

9.4: Probability Distribution Needed for Hypothesis Testing

Earlier in the course, we discussed sampling distributions. Particular distributions are associated with various types of hypothesis testing. The following table summarizes various hypothesis tests and corresponding probability distributions that will be used to conduct the test (based on the assumptions shown below):

Table 9.4.1

Type of Hypothesis Test	Population Parameter	Estimated value (point estimate)	Probability Distribution Used
Hypothesis test for the mean, when the population standard deviation is known	Population mean μ	Sample mean \bar{x}	Normal distribution, $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X}{\sqrt{n}}\right)$
Hypothesis test for the mean, when the population standard deviation is unknown and the distribution of the sample mean is approximately normal	Population mean μ	Sample mean \bar{x}	Student's t-distribution, t_{df}
Hypothesis test for proportions	Population proportion p	Sample proportion p'	Normal distribution, $P' \sim N\left(p, \sqrt{\frac{p \cdot q}{n}}\right)$

Assumptions

When you perform a hypothesis test of a single population mean μ using a normal distribution (often called a ztest), you take a simple random sample from the population. The population you are testing is normally distributed, or your sample size is sufficiently large. You know the value of the population standard deviation, which, in reality, is rarely known.

When you perform a hypothesis test of a single population mean μ using a Student's t-distribution (often called a t-test), there are fundamental assumptions that need to be met in order for the test to work properly. Your data should be a simple random sample that comes from a population that is approximately normally distributed. You use the sample standard deviation to approximate the population standard deviation. (Note that if the sample size is sufficiently large, a t-test will work even if the population is not approximately normally distributed).

When you perform a hypothesis test of a single population proportion p , you take a simple random sample from the population. You must meet the conditions for a binomial distribution: there are a certain number n of independent trials, the outcomes of any trial are success or failure, and each trial has the same probability of a success p . The shape of the binomial distribution needs to be similar to the shape of the normal distribution. To ensure this, the quantities np and nq must both be greater than five ($np > 5$ and $nq > 5$). Then the binomial distribution of a sample (estimated) proportion can be approximated by the normal distribution with $\mu = p$ and $\sigma = \sqrt{\frac{pq}{n}}$. Remember that $q = 1 - p$.

Hypothesis Test for the Mean

Going back to the standardizing formula we can derive the test statistic for testing hypotheses concerning means.

$$Z_c = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \quad (9.4.1)$$

The standardizing formula cannot be solved as it is because we do not have μ , the population mean. However, if we substitute in the hypothesized value of the mean, μ_0 in the formula as above, we can compute a Z value. This is the test statistic for a test of hypothesis for a mean and is presented in Figure 9.3. We interpret this Z value as the associated probability that a sample with a sample mean of \bar{X} could have come from a distribution with a population mean of H_0 and we call this Z value Z_c for "calculated". Figure 9.3 and Figure 9.4 show this process.

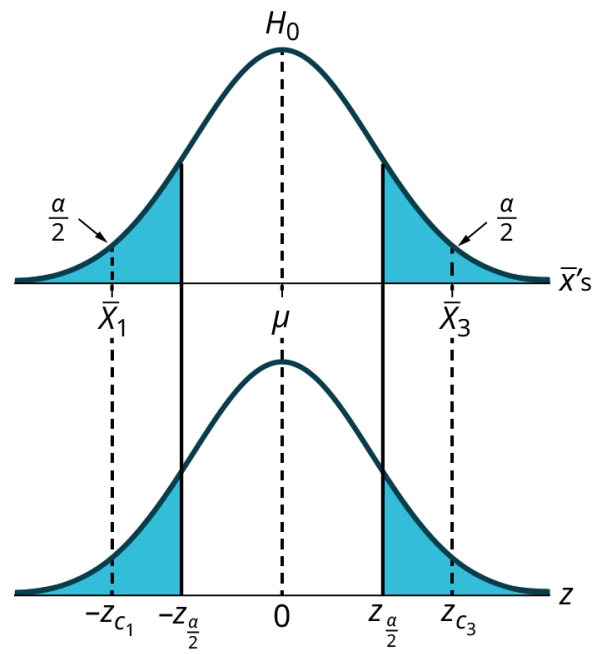


Figure 9.4.1

In Figure 9.4.1 two of the three possible outcomes are presented. \bar{X}_1 and \bar{X}_3 are in the tails of the hypothesized distribution of H_0 . Notice that the horizontal axis in the top panel is labeled \bar{X} 's. This is the same theoretical distribution of \bar{X} 's, the sampling distribution, that the Central Limit Theorem tells us is normally distributed. This is why we can draw it with this shape. The horizontal axis of the bottom panel is labeled Z and is the standard normal distribution. $Z_{\frac{\alpha}{2}}$ and $-Z_{\frac{\alpha}{2}}$, called the critical values, are marked on the bottom

panel as the Z values associated with the probability the analyst has set as the level of significance in the test, (α). The probabilities in the tails of both panels are, therefore, the same.

Notice that for each \bar{X} there is an associated Z_C , called the calculated Z , that comes from solving the equation above. This calculated Z is nothing more than the number of standard deviations that the **hypothesized** mean is from the sample mean. If the sample mean falls "too many" standard deviations from the hypothesized mean, given our pre-set required level of significance. It **could** have come from H_0 , but it is deemed just too unlikely. In Figure 9.3 both \bar{X}_1 and \bar{X}_3 are in the tails of the distribution. They are deemed "too far" from the hypothesized value of the mean given the chosen level of alpha. If in fact this sample mean it did come from H_0 , but from in the tail, we have made a Type I error: we have rejected a good null. Our only real comfort is that we know the probability of making such an error, α , and we can control the size of α .

Figure 9.4.2 shows the third possibility for the location of the sample mean, \bar{x} . Here the sample mean is within the two critical values. That is, within the probability of $(1-\alpha)$ and we cannot reject the null hypothesis.

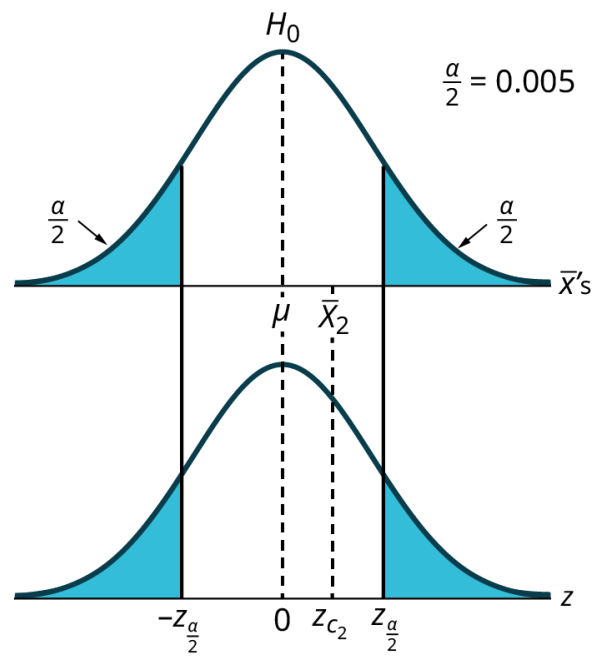


Figure 9.4.2

This gives us the decision rule for testing a hypothesis for a two-tailed test:

Decision rule: two-tail test
If $ Z_c < Z_{\frac{\alpha}{2}}$: then do not REJECT H_0
If $ Z_c > Z_{\frac{\alpha}{2}}$: then REJECT H_0

This rule will always be the same no matter what hypothesis we are testing or what formulas we are using to make the test. The only change will be to change the Z_c to the appropriate symbol for the test statistic for the parameter being tested. Stating the decision rule another way: if the sample mean is unlikely to have come from the distribution with the hypothesized mean we cannot accept the null hypothesis. Here we define "unlikely" as having a probability less than alpha of occurring.

P-Value Approach

An alternative decision rule can be developed by calculating the probability that a sample mean could be found that would give a test statistic larger than the test statistic found from the current sample data assuming that the null hypothesis is true. Here the notion of "likely" and "unlikely" is defined by the probability of drawing a sample with a mean from a population with the hypothesized mean that is either larger or smaller than that found in the sample data. Simply stated, the p-value approach compares the desired significance level, α , to the p-value which is the probability of drawing a sample mean further from the hypothesized value than the actual sample mean. A large p-value calculated from the data indicates that we should not reject the null hypothesis. The smaller the p value, the more unlikely the outcome, and the stronger the evidence is against the null hypothesis. We would reject the null hypothesis if the evidence is strongly against it. The relationship between the decision rule of comparing the calculated test statistics, Z_c , and the Critical Value, Z_{α} , and using the p-value can be seen in Figure 9.4.3.

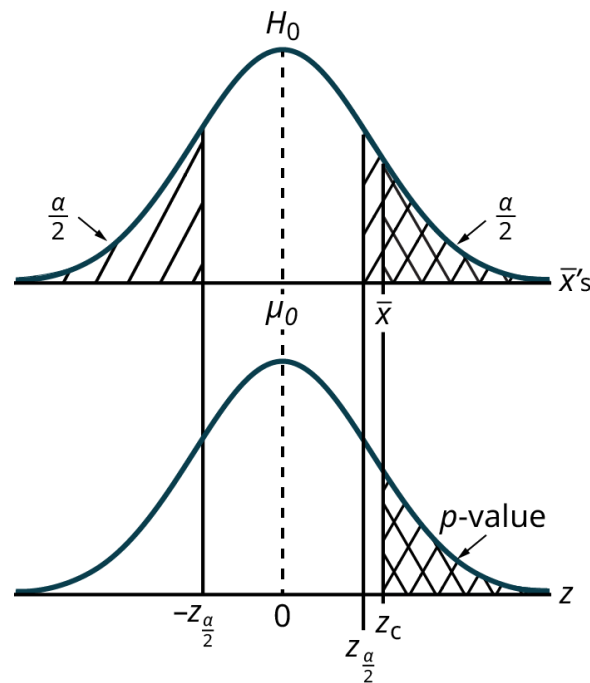


Figure 9.4.3

The calculated value of the test statistic is Z_c in this example and is marked on the bottom graph of the standard normal distribution because it is a Z value. In this case the calculated value is in the tail and thus we cannot accept the null hypothesis, the associated \bar{X} is just too unusually large to believe that it came from the distribution with a mean of μ_0 with a significance level of α .

If we use the p -value decision rule we need one more step. We need to find in the standard normal table the probability associated with the calculated test statistic, Z_c . We then compare that to the α associated with our selected level of confidence. In Figure 9.5 we see that the p -value is less than α and therefore we cannot accept the null. We know that the p -value is less than α because the area under the p -value is smaller than $\alpha/2$. It is important to note that two researchers drawing randomly from the same population may find two different P-values from their samples. This occurs because the P-value is calculated as the probability in the tail beyond the sample mean assuming that the null hypothesis is correct. Because the sample means will in all likelihood be different this will create two different P-values. Nevertheless, the conclusions as to the null hypothesis should be different with only the level of probability of α .

Here is a systematic way to make a decision of whether you cannot accept or cannot reject a null hypothesis if using the p -value and a preset or preconceived α (the "significance level"). A preset α is the probability of a Type I error (rejecting the null hypothesis when the null hypothesis is true). It may or may not be given to you at the beginning of the problem. In any case, the value of α is the decision of the analyst. When you make a decision to reject or not reject H_0 , do as follows:

- If $\alpha > p$ -value, cannot accept H_0 . The results of the sample data are significant. There is sufficient evidence to conclude that H_0 is an incorrect belief and that the alternative hypothesis, H_a , may be correct.
- If $\alpha \leq p$ -value, cannot reject H_0 . The results of the sample data are not significant. There is not sufficient evidence to conclude that the alternative hypothesis, H_a , may be correct. In this case the status quo stands.
- When you "cannot reject H_0 ", it does not mean that you should believe that H_0 is true. It simply means that the sample data have failed to provide sufficient evidence to cast serious doubt about the truthfulness of H_0 . Remember that the null is the status quo and it takes high probability to overthrow the status quo. This bias in favor of the null hypothesis is what gives rise to the statement "tyranny of the status quo" when discussing hypothesis testing and the scientific method.

Both decision rules will result in the same decision and it is a matter of preference which one is used.

One and Two-tailed Tests

The discussion of Figure 9.3-Figure 9.5 was based on the null and alternative hypothesis presented in Figure 9.3. This was called a two-tailed test because the alternative hypothesis allowed that the mean could have come from a population which was either larger or smaller than the hypothesized mean in the null hypothesis. This could be seen by the statement of the alternative hypothesis as $\mu \neq 100$, in this example.

It may be that the analyst has no concern about the value being "too" high or "too" low from the hypothesized value. If this is the case, it becomes a one-tailed test and all of the alpha probability is placed in just one tail and not split into $\alpha/2$ as in the above case of a two-tailed test. Any test of a claim will be a one-tailed test. For example, a car manufacturer claims that their Model 17B provides gas mileage of greater than 25 miles per gallon. The null and alternative hypothesis would be:

- $H_0 : \mu \leq 25$
- $H_a : \mu > 25$

The claim would be in the alternative hypothesis. The burden of proof in hypothesis testing is carried in the alternative. This is because failing to reject the null, the status quo, must be accomplished with 90 or 95 percent confidence that it cannot be maintained. Said another way, we want to have only a 5 or 10 percent probability of making a Type I error, rejecting a good null; overthrowing the status quo.

This is a one-tailed test and all of the alpha probability is placed in just one tail and not split into $\alpha/2$ as in the above case of a two-tailed test.

Figure 9.4.4 shows the two possible cases and the form of the null and alternative hypothesis that give rise to them.

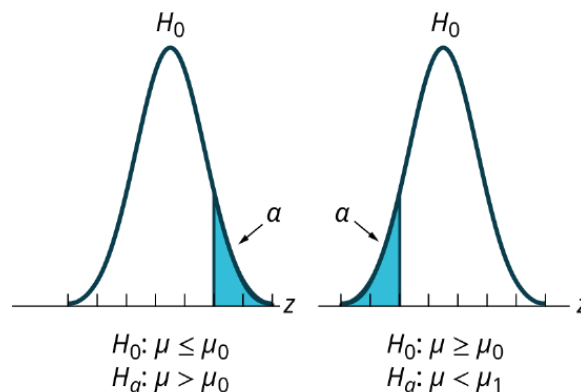


Figure 9.4.4

where μ_0 is the hypothesized value of the population mean.

Table 9.4.3: Test Statistics for Test of Means, Varying Sample Size, Population Standard Deviation Known or Unknown

Sample Size	Test Statistic
< 30 (σ unknown)	$t_c = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$
< 30 (σ known)	$Z_c = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$
> 30 (σ unknown)	$Z_c = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$
> 30 (σ known)	$Z_c = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Effects of Sample Size on Test Statistic

In developing the confidence intervals for the mean from a sample, we found that most often we would not have the population standard deviation, σ . If the sample size were less than 30, we could simply substitute the point estimate for σ , the sample standard deviation, s , and use the student's t -distribution to correct for this lack of information.

When testing hypotheses we are faced with this same problem and the solution is exactly the same. Namely: If the population standard deviation is unknown, and the sample size is less than 30, substitute s , the point estimate for the population standard deviation, σ , in the formula for the test statistic and use the student's t -distribution. All the formulas and figures above are unchanged except for this substitution and changing the Z distribution to the student's t -distribution on the graph. Remember that the student's t -distribution can only be computed knowing the proper degrees of freedom for the problem. In this case, the degrees of freedom is computed as before with confidence intervals: $df = (n-1)$. The calculated t -value is compared to the t -value associated with the pre-set level of confidence

required in the test, $t_{\alpha, df}$ found in the student's t tables. If we do not know σ , but the sample size is 30 or more, we simply substitute s for σ and use the normal distribution.

Table 9.4.3 summarizes these rules.

A Systematic Approach for Testing a Hypothesis

A systematic approach to hypothesis testing follows the following steps and in this order. This template will work for all hypotheses that you will ever test.

- Set up the null and alternative hypothesis. This is typically the hardest part of the process. Here the question being asked is reviewed. What parameter is being tested, a mean, a proportion, differences in means, etc. Is this a one-tailed test or two-tailed test?
- Decide the level of significance required for this particular case and determine the critical value. These can be found in the appropriate statistical table. The levels of confidence typical for businesses are 80, 90, 95, 98, and 99. However, the level of significance is a policy decision and should be based upon the risk of making a Type I error, rejecting a good null. Consider the consequences of making a Type I error.

Next, on the basis of the hypotheses and sample size, select the appropriate test statistic and find the relevant critical value: Z_{α} , t_{α} , etc. Drawing the relevant probability distribution and marking the critical value is always big help. Be sure to match the graph with the hypothesis, especially if it is a one-tailed test.

- Take a sample(s) and calculate the relevant parameters: sample mean, standard deviation, or proportion. Using the formula for the test statistic from above in step 2, now calculate the test statistic for this particular case using the parameters you have just calculated.
- Compare the calculated test statistic and the critical value. Marking these on the graph will give a good visual picture of the situation. There are now only two situations:
 - a. The test statistic is in the tail: Cannot Accept the null, the probability that this sample mean (proportion) came from the hypothesized distribution is too small to believe that it is the real home of these sample data.
 - b. The test statistic is not in the tail: Cannot Reject the null, the sample data are compatible with the hypothesized population parameter.
- Reach a conclusion. It is best to articulate the conclusion two different ways. First a formal statistical conclusion such as “With a 5 % level of significance we cannot accept the null hypotheses that the population mean is equal to XX (units of measurement)”. The second statement of the conclusion is less formal and states the action, or lack of action, required. If the formal conclusion was that above, then the informal one might be, “The machine is broken and we need to shut it down and call for repairs”.

All hypotheses tested will go through this same process. The only changes are the relevant formulas and those are determined by the hypothesis required to answer the original question.

This page titled 9.4: Probability Distribution Needed for Hypothesis Testing is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.