

Statistical Literacy

Problem Set 9

Rémi Viné

Due November 27th, 2023

1. State the hypotheses.

- (a) A local Non-Governmental Organization (NGO) states that a refugee camp welcomes 500 people every day. The staff counted the number of people entering the camp every day for 81 days and observed that there were 558 people entering per day on average.
- In the hypothesis testing procedure, what would be the μ_0 , *i.e.*, the value to confront?
 - In the hypothesis testing procedure, what would be the point estimate?
 - What would be the hypotheses to state if you want to test whether the NGO *understated* the number of entrances? What type of test would it be?
 - What would be the hypotheses to state if you want to test whether the NGO *mis-stated* the number of entrances? What type of test would it be?
 - In the context of the data, would it make sense to state the following alternative hypothesis $H_a : \mu < 500$?
- (b) You are a HR manager and want to verify that employees are working 8 hours per day. You do not want people to overwork because this is associated with more health issues and it lowers the productivity of workers over time. On the other hand, you do not want the employees to under-work because their contracts state 8 hours of daily work. An analyst in the company noted that in building A, randomly surveying 400 employees, people work on average 7.91 hours (.91 is equivalent to $0.91 \times 60 = 54.6$ min). In building B, randomly surveying 41 employees, people work on average 8.23 hours. Last, in building C, randomly surveying 16 employees, people work on average 5.83 hours.

Based on these results, you want to make inferences *per building*. The HR thinks that employees in Building A underwork (as $\bar{x} < 8$), that employees in building B overwork (as $\bar{x} > 8$), and that employees in building C underwork (as $\bar{x} < 8$), what would be, building by building, the correct one-tail tests (if any)?

2. Hypotheses testing. Conduct the appropriate hypothesis testing whenever possible.

- (a) There is a debate on retirement age. Some are in favor of an identical retirement age, some are in favor of different retirement ages based on life expectancy at retirement. The government in favor of the law argues that the gap in years of life expectancy at retirement age between white collars and blue collars is negligible, justifying a reform toward an identical retirement age. On the contrary, a labor union argues that the gap is different: blue collar workers have a lower life expectancy at retirement age than white collar workers.

Working for a poll agency, you conduct an analysis using a representative sample of 144 blue collar-white collar dyads (so one observation in your data is one dyad, the gap of life expectancy between a blue collar and a white collar). You observe a variance of 2.96 years and an average of 0.4 years. Can you confirm the labor union statement at $\alpha = 0.01$?

- (b) The current Head of State in a given country tries to defend his or her past mandate. The Head of State declares: “When I took the office, households had on average 8h of electricity per day, now households have on average electricity 18h per day”.

You are suspicious toward this statement. You believe that the Head of State is overstating access to electricity and you investigate using a representative sample of 41 students. You find that the average access to electricity in the households is 17h per day with a sample variance of 10 hours. Conduct a hypothesis testing related to the Head of State’s statement at $\alpha = 0.05$.

- (c) Using the same case of electricity access, what would be your conclusion in the case of a level of significance of 1%?

- (d) The organization in charge of the highways in Switzerland claims that the commuting time through the highway is 40min. It is assumed that the commuting time of the population is Normally distributed and that the population variance is 25min.

You do not know whether the organization is correct and you want to check for a wrong assessment of the commuting time. Taking a representative sample of 12 drivers, you find a commuting time of 39min. Would you conclude on a wrongly estimated commuting time at $\alpha = 0.03$?

- (e) The IHEID team working on the Applied Research Projects asks for a word limit, say 10000 words. You look at 18 representative projects and find an average of 11245 words with a sample variance of 900 words. At a level of significance of 1%, would you conclude that the students exceed the word limit?

3. Go beyond: types of errors and other issues.

- (a) Conducting a hypothesis testing, if you take a very small α , how likely are you to reject H_0 ?
- (b) Conducting a hypothesis testing, if you take a very small α , what type of error are you likely to commit?
- (c) Assume you want to test whether there is an income gap between men and women. You test the wage difference between pairs of workers: ($wage_{men} - wage_{women}$). The 100 men are taken from the blue collar sectors and the 100 women are taken from the white collar sectors. If the results obtained lead to conclude that men do not earn more than women, how relevant are such results?