

7.4: Distribution Needed for Hypothesis Testing

Earlier, we discussed sampling distributions. Particular distributions are associated with hypothesis testing. We will perform hypotheses tests of a population mean using either a normal distribution or a Student's t -distribution. (Remember, use a Student's t -distribution when the population standard deviation is unknown and the sample size is small, where small is considered to be less than 100 observations.) We perform tests of a population proportion using a normal distribution when we can assume that the distribution is normally distributed. We consider this to be true if the sample size is 100 or more. This is the same rule of thumb we used when developing the formula for the confidence interval for a population proportion.

Hypothesis Test for the Mean

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

$$z_{obs} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

The standardizing formula can not 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 7.4.1. 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_{obs} for “observed”. At times notation z_c for “calculated” is used (see figures below). Figure 7.4.1 and Figure 7.4.2 show this process.



Figure 7.4.1

In Figure 7.4.1 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 or z_c , called the observed or calculated z , that comes from solving the equation above. This observed z is nothing more than the number of standard errors that the sample mean is from the **hypothesized** mean. If the sample mean falls “too many” standard errors from the hypothesized mean we conclude that the **sample** mean could not have come from the distribution with 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 7.4.1 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 7.4.2 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 7.4.2

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

Decision rule: two-tailed test
If $ z_{obs} \leq z_{\frac{\alpha}{2}}$: then do not reject H_0
If $ z_{obs} > z_{\frac{\alpha}{2}}$: then reject H_0

Table 7.4.1

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_{obs} 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 must reject 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 of observing the sample mean \bar{x} if we **assume the null hypothesis is true**. This probability would, in other words, equal the tail probability that the t-test statistic is equal or greater than our observed test statistic z_{obs} from sample data (under the assumption that H_0 is true).

Here the notion of "likely" and "unlikely" is defined by the probability of drawing a sample with a mean \bar{x} 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, assuming H_0 is true. A large *p*-value calculated from the data indicates this is a likely outcome and 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 observed (calculated) test statistic, z_{obs} , and the critical value, z_α , and using the *p*-value can be seen in Figure 7.4.3.



Figure 7.4.3

The observed (calculated) value of the test statistic is z_c in this figure 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 must reject the null hypothesis. The associated \bar{x} is just unusually large (and thus too unlikely) 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 observed test statistic, z_{obs} . We then compare that to the α associated with our selected level of confidence. In Figure 7.4.3 we see that the *p*-value is less than α and therefore we must reject the null. We know that the *p*-value is less than α because the area under the *p*-value is smaller than $\alpha/2$. When our sample size is small (having less than 100 observations), we can find an approximate *p*-value from the Student's *t* table. In the relevant *df* row based on our sample size, we first locate the test statistic value(s) closest to our observed *t*-statistic, t_{obs} . We then find the one-tailed or two-tailed probabilities associated with those test statistic value(s) depending on the type of test we are conducting (see next section). In case of a two-tailed test, the one-tailed probabilities from the Student's *t* table need to be multiplied by two. The obtained *p*-value interval will give us the range that we will then compare to α and make our decision.

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.

Here is a systematic way to make a decision of whether you reject or do not 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 $p\text{-value} < \alpha$, reject H_0 . There is sufficient evidence to conclude that H_0 is an incorrect belief and that the **alternative hypothesis**, H_a , may be correct.
- If $p\text{-value} \geq \alpha$, cannot reject H_0 . There is not sufficient evidence to conclude that the alternative hypothesis, H_a , may be correct. In this case the status quo - the baseline set by the null hypothesis - 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 ; and we must remember that the null is only a baseline comparison point, set up for testing our alternative or research hypothesis, H_a .

One and Two-tailed Tests

The discussion of Figure 7.4.1 - Figure 7.4.3 was based on the null and alternative hypothesis presented in Figure 7.4.1. 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 an interest in whether the sample differs in a certain **direction** from (for example, is higher than) 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. 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 significance 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.

Figure 7.4.4 shows the two possible cases and the form of the null and alternative hypothesis that give rise to them, where μ_0 is the hypothesized value of the population mean.



Figure 7.4.4

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 100, 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 100, 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 observed 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 100 or more, we simply substitute s for σ and use the normal distribution.

Table 7.4.2 summarizes these rules.

Sample size	Test statistic
< 100 (σ unknown)	$t_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
< 100 (σ known)	$z_{obs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
> 100 (σ unknown)	$z_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
> 100 (σ known)	$z_{obs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

Table 7.4.2 Test Statistics for Test of Means, Varying Sample Size, Population Standard Deviation Known or Unknown

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 90, 95, 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:
 1. The test statistic is in the tail: Reject 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.
 2. 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 must reject the null hypothesis 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 [7.4: Distribution Needed for Hypothesis Testing](#) is shared under a [CC BY](#) license and was authored, remixed, and/or curated by .