CSCI 443: LECTURE 18 NON-BINOMIAL HYPOTHESIS TESTS, T-TESTS.

Professor David Harrison

OFFICE HOURS

Tuesday	4:00-5:00 PM
Wednesday	12:30-2:30 PM

BLACKBOARD & GITHUB

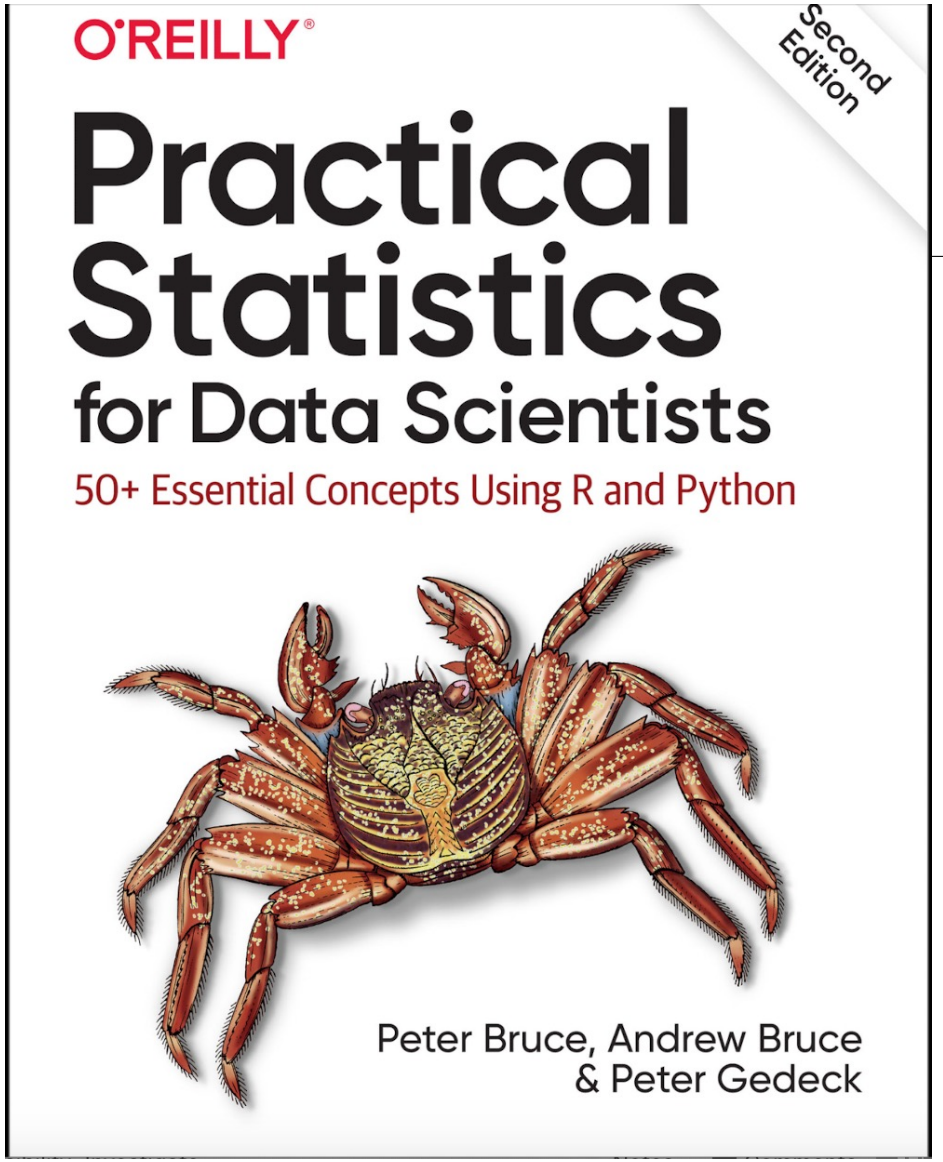
Slides up, handwritten notes AND a jupyter notebook for lecture 16 are on blackboard and in GitHub.

The project is at

https://github.com/dosirrah/CSCI443_AdvancedDataScience

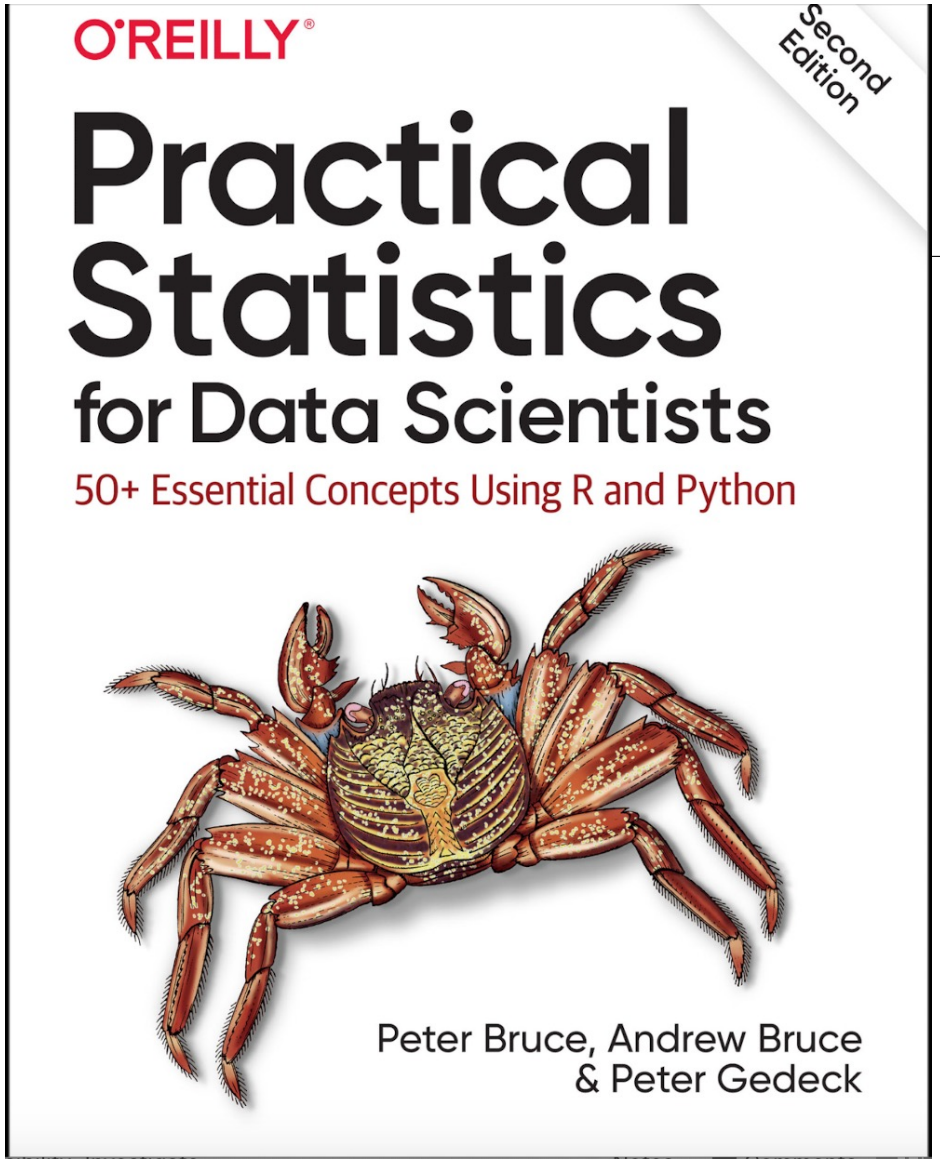
READ ABOUT

- chapter 3: experiments, hypothesis testing
  - Degrees of Freedom
  - Permutation Tests
  - T-tests
  - ANOVA
  - Chi-square



THINGS I WANT TO COVER TODAY

- Non-binomial Hypothesis Tests
- T-tests



PREVIOUS LECTURE: AD COMPARISON

A:

B:

1000 views of each.

0.1% click through rate for A.

0.5% click through rate for B.

This is a hypothesis test between binomial random variables.

MODIFIED FROM PREVIOUS LECTURE: AD COMPARISON

$\hat{p}_A$ = click through rate for A
= Bin(n_A, p_A) / n_A

$\hat{p}_B$ = click through rate for B
Bin(n_B, p_B) / n_B

A:

B:

NON-BINOMIAL HYPOTHESIS TESTS

What if hypothesis test for random variables that are NOT binomial.

Example: trial for Novo Lilly's new life-extension drug Zombivia.

X = systolic blood pressure for patient in control group.

Y = systolic blood pressure for patient in exposed group.

X and *Y* are continuous numerical random variables

$$H_0 : \mu_x = \mu_y$$

$$H_A : \mu_x \neq \mu_y$$

NOVO LILLY'S ZOMBIVIA TRIAL: T-TESTING

We will perform a two-sample t-test.

It is a two sample because we are comparing a statistic from one sample set (exposed) against another (control).

You can read about the one-sample t-test variant, which compares a statistic of a sample with a known or hypothesized population mean.

Step 0: Choose alpha

$$\alpha = 0.05$$

ZOMBIVIA TRIAL: STEP 1

Step 1: Collect Data

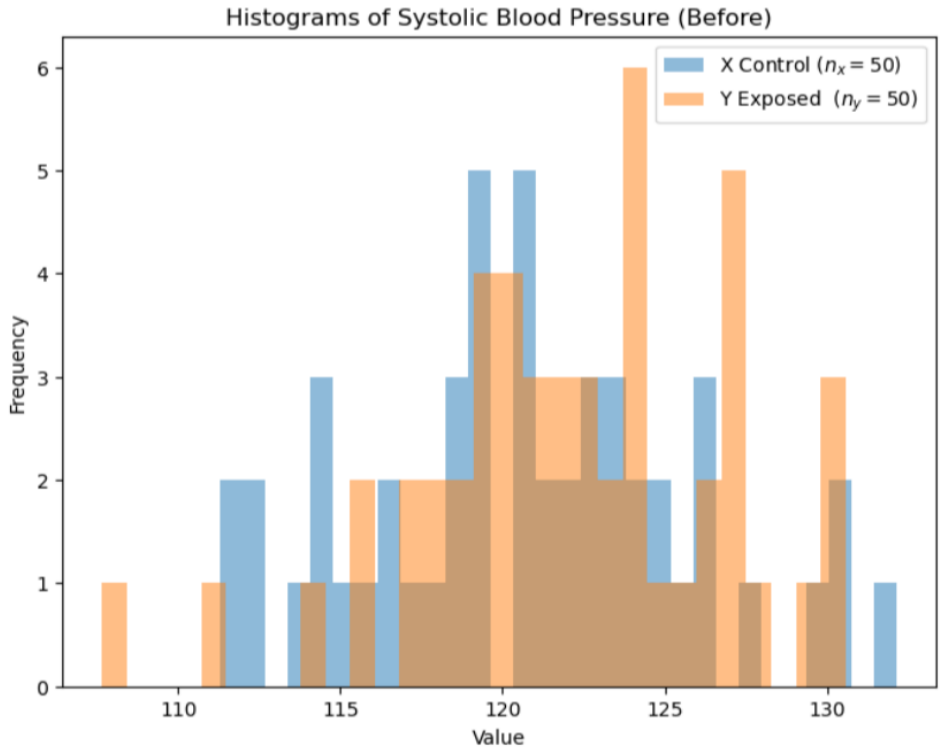
Gather systolic blood pressure measurements from patients in both the control group and the exposed group.

Ensure that the sample sizes (n_x for the control group and n_y for the exposed group) are sufficient to detect a significant difference

ZOMBIVIA TRIAL: STEP 1 COLLECT DATA

Measures blood pressure before.

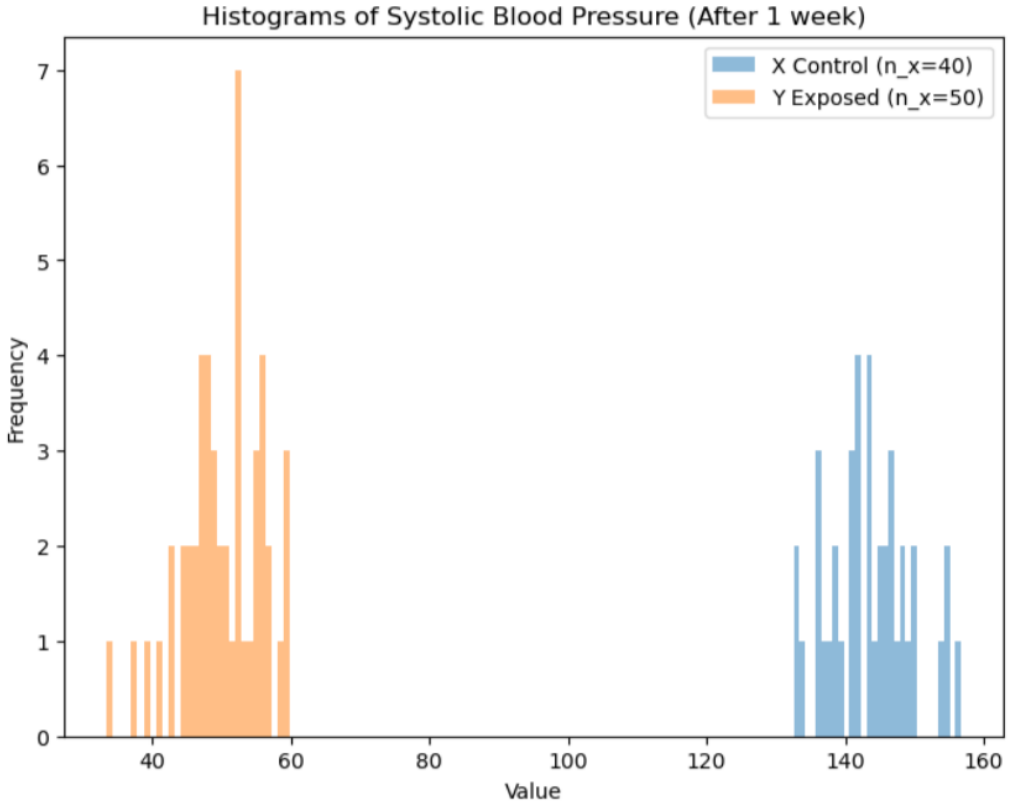
- Just to detect obvious sampling bias.



$$\bar{x} = 120.8, \quad \bar{y} = 122.1$$

ZOMBIVIA TRIAL: STEP 1 COLLECT DATA

Systolic blood pressure after 1 week.



ZOMBIVIA TRIAL: STEP 2

Check data for appearances of bias or methodological failures in collection.

WE NOTICE THREE POSSIBLE ISSUES:

1. The blood pressure in the control group seems to have risen.

BEFORE:

$$\bar{x} = 120.8, \quad \bar{y} = 122.1$$

AFTER 1 WEEK:

$$\bar{x} = 143.6, \quad \bar{y} = 50.0$$

- 143.6 is typical of stage 2 hypertension

ZOMBIVIA TRIAL: STEP 2: ISSUE 1

The increase in blood pressure could be

Performance bias:

- Performance bias refers to systematic differences in the care provided, or in the exposure to factors other than the interventions of interest, that participants receive in a study.
- emerges from participants or caregivers acting differently because they are aware of which treatment or intervention participants are receiving
    - But there is no way this could happen because the study staff and nurses are double blinded.

Nocebo Effect:

- A sense of disappointment or hypertension due to correctly deducing that they are members of the control group.
    - But how would they guess that they aren't in the exposed group?

ZOMBIVIA TRIAL: STEP 2 ISSUE 2

There is a marked reduction in blood pressure in the exposed group.

BEFORE:

$$\bar{x} = 120.8, \quad \bar{y} = 122.1$$

AFTER 1 WEEK:

$$\bar{x} = 143.6, \quad \bar{y} = 50.0$$

ZOMBIVIA TRIAL: STEP 2 ISSUE 2

BEFORE:

$$\bar{x} = 120.8, \quad \bar{y} = 122.1$$

AFTER 1 WEEK:

$$\bar{x} = 143.6, \quad \bar{y} = 50.0$$

But this is the exposed group. It could be a regular effect of the drug!

ZOMBIVIA TRIAL: STEP 2 ISSUE 2

Although this may be an effect of the drug, it is an extreme result.

When this occurs, the trial should be aborted.

This trial exhibits either missing or inadequate

stopping criteria

ZOMBIVIA TRIAL: STEP 2: ISSUE 3

3. Seems 10 people are missing from the control group.

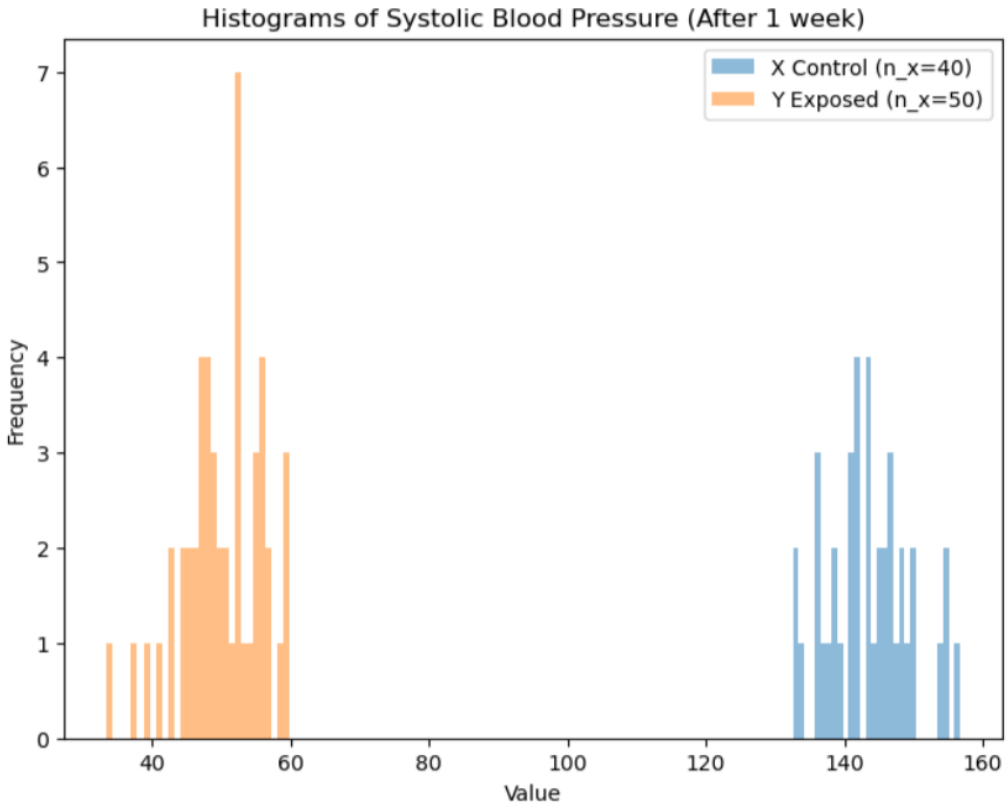
- This couldn't be a real problem, so we will ignore it.
- Just remove the 10 missing from the trial.
- **???**

ZOMBIVIA TRIAL: STEP 3

Step 3: Test Assumptions

Assumptions of a two-sample t-test.

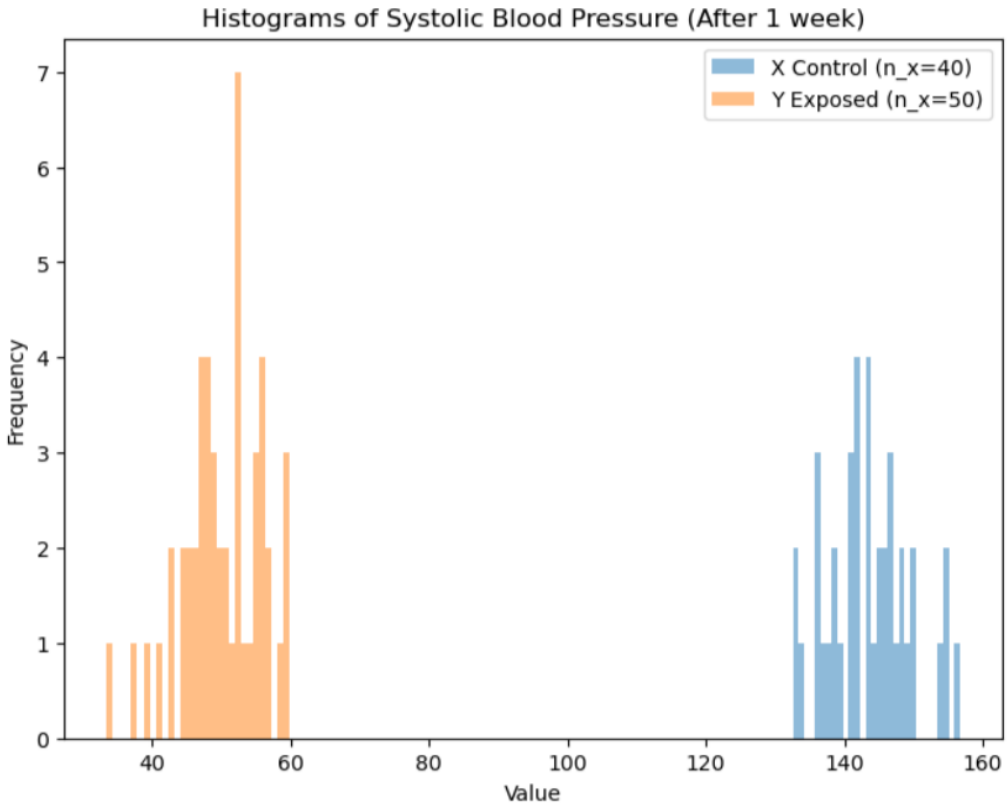
- **Independence:** Systolic blood pressure measurements of *X* and *Y* should be independent.
- **Normality:** Distributions of X and Y should be approximately normal. Look for skew. Look for heavy tails.
- **Equal variances**: The sample variances in both groups should be roughly equal.



ZOMBIVIA TRIAL: STEP 3

Step 3: Test Assumptions (cont.)

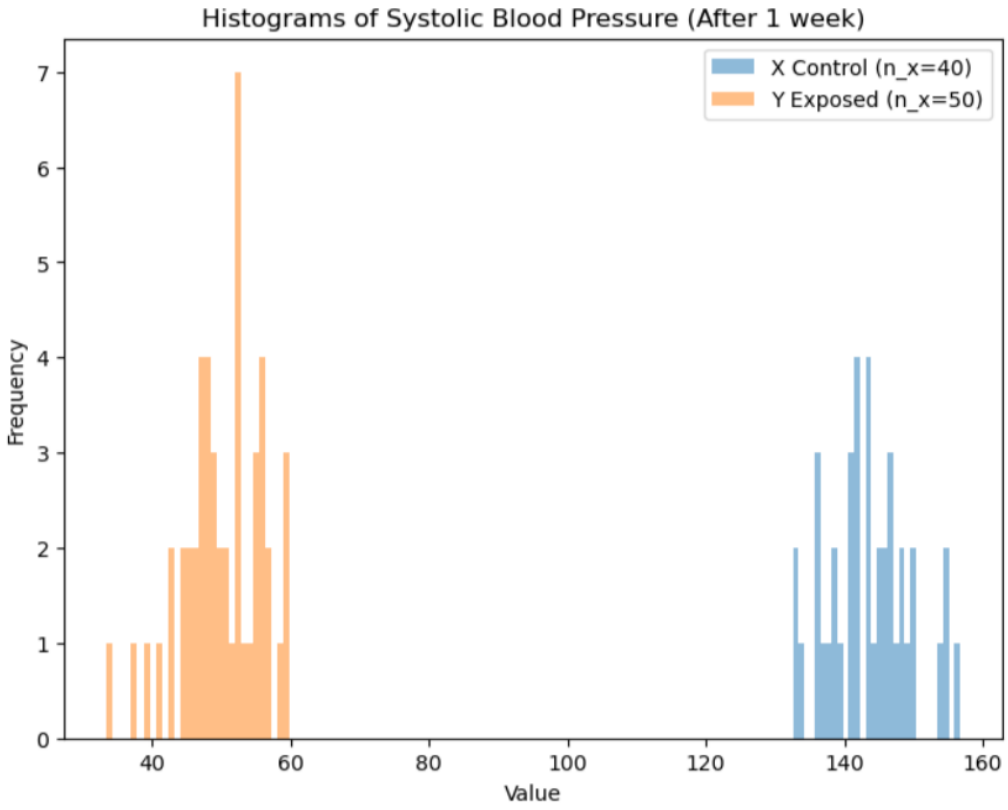
- **Independence:** Systolic blood pressure measurements of *X* and *Y* should be independent.
  - This assumption is true because trial is double blinded.
  - The patients don't know so nothing would affect their blood pressure in any way different between groups.
  - The personnel don't know so they treat control and exposed groups identically.



ZOMBIVIA TRIAL: STEP 3

Step 3: Test Assumptions (cont.)

- **Normality:** Distributions of X and Y should be approximately normal. Look for skew. Look for heavy tails This assumption is true because trial is double blinded.
  - I don’t see significant skew.
  - No heavy tails.
  - Assume normal.



ZOMBIVIA TRIAL: STEP 3

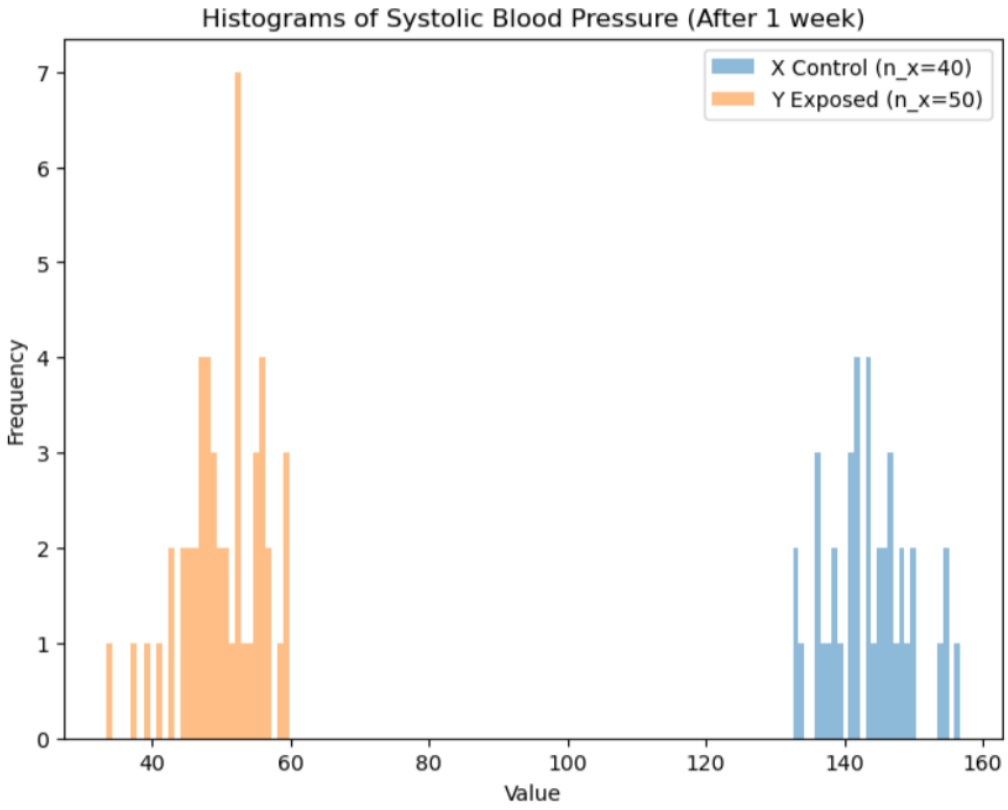
Step 3: Test Assumptions (cont)

- **Equal variances**: The sample variances in both groups should be roughly equal.

$$s_x = 6.0, \quad s_y = 5.8$$

- Pretty close.

- This is known as *homoscedasticity*



ZOMBIVIA TRIAL: STEP 4

Step 4: Calculate the test statistic.

First calculate the pooled variance

If the null hypothesis is true then the underlying distributions are identical between X and Y and so we create a single variance called a *pooled variance*.

> Let n_x = number of patients in the exposed group.
>
> Let n_y = number of patients in the control group.

When $n_x = n_y$,

$$s_p^2 = \frac{s_x^2 + s_y^2}{2}$$

ZOMBIVIA TRIAL: STEP 4

In the Zombivia trial, n_x and n_y started equal, but for some reason, members of the control group seemed to have disappeared.

Now $n_x < n_y$.

When $n_x \neq n_y$ then we use a weighted average weightd by the degrees of freedom in each sample set.

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

$$s_x^2 = 36.3, \quad s_y^2 = 33.8, \quad s_p^2 = 34.9$$

ZOMBIVIA TRIAL: STEP 4

$$s_x^2 = 36.3, \quad s_y^2 = 33.8, \quad s_p^2 = 34.9$$

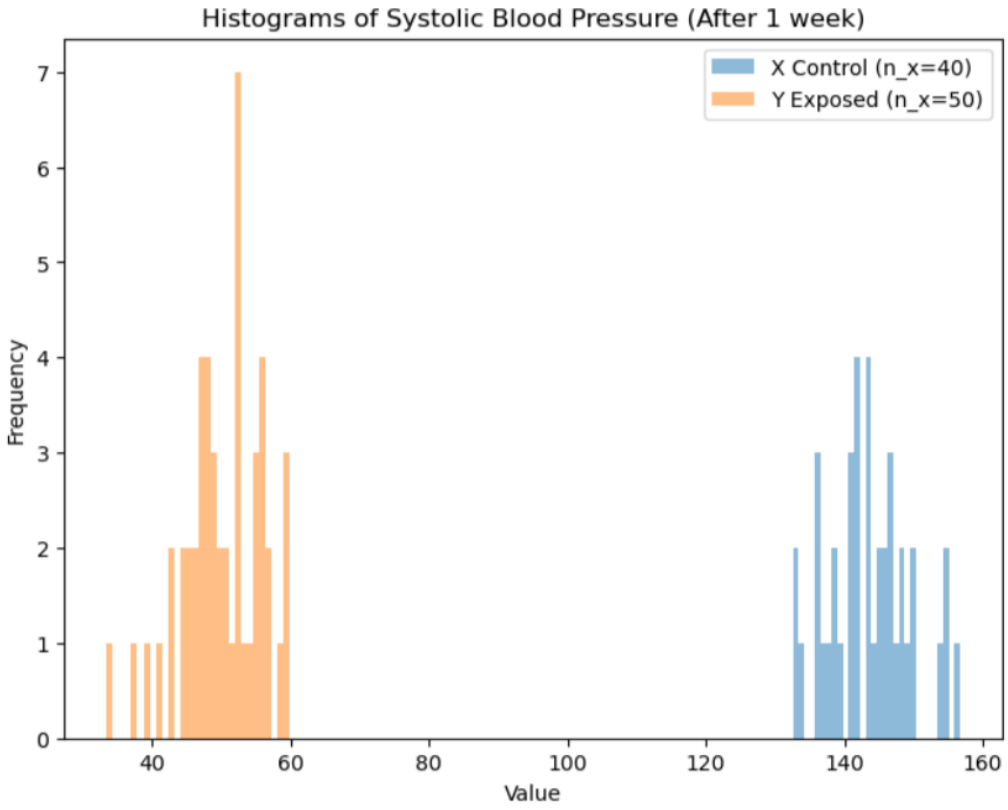
$$s_p = \sqrt{s_p^2}$$

$$s_x = 6.0, \quad s_y = 5.8, \quad s_p = 5.9$$

Standard Error of the difference between X and Y is

$$SE_{\bar{x}-\bar{y}} = \sqrt{s_p^2(\frac{1}{n_x} + \frac{1}{n_y})}$$

Test statistic: $$t = \frac{\bar{x} - \bar{y}}{SE_{\bar{x}-\bar{y}}}$$



ZOMBIVIA TRIAL: STEP 4

Test statistic:

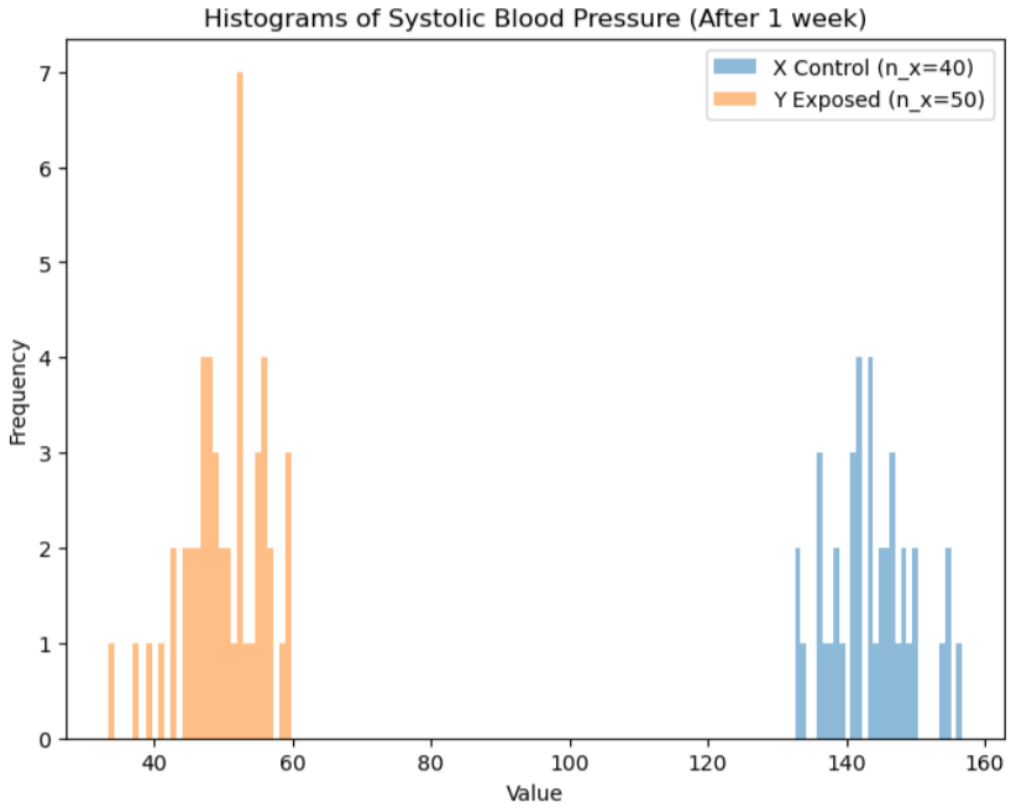
$$t = \frac{\bar{x} - \bar{y}}{SE_{\bar{x}-\bar{y}}}$$

$$\bar{x} = 143.6, \quad \bar{y} = 50.0$$

$$SE_{\bar{x}-\bar{y}} = 7.40$$

.

$$t = 12.65$$



ZOMBIVIA TRIAL: STEP 4

$$t = 12.65$$

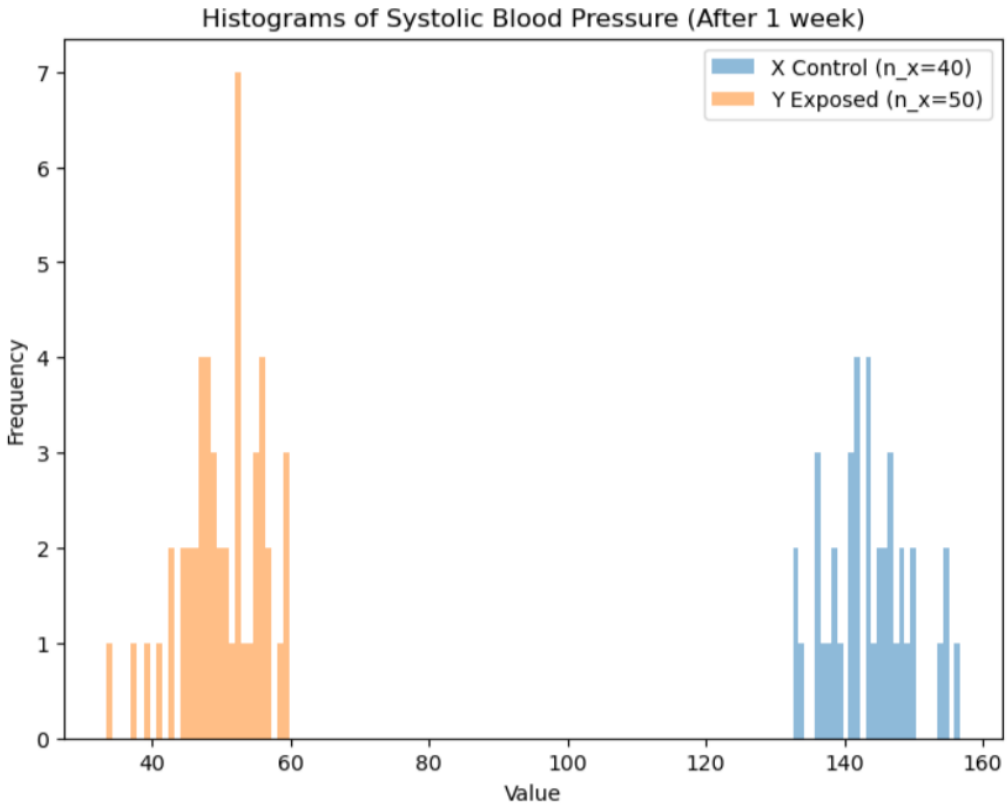
Compute degrees of freedom:

$$df = n_X + n_Y - 2$$

$$df = 88$$

Compute p-value.

$$p = 1.659307456321507e - 21$$



ZOMBIVIA TRIAL: STEP 4

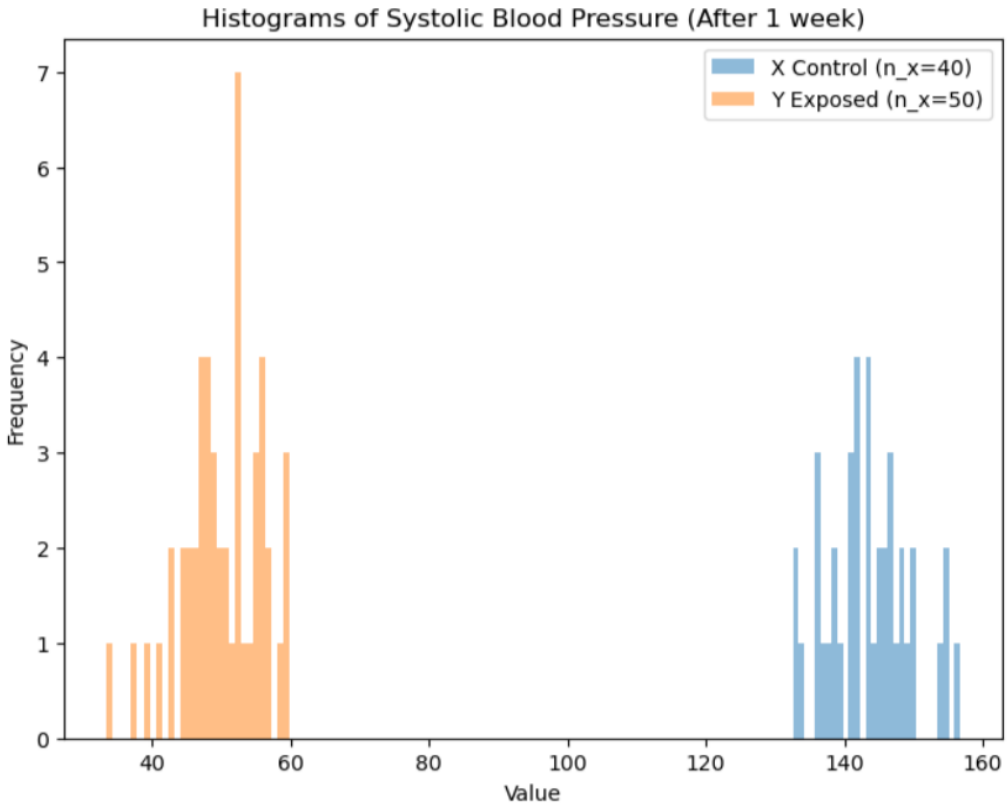
$$p = 1.6593074563215 07e - 21$$

Alpha was

$$\alpha = 0.05$$

p value is far below alpha so.

We <u>reject</u> the null hypothesis. The systolic blood pressure significantly differs between control and exposed groups.



ZOMBIVIA TRIAL: STEP 4

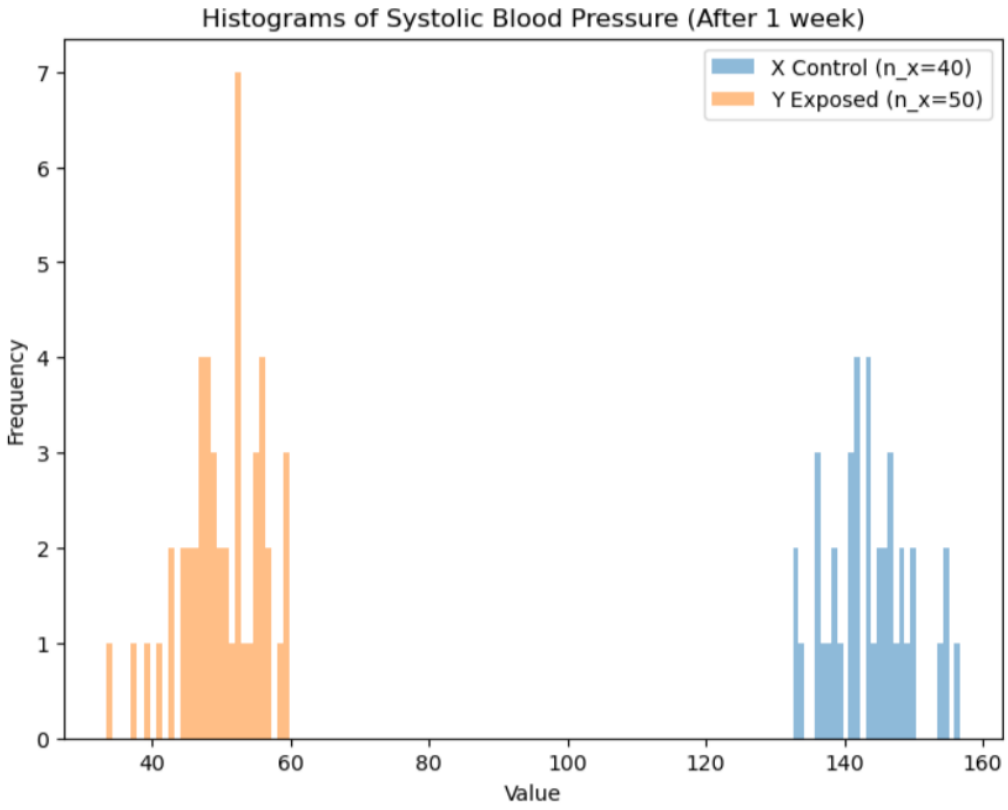
$$p = 1.6593074563215 07e - 21$$

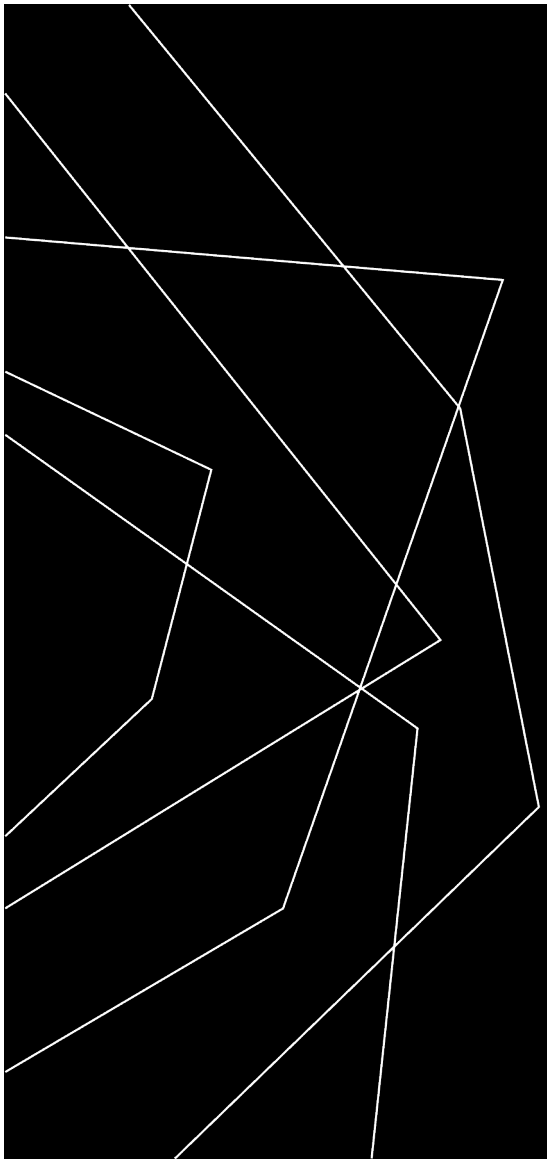
Alpha was

$$\alpha = 0.05$$

p value is far below alpha so.

We <u>reject</u> the null hypothesis. The systolic blood pressure significantly differs between control and exposed groups.





THANK YOU

David Harrison

Harrison@cs.olemiss.edu