Homework 2

Masoud Akbarzadeh

March 1, 2024

Estimated time to Completion: 10 hours
Maximum Allocated Time: 15 hours
Actual Time to Completion: 11 hours
Collaborators: Nari, Juan, Amel, Brandon, Jesse, Katurah

Problem 1

1.a

The average pressure: $846.33hPa$, with the standard deviation of: $5.62hPa$
The average pressure when it rains: $847.03hPa$

1.b

Since the number of data points is large enough, a z-test was used. Also, since the pressure can be anomalously higher or lower, a two-sided z-test was used.
Assumptions:

- Confidence level: 95% (two-tailed)
- $H_0 : \mu = 846.33$
- $H_1 : \mu \neq 846.33$
- Z-critical value: 1.96

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{N}} = \frac{847.03 - 846.33}{5.62/\sqrt{384}} = 2.44 > 1.96$$

Since the z-score is greater than the critical value, the null hypothesis is rejected. So the average pressure when it rains is significantly different from the average pressure.

1.c

Using bootstrapping, 100000 samples were generated. By calculating the fraction of the time the average pressure when it rains is greater than the average pressure, the p-value was calculated to be 0.007. So the likelihood of the average pressure being greater when it rains is about 0.7%. So the results is in agreement with the z-test.

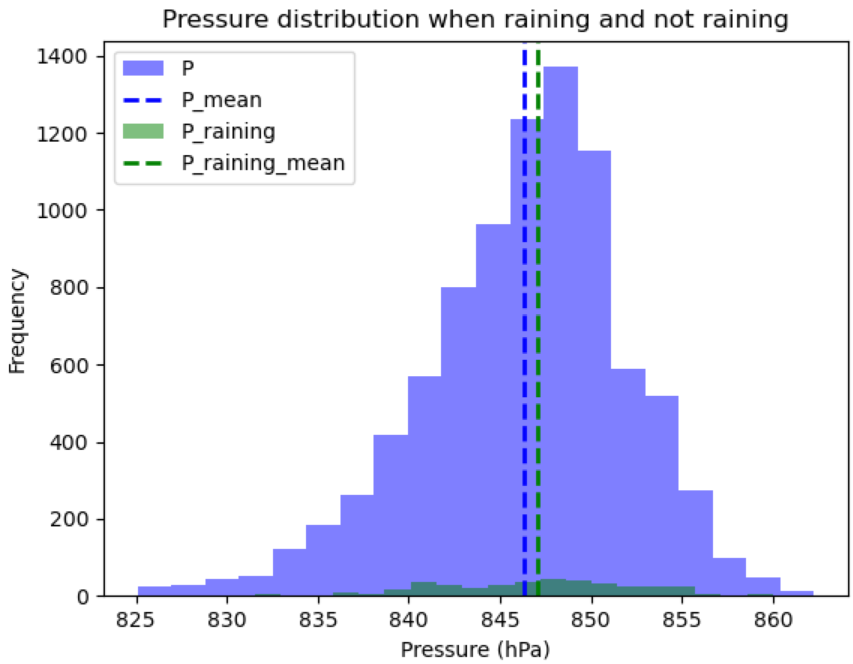

Figure 1: Problem 1-A

Since the number of data points is large enough, the central limit theorem applies, and the distribution of the sample means is approximately normal. So using the z-test is valid.

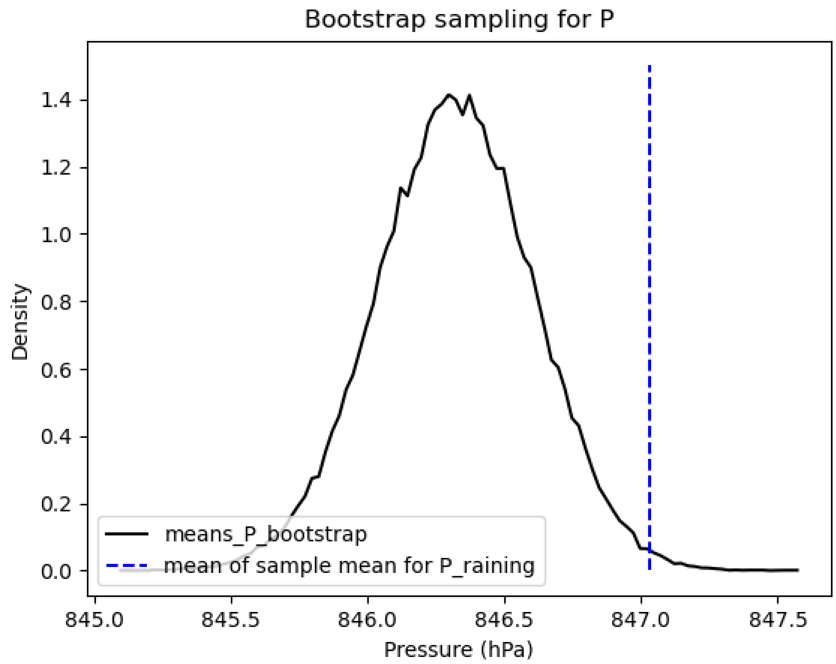

Figure 2: Problem 1-C

Problem 2

2.a Hypothesis testing

Hypothesis testing with these assumptions:

- $H_0 : \mu = 0$
- Two-tailed confidence interval of 95%
- Using z-score since the sample size is larger than 30
- A two-tailed test is used
- Z-critical value is 1.96

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{N}} = \frac{0.13 - 0}{1/\sqrt{100}} = 1.3 < 1.96$$

So we cannot reject the null hypothesis that the air is from the west.

2.b Bayesian Approach

Assuming:

- $\gamma = 0.4$ is the fraction of time the wind is from the east, and $\gamma = 0.6$ is the fraction of time the wind is from the west

- Calculating the probability of the concentration of a pollutant being between 0.12 and 0.14 instead of exactly 0.13

Using Bayes' theorem:

$$Pr(W) = 0.4$$
$$Pr(\bar{W}) = 0.6$$
$$Pr(0.12 < \bar{C}_{100} < 0.14|W) = 0.034$$
$$Pr(0.12 < \bar{C}_{100} < 0.14|\bar{W}) = 0.062$$
$$Pr(W|0.12 < \bar{C}_{100} < 0.14) = \frac{Pr(0.12 < \bar{C}_{100} < 0.14)Pr(W)}{Pr(0.12 < \bar{C}_{100} < 0.14)Pr(W) + Pr(0.12 < \bar{C}_{100} < 0.14|\bar{W})Pr(\bar{W})}$$
$$Pr(W|0.12 < \bar{C}_{100} < 0.14) = 0.268 < 0.5$$

Because $Pr(W|0.12 < \bar{C}_{100} < 0.14) < 0.5$ this approach says the wind is from the east.

2.c Sensitivity Analysis

Assuming γ values are from 0 to 1. For each of gamma, 10000 monte carlo simulations were run. Then the fraction of time each approach gives the correct answer was calculated. The results are shown in the figure **??**.
The results show that for all values of γ, the Bayesian approach is better than the hypothesis testing approach for this problem. This is because the Bayesian approach is using the prior information about the wind direction. Also, the Bayesian approach is least effective when $\gamma = 0.5$ while the frequentist approach is least effective when $\gamma = 0$. The model in the frequentist approach does not have any knowledge about the wind that is coming from the east, and its probability of happening so the model is only effective in γ values close to 1 where the most probable wind is from the west.

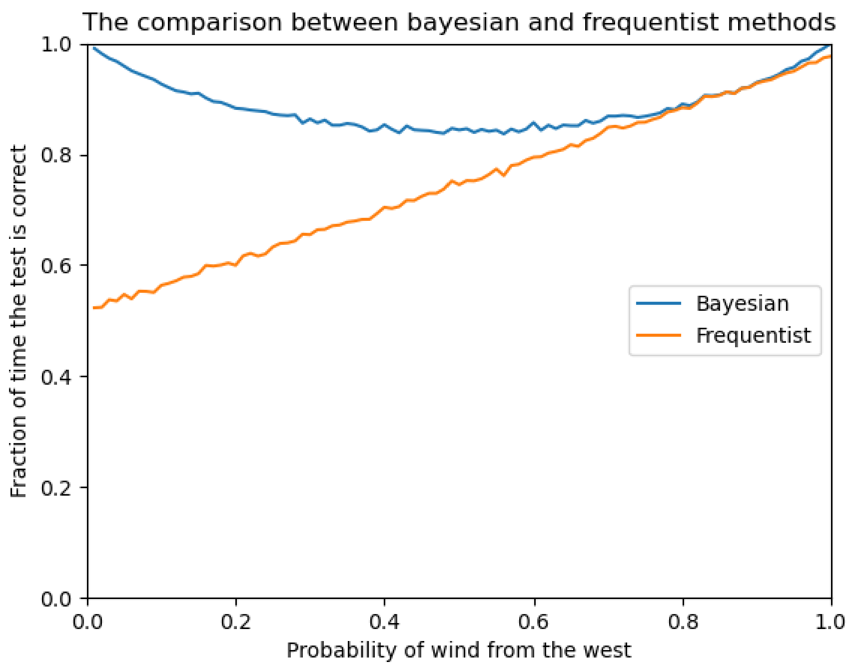

Figure 3: Problem 2-C