ANOVA (Analysis Of Variance)

Wyatt Beyers 080919

What test should you use? First, what type of data do you have?

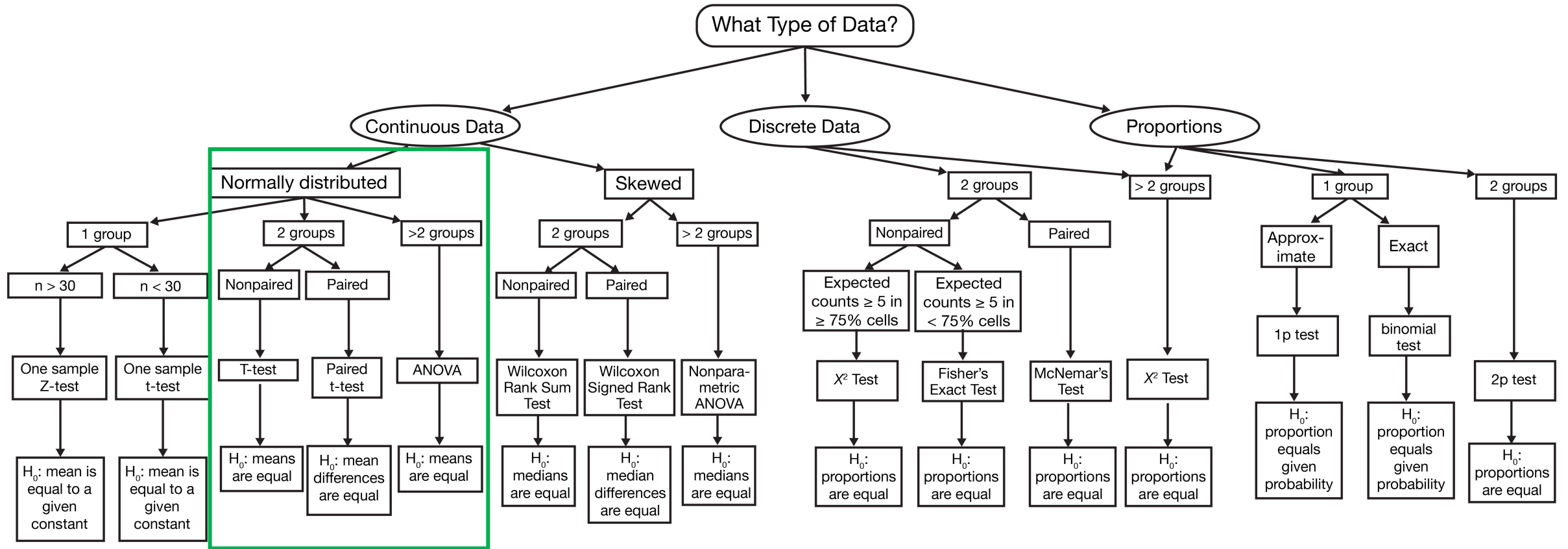## Examples of Continuous Data (can't be counted)

- Fluorescence Intensity

- Protein Concentration

- Cell Size

Examples of Discrete Data (can be counted)

- Sides on a dice

- Number of melanosomes in a melanocyte

Most data I work with is continuous data, meaning there are an infinite number of values that a measurement can take.

Flow chart: which test statistic should you use?

ANOVA (Analysis Of Variance)

- An ANOVA uses the F test – comparison of variance

- Significance test used to compare >2 means

- Reveals if a difference is present
  - Does NOT reveal which sample means are different
    - If there is no significance found, the test is concluded
    - If there is a difference → Post-hoc test (Tukey or Scheffe test)

Post-hoc pairwise comparison tests

- "Rule of thumb": Use Tukey test when samples are equal in size and Scheffe's test if samples differ in size.

- These tests are similar to a T test in that they do pairwise comparisons, but are adjusted so that they account for the fact that more than 2 means are compared.

One Way vs Two Way ANOVA

- One way used if you are only comparing one variable
  - Example: Bacteria strain influence on recombinant protein yield
- Two way used if you are testing two independent variables in the same experiment.
  - Example: Bacteria strain AND media type influence on protein yield
    - Does not only determine if each has an influence independently, but if there is *interaction effect*…Maybe strain has no influence on protein yield in LB Broth, but in 2XYT strain has a major influence. In other words, it tests whether the independent variables are truly independent or if they are somehow connected.

Why can't I just do multiple pairwise t tests?

- "With T test, only 2 means at a time are compared
  - F test compares all means at once
- The more T tests that are conducted, more likely to find significant differences by chance alone (Type I error)
- Many means to compare → more tests required
  - For 10 means, 45 t tests required"
- ANOVA greatly reduces Type I errors as the number of treatments to compare increases

  -Elementary Statistics, Allan Bluman

Conducting a One-Way ANOVA

- Not something you calculate by hand, but if you wanted to….

Finding the F Test Value for the Analysis of Variance

Step 1 Find the mean and variance of each sample:

$$(\bar{X}_1, s_1^2), (\bar{X}_2, s_2^2), \ldots, (\bar{X}_k, s_k^2)$$

Step 2 Find the grand mean.

$$\bar{X}_{GM} = \frac{\Sigma X}{N}$$

Step 3 Find the between-group variance.

$$s_B^2 = \frac{\Sigma n_i(\bar{X}_i - \bar{X}_{GM})^2}{k-1}$$

Step 4 Find the within-group variance.

$$s_W^2 = \frac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)}$$

Step 5 Find the F test value.

$$F = \frac{s_B^2}{s_W^2}$$

-Elementary Statistics, Allan Bluman

So, how to calculate ANOVA practically?

- Online ANOVA and Tukey calculator

https://astatsa.com/OneWay_Anova_with_TukeyHSD/

  - Input number of treatments
  - Copy and paste all of your raw data for each treatment that you want to compare

- Matlab is free for CSU students and can automate your statistics!

  - Built in Anova functions make it easy to compare your data

Example: script in matlab performs ANOVA and Tukey test on raw data and generates bar graph with significance stars

```
multcomparison =

    1.0000    2.0000   -0.0483    0.0384    0.1251    0.8497
    1.0000    3.0000   -0.0572    0.0342    0.1257    0.9273
    1.0000    4.0000    0.1050    0.1923    0.2795    0.0000
    1.0000    5.0000    0.5171    0.6107    0.7043    0.0000
    1.0000    6.0000    0.5285    0.6566    0.7848    0.0000
    1.0000    7.0000    0.4132    0.5027    0.5922    0.0000
    2.0000    3.0000   -0.1015   -0.0042    0.0932    1.0000
    2.0000    4.0000    0.0605    0.1539    0.2473    0.0000
    2.0000    5.0000    0.4729    0.5723    0.6717    0.0000
    2.0000    6.0000    0.4858    0.6183    0.7507    0.0000
    2.0000    7.0000    0.3688    0.4643    0.5598    0.0000
    3.0000    4.0000    0.0603    0.1580    0.2558    0.0000
    3.0000    5.0000    0.4729    0.5765    0.6800    0.0000
    3.0000    6.0000    0.4869    0.6224    0.7580    0.0000
    3.0000    7.0000    0.3687    0.4685    0.5683    0.0000
    4.0000    5.0000    0.3186    0.4184    0.5183    0.0000
    4.0000    6.0000    0.3316    0.4644    0.5971    0.0000
    4.0000    7.0000    0.2145    0.3105    0.4064    0.0000
    5.0000    6.0000   -0.0911    0.0460    0.1830    0.9566
    5.0000    7.0000   -0.2098   -0.1080   -0.0061    0.0294
    6.0000    7.0000   -0.2882   -0.1539   -0.0197    0.0128
```
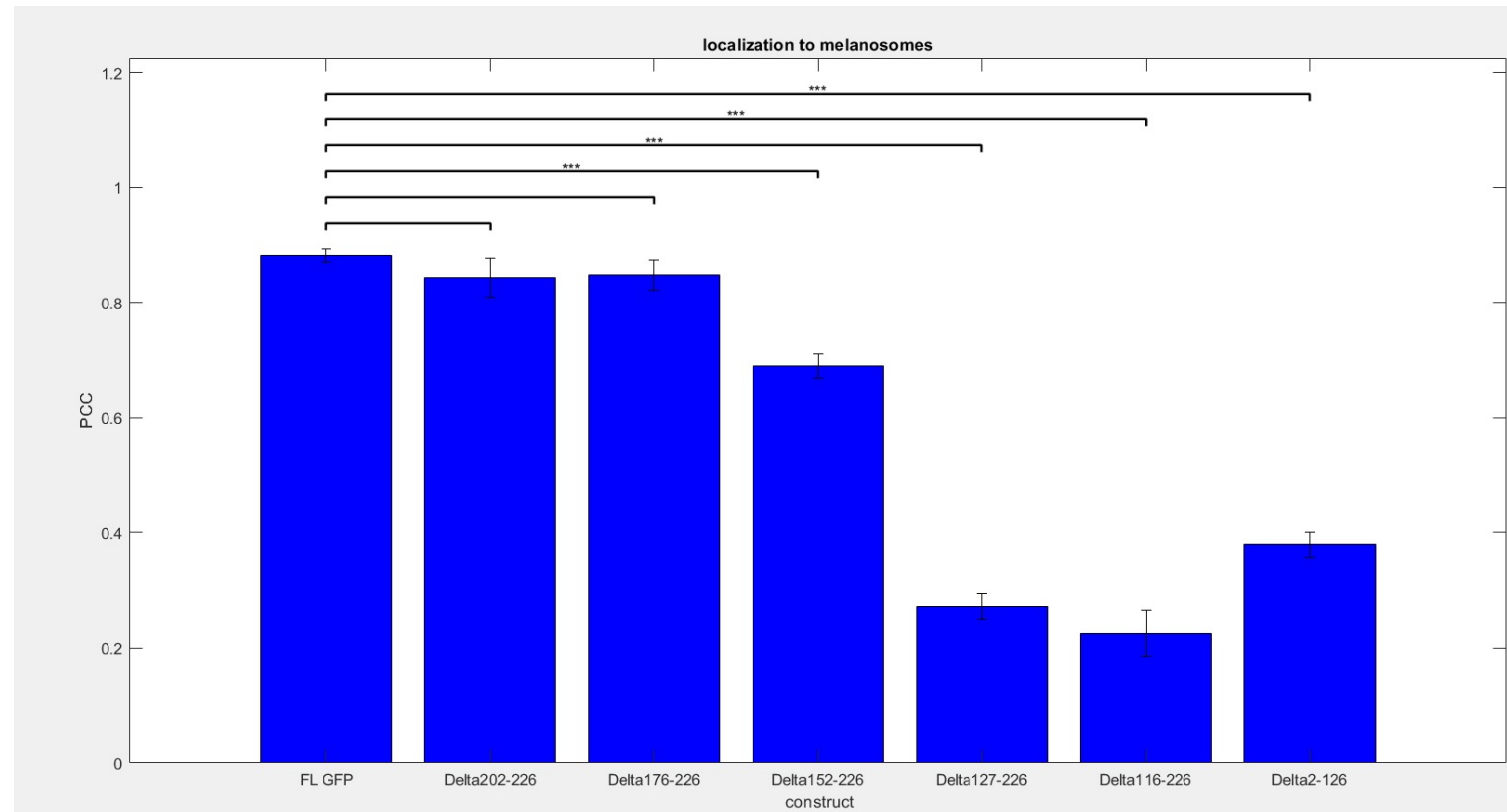
- Meet with me if you would like to learn how to use matlab for ANOVA!
- Wyatt.beyers@colostate.edu