Web Activity: Distribution of the P-values Under the Null Hypothesis of a One-Sample Z-test

Prerequisites:

- Be able to plot a histogram, either by hand or using a computer software.
- Know that a one-sample, two-sided z-test uses a random sample to determine whether the population mean is different from a specified value.
- Understand the terms *significance level* and Type I error in the context of a z-test, as well as their relationship with Type II errors.
- Be able to compute P-values of a z-test and explain how to use them to reject (or fail to reject) the null hypothesis at a given significance level.
- Have working knowledge of the Normal distribution and its two parameters, the mean and standard deviation, as well as the Uniform distribution in the interval [0, 1].

Learning Outcomes:

- Identify the probability distribution of the P-value of a one-sample, two-sided z-test when the null hypothesis is true.
- Explain in simple terms why a P-value follows a particular distribution under the null hypothesis.
- Recognize that the P-value of a hypothesis test is a random variable.

In this activity, we will develop an intuition about the distribution of the P-value under the null hypothesis of a two-sided, one-sample z-test.

> Recall that the P-value of a z-test is calculated as the probability of observing data more extreme than what was actually observed, assuming that the null hypothesis is true.

If you want to test whether the mean of a Normally-distributed population is equal to some value μ_0 or different from it, the test statistic is related to the sample mean $\bar{x}$. If the null hypothesis is true and your observations are independent, the sample mean follows a $N(\mu_0, \sigma^2/n)$ distribution, where μ_0 is the null population mean, σ is the population standard deviation, and n the sample size.

The P-value is calculated as the probability of exceeding the observed value of the sample mean. We reject the null hypothesis if the P-value is smaller than the significance of the test.

If μ_0 and σ are fixed, the P-value is uniquely determined by the sample mean, which in turn depends on the sample. As such, the P-value has a probability distribution. You will use two online interactive resources to understand this distribution, when the null hypothesis is true.

The first resource can be accessed at

`http://digitalfirst.bfwpub.com/stats_applet/stats_applet_14_signif.html`

This resource allows you to specify the values of the null population mean μ_0, the population standard deviation, the significance level, and the sample size. You can also specify the true value of the population mean, which can equal μ_0 or not. For this setting, you can simulate a sample from the true population and the resource calculates the P-value. Alternatively, if you have your own data, you can input the observed value of the sample mean and the resource calculates the P-value.

The second resource can be accessed at

`https://www.statcrunch.com/applets/type3&htmean`

This resource allows you to specify the same settings as the first resource, but it also allows you to generate multiple samples from the true population distribution, calculate the P-value for each sample, and create a histogram of the resulting P-values.

1 Recap of hypothesis testing

For this part of the activity, you will use the first resource. We want to test the null hypothesis that $\mu = 0$ against $\mu \neq 0$ at a significance level of $\alpha = 0.05$. The population standard deviation is known to be $\sigma = 2$, and you will obtain a sample of size $n = 10$. Use the controls on the left side of the resource to reflect this and check the option that says *"The truth about the population is $\mu = 0$."*

1) For these settings, what is the distribution of the sample mean?
2) What is the interpretation of the P-value that you observe in the resource? Try to use as few technical terms as possible.
3) At the current significance level, would you reject the null hypothesis based on the P-value that is displayed in the resource? Why or why not?
4) In general, if you are testing at a significance level of α, where α is a number between 0 and 1, what range of P-values would lead you to reject the null hypothesis?
5) Comparing the significance level and the probability of committing a Type I error, these two quantities are ___________.
   (*Fill in the blank with one of equal, different, or not comparable.*)

2 Understanding the P-value using the first resource

For this part of the activity, you will use the first resource. We want to test the null hypothesis that $\mu = 0$ against $\mu \neq 0$ at a significance level of $\alpha = 0.05$. The population standard deviation is known to be $\sigma = 2$, and you will obtain a sample of size $n = 10$. Use the controls on the left side of the resource to reflect this and check the option that says *"The truth about the population is $\mu = 0$."*

The Normal plot corresponds to the distribution of the sample mean, that is, a $N(\mu_0, \sigma^2/n)$ distribution. The resource also automatically generates a sample from the true population distribution every time you update any setting.

1) What do you think will happen to the location of the plot if you want to test the null hypothesis that $\mu = 2$ instead of $\mu = 0$? Don't use the resource yet—just think about this.
2) Modify the settings in the app so that we test the null hypothesis that $\mu = 2$. What happened to the location of the plot? Does this match what you expected to see in question 1)?
   (*Hint: observe the x-axis.*)

3) Modify the settings in the app so that we test the null hypothesis that $\mu = 0$ again. How is the significance level $\alpha = 0.05$ represented in the plot?

4) What do you think will happen to the spread of the plot if the true value of the standard deviation is 5? Don't use the resource—just think about this.

5) Modify the settings in the app so that $\sigma = 5$. What happened to the spread of the plot? Does this match what you expected to see in question 4)?
(*Hint: observe the x-axis.*)

6) Set $\sigma = 2$ again. What do you think will happen to the spread of the plot if the sample size were 20 instead of 10? Don't use the resource—just think about this.

7) Modify the settings in the app so that $n = 20$. What happened to the spread of the plot? Does this match what you expected to see in question 6)?

8) Set $n = 10$ again. How is the current sample represented in the plot? How is the sample mean represented in the plot?[1]

9) Check the option that says *"I have data, and the observed $\bar{x} = 0$."*. This option will allow you to set the sample mean. What are the sample mean and P-value for this sample? Is the test significant at the current level of 0.05?
(*Hint: observe the legend printed below the x-axis.*)

10) If you observed data with a sample mean of $\bar{x} = 1.4$, do you think the P-value would increase, decrease, or stay the same? Don't use the resource—just think about this.

11) Set the sample mean $\bar{x} = 1.4$ in the left tab. What is the new value of the P-value? Does this match what you expected to see in question 10)?

12) Is the test still significant at the current level of 0.05? Why do you think this is or is not so?

13) Did the location and/or spread of the plot change when changing from $\bar{x} = 0$ in question 9) to $\bar{x} = 1.4$ in question 11)? Explain why do you think they did or did not.

---

[1] CAUTION: since the sample is simulated, it will differ from a classmate's, or even from what you might see if you repeated the task. Try this! Click on the NEW SAMPLE button a couple of times and keep track of how the plot is modified each time a new sample is simulated.

3 P-value data collection

For this part of the activity, you will use the first resource. We want to test the null hypothesis that $\mu = 0$ against $\mu \neq 0$ at a significance level of $\alpha = 0.05$. The population standard deviation is known to be $\sigma = 2$, and you will obtain a sample of size $n = 10$. Use the controls on the left side of the resource to reflect this and check the option that says *"The truth about the population is $\mu = 0$."*

1) For these settings, the null hypothesis is ______.
   (*Fill in the blank with one of true or false.*)

2) Every time you click on the NEW SAMPLE button, a new sample is drawn from the true population distribution and the corresponding P-value is calculated. Generate at least 20 different samples. Use the table located at the end of this activity to write down the P-value of each.
   (*If you are working with teammates, you can each generate samples individually and then write the P-values in the table below.*)

3) Looking just at the list of P-values, which of the following distributions do you think best describes the distribution of the P-values? Are any of these distributions easily eliminated? Explain.

   (a) Normal.

   (b) Uniform.

   (c) Binomial.100

   (d) t.

4) Recall from part 1, question 4), that you reject the null hypothesis when the P-value is smaller than the significance level. From the P-values you obtained in question 2), what proportion led to a rejection of the null hypothesis?

5) Based on question 1), rejecting the null hypothesis in this scenario corresponds to ________.
   (*Fill in the blank with one of a Type I error or a Type II error.*)

6) Plot a histogram of the P-values you obtained in question 2). Use ten bins of equal width.
   (*If you know how to use a statistical programming language, such as R, you can save the P-values from question 2) in there and use the software to plot the histogram.*)

7) What distribution from question 3) does it resemble the most?

4 Distribution of the P-values

For this part of the activity, you will use the second resource. This resource allows you to simultaneously generate many samples like the samples you generated one by one in the previous part of this activity. It also plots the histogram of the P-values for you.

We want to test the null hypothesis that $\mu = 0$ against $\mu \neq 0$ at a significance level of $\alpha = 0.05$. The population standard deviation is known to be $\sigma = 2$, and you will obtain a sample of size $n = 10$. Assume that the true population mean is equal to 0. Use the controls on the uppermost part of the resource to reflect this and then click on the UPDATE APPLET button.

1) For these settings, the null hypothesis is ______.
(*Fill in the blank with one of true or false.*)

2) What is the significance level for this test?

3) On the resource, click on the P-VALUE button and press the 5 TESTS button. Each bar in the histogram corresponds to ____________________.
(*Fill in the blank with one of the P-value of a test or the sample mean of a sample.*)

4) Click on the 5 TESTS button until a red bar appears. What range of P-values does this bar represent? Why do you think it is colored in red, in relation to the hypothesis test?

5) Continue pressing the 5 TESTS button until you generate at least 50 samples. The distribution approximated by this histogram is ________ as/than the distribution being approximated in the histogram you constructed in part 3, question 6). Why?[2]
(*Fill in the blank with one of the same or different.*)

6) How many tests from those you generated were rejected at the 0.05 level? What proportion of tests were rejected at the 0.05 level?

7) Now click on the RESET button, and then press the 1000 TESTS button 10 times to generate 10,000 samples. Using the fact that the histogram has 20 bins, approximately what proportion of P-values is below the significance level?

8) If you were to test the same null hypothesis at a significance level of 0.1, would the proportion of P-values below the new significance level increase, decrease, or stay the same when compared to question 7)? Don't use the resource yet—just think about this.

[2] CAUTION: both histograms will differ because they are constructed with simulated data. The question is asking you to compare the distributions being approximated in each histogram rather than the histograms themselves.

9) Suppose that you test at a significance level of 0.1. Modify the appropriate setting in the resource to reflect this. What is, approximately, the proportion of P-values below the new significance level? Does this match what you expected to see in question 8)?

10) Repeat question 9) with the following significance levels: 0.2, 0.5, 0.9. For each significance level, write down the approximate proportion of P-values below the significance level.

11) Based on questions 7) through 10), what proportion of P-values is below α, where α is a significance value between 0 and 1?

12) Based on question 11), what do you think is the probability of the P-value being smaller than α if the null hypothesis is true?

13) Recall from part 1, question 5), that the probability of committing a Type I error is equal to the significance level α. Based on question 12), the probability of committing a Type I error, the significance level α, and the probability of observing a P-value below α are ________.
(*Fill in the blank with one of* *equal*, *different*, *or* *not comparable*.)

14) Use your answer from questions 12) and 13) to explain, using as few technical terms as possible, why the distribution of P-values is Uniform in the interval [0, 1].
(*Hint: the Uniform distribution in [0, 1] is the only one satisfying* $P(X \leq x) = x$ *for every number* x *between 0 and 1.*)

No.	P-value
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	