

Problem Set 1

TAs: Antto Jokelainen and Ellen Sahlström

Deadline - 19.1.2023 18:00

Submit your answers in **pdf** or **html** format with all the relevant code included in the file. In addition, return your **do-file** for replication purposes. Do not submit more than two files. Remember to include your student number! Note that we take honest effort into account, so at least try to answer all of the questions.

1. (18p) Compute the following probabilities:
 - (a) $P(7 \leq Y \leq 12)$ when $Y \sim N(9, 78)$. Use a standard normal distribution. Demonstrate how you do this.
 - (b) $P(Y \leq 14)$ when $Y \sim N(17, 35)$. Use a computer for the calculations and report how you do this.
 - (c) $P(Y > 1.75)$ when Y follows Student's t -distribution with 15 degrees of freedom. What kind of a variable has this distribution? What happens when the parameter of the Student's t -distribution gets large?
2. (18p) In this study we analyze the results of a survey by The Finnish Business School Graduates (FBSG), a central organization for graduates and students in economics and and business administration, on summer job pay. The survey targeted student members of the FBSG who were at least on their second year of studies and had begun their studies after 2009. The survey was conducted as an electronic survey that was e-mailed to the eligible members of the FBSG in the autumn of 2021. The above table is a translation of the Table 3 in the original report of the survey results by Venäläinen (2021). The table presents mean gross monthly salary from a summer job by the number of credits. You can assume that the observations are iid.

Credits		under 120	120-180	181-220	over 220
Observations		(N: 282)	(N: 251)	(N: 180)	(N: 179)
Gross monthly salary	Mean	2 384	2 436	2 514	2 419
	(Sd)	(1 151)	(1 289)	(1 100)	(804)
	Median	2 100	2 100	2 300	2 300

- (a) Is there evidence that students with over 220 credits earn less on average than those who have 181-220 credits? Use a t-test and report the test statistic and the p-value.
- (b) Calculate a 95% confidence interval for the mean gross monthly salary for the students with 120-180 credits. What does the confidence interval imply?
- (c) Compare means and medians in the table. What does this comparison tell you about the data?
3. (5p) What do you think about the following statement found when one asks Google "What confidence interval means?"

A confidence interval displays the probability that a parameter will fall between a pair of values around the mean.

4. (24p) The data on the body heights of ECON-C4100 participants yields the mean of $\bar{Y} = 176.49$ and standard deviation of $s_Y = 9.35$ with $n = 72$.
- (a) Construct a 95% confidence interval for the mean body height in the population.
- (b) What is the relationship between the 95% confidence level and a 5% significance level?

The sample average and standard deviation for the men are $\bar{X} = 181.04$ and $s_X = 6.60$. The sample average and standard deviation for the women are $\bar{Y} = 167.38$ and $s_Y = 7.13$. The number of men participants in the survey is 48 and the number of women participants is 24.

- (c) Is there statistically significant evidence that the average body height of men is higher than the average body height of women?

A course TA calculated the sample averages and standard deviations for the first half ($N = 36$) and for the second half ($N = 36$) of the survey participants separately. The results yield $\bar{X} = 177.10$, $s_X = 9.14$ and $\bar{Y} = 175.88$ and $s_Y = 9.64$.

- (d) Is there statistically significant evidence that the average body height of the first subgroup is higher than the average body height of the second group? Do you find your result plausible and why?
5. (20p) Suppose that X_i is an independent and identically distributed Bernoulli random variable with $p = 0.25$.
- (a) Use the Central limit theorem (CLT) to calculate approximations of
- (i) $P(\bar{X} > 0.30)$ when $n = 300$.
 - (ii) $P(\bar{X} \leq 0.15)$ when $n = 100$.
- (b) How large would n have to be to ensure that $P(0.20 \leq \bar{X} \leq 0.30) = 0.95$? Use the CLT. (Hint: What is the value of the t-statistic that corresponds to this significance level?)
6. (15p) Correlation and covariance
- (a) What is correlation? Provide a short definition.
 - (b) What is covariance? Provide a short definition.
 - (c) Give an example of two positively correlated phenomena, motivate why this is the case.
 - (d) Give an example of two negatively correlated phenomena, motivate why this is the case.
 - (e) Give an example of a case where the covariance is zero. Explain why.

References

Venäläinen, A. (2021). Kylterien kesätyökysely - tuloksia 2021. Suomen Ekonomit. Available at: https://www.ekonomit.fi/wp-content/uploads/2021/11/Kylterien_kesatyokysely_2021.pdf