Two-Sided Testing and C.I. s; Choosing the Levels of Significance

Chapter 8
Two-sided hypothesis testing and confidence intervals

• A **two-sided** significance test rejects the null hypothesis exactly when the claim falls outside the corresponding confidence interval for \( \mu \).

  – If the claim is in the CI \( \rightarrow \) fail to reject \( H_0 \)
  – If the claim is *not* in the CI \( \rightarrow \) reject \( H_0 \)

  – **NOTE:** must have “\( \neq \)” in \( H_a \)!
Figure 7.3 Capturing a central z curve area of .95
Let’s revisit a few examples

- **Metabolism**
  - C.I.? [-560.629, -169.171]
  - P-Value? 0.002
- **Module**
  - C.I.? [-.11, 3.31]
  - P-Value? .124

- A confidence interval can be used as a basis for testing hypotheses, and
- there is a confidence interval procedure (with $C = 1 - \alpha$) corresponding to any particular test procedure with significance $\alpha$.
- **Remark:** for two-sided test, use two-sided C.I.; for one-sided test, use C.I. in the same direction.
8.5 Choosing the level of significance

• \( \alpha = 0.05 \) is accepted standard, but...
• if the conclusion that \( H_a \) is true has “costly” implications, smaller \( \alpha \) may be appropriate
• not always need to make a decision: describing the evidence by P-value may be enough
• no sharp border between statistically significant and insignificant
Statistical vs. practical significance

- **Statistically significant effect may be small:**
  - Example ("Executive" blood pressure):
    - \( \mu_0 = 128 \)
    - \( \sigma = 15 \)
    - \( n = 1000 \) obs.
    - sample mean = 127
  - \( Z = (127-128)/ (15/\sqrt{1000}) = -2.11 \)
  - P-value for two sided \( H_a \) = 2\( \times \)0.0174 = 0.0348

Significant??
Stat. significance is not necessarily practical significance.
Significance vs. practical significance

• Plot your results and confidence interval, to see if the effect is worth your attention.

• Important effects may have large P-value if sample size too small. Converse also true.

• Outliers may produce or destroy statistical significance.
Lack of significance may be informative…

..for other researchers, as a warning not to invest more time in a given study.

But…, did your survey have a chance to detect the size of effect you were looking for?
(e.g. maybe too small a sample size?).
Statistical Inference—Not valid for all sets of data!

- Statistical Inference, no matter how well done, cannot fix basic flaws in the design
  - Bias due to:
    - Sampling (like voluntary response, etc)
    - Incorrect experimental design
    - Poorly worded questions
    - Etc.
Cautions (apply both to confidence intervals and tests of significance):

• Data: **assume SRS**
• Formulas for other randomized designs available
• Haphazard data = unreliable CI, tests

• Population need to be normal or not?
  – If \( n < 25 \), have to check normality (e.g. by making QQ-plot and check the straight-line pattern for each data set)
  – With \( n \geq 25 \), CLT can give us approximate normality in most situations.
Important topics in Ch. 7 and 8

- Identifying the type of problems
  - 1 or 2 samples, or paired data
  - means or proportions
  - z or t
- Large-sample C.I.s for different parameters; Small-Sample C.I.s
- Sample size for specific bounds (means and proportions)
- Unbiased and Consistent estimators
- Stating conclusions in "layman’s terms"
- Type I error
- Lots of concepts, e.g.
  - Meaning of “confidence level”
  - P-value
  - significance level
Continued

- Stat sig vs. practical sig
- assumptions we make (CI and sig tests)
- Relationship between CI and sig tests
- All of the steps of sig tests
- Interpreting CIs

- MUST know how to use
  - Table IV (Pg 566) for C.I.
  - and Table VI (Pg 568-570) For sig tests
After Class...

- Review Section 8.1, 8.2 and 8.5 (Page 387-388, and Page 393)
- Read ANOVA in Ch. 9

- Lab #4, this Wed

- Review Ch. 7 and 8