BIOSTAT 702: Exercise 3.2

One Sample Inference for a Binary Outcome

Fall 2025

Contents

Learning Objectives

1. Understand when 1-sided and 2-sided statistical tests are appropriate
2. Know how to conduct and interpret a hypothesis test for a binary outcome
3. Describe when and why a control group is needed in clinical research studies

How to Do This Exercise

We recommend that you read this entire document prior to answering any of the questions. If anything is unclear please ask for help from the instructors or TAs before getting started. You are also allowed to ask for help from the instructors or TAs while you are working on the assignment. You may collaborate with your classmates on this assignment—in fact, we encourage this–and use any technology resources available to you, including Internet searches, generative AI tools, etc. However, if you collaborate with others on this assignment please be aware that *you must submit answers to the questions written in your own words. This means that you should not quote phrases from other sources, including AI tools, even with proper attribution.*

Although quoting with proper attribution is good scholarly practice, it will be considered failure to follow the instructions for this assignment and you will be asked to revise and resubmit your answer. In this eventuality, points may be deducted in accordance with the grading rubric for this assignment as described below. Finally, you do not need to cite sources that you used to answer the questions for this assignment.

Grading Rubric

The assignment is worth 20 points (4 points per question). The points for each question are awarded as follows: 3 points for answering all parts of the question and following directions, and 1 point for a correct answer. Partial credit may be awarded at the instructor's discretion.

Resources

We will be using the article from Exercise 1.2 by Jabbour for this assignment.

Question 1

The clinical trial by Jabbour compares a first-generation tyrosine kinase inhibitor (imatinib) to a third-generation, ponatinib. In the discussion of the manuscript, it is stated that "While it is unclear how the current findings compare with second-generation tyrosine kinase inhibitors in the absence of a head-to-head randomized trial, the MRD negativity rate at the end of induction in the imatinib group in PhALLCON (22%) was comparable with previous reports for second-generation tyrosine kinase inhibitors (14%-18%).

We want to run a one-sided, one-sample proportion test to test if the MRD negativity rate at the end of induction for ponatinib is greater than previous reports for second-generation tyrosine kinase inhibitors (let's use the midpoint of 16%).

1. First, explain why it is acceptable in this situation for the test to be 1-sided.
2. From the manuscript, what is the MRD negativity rate at the end of induction for ponatinib from the study?
3. Run a one-sided one-sample test for the proportion, using the null hypothesis that $\pi \leq 16\%$. Do not use the continuity correction.
4. Write a definition of the 1-sided p-value in the context of this problem.

Question 2

1. Use the normal approximation to verify "by hand" that the function above gives the same test statistic and p-value.
2. Interpret the confidence interval you see from the output of the one-sided hypothesis test. If you were to run a two-sided test, what would you have to set the conf.level at to see the same lower bound? Run that test to check.

Note: You will learn more about 1-sided and 2-sided tests in the BIOSTAT 701 course. But since the topic is relevant to this exercise we will provide a little background information for you to explore further on your own. First, as you will find out in the 701 course, when you use a 1-sided test with alpha level of 5% you

gain a slight power advantage over a 2-sided test using the same alpha level. However, as the FDA points out in one of their clinical trial guidance documents:

"For one-sided hypothesis tests, the Type I error probability refers to the probability of concluding specifically that there is a beneficial difference due to the drug when there is not."

In a 2-sided test we allow for the possibility of detecting harm from the treatment in addition to benefit. Using normal approximation tests, we divide our Type I error probability evenly between these two possibilities (2.5% for each rejection region). By analogy, to preserve the same Type I error probability for 1-sided tests that are designed to identify only beneficial effects of treatment, we often set the Type I error probability to 2.5%. This also has the effect of eliminating the power advantage mentioned before; i.e., the test now has the same power as a 2-sided test at alpha of 5%.

Question 3

The study *now* includes 3 groups; first-, second-, and third-generation tyrosine kinase inhibitors. The investigators have the raw data from the first-generation drug, but we must rely on the literature to obtain summary results for the second-generation drug. What are the advantages and disadvantages of these two ways to make comparison groups? What biases might result?

Question 4

Now, we want to compare the observed MRD negativity rate by end of induction for imatinib with the historical summary estimate of 16% for second-generation tyrosine kinase inhibitors. Specifically, we are interested in determining if imatinib was a good control and exhibited the MRD negativity rate expected of earlier generation drugs.

1. Would it make more sense to perform this hypothesis test as a one-sided or two-sided test and why?
2. Run this test. Interpret the results.

Question 5

Finally, let's say we are interested in the testing the hypothesis that the proportional of loss-of-response events among those with MRD negativity before induction in the imatinib group is equal to 0.25.

1. What is the total sample size for this test and what is the sample proportion?
2. The normal approximation test is not useful in this case because we have such a small sample. So, you will use the exact binomial test instead. Show the probability distribution for 0 to 13 loss-of-response events under H0. Verify that the distribution sums to 1. Hint: you can use the dbinom function.
3. Find the two-sided p-value using the null hypothesis distribution you printed out in the question above.
4. Now find the p-value using the binom.test function. Both methods should give you the same answer.