Instructions. You should hand in your homeworks within the due date and time. Late deliveries will be penalized, please check homework delivery policies in the exam info.

Handing in. You should submit your work as *one zip file* containing i) *one folder* for Assignment 1, containing your code, clearly organized (see Assignment 1 for details), ii) the *pdf printout* of your results on HackerRank (see Assignment 1 for details) and iii) *one pdf* containing your answers and solutions to the theoretical exercises (i.e., Assignments 2 - 4). Answers to theoretical exercises should be either typed up or hand written **very clearly** and scanned. The first option is preferred.

Typesetting in Latex. Latex is very handy for typesetting (especially math), but you need to install it. If you do not want to install Latex, you can go for Overleaf, providing an integrated, Web interface, accessible for free in its basic version (which is enough for your needs). It allows you to both type in Latex using a Web interface, and compiling your code to produce a pdf document. Overleaf's documentation also contains a tutorial on Latex essentials.

Important note. Grading of your answers to the theoretical assignments will depend on i) correctness and solidity of the arguments, ii) clarity of the presentation and iii) mathematical rigour. E.g., ill-defined quantities, missing assumptions, undefined symbols etc. are going to penalize you. Rather than writing a lot, try to write what is needed to answer and write it well.

Assignment 1.

The goal of the first assignment focuses on checking/building up your skills in Python. The assignment is done electronically using the on line HackerRank.

You must create an account on HackerRank and complete as many challenges as you can from the list included below. When you are finished and want to submit your assignment, go go to your profile on HackerRank's Web site, tap on the `Submissions` link and produce a PDF printout of all your submissions. This is also directly available at the following url: `https://www.hackerrank.com/submissions/all` Your solution must include as attachments:

- The aforementioned pdf;
- A folder containing all your code, clearly organized.

The above two items will be included in the zip file containining your solutions for this homework (see general instructions above).

Important note. Please check the collaboration policy at the course web page. HackerRank provides solutions to the proposed challenges. You should not look at the solutions provided by HackerRank; if you do, please *explicitly mention* this in the corresponding response, possibly as a comment in the corresponding code. Below the assignment list:

(a) Data types (all – total: 6 - max points: 60) `https://www.hackerrank.com/domains/python/py-basic-data-types`

(b) Strings (all – total: 14 - max points: 220) `https://www.hackerrank.com/domains/python/py-strings`

(c) Sets (all – total: 13 - max points: 170) `https://www.hackerrank.com/domains/python/py-sets`

(d) Collections (all – total: 8 - max points: 220) `https://www.hackerrank.com/domains/python/py-collections`

(e) Built-ins (only 3 - max points: 80)

- `https://www.hackerrank.com/challenges/zipped`
- `https://www.hackerrank.com/challenges/python-sort-sort`
- `https://www.hackerrank.com/challenges/ginorts`

(f) Numpy (all – total: 15 - max points: 300) `https://www.hackerrank.com/domains/python/numpy`

Assignment 2.

Assume we collect n independent samples $X_1, \ldots, X_n$ from a distribution with expectation μ and variance σ^2.

(a) Consider $S = \frac{1}{n}\sum_{i=1}^{n} X_i$ (i.e., the sample mean). Compute $\mathbf{Var}(S)$.

(b) Assume we take the *sample variance*, defined as $Z = \frac{1}{n}\sum_{i=1}^{n}(X_i - S)^2$ and normally used to estimate the variance of a distribution from a sample. Compute $\mathbb{E}[Z]$ as a function of n and σ^2.

Note. These are all well-known facts in statistics and the goal here is that you know and understand them. If you help yourself with external references, please i) cite them and ii) be sure you understand them and are able to present the approach to me during the homework discussion.

Assignment 3.

Assume we want to estimate the fraction p of people that have been infected by a virus in a given population. To this purpose, we perform blood tests on n subjects, sampled *uniformly and independently at random* from the population.[1] Each test will return the correct answer (positive or negative to the virus) with 100% accuracy. Our estimator of p is the fraction $\hat{p}$ of subjects that tested positive. Given $0 < \varepsilon < 1$, $0 < \delta < 1$

[1] We are implicitly assuming that i) we are able to sample uniformly from the population (which is not obvious) and ii) that the sample collection process is fast enough that p does not change meanwhile.

and $0 < \theta < 1$, we want to compute the minimum value of n such that, if we collect n samples:

$$\mathbb{P}\left(|\hat{p} - p| > \varepsilon p\right) < \delta, \text{ whenever } p > \theta,$$

while we do not require any accuracy guarantees if p is sufficiently small, i.e., $p \leq \theta$.

Hints. i) Convince yourself that $\hat{p}$ is a random variable and compute its expectation; ii) Note that the Chernoff bounds we saw in class are more general and apply unchanged to the sum of independent random variables with domain $[0, 1]$. See for example Theorem 1.1 in Dubhashi and Panconesi's book on concentration of measure (here is a draft).

Assignment 4.

Professor Cooper and Professor Wolowitz collected the heights of 200 subjects. They found that i) all measured heights vary in the interval $[160, 190]$ (cm.) and ii) the average height in the group is 180 cm. Professor Cooper argues that these results are compatible with the hypothesis that the heights in the population are distributed *uniformly* in the interval $[160, 190]$, while Professor Wolowoitz disagrees firmly, arguing that these results are inconsistent with the "uniform" hypothesis in a statistically significant way.

(a) Formulate the problem as one of hypothesis testing, clearly defining the null hypothesis H_0;

(b) Provide quantitative and mathematically rigorous arguments in favour of Professor Cooper's or Professor Wolowitz's thesis.

Hints. There are different ways to solve this exercise. If at any points you thought you need to apply something like a Chernoff bound, note that if a variable ranges in $[a, b]$, you can rescale it to range in $[0, 1]$.