

BEE552 Biometry Week 6

Maria Feiler

03/02/2022

My Learning Journey

Over the last week, I participated in Biometry in the following ways:

- I asked / answered **2** questions posed in class.
- I asked **2** questions in Slack.
- I answered **0** questions posed by other students on Slack.
- I came to Heather's office hours: **Yes**
- I came to Jose's office hours: **No**
- I met with Heather or Jose separately from office hours: **No**

Anything not falling into one of the above categories?

No

On a scale of 1 (no knowledge) to 10 (complete expert), how would I rate my comfort with R programming after this week?

7

Any topics from last week that you are still confused about?

Doing fine for now

Problem Set

Part I

Listen to Planet Money's podcast Episode 677: The Experiment Experiment. <http://www.npr.org/sections/money/2016/01/15/463237871/episode-677-the-experiment-experiment>

List three take-home messages from the podcast (3 pts):

Part II

Hint: For many of these exercises, you may find it helpful to use R's $\text{dbinom}(x, \text{size}, \text{prob})$ function, where x corresponds to “ z ” (e.g., the number of heads; a vector from 0 to N) and size corresponds to N (the number of coin flips). R also has a function for the negative binomial, $\text{dnbinom}(x, \text{size}, \text{prob})$. Be careful if you use the negative binomial density, because the argument size corresponds to z (in the case of a negative binomial, this is a constant).

Suppose an experimenter plans to collect data on a coin-flipping experiment based on a two-tier stopping criterion (assume the coin is a fair coin). The experimenter will collect an initial batch of data with $N=30$ and then do a null hypothesis significance test. If the result is not significant, then an additional 15 subjects' data will be collected, for a total of 45. Suppose the researcher intends to use the standard critical values for determining significance at both the $N=30$ and $N=45$ stages. Our goal is to determine the actual false alarm rate (the Type I error rate α) for this two-stage procedure, and to ponder what the mere intention of doing a second phase implies for interpreting the first stage, even if data collection stops with the first stage.

A For $N=30$, what are the lower (z_{low}) and upper (z_{high}) limits of the 95th percentile confidence interval for z (z =number of heads)

$$p(z \leq z_{\text{low}} | N = 30, \theta = 0.5) < 0.025 \quad p(z \geq z_{\text{high}} | N = 30, \theta = 0.5) < 0.025$$

assuming a two-tailed Type I error rate of 0.05 or less?

B For $N=45$, what are the lower (z_{low}) and upper (z_{high}) limits of the 95th percentile confidence interval for z (z =number of heads)

$$p(z \leq z_{\text{low}} | N = 45, \theta = 0.5) < 0.025 \quad p(z \geq z_{\text{high}} | N = 45, \theta = 0.5) < 0.025$$

assuming a two-tailed Type I error rate of 0.05 or less?

For the next part of the exercise, consider the table provided.

Second 15 flips	First 30 flips														
	0	...	9	10	...	15	16	...	20	21	...	30			
0	✚✚		✚✚	✚		✚	–		–	✚		✚✚			
1	✚✚		✚✚	✚		–	–		–	✚		✚✚			
2	✚✚		✚✚	✚		–	–		–	✚		✚✚			
3	✚✚		✚✚	✚		–	–		–	✚		✚✚			
4	✚✚		✚✚	✚		–	–		–	✚		✚✚			
5	✚✚		✚✚	✚		–	–		–	✚		✚✚			
6	✚✚		✚✚	–		–	–		–	✚		✚✚			
7	✚✚		✚	–		–	–		–	✚		✚✚			
8	✚✚		✚	–		–	–		–	✚		✚✚			
9	✚✚		✚	–		–	–		–	✚✚		✚✚			
10	✚✚		✚	–		–	–		✚	✚✚		✚✚			
11	✚✚		✚	–		–	–		✚	✚✚		✚✚			
12	✚✚		✚	–		–	–		✚	✚✚		✚✚			
13	✚✚		✚	–		–	–		✚	✚✚		✚✚			
14	✚✚		✚	–		–	✚		✚	✚✚		✚✚			
15	✚✚		✚	–		✚	✚		✚	✚✚		✚✚			

Figure 1: Thirty first flips and fifteen second flips of a fair coin

Each cell of the table corresponds to a certain outcome from the first 30 flips of a fair coin and a certain outcome from the second 15 flips of the same fair coin. A cell is marked by a dagger, ✚, if it has a result for

the first 30 flips that would reject the null hypothesis. A cell is marked by a star, ★, if it has a result for the total of 45 flips that would reject the null hypothesis. For example, the cell with 10 heads from the first 30 flips and 1 head from the second 15 flips is marked with a ★ because the total number of heads for that cell, $10+1=11$, is less than 15 (which is z_{low} for $N=45$ [a hint for part B!]). That cell has no dagger, †, because getting 10 heads in the first 30 flips is not extreme enough to reject the null. If neither the first 30 coin flips, nor the second 15 coin flips, would reject the null hypothesis of a fair coin, then the cell is marked with a dash -.

C Denote the number of heads in the first 30 flips as z_1 , and the number of heads in the second 15 flips as z_2 . Explain why it is true that the z_1, z_2 cell of the table has a joint probability equal to $\text{dbinom}(z_1, 30, 0.5) * \text{dbinom}(z_2, 15, 0.5)$

D What is the sum of the probabilities of all the cells that contain a † (whether or not it contains a ★)? Explain how you got your answer!

E What is the sum of the probabilities of all the cells that contain a ★ (whether or not it contains a †)? Explain how you got your answer!

F What is the sum of the probabilities of all the cells that contain either a † or a ★? (Note: This is the Type I error rate for the two-stage design, because these are all the ways you would decide to reject the null even when it is true.) Explain how you got your answer!

G Suppose that the researcher intends to run an experiment using this two-stage stopping criterion. She collects the first 30 flips and finds 8 heads. She therefore rejects the null hypothesis and reports that $p < 0.05$. Is that correct? Explain.

H Whenever we run an experiment and get a result that trends away from the null experiment, but isn't quite significant, it's natural to consider collecting more data. We saw in the previous part that even intending to collect more data, but not actually doing it, inflates the Type I error rate. Doesn't the fact that we always consider collecting more data mean that we always have a much higher Type I error rate than we pretend we do? Doesn't the actual Type I error rate of an experiment depend on the maximal number of data points we'd be willing to collect over the course of our lifetimes? In 1-2 paragraphs, discuss this conundrum and decide whether or not you think this poses a fundamental problem with null hypothesis testing.

Bonus Whenever we run an experiment and get a result that trends away from the null experiment, but isn't quite significant, it's natural to consider collecting more data. We saw in the previous part that even intending to collect more data, but not actually doing it, inflates the Type I error rate. Doesn't the fact that we always consider collecting more data mean that we always have a much higher Type I error rate than we pretend we do? Doesn't the actual Type I error rate of an experiment depend on the maximal number of data points we'd be willing to collect over the course of our lifetimes? In 1-2 paragraphs, discuss this conundrum and decide whether or not you think this poses a fundamental problem with null hypothesis testing.