Week 7: Tests for Differences

Session 1

Spring 2020

iClicker Question 1

Scenario:

- You want to know whether the growth rates of oak trees differ in two habitats.
- You have two clones each of 37 different genotypes.
- One clone of each genotype is planted in a mesic prairie, the other in an old agricultural field.
- Which test should you use?

A `two-sample t-test`

B `Mann-Whitney paried test`

C `Mann-Whitney U test`

D `paired t-test`

iClicker Question 2

Scenario:

- You are testing whether Pennsylvania sedge (C. pennsylvanica) cover is greater under hardwoods or conifers.
- You have collected percent cover data from 50 hardwood and 50 conifer patches at Mount Toby.
- Which test should you use?

A `two-sample t-test`

B `Mann-Whitney paried test`

C `Mann-Whitney U test`

D `paired t-test`

iClicker Question 3

Scenario:

-You want to know whether the growth rates of oak trees differ in two habitats.

-You think they will grow more quickly in a old agricultural field than in a mesic prairie.

- Which is the best alternative hypothesis?

A `growth is faster in the prairie`

B `there is no difference in growth between habitats`

C `growth is faster in the field`

D `there is a difference in growth between habitats`

E `growth rates are the same`

iClicker Question 4

Scenario:

-You want to know whether the growth rates of oak trees differ in two habitats: an old agricultural field and mesic prairie.

- Which is the best alternative hypothesis?

A `growth is faster in the prairie`

B `there is no difference in growth between habitats`

C `growth is faster in the field`

D `there is a difference in growth between habitats`

E `growth rates are the same`

iClicker Question 5

Scenario:

- You calculate a t-value, and an associated p-value, for individual length in two populations of Daphnia.
- Which critical significance category for the t-value would give you the best evidence that the two populations are different:

A `1%`

B `10%`

C `5%`

D `95%`

E `0.1%`

If you're feeling stuck, remember my office hours are Tuesday/Thursday 1:00 - 2:00.

For Today

- Toward statistics
- Tests for differences

Follow-up questions from the Chapter 6 homework

- What questions do you have?

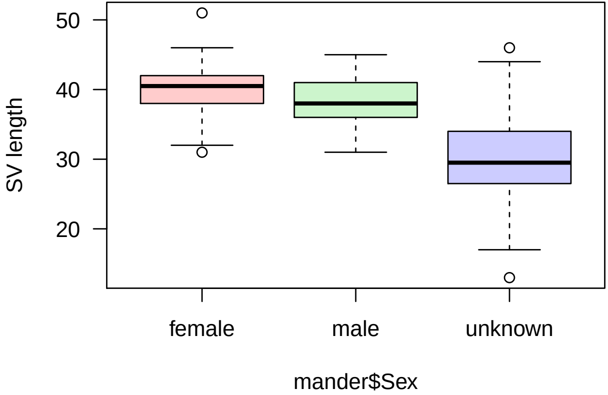
Beyond graphs, Towards statistics

- Graphs are powerful tools that provide insight and understanding of the patterns and relationships in the data.
- Graphs alone don't give us the complete answer. We need to **quantify** the relationships we see in our plots.
- What other tools do we have to **support** our conclusions?

?

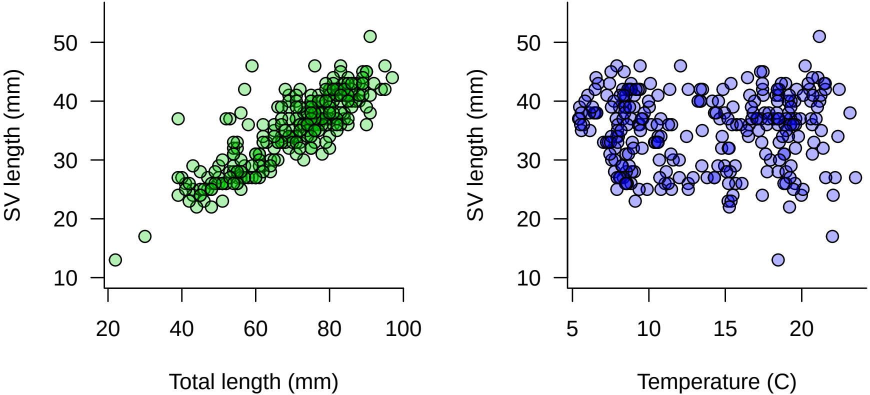
Beyond graphs, Towards statistics

- How can we **quantify** our evidence for relationships?
  - Are differences between groups *significant*?
  - Are differences between groups *meaningful*?



Beyond graphs, Towards statistics

- How can we **quantify** our evidence for relationships?
  - Are associations between 2 variables *significatnt*?
  - Are associations between 2 variables *meaningful*?



Beyond graphs, Towards statistics

- Statistics is the tool we use to formally answer these questions!
  - Are differences *are/are not* significant?
  - Are associations *are/are not* significant?

Wait a second... what do we mean when we say **significant**?

?

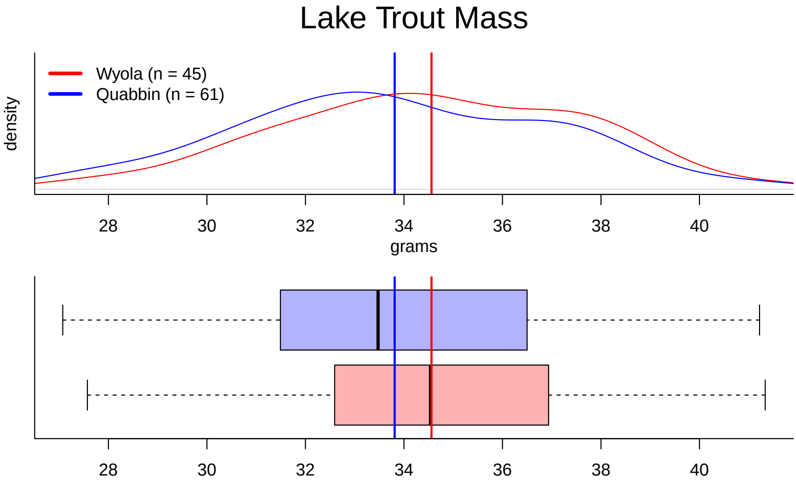
Let's examine some plots to gain intuition:

- Scenario: We want to know whether the size of 3-year-old bluegill (*Lepomis macrochirus*) are larger in some Massachusetts lakes than others.
- We have collected data for bluegill from Wyola Lake and the Quabbin Reservoir in Western Mass.

[1] Image credit: New York Fish and Game Commission

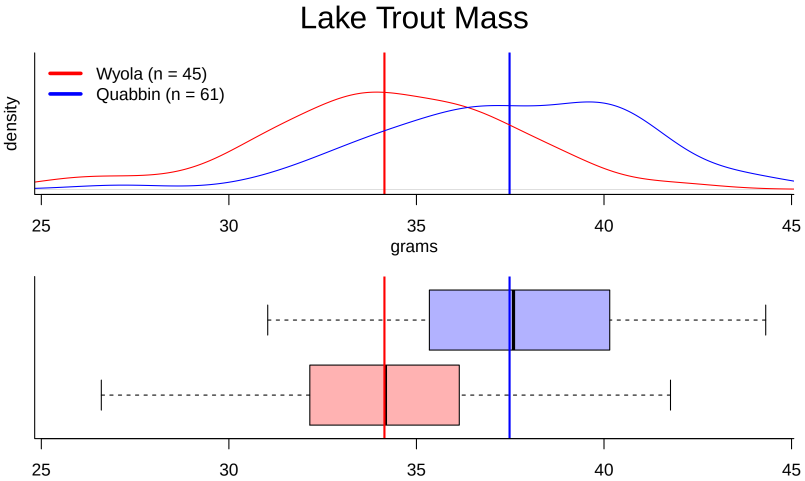
Bluegill Data I

- Are differences between lakes *significant*?
- Are differences between lakes *meaningful*?



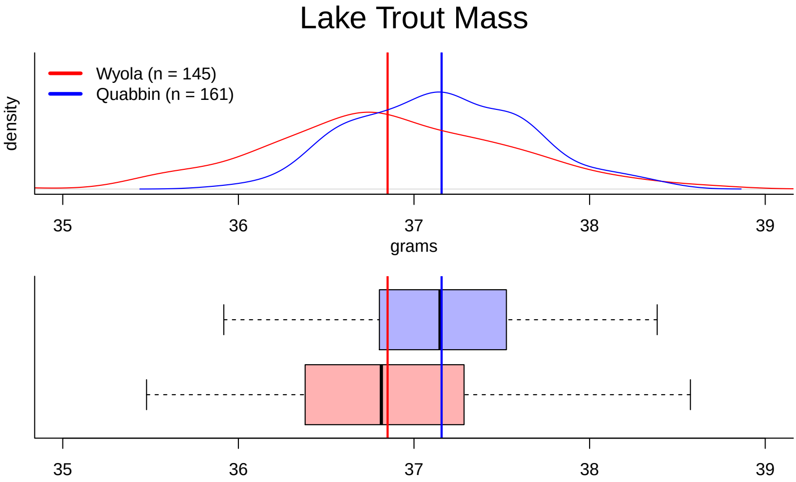
Bluegill Data II

- Are differences between lakes *significant*?
- Are differences between lakes *meaningful*?



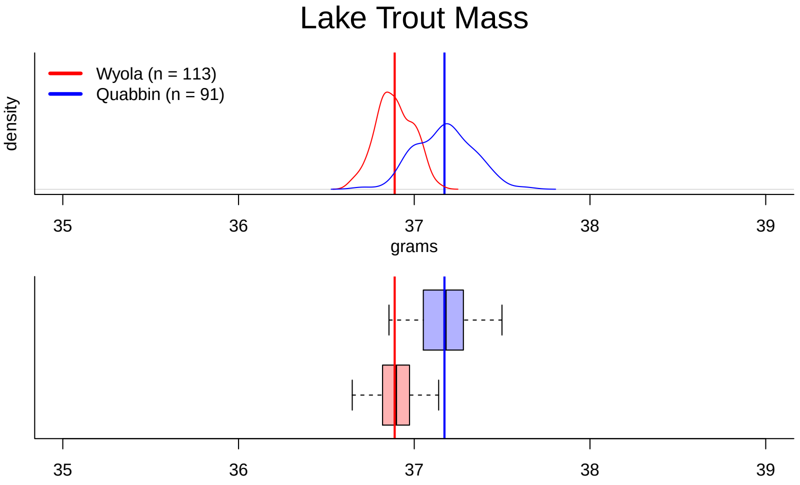
Bluegill Data III

- Are differences between lakes *significant*?
- Are differences between lakes *meaningful*?



Bluegill Data IV

- Are differences between lakes *significant*?
- Are differences between lakes *meaningful*?



Tests for differences

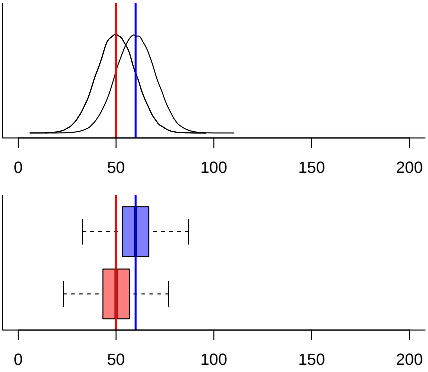
- How can we test if two groups of observations are different?
- What kind's of tests do we know about?
- Which statistic(s) do the tests test for?

?

Tests for differences

Often we want to know if two of more samples are different

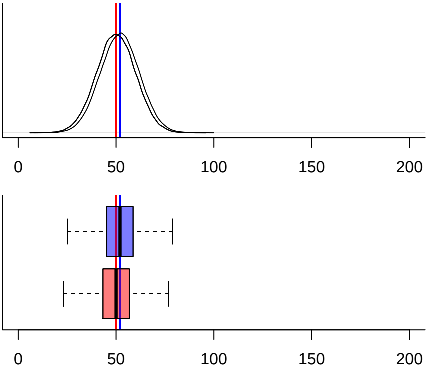
- are the sample *means* different?
- are the sample *medians* different?
- are the differences *statistically significant*?



Tests for differences

Often we want to know if two of more samples are different

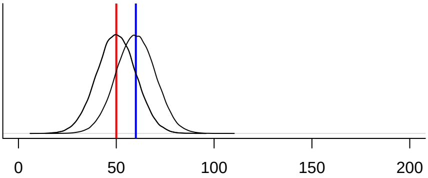
- are the sample *means* different?
- are the sample *medians* different?
- are the differences *statistically significant*?

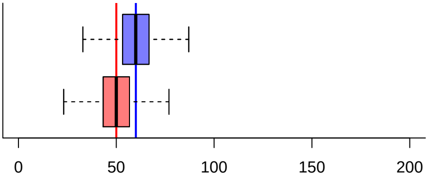


Tests for differences

Often we want to know if two of more samples are different

- are the sample *means* different?
- are the sample *medians* different?
- are the differences *statistically significant*?





Tests for differences

Often we want to know if two of more samples are different

- are the sample *means* different?
- are the sample *medians* different?
- are the differences *statistically significant*?

To determine the significance of differences between **two**, we need a statistical test

- *t-test*
- *U-test*

Tests for differences: intuition

- What information would we need to know?
- What kinds of evidence would support our conclusion?
- How do we define *different*?

- Let's draw some distributions:

Tests for differences: intuition with R

Differences: t-test

Purpose:

- compare the means of two samples (say a and b)

Assumptions:

- both samples normally distributed
- both samples have equal variances

Differences: t-test

Purpose:

- compare the means of two samples (say a and b)

Assumptions:

- both samples normally distributed
- both samples have equal variances

$$t = \frac{|\bar{x}_a - \bar{x}_b|}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$$

- t: the t-statistic
- $\bar{x}$: sample mean
- s: sample standard deviation
- n: sample size

Differences: t-test

Purpose:

- compare the means of two samples (say a and b)

Assumptions:

- both samples normally distributed
- both samples have equal variances

$$t = \frac{|\bar{x}_a - \bar{x}_b|}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$$

- if $|\bar{x}_a - \bar{x}_b|$ is large, then t is ????
- if $\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$ is large, then t is ????

Differences: t-test

Purpose:

- compare the means of two samples (say a and b)

Assumptions:

- both samples normally distributed
- both samples have equal variances

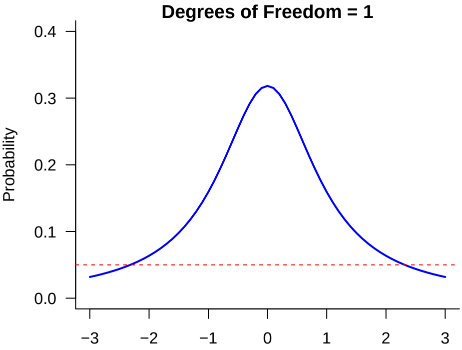
$$t = \frac{|\bar{x}_a - \bar{x}_b|}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$$

- if $|\bar{x}_a - \bar{x}_b|$ is large, then t is large

- if $\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$ is large, then t is small

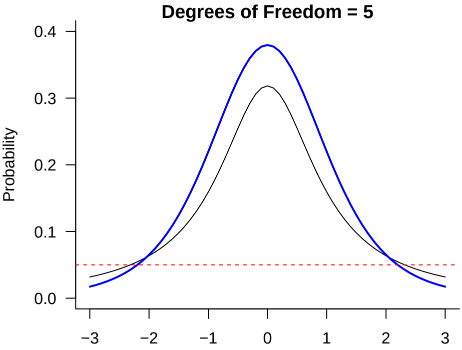
Differences: t-test

Understanding the *t-distribution*:



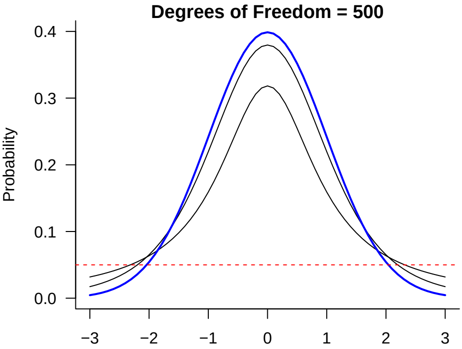
Differences: t-test

Understanding the *t-distribution*:



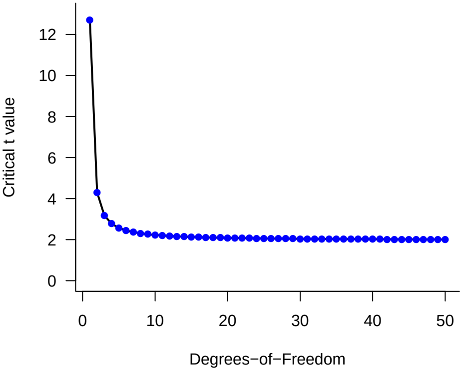
Differences: t-test

Understanding the *t-distribution*:



Differences: t-test

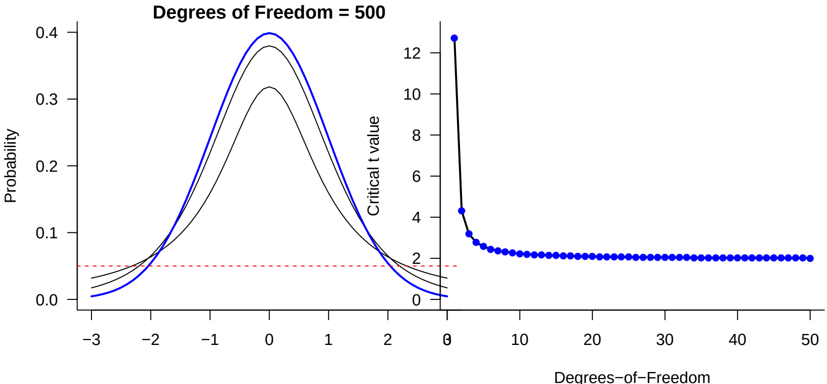
Understanding the *t-distribution*:



Differences: t-test

Understanding the *t-distribution*:

- whether a difference is significant depends on:
  - the *t-statistic*
  - degrees-of-freedom $(n_a - 1 + n_b - 1)$
- larger *t-statistics* more likely to be significant



Differences: t-test

Understanding the *p-value*:

- *p-value* is the probability of observing a *t-statistic* as high as we did by chance
- if *p-value* is lower than significance level (e.g. 5%):
  - difference is significant
  - reject the null hypothesis
  - accept the alternative hypothesis

Differences: t-test

Which *t-test*?

- standard *t-test*
  - compare two independent samples
  - both normally distributed
  - equal (similar) variances
  - samples sizes can be the same or not

$$t = \frac{|\bar{x}_a - \bar{x}_b|}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$$

- t: the t-statistic
- $\bar{x}$: sample mean
- s: sample standard deviation
- n: sample size

Differences: paired t-test

Sometimes samples are not independent

- compare pairs of samples
  - e.g., before-after
  - e.g., north-south
  - e.g., left-right
- both normally distributed
- equal (similar) variances
- samples sizes *must* be the

Differences: paired t-test

Which *t-test*?

- paired *t-test*
  - compare pairs of samples
  - both normally distributed
  - equal (similar) variances
  - samples sizes are __________ ?

$$t = \frac{\bar{D}}{\sqrt{\frac{s_D^2}{n}}}$$

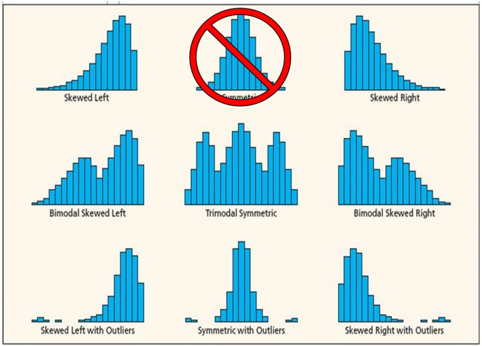
- t: the t-statistic
- $\bar{D}$: mean of the *differences*
- s: standard deviation of the *differences*
- n: number of *paired* samples

Differences: When might the t-test be inappropriate?

?

Differences: U-test

- compare two samples
- one or both *not* normally distributed
- based on *median*, *range*, and *ranks*
- rank all values as one sample, calculate group rank sums R
- calculate a U-value, a measure of overlap



Differences: U-test

- compare two samples
- both or differences *not* normally distributed
- based on *median*, *range*, and *ranks*
- rank all values as one sample, calculate group rank sums R
- calculate a U-value, a measure of overlap

$$U_a = n_a \times n_b + \frac{n_a(n_a + 1)}{2} - R_a$$

$$U_b = n_b \times n_a + \frac{n_b(n_b + 1)}{2} - R_b$$

- n_a: number of samples in sample a
- n_b: number of samples in sample b
- R_a: sum of the ranks of values in a
- R_b: sum of the ranks of values in b

Differences: U-test

- compare two samples
- both or differences *not* normally distributed
- based on *median*, *range*, and *ranks*
- rank all values as one sample, calculate group rank sums R
- calculate a U-value, a measure of overlap

$$U_a = n_a \times n_b + \frac{n_a(n_a + 1)}{2} - R_a$$

$$U_b = n_b \times n_a + \frac{n_b(n_b + 1)}{2} - R_b$$

- smallest is used to find the p-value
- unlike the t-statistic, lower U-values are more likely to be significant

Differences: Wilcoxon matched-pairs test

- both or differences *not* normally distributed
- based on ranked *differences*
  - first calculate the differences
  - second rank the differences
  - 0's not ranked
- sum and compare +ve and -ve ranks

$$W^+ = \sum R^+$$

$$W^- = \sum R^-$$

- W^+: the Wilcoxon test statistic for positive differences
- W^+: the Wilcoxon test statistic for negative differences
- R^+: the sum of the ranks of positive differences
- R^+: the sum of the ranks of negative differences

Differences: Wilcoxon matched-pairs test

- pairs or differences *not* normally distributed
- based on ranked *differences*
  - first calculate the differences
  - second rank the differences
  - 0's not ranked
- sum and compare +ve and -ve ranks

$$W^+ = \sum R^+$$

$$W^- = \sum R^-$$

- smallest is used to find the p-value
- lower W-values are more likely to be significant

Group Assignment

Using the whale count data, compare the differences between first and second abundance guesstimates using first `excel` and then `R`.

Submit a single written group report that outlines the following points:

1. state the null and alternative hypotheses being tested
2. the reason for choosing the statistical test you used
3. a summary of the results:
   - degrees-of-freedom, test statistic, p-values (at 5% level)?
   - did you accept or reject the null hypothesis?
   - is there a difference?
4. conduct the analysis in `R` and `excel` and submit:
   - a written report of points 1, 2 and 3 as **PDF**
   - an `excel` workbook showing your results
   - a saved `R` file showing results