18.0 More Significance Tests

- Answer Questions
- Review Confidence Intervals
- Review Significance Tests
- The t-test
- Paired Difference Tests

18.1 Review of Confidence Intervals

Suppose you set a 95% confidence interval on the average ppm of mercury in catfish from the Wacamaw river. You get the interval $[1.11, 1.44]$.

There are two wrong ways to read this:

- 95% of the catfish in the Wacamaw river have mercury levels between 1.11 and 1.44.
- The probability that the average mercury level in Wacamaw catsfish lies between 1.11 and 1.44 is .95.

The first answer is wrong because it confuses individual values (a catfish) with the population mean for all catfish.

The second answer is wrong in a more subtle way. Once the CI is calculated, it is no longer random. The true value is either between 1.11 and 1.44 or not—the probability is zero or one.

The correct interpretation is that 95% of confidence intervals constructed by the procedure used in this problem will contain the true mean. We don't know whether this is one of the ones that does.

But is this confidence interval really the one that you want? It is a two-sided interval, and perhaps all you really care about is an upper bound on the average ppm of mercury.

That is, you want to find U such that you are 95% confident that the true average mercury contamination is below U. Showing that the level is above a lower bound does not help you decide whether to eat catfish (unless, of course, the problem is so bad that the lower bound exceeds health guidelines).

Suppose you have a sample of 100 catfish, and find that the average ppm of mercury is 1.2, and the *sd* in the sample is .3. Suppose the EPA guideline says that with 95% confidence, the average ppm should be below 1.3.

From the handout, the formula for the upper bound is:

$$U = (pe) + (se)(cv_{1-C}) = \bar{X} + \frac{SD}{\sqrt{n}} z_{.95}.$$

In this case, the pe is 1.2, the sd is .3, and $z_{.95}$ is the value in the standard normal table that has area 95% under the curve and to the left. Thus $z_{.95} = 1.95$.

Putting all this together,

$$U = 1.2 + \frac{.3}{10} 1.95 = 1.2585.$$

So you can be 95% certain that the EPA guidelines are not exceeded, and thus you can eat the catfish.

18.2 Review of Significance Tests

Recall that a significance test is a way to decide whether the data strongly support point of view or another.

A significance test requires:

- a null and alternative hypothesis
- a test statistic
- a significance probability (P-value).

The handout gives an overview of many tests.

Example 1: Suppose it is known that 30% of students fall asleep in class. I want to show that my class is better. So I count; out of 120 students, only 35 fall asleep. Is this evidence that my teaching is more lively?

First, we must find the null and alternative hypotheses. How do we pick these?

.

For the null and alternative hypotheses below, we must find the test statistic.

H_0: $p \geq .3$

$H_A : p < .3$

From the handout, which test statistic should we use?

.

$$ts = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{35}{120} - .3}{\sqrt{\frac{.3*.7}{120}}} = -.1992.$$

Finally, we need to find the significance probability, or P-value. From the handout, this is:

$$\mathbf{P}[z < ts] = \mathbf{P}[z < -.1992] = .42075.$$

This result is not very unlikely. Just by chance, 42% of the time a typical class will give a result that is similar to what we found.

Example 2a: Your date wants to toss a coin to decide who pays for dinner. Before you agree, you toss the coin 20 times and want to know whether there is evidence that the coin is unfair.

Example 2b: You want to show that a new piece of legislation has increased the proportion of welfare recipients in Durham.

Example 2c: You want to show that a new drug decreases the proportion of children born with spina bifuda.

Whis is the p_0 in each of these examples? What is the alternative hypothesis?

For all three of these situations the test statistic is the same:

$$ts = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

Note that this looks very much like the form of the CLT for a proportion. How is it different?

The P-value depends upon the hypothesis pair:

- 2.a has P-value $= \mathrm{P}[z > |ts|]$
- 2.b has P-value $= \mathrm{P}[z > ts]$
- 2.c has P-value $= \mathrm{P}[z < ts]$

18.3 The t-Test

Our previous tests for the mean and the confidence intervals for the mean assumed that the sd of the population was known, or that we could estimate it accurately from a sufficiently large sample.

But in practice one often does not know the true sd of the population and n is too small ensure accurate estimation from the sample. (How small is "too small" is a matter of judgement; in this class we shall say that a sample of size 26 or less is too small.)

When we don't know the sd of the population, all we can do is to estimate it by the sd of the sample. But when n is too small, we also need to account for the additional uncertainty introduced by this estimation.

In particular, when n is too small, the distribution of the test statistic does not follow a $N(0, 1)$ distribution. It follows a t-distribution, which was derived by William Gossett, an executive at the Guinness Brewery.

As n gets large, the uncertainty introduced by estimating the population sd becomes negligible and so the t-distribution converges (quite quickly) to the standard normal distribution.

A t-distribution is indexed by $n-1$, which is called the degrees of freedom.

A t-distribution is centered at 0, and is symmetric about 0.

A t-distribution is fatter in the tails than the $N(0,1)$ distribution; this reflects the fact that uncertainy in estimating the sd means that you are more likely to get rather large or small values for your test statistic.

As $n \to \infty$, the distribution of a t random variable with $n-1$ degrees of freedom approaches that of a standard normal random variable.

The book gives t values only for sample sizes up to 26. That is why our class uses that cut-off. In principle, one can derive the t-distribution for any sample size.

And the values in the book's table are only for certain conventional values of significance: .25, .1, .05, .025, and .01. The book did not want to have a separate page for the t-distribution for $1, 2, \ldots, 25$ degrees of freedom.

Find the following:

- The t-value with 4 degrees of freedom that has area .05 under the curve and to the right.
- The t-value with 8 degrees of freedom that has area .1 under the curve and to the left.
- The t-value with 20 degrees of freedom that has area .99 in the middle.

Example 3. You are a Dean at Yale and want to show that the mean Duke IQ is less than 125. You take a random sample of four Duke students and find

$$120,\ 115,\ 110,\ 130.$$

What are your null and alternative hypotheses?

.

H_0: The mean Duke IQ is 125 or greater.

H_A : The mean Duke IQ is less than 125.

In notation, this is just

H_0: $\mu \geq 125$

H_A : $\mu < 125$

In general, I prefer to see hypotheses written out in words—the notation is really just a mathematical shorthand for the ideas.

The test statistic for this problem is:

$$ts = \frac{\bar{X} - \mu_0}{\frac{sd}{\sqrt{n-1}}} = \frac{118.75 - 125}{\frac{7.395}{\sqrt{3}}} = -1.4639.$$

We calculate $\bar{X}$ and the sd from the sample in the usual way.

To find the significance probability, or the P-value, we compare this test statistic to a t-distribution with 3 (= 4-1) degrees of freedom.

Using the Chinese menu, we see the the significance probability is

$$\mathbf{P}[t_3 \leq ts] = \mathbf{P}[t_3 \leq -1.46] > \mathbf{P}[t_3 \leq -1.64] = .1.$$

This means that the significance probability for the test is bigger than .1. By most standards, there is no reason to reject the null hypothesis.

And the Yale Dean is disappointed.