Hypothesis Testing for the Population Mean: t score

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

Recall that we were discussing the situation of testing for a mean, in the case when σ is unknown. We've seen previously that when σ is known, the test statistic is $z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ (note the σ (σ) in the formula), which follows a normal distribution. But when σ is **unknown**, the test statistic in the test for a mean becomes $t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$ (note the use of "s" in the formula, in place of the unknown σ). **Here** is where the t-distribution arises in the context of a test for a mean, because $t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$ (with "s" in the formula in place of the unknown σ) follows a t distribution.

Notice the only difference between the formula for the Z statistic and the formula for the t statistic: In the formula for the Z statistic, σ (the standard deviation of the population) must be known; whereas, when σ isn't known, then "s" (the standard deviation of the sample data) is used in place of the unknown σ. That's the change that causes the statistic to be a t statistic.

Why would this single change (using "s" in place of "σ") result in a sampling distribution that is the t distribution instead of the standard normal (Z) distribution? Remember that the t distribution is more appropriate in cases where there is more variability. So why is there more variability when s is used in place of the unknown σ?

Well, remember that σ (σ) is a parameter (it's the standard deviation of the population), whose value therefore never changes. Whereas, s (the standard deviation of the sample data) varies from sample to sample, and therefore it's another source of variation. So, using s in place of σ causes the sampling distribution to be the t distribution because of that extra source of variation:

In the formula $z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$, the only source of variation is the sampling variability of the sample mean $\overline{X}$ (none of the other terms in that formula vary randomly in a given study);

Whereas in the formula $t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$, there are **two** sources of variation: One source is the sampling variability of the sample mean $\overline{X}$; The **other** source is the sampling variability of sample standard deviation s.

So, in a test for a mean, if σ isn't known, then s is used in place of the unknown σ and that results in the test statistic being a t score.

The t score, in the context of a test for a mean, is summarized by the following figure:

In fact, the t score that arises in the context of a test for a mean is a t score with (n – 1) degrees of freedom. Recall that each t distribution is indexed according to "degrees of freedom." Notice that, in the context of a test for a mean, the degrees of freedom depend on the sample size in the study. Remember that we said that higher degrees of freedom indicate that the t distribution is closer to normal. So in the context of a test for the mean, the **larger the sample size**, the higher the degrees of freedom, and **the closer the t distribution is to a normal z distribution**. This is summarized with the notation near the bottom on the following image:

As a result, in the context of a test for a mean, the effect of the t distribution is **most important** for a study with a **relatively small sample size**.

We are now done introducing the t distribution. What are implications of all of this?

1. The null distribution of our t-test statistic: $t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$ is the t distribution with (n-1) d.f. In other words, when H_o is true (i.e., when $\mu = \mu_0$), our test statistic has a t distribution with (n-1) d.f., and this is the distribution under which we find p-values.

2. For a large sample size (n), the null distribution of the test statistic is approximately Z, so whether we use t(n - 1) or Z to calculate the p-values should not make a big difference. Here is another practical way to look at this point. If we have a large n, our sample has more information about the population. Therefore, we can expect the sample standard deviation s to be close enough to the population standard deviation, σ, so that for practical purposes we can use s as the known σ, and we're back to the z-test.
