Goals for the Week

Week 7 Content

R Questions

Progress Check

STAT 140: Week Six Catch Up

10/08/2021

My hope is that after Week 7 (10/13-10/15), you will be able to:

After this week, you should be able to:

- Derive the sampling distribution for a sample proportion.
- Build and interpret a confidence interval for a proportion.
- Understand how changing different characteristics of a confidence interval affects the interval.

Probability Review

Question: Could we clarify using `pnorm()` vs `1 - pnorm()`?

By default in R, `pnorm(q, mean, sd)` gives the probability that the random variable X is *less than or equal to* `q`. Then, because the area under the normal curve sums to 1, we can use the Complement Rule to find the probability that the random variable X is *greater than* `q` with `1 - pnorm(q, mean, sd)`.

Expected Value and Variance 1

Question: I'm confused about the equation that was used to solve the expected value for the Baggage section in lab. It was `exp_baggage <- sum(c(0, 25, 60)*c(0.54, 0.34, 0.12))`.

There are a lot of different ways we can find the expected value. For a discrete variable, probably given in a one-way table, we use:

$$E(X) = \sum_{i=1}^{n} x_i P(x_i).$$

All of the formulas you learned for expected values in the different distributions (i.e., Bernoulli–p, Geometric–$\frac{1}{p}$, Binomial–np, Poisson–λ) were derived using this formula! But once we know the general formulas, it's easier to use those directly. For normal distributions, you would have to be given $E(X) = \mu$, or if using the CLT, you'd have to find it in the problem (also typically given to you).

Expected Value and Variance 2

Question: I am still having trouble with variance.

Same thing! The variance is given by

$$Var(X) = \sum_{i=1}^{n} (x_i - E(X))^2 P(x_i).$$

Again, this formula is used to calculate variances for the discrete distributions and you can use the "shortcuts". For the Normal distribution, you would have to be given $Var(X) = \sigma^2$.

Sampling Distributions 1

Question: How do we know when to use sampling error?

Excellent question! The sampling error is the standard error in a sampling distribution. Remember, sampling distributions describe the distribution of a statistic! So, if your question is asking about the probability of an average, you should be using sampling/standard error $\frac{\sigma}{\sqrt{n}}$.

Sampling Distributions 2

Question: When should we use a proportion distribution as opposed to a traditional sampling distribution?

What we learned last week is that according to the Central Limit Theorem, for a random sample of variables $X_1, X_2, ..., X_n$,

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- If X is a numerical variable, like dollars or pounds, etc., you would use μ and σ as they were given in the problem (or potentially calculated from something else, like a one-way table).
- If X is a yes or no variable, like approval or disapproval or hesitancy or not hesitancy, you can use $\mu = \hat{p}$ and $SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Prepare

Question: Why do we usually use 95% as a normal confidence level?

Question: How can we determine what confidence level we use in the question? Will it tell us, or do we just default to 95% confidence?

A 95% confidence level is just by convention. Most of the time in statistics or in other fields, practitioners will use 95%. We will talk about more options later–but for now, you can either use 95% all of the time or pick one of 90%, 95%, or 99% unless I state otherwise.

Check

Question: Does the sample proportion only fail to meet the conditions for a confidence interval if the sample size is too small or the rate is too small?

Yes.

- If n is too small, the conditions will fail regardless of what p is. Consider $n = 10$ and $p = 0.5$
  - $np = 10 \times 0.5 = 5, n(1-p) = 10 \times (1 - 0.5) = 5$
- If n is large, but p is too small, the conditions will also fail. Consider $n = 200$ and $ p = 0.96$.
  - $np = 200 \times 0.96 = 192, n(1-p) = 200 \times (1 - 0.96) = 8$

Calculate 1

Question: what are the properties of point estimates?

A point estimate is just a single value that estimates some parameter, while an interval estimate (like a confidence interval) is a whole range of possible values for a parameter. We use point estimates as the center of the confidence interval.

Question: What does quantile mean?

We learned about percentiles, which splits the distribution by returning a value with a specific area (or percent) below it. A quantile is the same thing, just another word.

Calculate 2

Question: How does the standard error affect the confidence interval?

Let's re-express a confidence interval as

$$(\hat{p} - MoE,\ \hat{p} + MoE).$$

The Margin of Error, or MoE, is

$$z^* \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

If either of these pieces gets bigger (z^* from using a higher confidence level, SE from using a smaller n), the entire interval will be wider.

Conclude

Question: Does a confidence interval represent the plausible range for sample values or for real values?

The real values! Remember, we don't know what the real value is, but given a sample, we can directly calculate what $\hat{p}$ is. There is no need to provide an interval for it, because we already have the exact value (say, $\hat{p} = \frac{72}{300} = 0.24$).

Question: Can the confidence interval be thought of as the error bound?

Yes! Many times in scientific journal articles, researchers will report error bounds–these are another way to describe confidence intervals.

More on Confidence Intervals

Question: Does the sample size affect the width of the interval?

Yes! Remember that

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

When you have a fraction with a large number in the denominator, the entire quantity will get smaller. Alternatively, you can play around with different values of $\hat{p}$ and n to convince yourself in R!

If the standard error is smaller, the margin of error is smaller, and the interval will be narrower!

Confidence Intervals in Assigments

Question: How do we interpret the phrasing of questions about confidence?

Every time I ask you to carry out a confidence interval, make sure you are using the four steps in your book.

1. Prepare
2. Check
3. Calculate
4. Conclude

R Questions

Question: Could you explain when we should use the `prop.test()` function and when we should just calculate the CI on our own?

You can always use `prop.test()`! I only have you do it manually once so you know what R is doing and where all the parts are coming from.

Question: Can `prop.test()` only be used in R? What about other computer languages?

`prop.test()` is the command in R, but there are definitely ways to do this in other languages. You would just have to look them up!

Progress Check

Question: Can you go over Question 3 in the progress check?

3. This confidence interval was built based off of a sample of 2,050 people. If the sample had consisted of 1,025 people, what would happen to the width of the interval?

We know that if the sample size increases, the margin of error is smaller and the interval is narrower. If the sample size decreases (like 1,025 to 2,050), the interval will get wider.

Question: Does n equal 2050 in question 4?

Yes (although even if $n = 1025$, the conditions would still be met).