



M5-07: 1 Sample Z Test

Part of the "Polling, Confidence Intervals, and the Normal Distribution" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m5-07/>

Hypothesis Testing

Hypothesis Tests (or Significance Tests) are statistical tests to see if a difference we observe is due to chance.

We will look at many different types of hypothesis tests for different scenarios in Data Science DISCOVERY, but they all have the same basic ideas.

The One-Sample Z Test

The first type of hypothesis test we are going to look at is the one-sample z-test. You can do a z-test for means or for proportions. Let's look at one below!

Puzzle: Suppose a large midwestern university claims that the average ACT score of their incoming freshman class is 29, but we think the University may be inflating their average. To test the University's claim we take a simple random sample of 50 students and find their average to be only 28.3 with an SD of 4. Perform a hypothesis test to test the claim.

Hypothesis Test Steps and Set-Up:

1. Formulate a null hypothesis and an alternative hypothesis.

Ho- Null Hypothesis:

Ha- Alternative Hypothesis:

2. Think about what you would expect to get if you randomly sampled from the population, assuming the null is true. Compare your observed data and expected data and calculate the test statistic.



M5-07: 1 Sample Z Test

Part of the “Polling, Confidence Intervals, and the Normal Distribution” Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m5-07/>

3. Calculate the probability of getting the data you got or something even more extreme if the null were true. This is called the p-value.

4. Make your conclusion and interpret it in the context of the problem. If p is very low, we say that the data support rejecting the null hypothesis.

How low is “very low”?

The convention is to reject the null when $P < 5\%$ and call the result “significant”. There’s no particular justification for this value but it’s commonly used. (see [Historical Note on the Conventions](#) from Berkeley for more details- scroll to the bottom of the page)

The P-value cut-off is called the significance level, α .