Bayesian Statistics – Lecture 1

Thomas J. Faulkenberry, Ph.D.

May 21, 2024

A simple example

The NEO Personality Inventory-Revised (NEO PI-R) is a 240-item questionnaire that measures the "Big 5" dimensions of personality: Agreeableness (A), Conscientiousness (C), Neuroticism (N), Extraversion (E), and Openness to Experience (O).

Dolan et al. (2009) administered a Dutch version of the NEO PI-R to 500 first-year psychology students at the University of Amsterdam.

Let's take a look at that data in JASP

https://jasp-stats.org

Correlation

One problem of interest is to consider how these dimensions are associated with each other. This degree of association can be quantified by the **Pearson correlation**:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$$

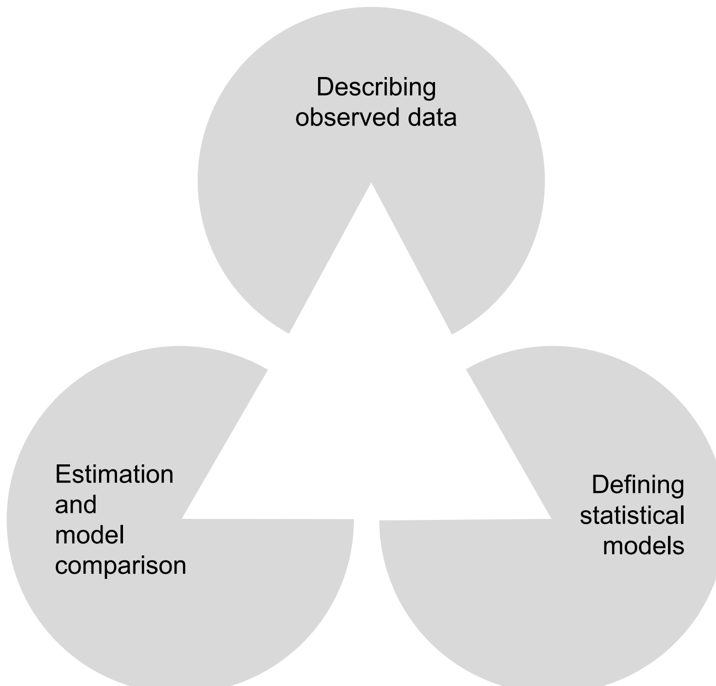
Recall:

- $-1 \leq \rho \leq 1$
- as $p \to \pm 1$, the degree of association **increases**

The problem of (scientific) inference

How do we decide whether two of the dimensions are **meaningfully associated**?

A framework for (statistical) inference



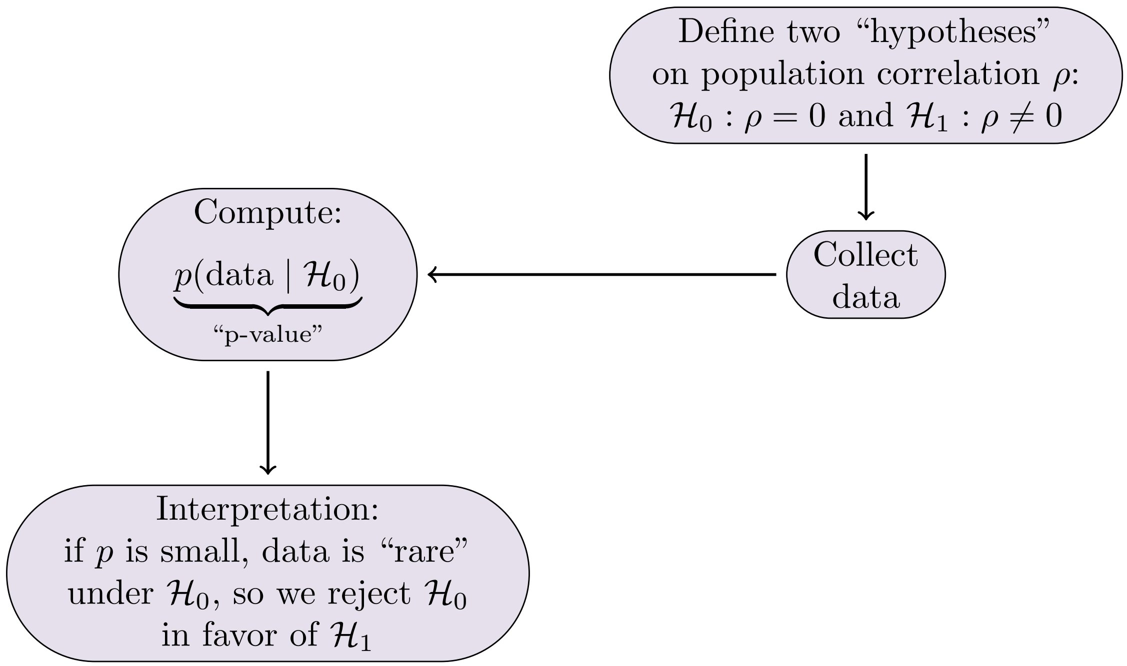
A framework for (statistical) inference

- Describing observed data
  - *Pearson correlation*

- Defining statistical models
  - $\mathcal{H}_0 : \rho = 0$
  - $\mathcal{H}_1 : \rho \neq 0$

- Model comparison and estimation
  - *hypothesis test*
  - *confidence interval for* ρ

The (frequentist) hypothesis test



The (frequentist) hypothesis test

Let's see how this works in JASP

JASP

A Fresh Way to Do Statistics

Some questions from Example 2

1. We saw $p = 0.149$, so we fail to reject $\mathcal{H}_0$. Does this mean that $\mathcal{H}_0$ is likely true? **Does the p-value tell us anything about that probability?**

2. We got a 95% confidence interval of $[-0.02, 0.15]$. **What is the likely true value of the (population) correlation ρ?**

Some questions from Example 2

The problem is that we cannot answer either of those questions from the information we have.

Problems with p-values

p-values depend on data that were never observed

- p is the probability of the observed data, **or more extreme**, under $\mathcal{H}_0$. That is, p requires knowing the probablity of all the (hypothetical) events that are more extreme that the data that were actually observed.
- this violates the **conditionality principle** (Berger & Wolpert, 1988), which states that statistical conclusions should only be based on data that have actually been observed.

Problems with p-values

p-values depend on possibly unknown subjective intentions

- constructing the sampling distribution of the test statistic under $\mathcal{H}_0$ requires also knowing how sampling is performed.

- there are multiple ways of sampling to get a particular set of observed data
  - sample until desired N is reached
  - sample until desired value of test statistic is reached
  - each results in a different sampling distribution, thus a different p-value
  - without knowing the intended sampling method, the p-value is mathematically undefined

Problems with p-values

p-values do not quantify evidence (or probability) in favor of one hypothesis over the other

- if p can be considered to index the evidence against $\mathcal{H}_0$, **then two experiments with identical p-values should be considered as providing equivalent evidence**.

- however, this is widely not true in practice. Consider the following two experiments:
  - Experiment 1 finds $p = 0.025$ after testing 10 subjects
  - Experiment 2 finds $p = 0.025$ after testing 100 subjects
  - Empirically, people do not consider these experiments as providing equivalent evidence (see Rosenthal & Gaito, 1963 or Bakan, 1966)

So what do we do?

In the next unit, we will consider an alternative to frequentist hypothesis testing that does not suffer from these problems. This alternative is **Bayesian inference**