# Homework 7: New York City Trees

## **Comparing a Numerical Variable Across Two Groups**

### **New York City Trees**

New York City's trees shade us in the summer, beautify our neighborhoods, help reduce noise, and support urban wildlife. Beyond these priceless benefits, our urban forest provides us a concrete return on the financial investment we put into it. This return includes stormwater interception, energy conservation, air pollutant removal, and carbon dioxide storage. Our publicly owned trees are as much of an asset to us as our streets, sewers, bridges, and public buildings.

This dataset includes a record for every tree in New York City and includes the tree's location by borough and latitude/longitude, species by Latin name and common names, size, health, and issues with the tree's roots, trunk, and branches.

```
head(nyc_trees)
```

200170

```
# A tibble: 6 x 6
 tree_id tree_dbh status spc_common sidewalk orig.id
   <dbl>
             <dbl> <chr>
                         <chr>
                                      <chr>
                                               <chr>
  269843
               11 Alive honeylocust Damage
                                               51727
1
2
  241665
               15 Alive honeylocust NoDamage 86978
3 585168
               12 Alive honeylocust NoDamage 159
  611956
                2 Alive honeylocust NoDamage 49967
  235391
               20 Alive honeylocust NoDamage 65210
```

**Research question** Is the average tree diameter of a Norway Maple different than the tree diameter of a Honey Locust?

13 Alive honeylocust NoDamage 41318

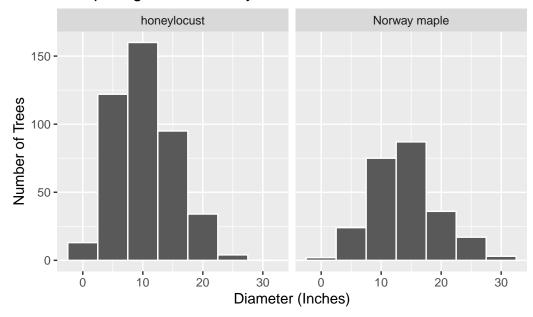
```
favstats(tree_dbh ~ spc_common, data = nyc_trees)
```

```
      spc_common min
      Q1 median
      Q3 max
      mean
      sd
      n missing

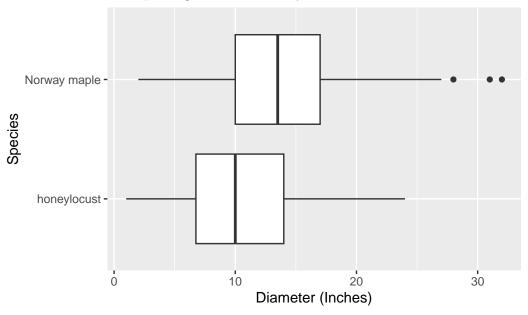
      1 honeylocust
      1 6.75
      10.0
      14 24 10.25234 4.983990 428
      0

      2 Norway maple
      2 10.00
      13.5
      17 32 13.94672 5.595526 244
      0
```

# Comparing New York City Trees

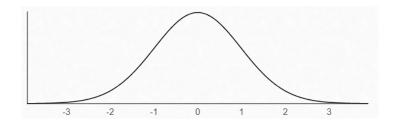


# Comparing New York City Trees



- 1. Identify the variables and data type.
- Explanatory:
- Response:
- 2. Is this an experimental study or observational study? Explain.
- 3. Discuss the visual comparison of the two tree species. Describe their distributions (recall: 4 characteristics). Which tree species appears to have a higher average diameter?
- 4. Identify the observed mean, standard deviation, and sample size of each species in the study.

5.	Calculate the observed difference in average diameter (Norway Maple - Honeylocust). Assign an appropriate symbol.
6.	State the parameter in words (\$ $Norway\ Maple$ - Honeylocust).
7.	Write the null and alternative hypotheses using the appropriate mathematical symbols.
8.	Check the assumptions necessary to conduct a two-sample independent t-test.
9.	Calculate the t-statistic.
10.	Using the t-distribution below, show how you would calculate/estimate the p-value. Yes, this is going to be a little funky given the magnitude of the t-statistic.



11. The following R code conducts a two-sample t-test for the data. What do you think the order = argument does? What do you hypothesize would happen to the results if we used order = c("honeylocust", "Norway maple") instead?

12. Using the output above, write a conclusion in the context of the problem.

13. Where can we generalize our results (i.e. what is our population in which we obtained the observed data)?

14. Construct a 95% confidence interval for the true difference in mean diameter between Honeylocust and Norway Maple trees.

### Paired t-test vs Two-Sample Independent t-test

For each of the following scenarios, indicate whether a paired t-test analysis (dependent groups) or a two-sample independent t-test analysis (independent groups) would be appropriate based on the details given.

a. You wonder whether undergraduate or graduate students tend to go to more football games. You take a random sample of 20 undergraduate students and then another random sample of 20 graduate student, and ask how many games they attended last fall.

PAIRED TWO-SAMPLE

b. On average, is there a significant difference in one's weight before and after the holidays?

PAIRED TWO-SAMPLE

c. Is autism marked by different brain growth patterns in early life, even before an autistic diagnosis is made? Studies have linked brain size in infants and toddlers to a number of future ailments, including autism. One study looked at the brain sizes of 30 autistic boys and 13 non-autistic boys who all had received an MRI scan as toddlers. The whole-brain volume (in ml) was recorded for each boy.

PAIRED TWO-SAMPLE

d. Woo and McKenna (British Journal of Dermatology, 2003) investigated the effect of broadband ultraviolet B (UVB) therapy and topical calcipotriol cream used together on area of psoriasis. One of the outcome variables from the study was the Psoriasis Area and Severity Index (PASI). The PASI score of twenty subjects were measured at baseline and after eight treatments.

#### **PAIRED**

#### TWO-SAMPLE

e. A researcher interviews 64 students in their freshman year and then again in their senior year to determine if their opinion has changed regarding their ideal number of children.

PAIRED

TWO-SAMPLE

f. Some archaeologists theorize that ancient Egyptians interbred with several different immigrant populations over thousands of years. To see if there is any indication of changes in body structure that might have resulted, researchers measured the breadth (measured in mm) of a random sample of male Egyptian skulls dated from 4000 B.C. and from 200 B.C. (Ancient Races of the Thebaid, 1905).

PAIRED

TWO-SAMPLE



Canvas Quiz

Make sure to complete the Homework Quiz on Canvas.