4/19/2022

Confidence Intervals
Hypothesis Tests Intro

Last time, we talked about this idea of sampling distributions. As you sample, we get better estimates for the mean (or proportion) of the population when we take larger samples. Samples vary less than individual members of a population. Can use the properties of sampling distributions to estimate how big the difference is between our estimate and the true value.

We want to calculate a margin of error, that summarizes how good our estimate is.

What is the difference between a point estimate and a confidence interval? And why the confidence is better?

A point estimate is a single value (statistic) that is used to estimate the population parameter. That value is an estimate and so kinda by definition is wrong. Estimates are approximations, and we can think of them in some absolutist sense as necessarily being wrong. But, when we provide confidence intervals (values with margins of error), this gives some additional sense of HOW wrong they are, or how right they are.

You want to get close enough to the correct value so that your answer/estimate is meaningful and isn't a significant impediment to using that value as though it were the truth.

Confidence intervals in general are constructed in a consistent pattern.

Public version: Value +/- the margin of error
Math class version: (Value – margin of error , Value + margin of error)

Suppose and the mean is 2.4, and the standard error is 0.1
Margin of error is the z-score for the confidence level * standard error.
The most common confidence level is 95% ~ 2 (from the empirical rule, two standard deviations from the mean is 95% of the data). The exact value for the normal distribution is 1.96.

Margin of error for this example is approximately 1.96*0.1 = 0.196
Public version: 2.4 ± 0.2
Math version: $(2.2, 2.6)$

In English, we mean we are 95% confident that the true population mean is between 2.2 and 2.6.

I will ask you to interpret confidence interval in plain language on the exam.

Examples from statistics.
Means:
Confidence interval in terms the basic statistics:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

For proportions:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

α is the complement of the confidence level, so if the confidence level is 95%, then $\alpha = 5\%$, and $\frac{\alpha}{2} = 2.5\%$, this is the value we put into the Norm-Inv function to get the z-score.

Example for means.
Suppose a particular high school has 100 students take the ACT and they obtain a mean score of 22. We know the standard deviation of the test is 5.2. Does this give us good reason to think the score of the school is significantly higher than the national average? Construct a confidence interval to answer this question. (Use a 95% confidence interval.)

$$\bar{x} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 22 \pm 1.96 \times \frac{5.2}{\sqrt{100}} = 22 \pm 1.0192$$

$$(20.9808, 23.0192)$$

This interval means that we are 95% confident that the true mean of the school is between 20.98 and 23.02. If the population mean is 21, then this value for the school is significantly (statistically) different if it falls outside the confidence interval. If the confidence interval included this population value, then we conclude we don't have enough evidence to say the school is statistically better than average.

Example with proportions.
Two Candidates are running for office, Candidate A and Candidate B. In a recent poll of 450 residents, Candidate A received 54% of the vote, and Candidate B received 46% of the vote among likely voters. Can we establish, using a 95% confidence interval, that Candidate A is ahead in the polls right now?

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.54 \pm 1.96\sqrt{\frac{0.54(0.46)}{450}} = 0.54 \pm 0.046\ldots$$

$$(0.494, 0.586)$$

We are 95% confident that the true support for Candidate A is between 49.4% and 58.6%. So, we cannot be certain that A is ahead.

We want to be able to calculate these in Excel and from raw data.
See Excel for examples.

When to use the z-score versus when to use the t-distribution (student-T distribution)?

- When doing a proportion confidence interval—always use the z-score
- For means, check some conditions – if ALL apply, then use the z-score. If all don't apply, or you aren't sure, use the t-score (from T-distribution)
    - The distribution is known to be normal
    - The population standard deviation is known (σ)
    - A sample bigger than 40

If in doubt on a means problem, use the t-distribution. It is safer/more conservative.

Go to Excel for the t-distribution examples.
Works much the same way as the standard normal distribution (assumes a mean of 0 and a standard deviation of 1). But we have to specify the "degrees of freedom". And the degrees of freedom in this case is the sample size minus 1.

$$df = n - 1$$

Calculating Sample Size

Before you conduct a study, you want some idea of how accurate you need your analysis to be. What kind of margin of error do you want? We need/want to calculate the number of samples needed to obtain the margin of error we desire.

$$E = \frac{z\sigma}{\sqrt{n}}$$

$$E^2 = \frac{z^2\sigma^2}{n}$$

$$n = \frac{z^2\sigma^2}{E^2}$$

For means.
For proportions:

$$E = z\sqrt{\frac{p(1-p)}{n}}$$

$$E^2 = \frac{z^2 p(1-p)}{n}$$

$$n = \frac{z^2 p(1-p)}{E^2}$$

Think of these formulas a threshold: minimums. If you need to round, you must round n UP.

Suppose you want to be sure that the margin of error for a political poll is less than 3% with a 95% confidence level. What size sample is needed?
(If the problem does not specify an expected proportion, assume it is around 50%--the reason is because this is where the margin of error is the widest).

$$n = \frac{1.96^2(0.5)(0.5)}{(0.03)^2} = 1067.1111 \ldots$$

The smallest value of n is 1068.

Suppose that I want to estimate the height of women in the US, and with 95% confidence, the standard deviation is estimated to be about 3 inches. I would like a margin of error to be under 0.1 of an inch. What is the minimum sample size needed?

$$n = \frac{1.96^2(3)^2}{(0.1)^2} = 3457.44$$

So, the smallest value of n is 3458.

To be within 1 inch:

$$= \frac{1.96^2(3)^2}{(1)^2} = 34.5744$$

The minimum number would be 35.

Hypothesis Testing.
We are asking whether the data we have is sufficiently strong to establish (to conclude) whether or not our assumptions are correct or can be replaced.

The legal system essentially employs a hypothesis test in criminal cases.
Innocent until proven guilty:

$$H_0: the\ accused\ person\ is\ innocent$$
$$H_a: what\ we\ are\ trying\ to\ prove: they\ are\ guilty\ of\ some\ crime$$

H_0 is the notation for the null hypothesis, it's what we think about the world without considering the new evidence, it's our default position, our fallback.
H_a is the alternative hypothesis is the claim we want to prove. Our evidence needs to be sufficiently strong to prove that we aren't likely to have gotten this result by chance.

In the case of hypothesis testing in statistics, we set a threshold called α (alpha) that is a number, usually around 5%.

One type of error is to allow a guilty person to go free. (as failing to reject the null hypothesis even though it is wrong – evidence is weak). Called a Type 1 error, and the α is an expression of how likely we are to make this mistake.

Another type of error is to allow an innocent person to go to jail. (rejecting the null hypothesis when it is true). This is called a Type II error. The probability associated with this is called beta β, and $1-\beta$ is called the power of a test.

	H_0 is true	H_0 is false
Conclude H_0 is true (failing to reject the null)	GOOD	Type 1 Error
Conclude H_0 is false (reject the null)	Type II error	GOOD

How to express the null and alternative hypotheses in correct notation.

The null hypothesis and so it takes all the equal signs.

$$H_0: \mu = 0$$

Or

$$H_0: p = 0.5$$

The alternative takes the inequality or not-equal-to signs.

$$H_a: \mu > 0$$
$$H_a: \mu < 0$$
$$H_a: p \neq 0.5$$

The number in the statement is the SAME, but the sign (inequality) changes.
The null is based on previous knowledge about the world (assumptions or previous data). The sample statistics that we obtain from the never appear in a hypothesis.

Hypothesis tests can be one-tailed or two-tailed:
One-tailed tests have inequalities > or < in their alternatives (we only care about being less or greater, and we know which, not both).
Two-tailed tests could be either greater or less and are stated with ≠ sign. “different”