Project content:

1) computational aspects,

2) analytical derivation

3) data analysis including graphical representation.

Projects may be completed **individually** or in **groups of two students**.
Projects must be submitted in **written form** and afterwards presented **orally**.

Evaluation criteria:

1. Accuracy of the provided solutions (maximum 20 points)
2. Structure of the report (maximum 5 points)
3. Accuracy and structure of the attached programming codes (maximum 5 points)
4. Originality and creativity of the approaches (maximum 5 points)
5. Presentation (maximum 5 points)

Specific details about each component of the project:

1) **Computational component + analytical (?):**

1.1) Computer projects provided as exercises at the end of Chapter 5 of the textbook, Baron M. (2019), 3rd edition

1.2) Programming Exercises provided in the book of Forsyth D. (2018). *Probability and statistics for computer science*

1.3) Come up with your own topic

1.4) Suggested topics:

1.4.1) Select random variables from a specific distribution and derive, theoretically, how they relate to each other or to another distribution. Generate simulation code to confirm your theoretical results.

Examples:

a.1) If X_1 and X_2 are **Poisson** random variables with means μ_1 and μ_2 respectively, then $X_1 + X_2$ is a **Poisson** random variable with mean $\mu_1 + \mu_2$. Show analytically and computationally

a.2) The sum of n **geometric** random variables with probability of success p is a **negative binomial** random variable with parameters n and p

a.3) Approximate the binomial distribution by a Poisson distribution. Show this analytically and computationally (reference: Walpole (2012) Probability and Statistics for Engineers and Scientists)

1.4.2) Derive analytically the maximum likelihood estimator (MLE) for selected distribution. Confirm your result computationally. *Bonus question*: Is the derived estimator biased or unbiased?

2) Data analysis. Obtain dataset yourself or use the datasets from the following literature:

- Baron M. (2019), Chapter 9
- Rosner B. (2010). Fundamentals of Biostatistics, Chapters 6-8
- Forsyth D. (2018). *Probability and statistics for computer science,* Chapters 6-8
- Other sources

Topics:

- Testing means (one sample inference, two sample inference)
- Testing proportions (one sample, two sample)
- Testing variances
- Identify type of distribution in real data.

Alternative option:
Select a topic and study about that topic from theoretical/analytical, computational and data analysis point of view.
Topic 1: measures of dispersion (standard deviation and others), show their properties mathematically and graphically, show their use in practice
Topic 2: measures of location (median, mode, mean and others), show their properties mathematically and graphically, show their use in practice
Topic 3: compare simulation methods using 1) R built-in functions, 2) inverse transformation method for generating continuous or discrete probability distribution from the uniform distribution, 3) uniform random variable or algorithm for generating arbitrary discrete distributions
Topic 4: When we use the Inverse Transformation Method, the cdf F(x) should have reasonably simple form so that it allows direct computation of $X = F^{-1}(U)$. When F(x) doesn't have a simple form but the pdf f(x) is available, random variables with density f(x) can be generated by the rejection method**.** Compare results from rejection method with the one from inverse transform method.
Topic 5: consequence of the central limit theorem (Cohen Y& Cohen J.Y. (2018), Chapter 7
Topic 6: topics from Cohen Y& Cohen J.Y. (2018) *Statistics and Data with R: An Applied Approach Through Examples.* (Wiley & Sons), 1st Edition