Exam #1 Review

Ryan Miller

Topic #1 - Describing Data

What to know:

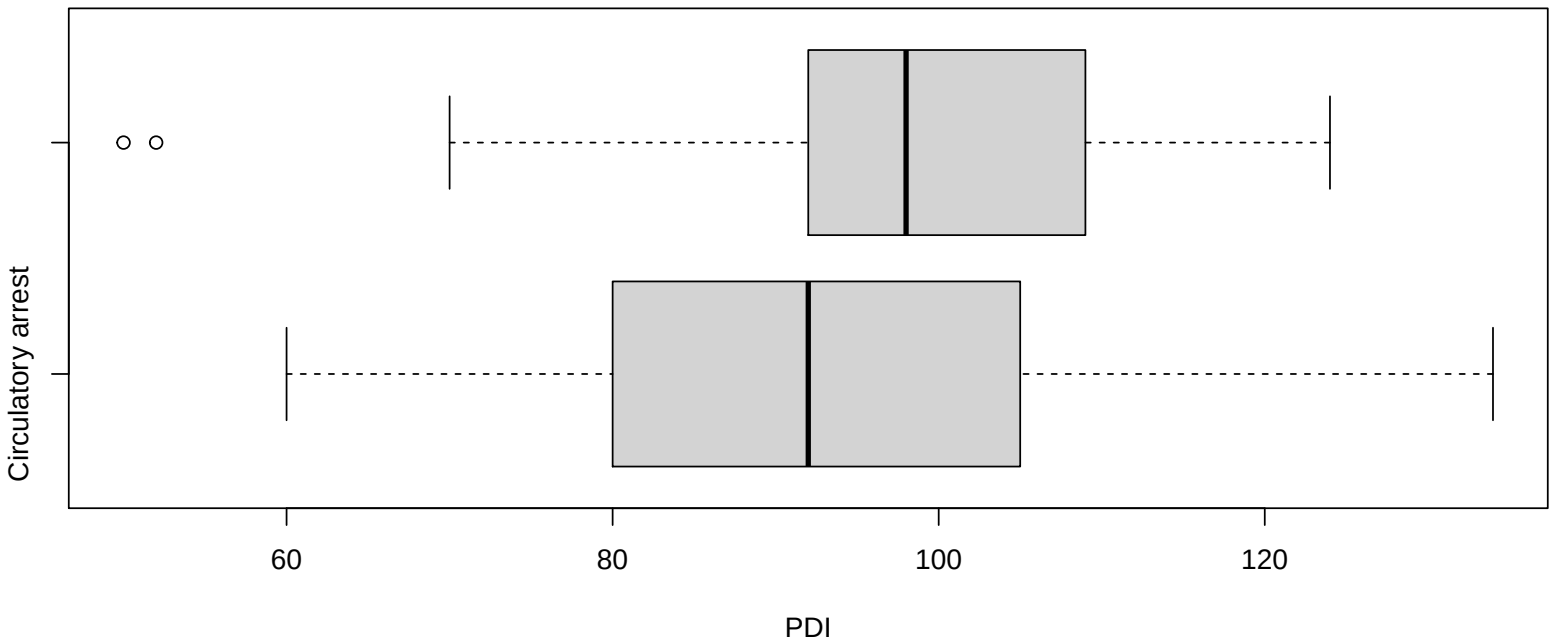
- Identifying cases and variables
- Appropriate graphs and descriptive statistics for different types of variables
- Describing associations between variables

Practice

- Researchers at Harvard Medical School performed an experiment on infants in need of heart surgery comparing two procedures: "low-flow bypass", a new treatment, or "circulatory arrest", an existing treatment
- Two years after surgery, they followed up on each infant to measure their "PDI" (a numeric developmental score) and "MDI" (a numeric cognitive performance score)

1) Identify the explanatory variable in the study and describe whether it's categorical or quantitative
2) Describe a graph you'd use to relate the explanatory variable and "PDI"

Practice (cont)



3) Does there appear to be an association between surgical approach and PDI score?
4) Is there more variability in PDI scores in the low-flow group or the circulatory arrest group?

Topic #2 - Sampling from a Population

What to know:

- Sampling bias vs. sampling variability and the role of sample size
- Difference between standard error and standard deviation
- Statistical notation for population parameters vs. sample estimates

Practice (cont)

The table below summarizes PDI scores for each surgical type:

Treatment	mean	sd	n
Circulatory arrest	91.9	16.5	73
Low-flow bypass	97.8	14.7	70

Suppose the researchers want to generalize the differences found in this experiment to *all* infants in need of heart surgery.

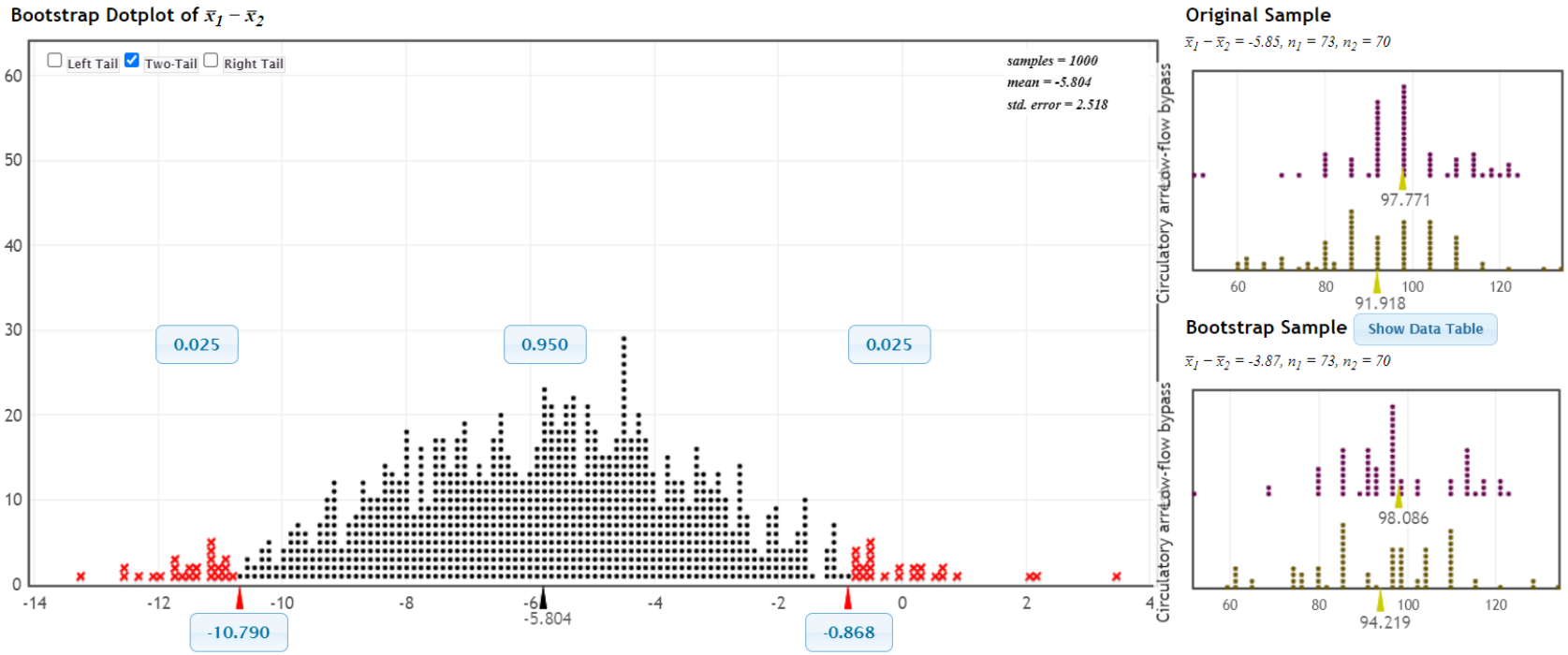
5) Provide notation for the relevant *population parameter*.
6) Provide a *point estimate*, including proper notation, for this population parameter.
7) Briefly explain why this point estimate might differ from the population parameter.

Topic #3 - Confidence Intervals and Bootstrapping

What to know:

- Goals of bootstrapping (ie: understanding the sampling variability of an estimate)
- How to read StatKey output (ie: dots in the bootstrap distribution, original sample, SE, etc.)
- Definition of a Confidence Level
- 2-SE method and percentile method

Practice (cont)



8) What does dot in the bootstrap distribution represent?
9) Find a 95% CI using the 2-SE method. How does this CI compare to the 95% percentile bootstrap CI?
10) Can you confidently conclude that one these surgical approaches yields superior outcomes?

Topic #4 - Central Limit theorem

What to know:

- How to use SE formulas and a given value of "c" to form a confidence interval
- When the Normal curve is used to find "c" (ie: proportions) and when the t-distribution is needed (ie: means)

Practice

The table below summarizes PDI scores for each surgical type:

Treatment	mean	median	sd	n
Circulatory arrest	91.9	92	16.5	73
Low-flow bypass	97.8	98	14.7	70

11) Using $c = 1.99$, find a 95% CI estimate for the difference in mean PDI scores using the appropriate CLT formula
12) If the researchers had performed their experiment on a larger sample, would you expect the 95% CI estimate from that sample to be wider or narrower than the one you found in #11?