Course > Inference: Hypothesis Testing for the Population Mean > z-test for the Population Mean > Statistics Package Exercise: Examining the Conditions for the z-test for the Population Mean (μ)

Bookmark this page

Statistics Package Exercise: Examining the Conditions for the z-test for the Population Mean (μ)

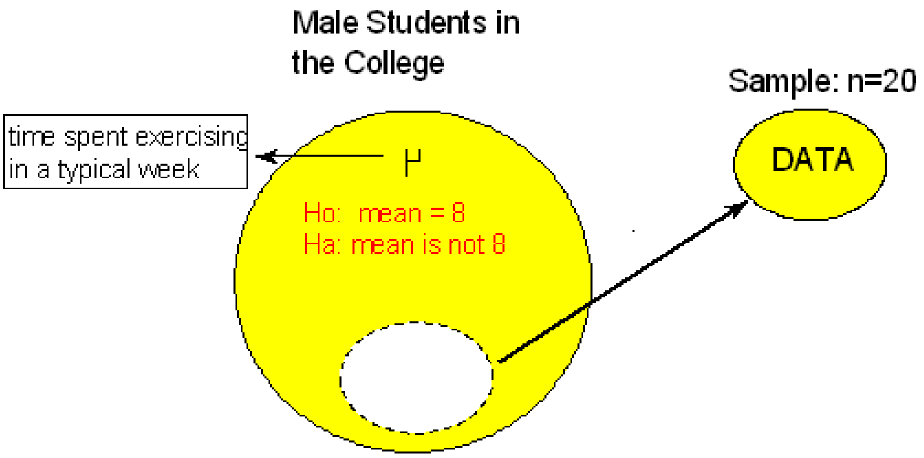
Learning Objective: Carry out hypothesis testing for the population proportion and mean (when appropriate), and draw conclusions in context.

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.


Male Students in the College
time spent exercising in a typical week
μ
Ho: mean = 8
Ha: mean is not 8
Sample: n=20
DATA

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.

- R StatCrunch TI Calculator Minitab Excel

R Instructions
To open R with the data set preloaded, right-click here and choose "Save Target As" to download the file to your computer. Then find the downloaded file and double-click it to open it in R.

The data have been loaded into the data frame

```
time
```

. Enter the command

```
time
```

to see the data. There are 4 columns (named

```
time1
```

through

```
time4
```

) representing 4 different samples of size 20 in the data frame.

To create a histogram of the column

```
time1
```

with R, enter the command:

-
  ```
  hist(time$time1,xlab="Time Exercising Per Week",
  main="Sample 1")
  ```

Now change

```
time1
```

to

```
time2
```

,

```
time3
```

, and

```
time4
```

in the code above to see the histogram for each time column.

Note: You can modify the x-label and title as you choose.

Learn By Doing (1/1 point)

For each sample, create a histogram and 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:

1 is perfectly fine!
2 is also fine
3 is okay pa, no outliers, still semblance of normality
4 is definitely skewed

Our Answer:

Time 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. Time 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. Time 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. Time 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 $\bar{x}$ has a normal distribution.

Resubmit Reset

Comments:

1. 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 module—inference for relationships—doing exploratory data analysis before inference will be an integral part of the process.