Practical statistics for data science - 2022

List 1 for the 2nd lab.

Marek Klonowski

I. Simple Probability

1. Simulate 20 rollings of a biased dice (i.e., the face with 6 dots is twice more probable than other outcomes).

2. Choose 10 numbers from the set $\{5, 6, \ldots, 100\}$ without replacement.

3. Generate 2000 Bernoulli trails with success probability equal $p = 0.5$ (fair coin tosses). Compare the number of successes with the expected value.

4. Generate 10^i samples from a normal distribution with parameters $\mu = 0$ and $\sigma = 1$. Draw histogram for each $i \in \{0, \ldots, 8\}$.

5. Generate 2000 Bernoulli trails with success probability equal $p = 0.15$.

   - Find the number of successes.
   - Find the maximal number of consecutive successes (i.e., sequences "1"in a row)
   - Find the probability of getting less than 250 or more than 350 successes.
   - Approximate the minimal p such that we will get at least 30 consecutive with high probability.

6. We toss a coin 1000 times. Let X be the number of heads. Find the probability that X is a number divisible by 11.

7. Assume that the height of a male is normally distributed with a mean equal to 177 cm and a standard deviation equal to 8 cm.

   - Find the probability that a randomly chosen man is higher than 200 cm.
   - Find the probability that a randomly chosen male is between 160 and 180 cm.
   - Approximate the probability that in a random group of 20 males, no one is higher than 165.

8. Generate a V_1: vector with 2000 random natural numbers from the set $\{50, \ldots, 100\}$. Generate V_2 - a vector such that for every entry in V_1, we add an independent random number from a normal distribution with mean 0 and variance 20. V_3 - a vector such that for every entry in V_1, we add an independent random number from a normal distribution with mean -1 and variance 100. Find the correlation between

   - V_1 and V_1,
   - V_1 and V_2,
   - V_1 and V_3.