# Learn By Doing Activity

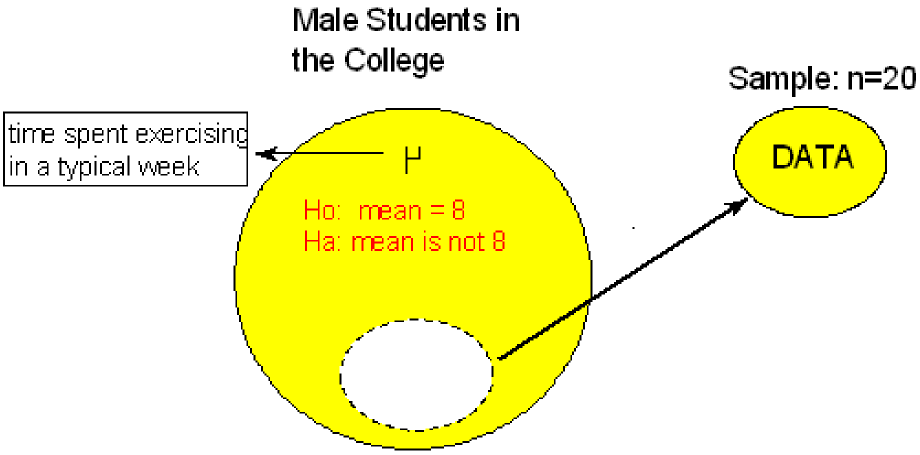
## Scenario: Exercise Habits of Male College Students

The purpose of this activity is to discuss how in some cases exploratory data analysis can help you determine whether the conditions that allow us to use the z-test for the population mean (μ) are met.

**Background:** In the Exploratory Data Analysis unit, we stressed that in general, it is always a good idea to **look at your data** (if the actual data are given). Moreover, related to our discussion now, looking at the data can be very helpful when trying to determine whether you can reliably use the test. In both of our leading examples, the data summaries (sample size, sample mean) were given rather than the raw data, but in practice, you are often working with the raw data. In example 1, we were told the SAT-M scores vary normally in the population, so even though the sample size ($n = 4$) was quite small, we could proceed with the test. In example 2, the sample size was large enough ($n = 100$) for us to proceed with the test even though we do not know whether the concentration level varies normally.

Now imagine the following situation: A health educator at a small college wants to determine whether the exercise habits of male students in the college are similar to the exercise habits of male college students in general. The educator chooses a random sample of 20 male students and records the time they spend exercising in a typical week. Do the data provide evidence that the mean time male students in the college spend exercising in a typical week differs from the mean time for male college students in general (which is 8 hours)?

**Comment:** Whether σ is known or not is really not relevant to this activity.




Here is a situation in which we do not have any information about whether the variable of interest, "time" (time spent exercising in a typical week) varies normally or not, **and** the sample size (n = 20) is not really large enough for us to be certain that the Central Limit Theorem applies. Recall from our discussion on the Central Limit Theorem that unless the distribution of "time" is extremely skewed and/or has extreme outliers, a sample of size 20 should be fine. However, how can we be sure that is, indeed, the case?

- If only the data summaries are given, there is really not a lot that can be done. You can say something like: "I'll proceed with the test assuming that the distribution of the variable "time" is not extremely skewed and does not have extreme outliers."
- If the actual data are given, you can make a more informed decision by looking at the data using a histogram. Even though the histogram of a sample of size 20 will not paint the exact picture of how the variable is distributed in the population, it could give a rough idea.

---

There are 4 different samples of size 20 in the data set. We created a histogram for each sample and these are shown below.

---

## Learn By Doing (1/1 point)

Looking at the four histograms above, comment on whether you think it would be safe to proceed with the test had those been the actual data in the problem above.

**Your Answer:**

It wouldn't be, because the trends were all different

1 = roughly normal?
2 = roughly normal? There's one outlier though
3 = roughly normal? Now there are two outliers on the left
4 = Definitely skewed to the right so I wouldn't push for doing the study

**Our Answer:**

Sample 1—The histogram displays a roughly normal shape. For a sample of size 20, the shape is definitely normal enough for us to assume that the variable varies normally in the population and therefore it is safe to proceed with the test. Sample 2—The histogram displays a distribution that is slightly skewed and does not have any outliers. The histogram, therefore, does not give us any reason to be concerned that for a sample of size 20 the Central Limit Theorem will not kick in. We can therefore proceed with the test. Sample 3—The distribution does not have any "special" shape, and has one small outlier which is not very extreme (although it is arguable whether you would classify it as an outlier). Again, the histogram does not give us any reason to be concerned that for a sample of size 20 the Central Limit Theorem will not kick in. We can therefore proceed with the test. Sample 4—The distribution is extremely skewed to the right, and has one pretty extreme high outlier. Based on this histogram, we should be cautious about proceeding with the test, because assuming that this histogram "paints" at least a rough picture of how the variable varies in the population, a sample of size 20 might not be large enough for the Central Limit Theorem to kick in and ensure that x-bar has a normal distribution.

Resubmit Reset

## Comment

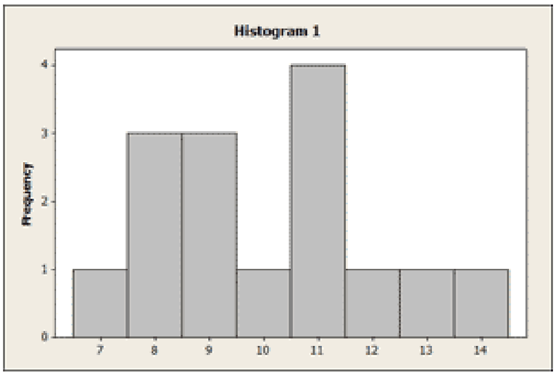
It is always a good idea to look at the data and get a sense of their pattern regardless of whether you actually need to do it in order to assess whether the conditions are met.

This idea of looking at the data is not only relevant to the z-test, but to tests in general. In particular, we'll see that in the case where σ is unknown (which we'll discuss next) the conditions that allow us to safely use the test are the same as the conditions in this case, so the ideas of this activity directly apply to that case as well. Also, as you'll see, in the next section—inference for relationships—doing exploratory data analysis before inference will be an integral part of the process.

## Did I Get This

1/1 point (graded)

Use the following histogram to answer the question below.




The sample size is 25. Using the histogram, should we proceed with a hypothesis test for the population mean?

- Yes, the sample size is large enough.

- Yes, although the sample size < 30, there are no outliers.
- Yes, although the sample size < 30, the distribution is not very far from normal in shape, with no outliers. ✔
- No, the sample size is not large enough.
- No, the sample size is < 30 and there are outliers.

**Answer**
Correct:
Even though the sample size < 30, we can proceed if certain conditions are met such as the shape is nearly normal and there are no outliers.

Submit

## Did I Get This

1/1 point (graded)
Normal body temperature for healthy, at-rest human beings has always been said to be 98.6°F. A doctor has seen a lot of patients who had a lower or higher body temperature when they were not ill. He has read research that says it is actually lower. So, he collected 50 randomly selected temperatures that had a mean of 98.4°F. The standard deviation is known to be 0.35°F.

$H_o$: $\mu = 98.6$

$H_a$: $\mu \neq 98.6$

Calculate the test statistic for this sample.

- -0.34
- -0.57
- -2.39
- -4.04 ✔
- 4.04

**Answer**

Correct: The test statistic is (98.4 - 98.6) / (0.35 / sqrt(50)).

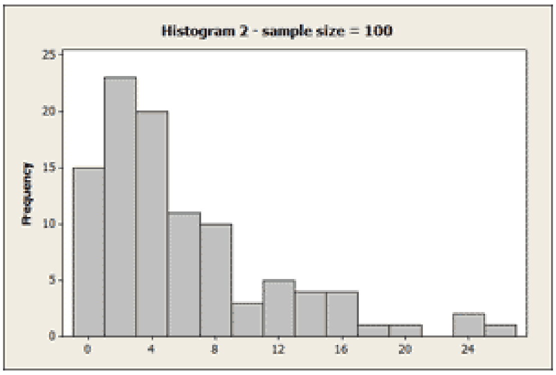
Submit

Each histogram in the questions below represents a random sample. We do not know if the variable is distributed normally in the population, but we want to be reasonably sure that the distribution of sample means will be normal so that we can use the z-test for testing claims about the population mean.

For each histogram, choose the option that best describes how to proceed with a hypothesis test for a population mean.

## Did I Get This

1/1 point (graded)



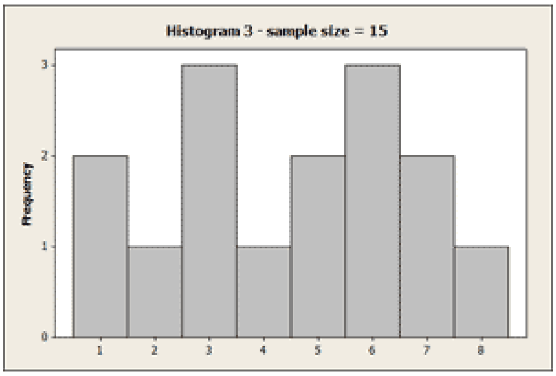

- (selected) Conditions are met; it is safe to proceed with the z-test. ✔
- Conditions may not be met; conduct the z-test but include a disclaimer with the results.
- Conditions are not met; do not use the z-test.

**Answer**
Correct:
The sample size is greater than 30, so sample means will be normally distributed. Therefore, the z-test can be used.

Submit

## Did I Get This

1/1 point (graded)



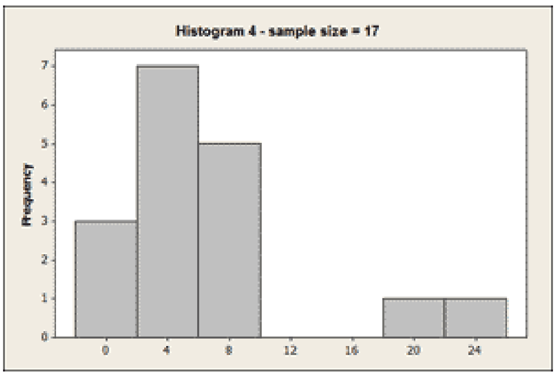

- Conditions are met; it is safe to proceed with the z-test.
- Conditions may not be met; conduct the z-test, but include a disclaimer with the results. ✔
- Conditions are not met; do not use the z-test.

**Answer**
Correct:
This is a borderline case. The sample size is only 15, so we can't be sure that the Central Limit Theorem applies. However, the sample is not heavily skewed and there are no outliers, so we will assume that this sample could have come from a population in which the variable is normally distributed. We will use the z-test with caution.

Submit

## Did I Get This

1/1 point (graded)



- ○ Conditions are met; it is safe to proceed with the z-test.
- ○ Conditions may not be met; conduct the z-test but include a disclaimer with the results.
- ◉ Conditions are not met; do not use the z-test. ✔

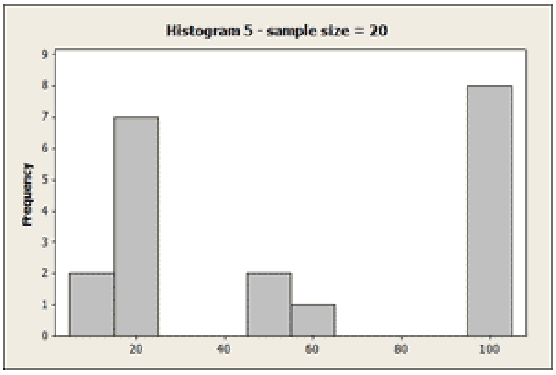
**Answer**
Correct:
The sample size is only 17, and it contains outliers. This suggests that the variable in the population might be skewed. So the Central Limit Theorem does not apply. We cannot assume that the sample means are normally distributed.

Submit

---

## Did I Get This

1/1 point (graded)




- ○ Conditions are met; it is safe to proceed with the z-test.
- ○ Conditions may not be met; conduct the z-test but include a disclaimer with the results.
- ◉ Conditions are not met; do not use the z-test. ✔

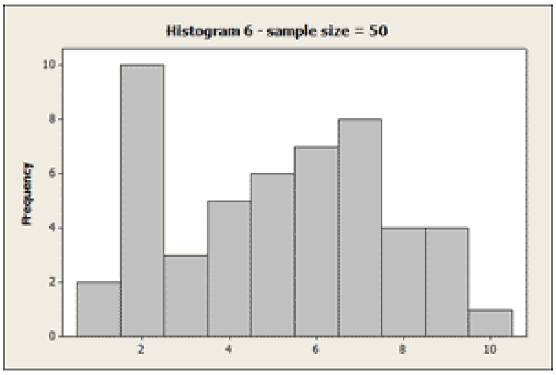
**Answer**

Correct:
The sample size is only 20 and the data are not normally distributed in the sample. This suggests that the variable in the population is not normally distributed. So the Central Limit Theorem does not apply. We cannot assume that the sample means are normally distributed.

Submit

## Did I Get This

1/1 point (graded)




- (selected) Conditions are met; it is safe to proceed with the z-test. ✔
- Conditions may not be met; conduct the z-test but include a disclaimer with results.
- Conditions are not met; do not use the z-test.

**Answer**
Correct:
The sample size is greater than 30, so sample means will be normally distributed. Therefore, the z-test can be used.

Submit