Lab 8

Statistics 10 S2022

Statistical Inference
Hypothesis Test of Population Proportion

Dave Zes
April 15, 2022

1 Introduction

1.1 Hypothesis Testing

Formal statistical hypothesis testing facilitates the following scientific scenario.

We need to make a binary decision concerning the value, or range of values, of a parameter within a population that can be only partially observed.

The binary decision is guided by two hypotheses concerning the population parameter, the null hypothesis, which in statistics circles should always be a strict equality, and an alternative hypothesis, which will be a space of values that are in opposition to the equality expressed in the null.

The null and alternative hypotheses are stated prior to analysis and preferably prior to collecting data. This step also includes stating a significance level, α. The significance level can be thought to be related to our perceived cost of making the wrong decision when we're done. If the cost of wrongly concluding the alternative is true (called a "Type I Error") is high, we may want to set our significance level very small, say, $\alpha = 0.001$.

We call the "partially observed" portion of our population, of course, our "sample".

By the way, a "parameter" can be a vector — that is, we can test many scalar parameter values all at the same time.

In an instructional setting, we may sample from a known population just to illustrate technique and results. However, in practice — in real application — we use statistical hypothesis tests when the entirety of the population is not practically accessible. After all, if we have access to the full population, there would be no need to undertake a formal hypothesis test since we can directly calculate the parameter values of interest.

1.2 Hypothesis Testing of Population Proportions

1.2.1 Coin Toss

The classic example is the “fair coin”. If a coin is fair, then heads or tails are equally probable, otherwise, not.

$$H_0 : p_H = \frac{1}{2}$$
$$H_a : p_H \neq \frac{1}{2}$$

What about testing the claim that a coin favors heads?

$$H_0 : p_H = \frac{1}{2}$$
$$H_a : p_H > \frac{1}{2}$$

Notice in both examples, the null and alternative are in *opposition*, that is, up against each other; the values or regions they define are contiguous, adjacent.

What’s the implied target population here? That’s a little tricky. It is tacitly suggested that, given some particular coin, it may be every possible toss and landing of this coin.

1.2.2 Treatment

Suppose we have some treatment that mitigates the adverse affects of disease X. Suppose the best current treatment mitigates the adverse effects of disease X in 74% of patients.

$$H_0 : p_{\text{getbetter}} = 0.74$$
$$H_a : p_{\text{getbetter}} > 0.74$$

What’s the implied target population here? Sufferers of Disease X. We might also define our population more particularly: Sufferers of Disease X who seek treatment and medicate with our treatment.

1.3 Etc.

Note that in RStudio we can directly execute code from the editor by selecting the code and, on Mac, pressing Command+Return; or, on Windows, Shift+Alt+T.

Please also consult our lab textbook by Peter Dalgaard, [1].

Avoid plagiarism! It is very important that if you borrow code from examples from other sources, like the internet (this is very common, even for skilled programmers) or from anyone else, you should give attribution.

2 For Your Lab 8 Submission

In the folder "yourLab", examine the .Rmd (Rmarkdown) file. When this file is knit to PDF, the corresponding .pdf file is created.

Follow instructions provided in Section "Your Work".

References

[1] Dalgaard Peter. *Introductory Statistics with R*. Springer, 2008.