace Model

- Two different (blue and red) dice are rolled
- What is probability that eye sum is 7?
- Elementary event describes numbers on both dice
- Result in form 14
- Result 14 is *not* equal to 41
- Convention: First digit result of blue die, second digit red die
- All elementary events:

$$\Omega = \{11, 12, \ldots, 16, 21 \ldots, 65, 66\}$$



- Number of elementary events:

$$|\Omega| = 36$$

- Event $E$: Rolling eye sum 7
- There are 6 elementary events:

$$E = \{16, 25, 34, 43, 52, 61\}$$

- All elementary events have equal probability: Probability for event $E$:

$$P(E) = \frac{|E|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$$



## Stochastic Independence

- Already seen:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- Question: How to calculate $P(A \cap B)$?
- No general rule: If $P(A)$ and $P(B)$ are known, value $P(A \cap B)$ cannot be calculated generally from $P(A)$ and $P(B)$
- Important special case: Calculation of $P(A \cap B)$ from $P(A)$ and $P(B)$ with product formula:

> If events $A$ and $B$ are *stochastically independent*, then
>
> $$P(A \cap B) = P(A) \cdot P(B)$$

- But what does „stochastically independent" means?
- Outcome of event $A$ has no influence on outcome of event $B$ and vice versa



## Example

- Event $A$: Roll 1 or 2 with a fair die: $P(A) = \frac{1}{3}$
- Event $B$: Head when tossing a fair coin: $P(B) = \frac{1}{2}$
- Tossing a coin has no influence on outcome of rolling a die
- Use formula above:

$$P(A \cap B) = P(A) \cdot P(B) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$



## Example

- Event $E$: Tokyo is shaken by an earthquake on certain day
- Event $F$: On this day a typhoon sweeps over the city
- Unlikely that earthquake would have any influence on occurrence of typhoon
- Hence both events are stochastically independent



## Example

- Tossing a coin twice
- Outcome of first toss has no influence on result of second toss
- However: Only correct if coin is ideal
- Real coin: Minimal changes due to impact
- These have influence on probability of tossing head (or tail) for next toss
- But changes so small that they are negligible

## Example

- 20 lottery tickets with 5 winning tickets
- Draw ticket twice *without replacing*
- Event $A$: Win in first draw
- Event $B$: Win in second draw
- These two events are *not* stochastically independent
- Draw a winning ticket in first draw: Probability that $A$ occurs:

$$P(A) = \frac{5}{20}$$



- If win in 1st draw: Probability to win in 2nd draw:

$$P(B) = \frac{4}{19}$$

- Drawing first a blank: Probability to win in 2nd draw:

$$P(B) = \frac{5}{19}$$

- Depending on whether event $A$ occurs or not, probability of $B$ occurring is different
- Events are not stochastically independent



## Example

- Event $A$: Tomorrow is fine weather
- Event $B$: Person is in a good mood tomorrow
- Most people more cheerful in good weather than in bad weather
- Occurrence of $A$ has influence on occurrence of $B$
- Events are not stochastically independent



## Caution

- Formula

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

  applies *only* if events $A$ and $B$ are stochastically independent
- If events are *not* stochastically independent, there is no general formula to calculate $P(A \cap B)$