BASICS OF STATS, R STUDIO

Jack Hester, MPH

CEPC 0904

Summer 2022

WHAT ARE STATISTICAL TESTS?

- Numerical data analysis
- Between-group comparison
  - 2+ populations, intervention + control, etc.
- Is there a difference? How much of a difference? How much data supports our conclusion? What can we predict?
  - Depends on which test you run
  - The ones we cover here mostly determine if there's a difference, not how much of a difference

HOW DO WE APPLY THEM?

- Generate hypotheses
- Test them with the appropriate test
- Usually you will have a null (H_0) and Alternative (H_A)
- There are general research questions/hypotheses, there are also specific hypotheses depending on the statistical method

TERMINOLOGY

- Standard deviation = measure of spread
- Sample = set of observations
- Population = whole group, usually you will take a sample of a population
- Sample size, (n), how many observations (data points) in the sample

NOTES ON THESE TESTS

- Covering basic stats tests
- All predictor (independent) variables are categorical
  - Categorical = a label vs a number
- All outcome (dependent) variables are continuous (numerical, can be any value or any value within range)

ONE SAMPLE TESTS

- Comparing sample to expected mean
  - E.g., comparing sample of heart rates to expected population avg heart rate
- Population standard deviation known
  - No → one sample t-test
  - Yes
    - n > 30 → one sample z-test
    - n <30 → one sample t-test

TWO INDEPENDENT SAMPLE TESTS

- Comparing two sample means with no relationship
  - E.g., test scores between people in different countries
- Population standard deviation known
  - No → two sample t-test
  - Yes
    - n > 30 → two sample z-test
    - n < 30 → two sample t-test

TWO PAIRED SAMPLE TESTS

- Comparing two sample means with relationship
  - E.g., cholesterol rates before and after taking a drug
- Population standard deviation known
  - No → two sample paired t-test
  - Yes
    - n > 30 → two paired sample z-test
    - n < 30 → two paired sample t-test

MORE THAN TWO GROUP COMPARISON

- Multiple categories/groups
- ANOVA test
- One independent variable with > 2 sub-groups (levels) and one dependent variable → one-way ANOVA
  - E.g., Brand of car vs life satisfaction
- More than two independent variables with one dependent variable → two-way ANOVA
  - E.g., brand of car, location of house, martial status vs life satisfaction

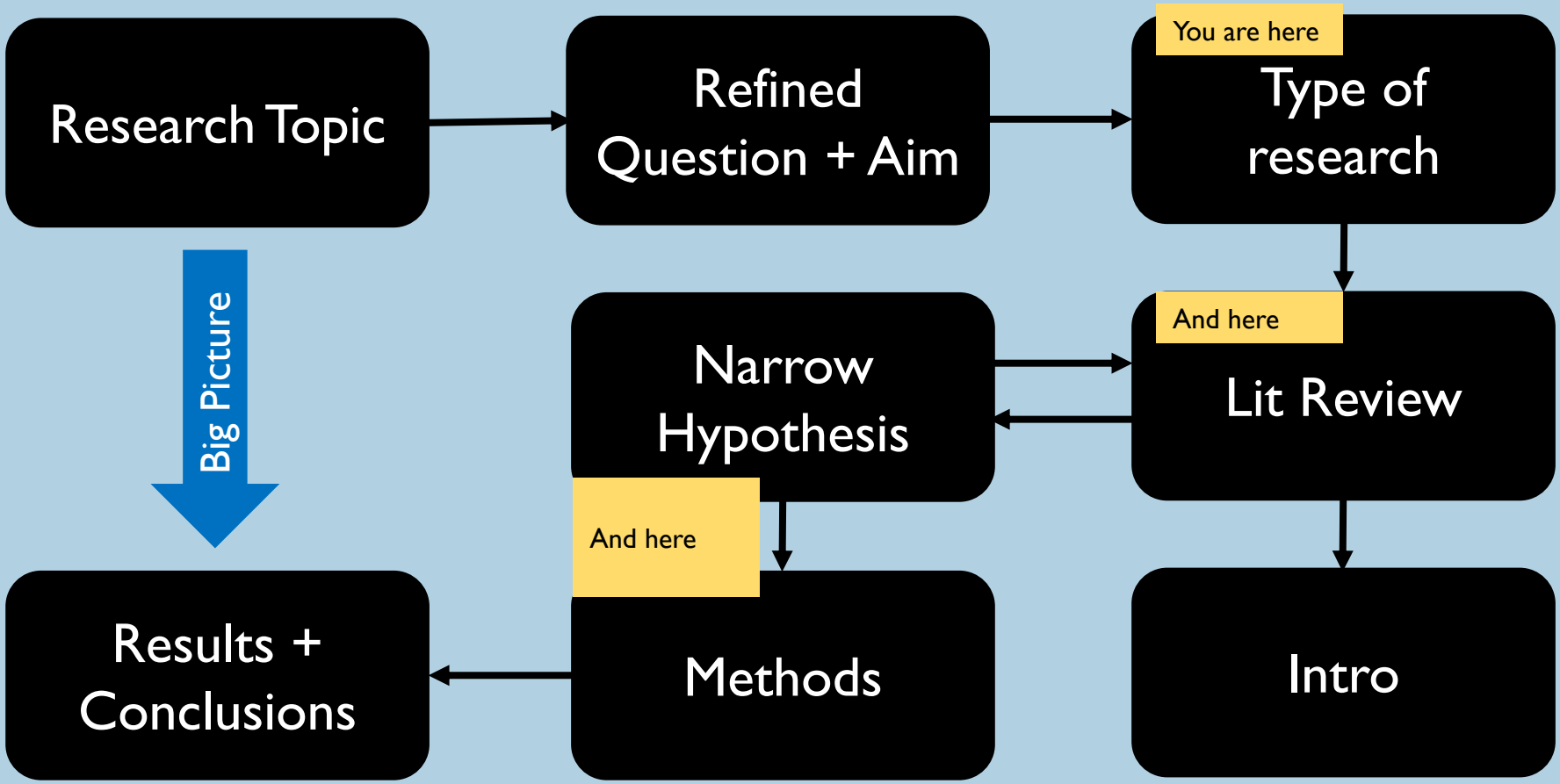
REFERENCE GUIDE

- https://cepc0904-22.jackhester.com/documents/stats_guide.pdf

NEXT STEPS

- Watch probability video (TBA on schedule next week)
- Review reference guide
- Practice this using R HW
- Take quiz 2

RESEARCH PROJECT – BIG PICTURE



UPCOMING ASSIGNMENTS

- Quiz 2: Fri/Sat, Quiz notes due tonight (opt.)
- Datacamp: Due tonight
- ~~HW 4: Due Monday (R studio, group work ok)~~
  - Delaying to next week
- HW 3: Already due, if delayed OK
  - All HW 2 graded by tomorrow

R MARKDOWN OVERVIEW

QUIZ/OTHER QUESTIONS?

HW/QUIZ PREP TIME