Stat 145, Wed 13-Oct-2021 -- Wed 13-Oct-2021
Biostatistics
Spring 2021

Wednesday, October 13th 2021

Due:: Moodle Quiz Ch. 3 at 11 pm

Wednesday, October 13th 2021

Wk 7, We
Topic:: Insight into inference
Read:: Lock5 4.5
HW:: Moodle Quiz Ch. 4 ends Wed.

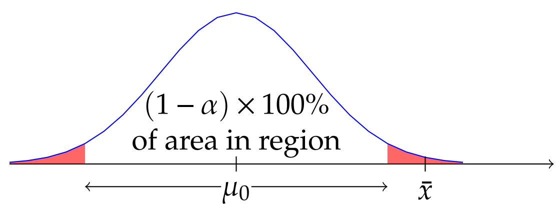
Exercises to look at:
  4.14
  4.16
  4.54
  4.64

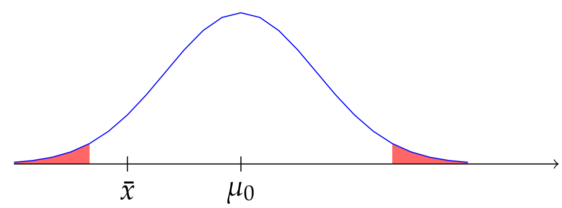
Some final words about hypothesis testing

Confidence intervals and hypothesis tests: two sides of the same coin?

In hypothesis testing, we construct a null distribution.

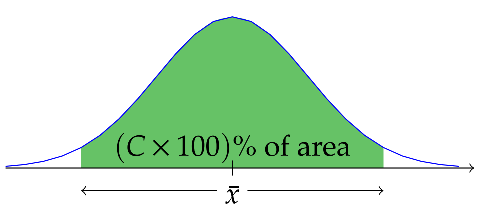
- often it has appeared to be symmetric, bell-shaped (normal?).
- null value is the mean/center
- in setting α we fix the area of the **rejection region** (two tails, colored red)
- Our test statistic may be in a tail (nonrejection region), or in the nonrejection region.



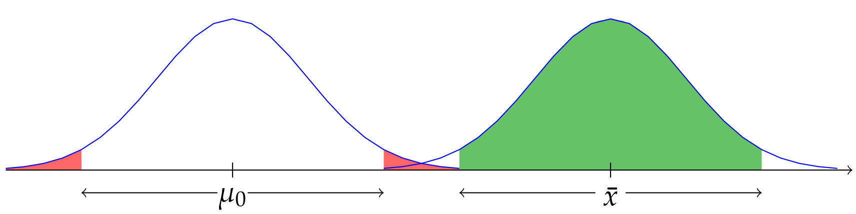


In confidence interval construction (centered interval method), we

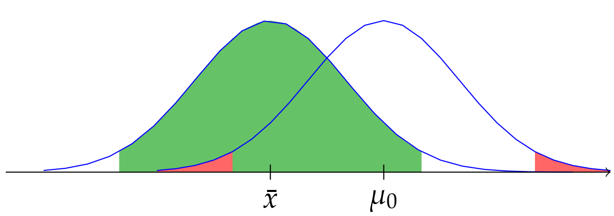
- use the point estimate as the center of a region (green).
- use the confidence level C to decide what portion to include in the region.



Now, in theory, whenever $C = 1-\alpha$, the width of the nonrejection (uncolored) region of the null distribution is approximately the same as the width of the confidence interval. In this case, the composite picture might be like this



or like this



So, say a random sample is collected. The sample statistic, perhaps $\bar{x}$, could be used to construct a confidence interval for the parameter, perhaps μ, but it also could be used as a test statistic in the 2-sided test of hypotheses

$$\mathbf{H}_0: \ \mu = \mu_0, \qquad \mathbf{H}_a: \ \mu \neq \mu_0.$$

But no matter which of these it is, confidence interval or hypothesis test, the one informs the other. Examples of the ways include these:

- If μ_0 is not in a 95% confidence interval, then the P-value of the hypothesis test is smaller than 0.05.
- If μ_0 is in a 90% confidence interval, then the P-value of the hypothesis test is larger than 0.1.
- If the P-value from the hypothesis test is 0.07, then μ_0 is in the 95% confidence interval, but not in the 90% confidence interval.

Cautions about multiple testing

Type I error can (and does) occur.
Your significance level α predicts how frequently.

Remember what was said earlier: setting $\alpha = 0.05$ for all your hypothesis tests means that, in cases where the null hypothesis is true, you will commit Type I error 5% of the time, *mistakenly* rejecting $\mathbf{H}_0$. That's a Type I error rate of 1-in-20. Any researcher conducting numerous statistical tests with $\alpha = 0.05$ should keep this in mind, and should maintain a healthy suspicion if about 5 percent of her tests are yielding statistically significant results. If, over the last year, 40 hypothesis tests have been conducted and 3 have been statistically significant, that is right near what we might expect to happen even if *none* of the null hypotheses in those tests of significance have been false.

Statistical significance is different from practical importance

Statistically significant result = Sample (test) statistic landed in rejection region

But, if the non-rejection region isn't very wide, the true parameter value may not be very far from the null value.

Referring to the picture above, statistical significance amounts to our test statistic being far enough from the null value (μ_0) that it lands in the rejection region, nothing more. This may be evidence enough to reject the null hypothesis in favor of the alternative $\mathbf{H}_a:\ \mu \neq \mu_0$, but it does not necessarily follow that the true value of μ is far away. You may have evidence that is statistically significant for showing that a drug does not leave blood pressure unchanged in those who suffer from high blood pressure even if its effect is only to decrease systolic pressure by 1.

Ellenberg has useful illustrations of hypothesis testing in Chapters 6 and 7 of his book, "How Not to Be Wrong: The Power of Mathematical Thinking." While I mention here his point that the word *significance* in statistics is meant in a technical sense that English language speakers are likely to misconstrue, I will let Ellenberg have the burden of hammering it home.