

Assignment 1: data collection & description, probability distributions

Statistics 2023 – 2024

Deadline: Friday 27 October, 23:59h

- Make sure your solutions are your own. Collaborating on the assignments (including the use of generative tools such as ChatGPT) is not allowed.
- For exercises containing a calculation, giving only the solution is not sufficient: show how you obtained the solution.
- The maximum number of points is indicated for each exercise. You can lose points for writing irrelevant or incorrect information. Your grade will be computed as $points/16 \cdot 10$.
- Hand in a PDF version of your solutions via Brightspace, created using your favorite text editor. Scans of hand-written solutions are not allowed and will not be graded.
- Pay attention: if you don't get a digital submission receipt (on your screen after submission and via e-mail), your solutions were not submitted correctly.

Good luck!

- (5) 1. The Law School Admission Test (LSAT) is a standardized test for prospective law school candidates in the US. It consists of 101 questions and has a score range of 120 to 180 (integers only). LSAT scores are approximately normally distributed with $\mu = 150$ and $\sigma = 10$.
- (2) (a) Some universities, such as Harvard, only admit students with LSAT scores in at least the top 4%. What is the minimum LSAT score required to be admitted to Harvard?
- (2) (b) Given that 70000 students have taken the LSAT this year, how many do you expect to have scored between 145 and 160 (inclusive)?
- (1) (c) A group of second-year law school students is given the LSAT. What are your expectations on the distribution, mean and median of their scores?

- (5) 2. Consider a random variable X that represents a fair die throw (with a regular die with six sides). We take a sample of $n = 2$ dice throws and obtain the sample $\{3, 5\}$. The parameter of interest is the mean of the dice throws.
- (2) (a) Sketch the population distribution, sample distribution, and sampling distribution. Make sure to label your axes! (For this question, including a scan or clear photo of a hand-drawn sketch is accepted.) Explain what the distributions describe.
- (2) (b) Instead of taking a sample of two dice throws, we now take a sample of $n = 10$ and are interested in the mean of the ten throws. Which of the distributions now change, and how?
- (1) (c) What happens to the sampling distribution if we increase the sample size further? Explain why this happens.
- (2) 3. The following table lists the top 10 chess players in September 2023 based on Elo rating:

Name	Elo rating
Magnus Carlsen	2839
Fabiano Caruana	2786
Hikaru Nakamura	2780
Ding Liren	2780
Alireza Firouzja	2777
Ian Nepomniachtchi	2771
Anish Giri	2760
Gukesh D	2758
Viswanathan Anand	2754
Wesley So	2753

Calculate the following: mean, median, mode, standard deviation, range, lower and upper quartile. You may round the mean Elo to the nearest integer to calculate the standard deviation.

- (4) 4. The municipality of Leiden is interested in the coffee consumption per household. The municipality is divided into 200 neighborhoods, each containing approximately 500 households, and no big differences between neighborhoods are expected. The population is defined as all households in the municipality.
- (2) (a) For each of the following methods, describe how you would obtain a good sample of $n = 1000$: simple random sampling, cluster sampling, and stratified sampling.
- (1) (b) Unfortunately, there is no budget to send the surveys by mail, so you will have to hand-deliver the surveys to the households in your sample. Which type of sampling would make more sense in this situation: cluster sampling or stratified sampling? Explain your answer.
- (1) (c) Someone suggests that instead of delivering the surveys on paper, you could send the survey by e-mail. Which type(s) of bias should you look out for if you send the survey by e-mail, that is not a risk if you deliver the surveys on paper? Explain your answer.