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 1 solution: The p-value of the frequentist approach reflects the probability that the observed sample came from the null distribution, given that H_0 is true. Therefore, no claim about the truthfulness of H_0 can be made – only about the probability that the data occurred if H_0 was true. The absence of proof is not the proof of absence. Furthermore, the p-value depends on the sample size. Therefore, if the researcher would have selected 20 000 Swiss and 20 000 French women, he/she would probably observe a significant p-value ($p < 0.05$). Based on the p-value, the researcher would have concluded that Swiss women are taller than French women. Interpreting the p-value alone is missing information: we should look at the effect size too.

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

Question 2 solution: To claim that the H_0 is true, the researcher should have adopted a Bayesian approach and calculated the Bayes Factor. The Bayes factor provides the ratio of evidence of the data under the null hypothesis (H_0) over the evidence of the data under the alternative hypothesis (H_1). Out of the Bayes Factor, we can therefore conclude whether the H_0 or H_1 have more support given the data.

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 3 solution: The probability density function gives a relative likelihood of observing values around points.

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 4 solution: The probability density function does not directly provide a probability, but instead a relative likelihood of observing values around points, and as such the absolute value of the function evaluated at any point is meaningless in the absence of context. To obtain the probability of observing some values in some range (e.g., between 1.9 and 2.1), we must consider the area under the probability density function in that range.

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.

Question 5 solution: The cumulative distribution function is the integral of the probability density function (i.e., the area under its curve). It tells us about the probability of observing any value that is at least as small as the input value. For example, the cumulative distribution function of the standard normal distribution $\mathcal{N}(0,1)$ evaluated at 2 is 0.97724, meaning that there is a 97.724% probability of observing a value of 2 or smaller when drawing from the standard normal distribution. Similarly, to obtain the inverse (i.e., the probability of drawing a value that is at least 2), we can simply subtract the cdf from 1:

$$P(x \geq 2) = 1 - \text{cdf}_{\mathcal{N}(0,1)}(2) = 1 - 0.97724 = 0.02276 \tag{1}$$

Thus, the probability of observing a value that is at least 2 is approximately 2.276%.

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 1 solution: Since our null hypothesis is that Swiss and French women have the same height, we should compute the likelihood of the null as both of our samples (Swiss and French women's heights) coming from the same grand mean. Similarly, our alternative hypothesis should be that each sample has likelihoods that come from the distributions described by their within-sample means (Swiss and French women's heights respectively).

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 2 solution:

```
swiss_heights = [165, 167, 164, 166, 168, 165, 167, 166, 168, 164]
french_heights = [164, 166, 165, 167, 165, 166, 164, 165, 167, 166]

# Mean and standard deviation under the null hypothesis
all_heights = np.concatenate([swiss_heights, french_heights])
null_hypothesis_mean = np.mean(all_heights)
# delta degrees of freedom (ddof) = 1 since it's a sample
null_hypothesis_std = np.std(all_heights, ddof=1)

# Means and standard deviations under the alternative hypothesis
mean_swiss_h1 = np.mean(swiss_heights)
std_swiss_h1 = np.std(swiss_heights, ddof=1)
mean_french_h1 = np.mean(french_heights)
std_french_h1 = np.std(french_heights, ddof=1)

>>>print(f"Null hypothesis mean: {null_hypothesis_mean}"
>>>print(f"Swiss mean: {mean_swiss_h1}, French mean: {mean_french_h1}")

    Null hypothesis mean: 165.75
    Swiss mean: 166.0, French mean: 165.5
```

We obtain the following mean values for our null and alternative hypotheses. We will use the aforementioned mean (and standard deviation) values to compute the relative likelihood of our hypotheses. To streamline the process, let us first define a function:

```
def compute_likelihood(heights, mean, std):
    # Compute relative likelihood value first
    pdf = norm.pdf(heights, mean, std)
    # Multiply values together to obtain joint
    # relative likelihood of the sample
    likelihood = np.prod(pdf)
    return likelihood
```

Next, we will use this function to compute the relative likelihoods of the two hypotheses.

```
# Compute null hypothesis likelihood
# Notice that we use the same total mean for both likelihoods
likelihood_swiss_h0 = compute_likelihood(swiss_heights,
                                         null_hypothesis_mean,
                                         null_hypothesis_std)
likelihood_french_h0 = compute_likelihood(french_heights,
                                          null_hypothesis_mean,
                                          null_hypothesis_std)
likelihood_h0 = likelihood_swiss_h0 * likelihood_french_h0

# Compute the alternative hypothesis likelihood
likelihood_swiss_h1 = compute_likelihood(swiss_heights,
                                         mean_swiss_h1,
                                         std_swiss_h1)
likelihood_french_h1 = compute_likelihood(french_heights,
                                          mean_french_h1,
                                          std_french_h1)
likelihood_h1 = likelihood_swiss_h1 * likelihood_french_h1
```

Now that we have our two likelihoods, let us compute the Bayes Factor to understand how much more likely our data are under H_1 versus H_0:

```
# Bayes Factor: BF = likelihood of H1 / likelihood of H0
bayes_factor = likelihood_h1 / likelihood_h0

>>>print(f"Bayes factor: {bayes_factor:.4f}")
```

```
    Bayes factor: 2.3899
```

Thus, the Bayes Factor is 2.3899.

Bonus Question 3: Interpret the Bayes Factor.

Bonus Question 3 solution: We interpret the Bayes factor here to mean that the data we observe are 2.3899 times as likely under the alternative hypothesis than under the null hypothesis. It is important to note that this does not mean that the alternative hypothesis is twice as likely, but rather that the data were twice as likely to occur under the alternative hypothesis, than under the null hypothesis. Typically, a Bayes factor of 2.3899 is considered anecdotal evidence, and no strong conclusion about a difference between the groups can be drawn on that basis.