Understanding Statistics and Experimental Design Exercises: Bayesian Statistics

2023

1 Interpretation of Frequentist Statistics

A researcher is investigating whether the average Swiss woman is taller than the average French woman. Here are their hypotheses:

- H_0: Swiss and French women have the same height
- H_1: Swiss women are taller than French women

To investigate this, the researcher measures the height of 10 French and 10 Swiss women selected at random. After checking that all assumptions are valid, the researcher decides to perform a one-tailed independent sample t-test between the height of Swiss and French women. No significant difference is found ($p = 0.20$). Therefore, the researcher concludes that the null hypothesis is true.

Question 1: Why is this conclusion not valid?

Question 2: Which other statistical approach should the researcher have taken to investigate whether the null is true? Why?

2 Bayesian Statistics

The following exercises are intended to dispel misunderstandings about distribution-related functions and probability.

Question 3: What is a probability density function?

Question 4: The probability density function of the standard normal distribution $\mathcal{N}(0, 1)$ evaluated at 2 is 0.05399. Is this the probability that when sampling from the standard normal distribution, we observe the sample 2? Why, or why not?

Question 5: What is a cumulative distribution function? Give an example of the cumulative distribution of a distribution of your choice evaluated at some value, and interpret what it means. Feel free to use an online cumulative distribution function calculator, such as this cdf calculator.

Bonus exercise (difficult). Use of programming software such as `python` or `R` is highly recommended. In class we saw how to evaluate the

Bayes factor when we had two specific hypotheses about distributions in mind. Now, however, the researcher wants to understand the relative evidence for the hypotheses H_0 and H_1 given the following samples of data:

- Swiss women heights (cm): $165, 167, 164, 166, 168, 165, 167, 166, 168, 164$
- French women heights (cm): $164, 166, 165, 167, 165, 166, 164, 165, 167, 166$

Bonus Question 1: How can we express H_0 and H_1 in a way that allows us to compute their relative evidence (Bayes Factor)?

Bonus Question 2: We will now compute the Bayes factor step by step. Hint: if using `python`, use the packages `numpy` and `scipy.stats` to help. If using `R`, the native installation should suffice.

- Step 1: Define the parameters (mean, std) of the H0 distribution and the H1 distributions
- Step 2: Compute the likelihoods of the H0 and the H1 hypotheses separately
- Step 3: Compute the Bayes Factor from these likelihood

Bonus Question 3: Interpret the Bayes Factor.