Post 1: Basics of Hypothesis Tests and One sample Z-test

Sudarshan Srirangapatanam

September 11, 2017

Introduction

A hypothesis test is simply a test we perform on our data to evaluate our hypothesis (or a claim). Suppose we just collected some data of our own and want to compare it someone else's data, we could try and look at them using visual tools and try to describe the comparison. This is perfectly fine as our first step, but say we want to answer a specific question from our data. How could we do this **objectively** and **reproducibly**? We do this by conducting a hypothesis test.

The utility of hypothesis tests:
Say there was a claim that on average students at Cal weigh 150 lbs. We collect data to evaluate this claim (our hypothesis) and our data might tell us that in our sample of students, the average weight was 200 lbs. Does this mean we reject our claim and set a new claim to be 200 lbs, or do we say that this difference is not actually true and is a result of how we sample (maybe the sample from Center St. is not representative of Cal as a whole)? and how would we do this with as little subjectivity as possible? This is where hypothesis tests come into play. They objectively tell us what we should do in such cases.

What do we need to conduct a hypothesis test? *(Ingredients)*

At the very least we need the following information to successfully run a test and be able to draw reliable interpretations:

1. Information regarding the data we collected:
    - Full data, or
    - Summary of the sample, or
    - Select statistics from the data
2. Hypothesis:
    - A claim we wish to evaluate, or
    - A hypothesis that we believe is true based on our data, or
    - A hypothesis that we would want to reject
3. A test:
    - Test that would allow us to accurately interpret results (taking into consideration the limitations of the sample)
4. Acceptable error:
    - maximum allowed error, or
    - interpret based on p-value

Testing scheme

This is a general scheme that can be used to perform any test. Some adjustments may be necessary for the different tests.

Step 1: Construct hypothesis

We must have two hypotheses (null and alternative). What we are testing, in testing terminology, is a function of our data and is called test-statistic.

- **Null Hypothesis:** This is the hypothesis we assume is true unless proven otherwise. This is also our claim that we wish to evaluate.
- **Alternative hypothesis:** This is precisely the opposite of our null. In the testing space, this could mean different things based on how we define it.
  If we do not have an alternative hypothesis, we can set this to be anything other than null.
  We can also test our null against a specific set of conditions, the condition we are testing against is the alternative.

We must define our alternative so that there is some reason to believe it is true given that the evidence doesn't support our null.

To better understand what hypothesis means in the testing world, think about the statement ***"Innocent until proven guilty."***
Here we claim that the person is innocent and assume that it is true. If we had enough proof that a person is NOT innocent, then we say that he/she is guilty. In this case, the fact that he/she is innocent is our null hypothesis and the fact that he/she is guilty is our alternative hypothesis.

Step 2: Decide on a test that is appropriate for the sample

We have many tests that we can use but not all tests are created equal, so we will have to be wary when deciding on a test that fits our data. The test we will be talking about in this post is the **one sample Z-test**. I will introduce you to other tests in the coming posts.

Each test has its own set of assumptions and its compatibility to our sample depends on whether the assumptions are applicable.

Step 3: Gather necessary statistics from the sample

Based on the test we decide from *Step 2*, we have to gather important/necessary statistics from our data/sample. Think of tests as black boxes, they take an input and spit out an output. Our inputs are the statistics we supply, and output is our conclusion.

Step 4: Decide on an acceptable error

For this step, we will have to use our judgment and decide on how much of error are we willing to sacrifice. A high acceptable error might lead to rejecting the claim even when it is true, and the converse is also true.

Consider the following scenarios to see how one would decide on acceptable error:

- **Case 1:** “Innocent until proven guilty”
  Ideally, we do not want any error, but with an error of zero our tests would always accept our claim. Because of this, we allow for 5% error, standard in most cases.
- **Case 2:** Identifying blood type
  A low error would again be ideal, but it also warrants a high maintenance cost for the company manufacturing the machine. So we settle for 5% error given that the case if not life-threatening.
- **Case 3:** HIV test
  In this case, We wouldn’t want our report to tell an individual with HIV that he/she is negative because this would mean that he/she will go untreated. Here we set the error to 1% or .01%.

Alternatively, we can think about the error as a threshold and define it as the lowest probability we are willing to accept as purely random (due to sampling). Any value lower than this would be considered to have occurred due to some underlying reason and NOT by chance.

Acceptable error in testing terminology is said to be *significance level* or denoted with Greek alpha α. Some like to use the term *confidence level* instead to define error and confidence level is simple (1 - α)*100%.

Step 5: Execute the test

This is the easiest step of all.
We just take our understanding of the test from *Step 2* and run a test using statistics we calculate from *Step 3*. Then, we interpret out results based our acceptable error from *Step 4* and decide whether we are willing to accept our null from *Step 1*.

One Sample Z-test

We will start with one sample Z-test since it is the simplest of all tests.

Assumptions

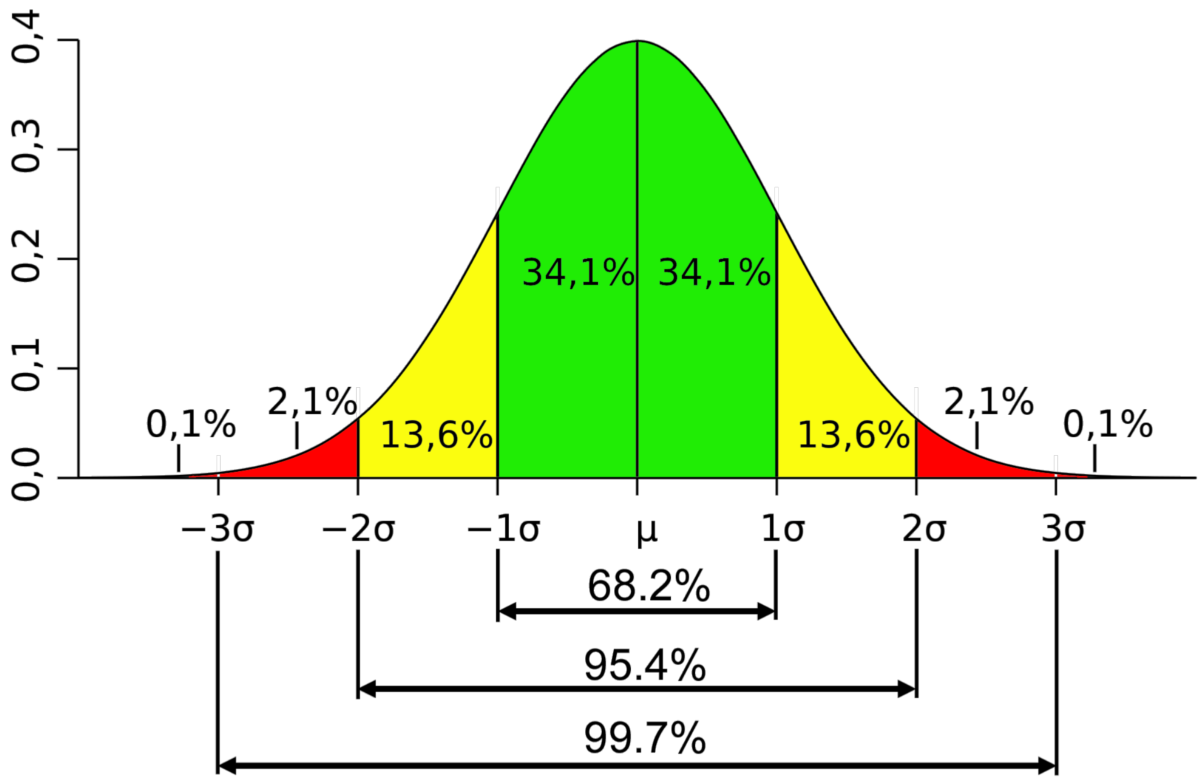
- The sample is approximately normal in distribution, or the test-statistic is approximately normal.
- We know the population variance in the context of our data (in other words, the sampling variance).

Basics of Z-test

When we do a Z-test we assume that the sample (or the test-statistic) is normally distributed, or at least is approximately normal.

A normal distribution is given by two parameters: **mean** and **variance.** We express our test statistic as a linear function of mean when possible to extend the analysis of mean to our test statistic.

As a result of this, we get a normal distribution of our test-statistic with a center (mean) and a spread (variance). With this information we can draw the following density curve:



adapted from http://www.muelaner.com

From this distribution we can make the following statements:

- Probability of observing a result $\leq X$ purely due to chance (sampling error) is given by $\int_{-\infty}^{X} Norm(\mu, \sigma^2)$
- Similarly, probability of observing a result $\geq X$ purely due to chance (sampling error) is given by $\int_{X}^{\infty} Norm(\mu, \sigma^2)$
- Probability of observing a result in the interval [a,b] or (a,b) purely due to chance (sampling error) is given by $\int_{a}^{b} Norm(\mu, \sigma^2)$

The notion of p-value: Notion of p-value is very important for any hypothesis test and is very simple to understand in terms of the notion of probability listed above. It is the probability of observing values *as or more extreme* than what was observed.

If the p-value (probability of observing result purely due to chance) is very high then the newly observed result can't be relied upon, so we don't reject our claim.
If on the other hand, the p-value is very low then the newly observed value must take some weight in our analysis so we should be skeptical about the original claim.

Example

```
claim <- 85 # null or claim
sderror <- 11.6 # standard deviation of the population

n <- 25 # sample size
x <- 80.94 # our study results

alpha <- .01 # significance leve (or acceptable error)

p_value <- pnorm(x, mean = claim, sd = sderror/sqrt(n))
```

Boys of a certain age are known to have a mean weight of μ = 85 pounds. A complaint is made that the boys living in a municipal children's home are underfed. As one bit of evidence, n = 25 boys (of the same age) are weighed and found to have a mean weight of $\bar{x}$ = 80.94 pounds. It is known that the population standard deviation σ is 11.6 pounds (the unrealistic part of this example!). Based on the available data, what should be concluded concerning the complaint?

Example adapted from Penn State early college of science.

Our main goal is to verify if the kids are actually underfed or if the low value we got from measuring children (80.94) is purely due to chance.

- **Null Hypothesis**: On average, the kids of known age group are known to weigh 85 pounds. That is $H_0 : \mu$ = 85.
- **Alternative Hypothesis**: Since our study showed a lower weight from the kids measured and we would like to answer if they are underfed. We would want to conclude that they are actually underfed if their weight is, in fact, less than what is claimed. That is $H_a : \mu$ < 85.

Since Z-test relies on normal distribution and measures deviation according to a normal distribution, we need mean and standard deviation. We can set the claim as our mean and sampling error of sample size 25 as our deviation for the distribution.

Now we will have to decide on an acceptable error. Since large error, in this case, would lead to undesirable consequences lets say we will allow for 1% error.

What does the error formally mean?
The error is simply the probability that we reject our null given that it is in fact true. $Pr(reject\ H_0 \mid H_0\ is\ true)$

There are two major ways to carry out a test. One could use the significance level to *find critical values* (values beyond which the test would reject null), and then compare these to the actual values to arrive at a conclusion. Or

One could carry out a test using only the actual values and find the p-value, and then **compare p-value to significance level** to arrive at a conclusion.

We will be using the latter method since it is more straightforward to understand and allows us to start a test without any defined error. This could be very useful when one cannot decide on the acceptable error at first.

Our first task is to find the probability of observing the extreme results (results of our study) given that the actual weight is same as the population mean. Remember that we do not reject our null hypothesis unless proven otherwise.

For this, we use a normal distribution with **mean 85** and **variance 5.3824**.

$Pr(x \leq 80.94 \mid \mu = 85, \sigma = 2.32) = 0.0400592$

Next, we do a simple comparison of allowed error and the value we get above (p-value). If the value is less than or equal to our error, we can conclude our null hypothesis is false. That is, we reject our claim since the probability value shows proof that this value is, in fact, significant. Otherwise, we assume that initial claim is still true.

In our case since 0.04 is greater than 0.01, we cannot reject our null, so we say that the claim is true is still true and the children are not underfed.
Note: If we used a different alpha (allowed error) we could have ended up deciding that we should reject our null

To see this test on a Z-curve, test the values above on our calculator (see app page). You can also see how the test results change when the values are changed. See below for detailed instructions on how to use the calculator.

More examples

Interested in more examples?

Click here for a word example.
Here is one from youtube as a video example.

Take home message

With hypothesis tests at your disposal, you can compare your study results to another study, verify claims (your own or someone else's), and much more in an objective and reproducible manner.

One Sample Z-test is one particular type of the test whose details are outlined above but the basic testing paradigm can be applied to any test. In the coming post, I will introduce another test which is much more applicable.

The calculator below is an effort to visualize One Sample Z-test and a convenient way to quickly verify test results.

Calculator

See `../shiny/z-test/app.R` for **One Sample Z-test Calculator** code or visit the app page for the web app.

What do the inputs mean?

- **Claim**: This is the number from your claim (the number from your null hypothesis).
- **Variance**: This is the population variance or the sampling variance.
  Note: Standard Error = Standard Deviation = Squared root of Variance.
- **Test Value**: This is the value you are testing your claim against, usually the value from your study.
- **Alternative hypothesis**: This is the sidedness of your alternative hypothesis (Not equal to the claim, less than, or greater than).
- **Significance level**: This is your allowed error or your alpha value.
  By default, the value is set to most used alpha (.05) and a slider that allows for values by .01 is given.
  If you need to enter a number manually select *"Yes"* from **"Enter a custom alpha value?"**

Entering the above example?

- Your *claim* would be the mean as noted in the null hypothesis so it would be **85**.
- Your *variance* is somewhat tricky. You will have to enter sampling variance but you are given variance for the full population. To convert population variance to sampling variance, you simply divide the population variance by sample size. So Variance you would enter is 134.56 / 25 = **5.3824**.
- *Test Value* is the value from your study so it would be **80.94**.
- *Alternative hypothesis* is the sidedness of our alternate above. We choose **<** based on the question, so we will choose the same one here.
- For alpha (*significance level*), we will have to change the value to **0.01**, either by moving the slider or by selecting "Yes" first and then entering 0.01.

References

1. Mathematical Statistics and Data Analysis by John A. Rice
2. The modern paradigm for hypothesis testing
3. The Neyman-Pearson approach
4. Significance Levels (Alpha) and P values in Statistics
5. Example Question
6. One-sample Z-test detailed example
7. Video Example of One-sample Z-test
8. Normal Distribution Image