Confidence Intervals

Lecture 20

Readings: GW 12 & SS 8.4

Motivation for confidence intervals

- Up to now, we have gotten the probability of observing a statistic (*t* or *z*) as or more extreme than observed pretending that the *null mean* is true (the *p*-value).
- But we might also want to know what the likely values of the *true mean* in the population are.
- Confidence intervals give you likely values of the true mean in the population
  - There is some population mean out there, but we don't know what it is.
  - A confidence interval tries to tell you between which two values the true mean is likely to lie

Confidence intervals

- Confidence intervals are an alternative way to think about statistics (compared to hypothesis testing)
- Namely, a confidence interval is an interval that contains all the potential null hypotheses that would not be rejected (that are plausible) given your observed mean (or observed mean difference)
- Luckily, almost all the concepts you need to understand CIs have already been learned.

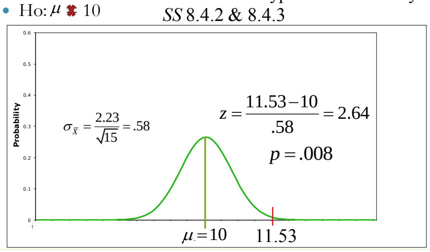
Confidence interval examples

- The mean approval of President Obama in a sample is 47%. That is unlikely to be the exact value if we had asked everyone in the US population. Between what two values is that true mean approval rating likely to lie? (e.g., between 45% and 49%, or +/- 2%).
- I observe a mean difference in height between males and females of 3.2 inches in some sample. That is unlikely to be the exact, true difference in the population. What are the likely values of the true mean difference (e.g., the true mean difference is likely to be between 2.8 & 3.6).

z-test hypothesis test example

- Say that:
  - $\sigma = 2.23$
  - $\bar{X} = M = 11.53$
  - n = 15
  - Ho: μ ✱ 10

The hypothesis that the mean is actually 10 is very unlikely given $\bar{X}$. What other hypotheses are unlikely? Which other hypotheses are likely? *SS* 8.4.2 & 8.4.3



z-test 95% confidence interval example

- Which "null" hypotheses would *not* be rejected (are not unlikely given the data)?
- Answer: all those between these two ? values:

$$z = \frac{11.53-?}{.58} = 1.96 \qquad z = \frac{11.53-?}{.58} = -1.96$$

$$11.53 - ? = 1.96 * .58 \qquad 11.53 - ? = -1.96 * .58$$

$$-? = 1.96 * .58 - 11.58 \qquad -? = -1.96 * .58 - 11.58$$

$$? = 11.58 - 1.96 * .58 \qquad ? = 11.58 + 1.96 * .58$$

$$? = \bar{X} - z_{crit} * \sigma_{\bar{X}} \qquad ? = \bar{X} + z_{crit} * \sigma_{\bar{X}}$$

$$CI = [\bar{X} - z_{crit} * \sigma_{\bar{X}}, \bar{X} + z_{crit} * \sigma_{\bar{X}}]$$

z-test 99%, 90%, etc. confidence intervals

- We usually use a 95% CI. But we could change alpha and get a 90% or 99% (e.g.) CI.
- To do this, merely change z_{crit}:

$$CI = [\bar{X} - z_{crit} * \sigma_{\bar{X}}, \bar{X} + z_{crit} * \sigma_{\bar{X}}]$$

$$99\% CI = [\bar{X} - 2.33 * \sigma_{\bar{X}}, \bar{X} + 2.33 * \sigma_{\bar{X}}]$$

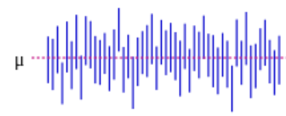
$$95\% CI = [\bar{X} - 1.96 * \sigma_{\bar{X}}, \bar{X} + 1.96 * \sigma_{\bar{X}}]$$

$$90\% CI = [\bar{X} - 1.64 * \sigma_{\bar{X}}, \bar{X} + 1.64 * \sigma_{\bar{X}}]$$

See *SS* 8.4.4

Confidence interval interpretation

- Can say "with 95% (or 90% or 99%, etc.) confidence, the true mean lies within the CI".
- This means that 95% (or 1-alpha %) of similarly constructed CIs would contain the true mean.



- If the original null hypothesis mean is within the CI, you would not have rejected the null hypothesis

t-test 95% confidence interval example

- Usually, we don't know the true variance (or standard deviation) of the population.
- To construct a CI in this case, use t_{crit}, not z_{crit}:

$$CI = [\bar{X} - t_{crit} * s_{\bar{X}}, \bar{X} + t_{crit} * s_{\bar{X}}]$$

- For a mean difference:

$$CI = [(\bar{X}_1 - \bar{X}_2) - t_{crit} * s_{(\bar{X}_1 - \bar{X}_2)}, (\bar{X}_1 - \bar{X}_2) + t_{crit} * s_{(\bar{X}_1 - \bar{X}_2)}]$$