Intelligent Data Analysis C. Braune, P. Held, A. Dockhorn

Exercise Sheet 1

Exercise 1 Probabilities: Rolling Dice

Samuel Pepys once wrote to Sir Isaac Newton to ask which of the following events is the most likely: that a person gets

- a) at least one six when six dice are rolled,
- b) at least two sixes when twelve dice are rolled,
- c) at least three sixes when eighteen dice are rolled?

Exercise 2 Geometric Probabilities

- a) A pole with a length of one meter is cut at two randomly chosen points. What is the probability that one can form a triangle out of the three parts? (Hint: View the lengths of the parts as coordinates of points in a three-dimensional space and exploit the triangle inequality in all three possible forms.)
- b) A secant (i.e., a straight line between two points on the perimeter) is drawn randomly into a circle. What is the probability that this secant is longer than a side of the largest equilateral triangle that can be inscribed into the circle?

Exercise 3 Conditional Probabilities

- a) Two fair dice are rolled. What is the probability that the sum of the points is ten if it is known that the sum is at least eight?
- b) Let it be known that a certain family has two children. What is the probability that both children are girls if it is known that at least one child is a girl?
- c) What is the probability in part b) if it is known that the younger child is a girl?

Exercise 4 Data Analysis and Statistics in Everyday Life

In newspapers and magazines one often reads statements that are derived from statistical data. However, these statements are not always meaningful or relevant in the context in which they appear. Here are some examples:

- In the fifties the *Time Magazine* remarked in a comment on an article of the New York *Sun*: "The average Yaleman, Class of '24, makes \$25,111 a year." (Yaleman: graduate of the American elite university Yale).
- The German car magazine *ADAC-Motorwelt* once printed the following headline: "Death rides with you! Four out of ten car drivers that died in accidents did not buckle their seat belts!"
- A snippet from the German magazine *Stern*: "Soccer players are [...] pure crash pilots. They cause more than half of the roughly one million sport accidents that happen each year."
- Under the headline "In old age you will get happier" a German newspaper argued: the fraction of suicides among all deaths is highest among teenagers, namely about 25% among people below 20 years of age, compared to 10% among those between 30 and 40, and less than 2% among those beyond 70. Therefore: "The tendency to commit suicide goes down the older you get."

Consider which statement is intended, whether the provided evidence is convincing, how the data may look like from which the statements where derived, how this data may have been collected etc. Based on these considerations, how would you assess these statements? Is it necessary to take other data or other statistical statements into account in order to arrive at meanigful, relevant, and correct statements?