Confidence Intervals

An expanded version of the notes from last week.

Confidence intervals

- A p-percent confidence interval is a recipe like

$$\overline{X} \; - \; z \cdot SE \;\text{ to }\; \overline{X} \; + \; z \cdot SE$$

  where the z comes from a fact like the 68–95–99.7 rule, or from a table. When you plug in:

  - $\overline{X}$ from the data;
  - $SE = SD/\sqrt{n}$;
  - z from the table

  you get two numbers between which you are confident the population mean lies.

- The SD "should be" for the population, but sometimes you have to estimate it from the data.

- The z value is from a table. When the Central Limit Theorem applies (averages and large samples!):

  - Use the 68-95-99.7 rule when you can
  - Use a Normal [or other] table otherwise.

Example.

The SAT is constructed so that scores have a national average of of 500 and a national standard deviation of 100. A random sample of 64 students from a recent entering class at Carnegie Mellon have an SAT verbal average of 555.

- A 95% confidence interval for SAT verbal score is

$$\begin{aligned} 555 - 2 \cdot 100/\sqrt{64} \quad &\text{to} \quad 555 + 2 \cdot 100/\sqrt{64} \\ 555 - 25 \quad &\text{to} \quad 555 + 25 \\ 530 \quad &\text{to} \quad 580 \end{aligned}$$

- Are these students like a random sample from the national average? Why or why not?
- A 99.99% confidence interval for SAT verbal score is

$$\begin{aligned} 555 - 3.89 \cdot 100/\sqrt{64} \quad &\text{to} \quad 555 + 3.89 \cdot 100/\sqrt{64} \\ 555 - 48.63 \quad &\text{to} \quad 555 + 48.63 \\ 506.37 \quad &\text{to} \quad 603.63 \end{aligned}$$

- Could we have made the same conclusion from a sample size of only 36?

$$\begin{aligned} 555 - 3.89 \cdot 100/\sqrt{36} \quad &\text{to} \quad 555 + 3.89 \cdot 100/\sqrt{36} \\ 555 - 64.83 \quad &\text{to} \quad 555 + 64.83 \\ 485.17 \quad &\text{to} \quad 614.83 \end{aligned}$$

Remember that the confidence is in the *recipe:* for 95% confidence:

- If you do this sample of 64 Carnegie Mellon students 100 times, you will get 100 different intervals

- 95 out of the 100 intervals will contain the true mean.

This example illustrates several features of confidence intervals:

1. To attain greater confidence, you have to widen the interval.

2. If you make the sample size smaller, you widen the interval.

3. Confidence intervals can be used to test hypotheses about data.

    - A hypothesis about a population gives a prediction about the sample mean in a sample from the population

    - Compute a confidence interval from the sample.

    - If the confidence interval makes the prediction look wrong, this undermines the hypothesis.

Variation I: Sample Size Calculations

We wish to conduct an SRS of Allegheny County residents to see if they like or dislike the design for remodeling the David L. Lawrence Convention Center. *How large a sample do we need to have a 95% confidence margin of error of* $\pm 4\%$*?*

- Margin of error $= 2 \cdot \sqrt{p(1-p)/n}$.

- don't know p, use $p = 0.5$:
$$\text{MOE} = 2 \cdot (0.5)/\sqrt{n} = 1/\sqrt{n}$$

- Solve for n:

$$\begin{aligned} 4\% = 0.04 &= 1/\sqrt{n} \\ \sqrt{n} &= 1/0.04 \\ n &= (1/0.04)^2 = 625 \end{aligned}$$

- Now do poll, discover $438/625 = 70\%$ like the expansion design. Then 95% CI is 66% to 74%

Variation II: The T-Distribution

A researcher gives a new sleeping drug to each of 36 patients who complain of difficulty sleeping, and discovers that on average these 36 patients increased their nightly sleep by 0.5 hours, with a standard deviation of 1 hour. Is this a “significant” gain?

- Back-of-the-envelope calculation:
  95% CI is $0.5 \pm 2 * 1/\sqrt{36}$, or 0.17 to 0.83.

  However, *this recipe will produce intervals that miss more than 95% of the time, since the SD is estimated from the data too.*

- An adjustment (widening) is needed, to get a “95%” recipe.

- The amount of adjustment depends on the sample size—with more than about 50 or 60 observations, there is no adjustment needed.

- Adjustment: t-distribution on $df = n - 1$ “degrees of freedom (df)”:
  95% CI is $0.5 \pm 2.03 * 1/\sqrt{36}$, or 0.16 to 0.84.

  (where “2” was replaced with $t = ...$” from the $df = 35$ t table).

Variation III: Two Samples

A better experimental design would have been

- give 36 subjects the sleeping pill: mean extra sleep 0.5h, SD=1h
- give 20 more subjects a sugar pill: mean extra sleep 0.25h,SD=.25h

- Back-of-the-envelope calculation:
  - in "drug" group, 95% CI is $0.5 \pm 2 * 1/\sqrt{36}$, or 0.17 to 0.83.
  - in "placebo" group, 95% CI is $0.25 \pm 2 * 0.25/\sqrt{20}$, or 0.14 to 0.36.

  Since the intervals *do* overlap we *don't* have evidence to contradict a "no change" hypothesis.

- If we ask Minitab to do this...

```
Two Sample T-Test and Confidence Interval

Two sample T for drug vs placebo
            N      Mean     StDev   SE Mean
drug       36      0.50      1.00      0.17
placebo    20     0.250     0.999      0.22

95% CI for mu drug - mu placebo: ( -0.31,  0
T-Test mu drug = mu placebo (vs not =): T= 0
```

The confidence interval is a 95% interval for the *difference* in true means between the two groups.

Hypothesis Testing: Some Vocabulary

(1) A Hypothesis in statistics refers to a statement about a feature of the population. Specifically, a hypothesis is a statement about the value(s) of a population parameter.

(2) The Null Hypothesis, (H_0), is a statement that the parameter is equal to some specified number. Some examples are:

"$p = 1/2$",

"The average CMU entering freshman SATV score is equal to the national average"

"The mean increase in sleep for patients on a new drug is the same as for patients who get a placebo."

(3) The Alternative Hypothesis, (H_a), is also called the *motivating or research hypothesis*. It is the opposite of the null hypothesis. Some examples are:

"$p \neq 1/2$",

"The average CMU entering reshman SATV score is different from the national average"

"The mean increase in sleep for patients on a new drug is NOT the same as for patients who get a placebo."