Final Exam Review

PH142 Fall 2025

Logistics

- **Date:** Wednesday, December 17th
- **Time:** 11:40AM–2:30PM, arrive no later than 11:30AM
- **Location(s):** Dwinelle 155, Cory 241 (DSP)
- **Material Covered:** Lectures 23–37, Lab 8–11
  - The exam is cumulative with an emphasis on Part 3
  - The exam is designed to be completed within 2hrs

What to Bring

- **Student ID**
- **Pencil/Pen**
- **Cheat Sheet** (double sided, handwritten, 8.5x11”)
- **Scientific Calculator** (non-graphing)

T-tests

T-tests

T-test types

- One-sample
- Two-sample
- Paired

T-tests

One Sample t-test

Goal: To test whether the mean of the sample is different than the mean of the population

Assumptions:

- SRS with independent observations
- The sampling distribution of the sample mean is approximately Normal

T-tests

One Sample t-test

Hypotheses:

$$H_0 : \mu = \mu_0$$

$$H_A : \mu \neq \mu_0, \text{ or } \mu < \mu_0, \text{ or } \mu > \mu_0$$

Test Statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

T-tests

Two Sample t-test

Goal: To test if two population means are the same or different

Assumptions:

- SRS with independent observations
- Each sample mean has an approximately Normal sampling distribution

T-tests

Two Sample t-test

Hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 \neq 0 \text{ (or } < \text{ or } > \text{)}$$

Test Statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

T-tests

Paired t-test

Goal: To test if there is a difference in means for the *same* individual at *different* points in time

Assumptions:

- SRS with matched observations
- The distribution of paired differences is approximately Normal

T-tests

Paired t-test

Hypotheses:

$$H_0 : \mu_d = 0$$

$$H_A : \mu_d \neq 0, \text{ or } \mu_d < 0, \text{ or } \mu_d > 0$$

Test Statistic:

$$t = \frac{\bar{x}_d - \mu_d}{s_d / \sqrt{n}}$$

One Sample Z Test	One Sample T Test	Two Sample T Test	Paired T Test
- SRS, independent observations - Sampling Distribution is Normal - Population SD is known	- SRS, independent observations - Sampling Distribution is Normal	- SRS, independent observations - Sampling Distribution of *both* treatments is Normal	- SRS with matched observations - Sampling Distribution of difference is Normal
$H_0 : \mu = 154$ $H_a : \mu \neq 154 \text{ (or } < \text{ or } >)$	$H_0 : \mu = 154$ $H_a : \mu \neq 154 \text{ (or } < \text{ or } >)$	$H_0 : \mu_1 - \mu_2 = 0$ $H_a : \mu_1 - \mu_2 \neq 0 \text{ (or } < \text{ or } >)$	$H_0 : \mu_d = 0$ $H_a : \mu_d \neq 0 \text{ (or } < \text{ or } >)$
$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ z follows a Normal Distribution with Mean 0 and SD 1	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ t follows a t-distribution with n-1 degrees of freedom	$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ t follows a t-distribution with a COMPLICATED degrees of freedom	$t = \frac{\bar{x}_d - \mu_d}{s_d/\sqrt{n}}$ t follows a t-distribution with n-1 degrees of freedom (where n is the number of pairs)
$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$	$(\bar{x_1} - \bar{x_2}) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$	$\bar{x}_d \pm t^* \frac{s_d}{\sqrt{n}}$

Inference for Proportions

Inference for Proportions

One Sample Proportion Test

Goal: To test if a population proportion equals a claimed value

Assumptions:

- SRS with independent observations
- Sampling distribution is normal ($np_0 \geq 10 \text{ and } n(1-p_0) \geq 10$)

Inference for Proportions

One Sample Proportion Test

Hypotheses:

$$H_0 : p = p_0$$
$$H_a : p \neq p_0 \ (\text{or} < \text{or} >)$$

Test Statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Inference for Proportions

Two Sample Proportion Test

Goal: To test if two population proportions are the same or different.

Assumptions:

- SRS from each independent population
- Sampling distribution is normal (observed successes and failures are > 10 for both samples)

Inference for Proportions

Two Sample Proportion Test

Hypotheses: $H_0 : p_1 - p_2 = 0$
$H_a : p_1 - p_2 \neq 0 \text{ (or } < \text{ or } >)$

Test Statistic: $$z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Inference for Proportions

Confidence Intervals for Proportions

1. Large Sample method
2. Plus Four method
3. Wilson Score method
4. Exact (Clopper Pearson) method

Inference for Proportions

1. Large Sample Method CI

- Generally has low coverage
- If we do not have a sampling distribution that approaches Normal, this confidence interval does not perform well

For One Sample Proportion: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

For Two Sample Proportion: $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Inference for Proportions

2. Plus Four Method CI

- Use when n ≥ 10 and confidence level ≥ 90%
- Add 2 imaginary successes and 2 failures to the dataset (increasing the sample size by 4 imaginary trials)

For One Sample Proportion:

$$\tilde{p} = \frac{\text{number of successes} + 2}{n+4}$$

$$\tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$$

For Two Sample Proportion:

$$\tilde{p}_1 = \frac{\text{no. of successes in pop1} + 1}{n_1+2}$$

$$\tilde{p}_2 = \frac{\text{no. of successes in pop2} + 1}{n_2+2}$$

$$(\tilde{p}_1 - \tilde{p}_2) \pm z^* \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}}$$

Inference for Proportions

3. Wilson Score Method CI

- For the Wilson Score method, use the prop.test() function
- The default with the prop.test() function is two-sided

Example: `prop.test(x = 100, n = 500, conf.level = 0.95)`

Inference for Proportions

4. Clopper Pearson Method CI

- For the Clopper Pearson method, use the binom.test() function
- This method is statistically conservative and it gives better coverage than it suggests

Example: `binom.test(x = 100, n = 500, conf.level = 0.95)`

Inference for Proportions

Sample Size for a Proportion

- We use the following formula to estimate a sample size for a proportion within a given margin of error

$$N = \left(\frac{z}{m}\right)^2 p * (1 - p)$$

- Z = critical value
- m = the desired margin of error
- p* = your best estimate for the underlying proportion

One Sample Proportion Test	Two Sample Proportion Test
- SRS with independent observations - Outcome variable is binary - Sampling dist. is Normal $np_0 \geq 10$ and $n(1-p_0) \geq 10$)	- SRS from each independent population - Outcome variable is binary - Sampling distribution is Normal (the number of observed successes and failures are > 10 for both samples)
$H_0: p = p_0$ $H_a: p \neq p_0$ (or $<$ or $>$)	$H_0: p_1 - p_2 = 0$ $H_a: p_1 - p_2 \neq 0$ (or $<$ or $>$)
$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ z follows a Normal Distribution with Mean 0 and SD 1	$z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ z follows a Normal Distribution with Mean 0 and SD 1
Large Sample Method CI: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	Large Sample Method CI: $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
Plus Four Method CI: $\tilde{p} = \frac{\text{number of successes} + 2}{n+4}$ $\tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$	Plus Four Method CI: $\tilde{p}_1 = \frac{\text{no. of successes in pop1} + 1}{n_1+2}$ $\tilde{p}_2 = \frac{\text{no. of successes in pop2} + 1}{n_2+2}$ $(\tilde{p}_1 - \tilde{p}_2) \pm z^* \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}}$

Bootstrapping for Confidence Intervals

Bootstrapping for CIs

Bootstrapping Overview

Goal: To estimate parameters, SEs, and CIs without making strong assumptions about the underlying data distribution

- The Bootstrap does *not* require the underlying distribution to be Normally distributed
- The Bootstrap can be made for any parameter based on its sample statistic

Bootstrapping for CIs

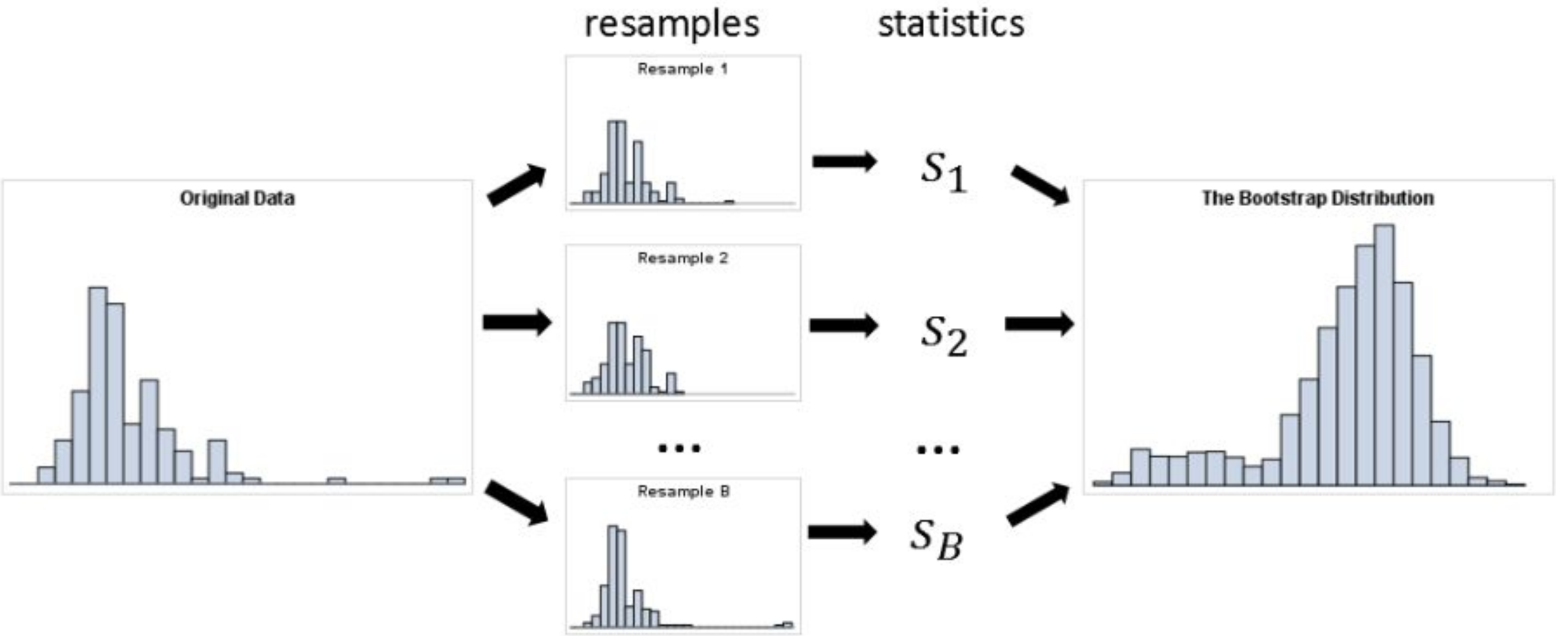
Bootstrapping Steps

Setup: We have a sample of size *n*. We want to estimate some statistic θ. For this example we are interested in the CI for the median, denoted by θ.

Steps:

1. Take a sample of size *n* **with replacement** from our sample. Calculate θ(hat) of this sample
2. Do this a large number of times (1000, 10000, etc). Each time calculate θ(hat) and denote it
3. Create a histogram of all your θ(hat)s
4. Use the histogram to generate a confidence interval using the quantile functions in R

Bootstrapping for CIs



Chi-Square

Chi-Square Tests

Chi-Square Tests

Goal: To determine whether your data are significantly different from what you expected.

Chi-Square tests are used for categorical data.

There are two types:

- Chi-Square Goodness of Fit
- Chi-Square Test for Independence

Chi-Square Tests

Chi-Square Tests

Chi-Square Test Statistic: $$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

k = the number of cells in the table
O = the observed count for the i th group
E = the expected count for the i th group

Code: `pchisq(q, df, lower.tail = F)`

Chi-Square Tests

Chi-Square Goodness of Fit

Goal: To test how well the observed counts "fit" the expected counts.

H_0: Proportion in observed is the same as the proportion from a given population

H_A: At least one of p_k is not equal to the proportion stated in the null hypothesis

$p_1 = \#_1, p_2 = \#_2, \ldots, p_k = \#_k$

where $\#_1, \#_2, \ldots, \#_k$ are provided or can be derived from percentages provided

Chi-Square Tests

Chi-Square Goodness of Fit Conditions

- Fixed n of observations
- All observations are independent
- The probability of a given outcome is the same for each observation
- At least 80% of the cells have 5 or more observations ($O_i \geq 5$ for $\geq 80\%$ of the cells)
- All expected values greater than 1 ($E_i > 1$)

Chi-Square Tests

Chi-Square Test for Independence

Goal: To test whether two categorical variables are related to each other.

H_0: Response and explanatory variables are independent

H_A: Response and explanatory variables are dependent

Chi-Square Tests

Chi-Square Test for Independence Conditions

- Independent SRSs from ≥ 2 populations, with each individual classified according to one category
- A single SRS, with each individual classified according to each of two categorical variables
- At least 80% of the cells have 5 or more observations ($O_i \geq 5$ for ≥ 80% of the cells)
- All expected values greater than 1 ($E_i > 1$)

Chi Square Test for Goodness of Fit	Chi Square Test for Independence
- Fixed number n, independent of observations - Each observation falls into just one of the k mutually exclusive categories - The probability of a given outcome is the same for each observation. - At least 80% of the cells have 5 or more observations expected AND all cells have expected counts > 1	- Independent SRSs from ≥ 2 population, with each individual classified according to one category - A single SRS, with each individual classified according to each of two categorical variables. - At least 80% of the cells have 5 or more observations expected AND all cells have expected counts > 1
$H_0 : p_1 = ?,\ p_2 = ?,\ \ldots\ ,\ p_k = ?$ H_a : At least one of the p_i is not as stated above	H_0 : Response and explanatory variables are independent. H_a : Response and explanatory variables are dependent.
$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ X^2 follows a Chi-Squared distribution with k-1 degrees of freedom (k categories)	$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ $\quad E_i = \frac{\text{row total} \times \text{col total}}{\text{overall total}}$ X^2 follows a Chi-Squared distribution with (r-1)*(c-1) degrees of freedom (r rows and c columns)
P-value is always area to the RIGHT of our test statistic for Chi-squared tests	P-value is always area to the RIGHT of our test statistic for Chi-squared tests

Permutation Tests

Permutation Tests

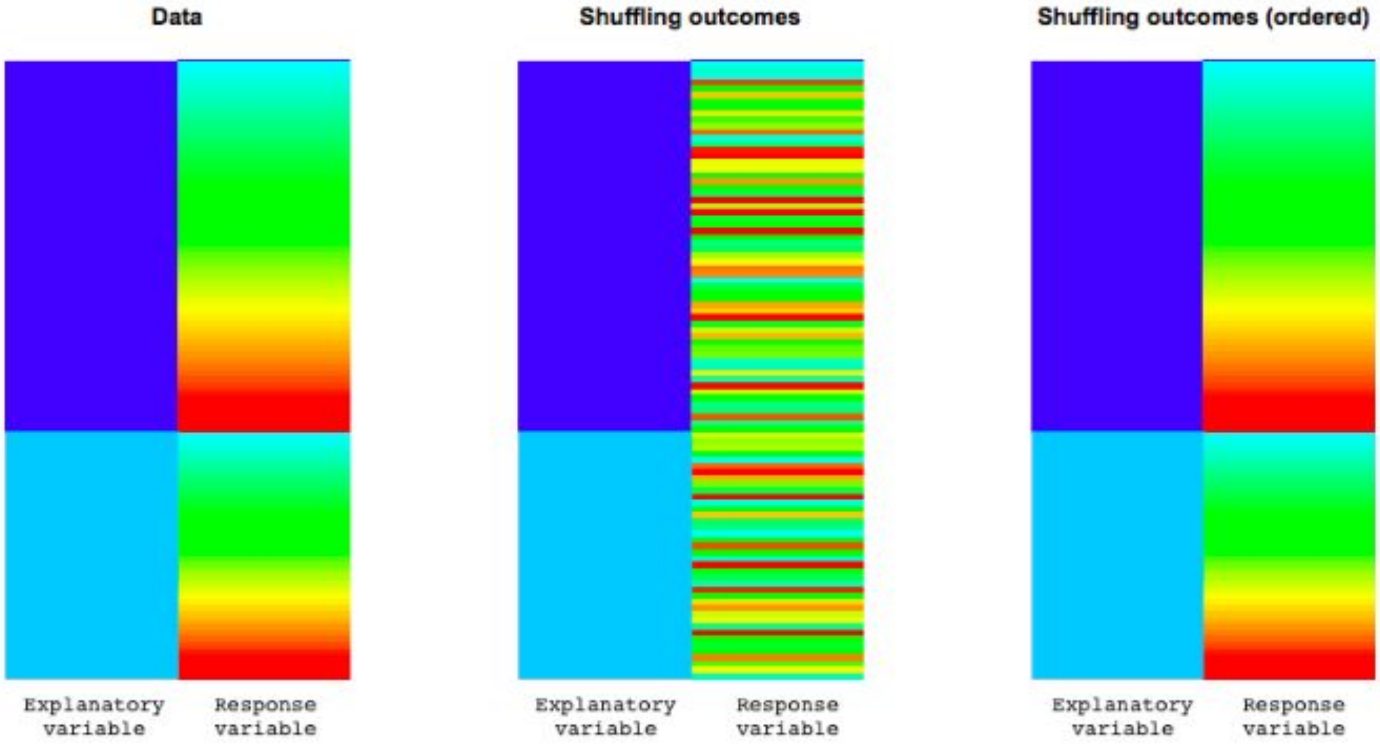
Permutation Test (Hypothesis Test)

Recall: the z test, t test, and chi-square test require large sample size and SRS assumptions. Permutation test does *not* require those assumptions.

Goal: Use random shuffling of the labels to create a null distribution to test our hypothesis

H0: $\mu 1 = \mu 2$ (there is no difference between the groups)

Permutation Tests



Permutation Tests

Permutation Test Steps

1. Compute the mean and take the difference between groups
2. Shuffle the labels between the groups
3. Re-compute the mean and difference between groups
4. Repeat steps 2-3 to get multiple test statistics
5. Calculate the p-value

Inference for Regression

Inference for Regression

Inference for Regression Assumptions

1. **Linearity:** The relationship between x and y is linear in the population
2. **Normality of Residuals:** Variable y varies Normally around the line of best fit
3. **Independence:** Observations are independent
4. **Constant Variance:** The standard deviation of the responses is the same for all values of x

Inference for Regression

Evaluating Assumptions

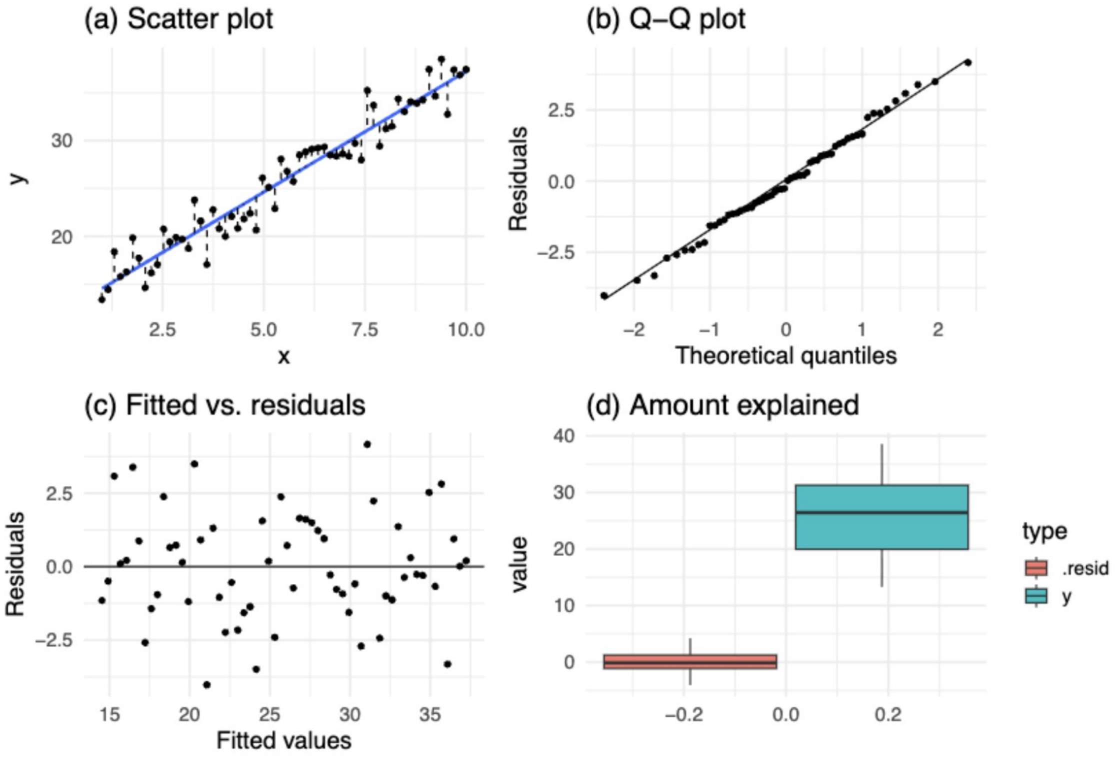
We can use different diagnostic plots to test our assumptions:

	What to Look For
Scatter plot of data w/ regression line	No trend, residuals are sometimes positive and sometimes negative
Q-Q plot of residuals	Residuals closely follow the line
Scatter plot of fitted vs. residuals	No trend, random scatter
Box plot of y vs. residuals	IQR of residuals box smaller than y box

Inference for Regression

Good Example

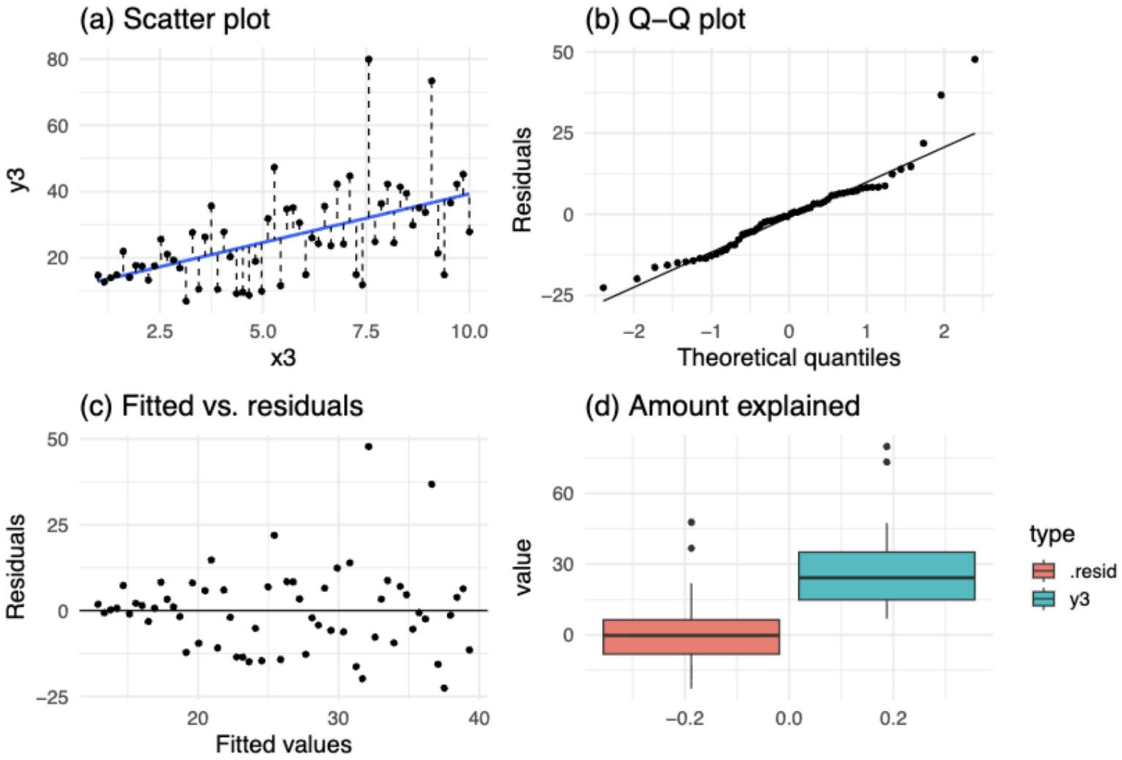
(Assumptions met)



Inference for Regression

Bad Example

(Assumptions not met)



Inference for Regression

Regression Standard Error

Squared Estimates of Error (SSE): used to determine the sum of differences between each y value and the fitted value based on the line of best fit.

$$SSE = \sum_i^n (y_i - \hat{y}_i)^2$$

Higher SSE = worse model fit.

Regression Standard Error: measures the spread of residuals around the regression line.

$$s = \sqrt{\frac{1}{n-2} \times SSE}$$

Inference for Regression

Hypothesis Testing for Regression

H0: b = 0 (the slope is equal to zero)
HA: b ≠ 0 (the slope is not equal to zero)

Compute in R

Test statistic: $t = \frac{\hat{b}}{SE_b}$

SE: $SE_b = \frac{s}{\sqrt{\sum(x-\bar{x})^2}}$

Regression SE: $s = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y-\hat{y})^2}$

Inference for Regression

Confidence Intervals for Regression

We can use the output from tidy(your_lm) to create a 95% confidence interval for the slope coefficient:

CI formula: $\hat{b} \pm t^* SE_b$

*t** is the critical value for the t distribution with *n* – 2 degrees of freedom with area C (e.g., 95%) between –*t** and *t**.

Inference for Regression

Mean vs Individual Response

If you want to make inference for the **mean response** μ_y when *x* takes the value *x**:

$$\hat{y} \pm t * SE_{\hat{\mu}}, \text{where } SE_{\hat{\mu}} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x - \bar{x})^2}}$$

If you want to make inference for a **single observation** *y* when *x* takes the value *x**:

$$\hat{y} \pm t * SE_{\hat{y}}, \text{where } SE_{\hat{y}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x - \bar{x})^2}}$$

ANOVA & Tukey’s HSD

ANOVA & Tukey's HSD

ANOVA

Goal: To determine if there are differences between the means of three or more independent groups.

H0: *$\mu 1 = \mu 2 = \ldots = \mu K$, where K is the number of levels of the grouping variable*

HA: *not all $\mu 1, \mu 2, \ldots, \mu K$ are equal (at least one mean differs from the rest)*

R Code: `aov(formula = cont_var ~ cat_var, data = your_data)`

ANOVA & Tukey's HSD

ANOVA

Test Statistic: $F = \frac{\text{mean squares for groups}}{\text{mean squares for error}} = \frac{MSG}{MSE}$

MSG: $\frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + ... + n_k(\bar{x}_k - \bar{x})^2}{k-1}$

MSE: $\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + ... + (n_k-1)s_k^2}{N_{Total} - k}$

*you will not be asked to calculate this by hand, but you should be able to understand/interpret the R output

ANOVA & Tukey's HSD

ANOVA R Interpretation

```
> cancer_anova <- aov(formula = tumor_volume ~ treatment, data = cancer_data)
> tidy(cancer_anova)
## # A tibble: 2 X 6
##   term              df      sumsq      meansq      statistic      p.value
##   <chr>           <dbl>     <dbl>      <dbl>         <dbl>         <dbl>
##   treatment          3        8060       2687          6.70        0.00313
## Residuals           18        7218       401             NA             NA
```

- `df` is the numerator and denominator degrees of freedom for this dataset
- `sumsq` is the treatment sum of squares and residual sum of squares
- `meansq` is the MSG (treatment) and MSE (residuals)
- `statistic` is the F test statistic, the ratio between MSG and MSE. This one tells us the variation between means is nearly 7 times as large as the variation within the groups
- `p.value` is the p-value for this test. Here, it is 0.3% which means there is a 0.3% chance of observing a test statistic at least as extreme as 6.70 under the null hypothesis of all means being the same. Therefore, we reject the null hypothesis

ANOVA & Tukey’s HSD

Tukey’s HSD

Goal: *After* obtaining a significant ANOVA result, we can use Tukey’s HSD to determine which pairs of group means are significantly different.

Use R to conduct the test:

```
model <- aov(outcome~groups, data = dataset)
tukey <- TukeyHSD(model)
tidy(tukey)
```

Non-Parametric Testing Alternatives

Non-Parametric Testing Alternatives

Non-Parametric Testing

If our data does not meet the assumptions to run a parametric test, we can turn to their nonparametric equivalents:

Parametric Test	Non-Parametric Alternative
Independent two sample t-test	Wilcoxon Rank Sum Test
Paired t-test	Wilcoxon Sign Rank Test
ANOVA	Kruskal-Wallis Test

*you should know how to calculate the first two by hand, and the R code for all three

Non-Parametric Testing Alternatives

Wilcoxon Rank Sum Test

Steps By-Hand:

1. Order observations from lowest to highest
2. Rank them
3. If 2+ observations are equal, assign them the average of the *ranks*
4. Sum ranks of group 1 and group 2
5. Calculate test statistic, p-value

Null: two population distributions are identical

$$Z_w = \frac{W - \mu_w}{\sigma_w}$$

$$\mu_w = \frac{n_s(n_s + n_l + 1)}{2}$$

$$\sigma_w = \sqrt{\frac{n_s n_l (n_s + n_l + 1)}{12}}$$

In R: `wilcox.test(data %>% pull(group1), data %>% pull(group2), paired = F)`

or `wilcox.test(outcome ~ groups, data = data,  paired =F)`

Non-Parametric Testing Alternatives

Wilcoxon Sign Rank Test

Steps By-Hand:

1. Calculate the difference between each pair
2. Rank the *absolute value* of the differences (throw away 0s and average any ties)
3. Assign a + or - to each **rank** depending on whether the difference was positive or negative
4. Sum positive ranks and negative ranks
5. Calculate test statistic, p-value

$$Z_T = \frac{T - \mu_T}{\sigma_T}$$

$$\mu_T = \frac{n(n+1)}{4}$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

In R:

```
wilcox.test(data %>% pull(group1), data %>% pull(group2), paired = T)
```

or

```
wilcox.test(outcome ~ groups, data = data,  paired =T)
```

Non-Parametric Testing Alternatives

Kruskal-Wallis Test

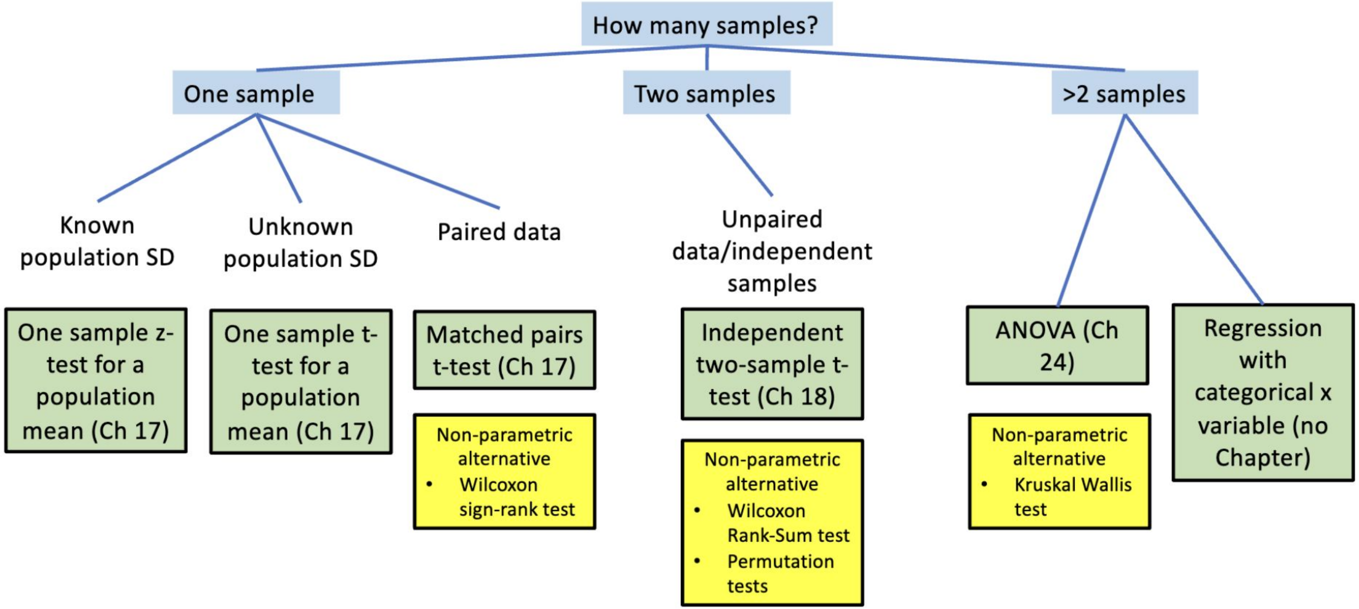
The Kruskal-Wallis test evaluates whether the distributions of the groups differ.

In R: `kruskal.test(outcome ~ groups, data = dataset)`

You do not need to know how to calculate this by hand.

Testing Summary

If you have continuous data



Regression with Categorical Exposure

Regression with Categorical Variables

Regression with Categorical Vars Overview

In previous lectures, we ran regression models with continuous explanatory variables

- With categorical variables, the model creates separate indicator terms for each category
- Each coefficient represents the mean difference from the reference group

The model takes the form: $\hat{y}_i = \hat{a} + \hat{b}_1 * I(\text{category}_1) + \hat{b}_2 * I(\text{category}_2)$

Regression with Categorical Variables

Regression Reference Groups

The reference group is the baseline category that others are compared to. Its mean becomes the **intercept** of the model.

- Changing the reference group changes the interpretation
- Use `fct_relevel()` to set the reference group before modelling

Regression with Categorical Variables

Interpreting the Model Output

- **Intercept** = mean outcome in the reference category
- **Coefficients** = mean difference vs the reference category
- **P-values** test whether each category’s mean equals the mean of the reference group
- **Predictions and CIs** estimate category means and uncertainty

Practice Problems

T-test Example

You have a sample of blood sugar levels from 100 adults. You find that their mean blood sugar level is 92 mg/dL with a sample SD of 8 mg/dL. You want to test if this average is different from the population average of 100 mg/dL.

1. State the null and alternative hypotheses
2. Find the t-test statistic by hand
3. Use R to compute the probability of observing this test statistic or more extreme
4. Interpret the p-value

T-test Example (Key)

You have a sample of blood sugar levels from 100 adults. You find that their mean blood sugar level is 92 mg/dL with a sample SD of 8 mg/dL. You want to test if this average is different from the population average of 100 mg/dL.

1. H_0: μ = 100 mg/dL ; H_A: μ ≠ 100 mg/dL
2. t = (92-100)/(8/√100) = -10
3. p value = pt(-10, df = 100 - 1) *2 = 1.093976e-16
4. Our p-value of 1.093976e-16 is much smaller than our cutoff value of 0.05, so we reject the null hypothesis in favor of the alternative that the mean blood sugar level of our sample is different than the population.

Two Sample T-Test Example

Answer the following questions about the two-sample t-test:

1. How does it compare with the one-sample t-test?
2. When the sizes of the two samples are equal and the two populations being compared have similar shapes, the two-sample t-test will work well for sample sizes as small as ________________.
3. When the two populations have ________________, larger samples are needed.

Two Sample T-Test Example (Key)

Answer the following questions about the two-sample t-test:

1. Two sample t-test procedures are more robust than the one-sample t-test, especially if the data are skewed
2. n1 = n2 = 5
3. Different shapes or unequal variances

Paired T-Test Example

Suppose 15 students took two exams: one in English and one in Math. We are interested in comparing the difference in student performance for the two exams. You are given this information about the mean difference in scores between Math and English for each student:

mean_diff: 8.4862
sd : 22.47689

1. State the null and alternative hypotheses (in context)
2. Calculate the test statistic by hand.
3. Write code for finding the p-value for this test statistic.
4. Write code to find t* for a 95% CI. Construct a 95% CI by hand.

Paired T-Test Example (Key)

Suppose 15 students took two exams: one in English and one in Math. We are interested in comparing the difference in student performance for the two exams. You are given this information about the mean difference in scores between Math and English for each student:

1. H_0: The mean difference between Math and English exam scores is 0

   H_A: The mean difference between Math and English exam scores is not 0.
2. t = (8.4862 - 0) / (22.47689 / √15) = 1.462254
3. pt(1.462254, df = 15 - 1, lower.tail = FALSE)*2 = 0.1657553
4. qt(.975, df = 14) = 2.145

   Lower bound = 8.4862 - 2.145*(22.47689/√15) = -3.96232

   Upper bound = 8.4862 + 2.145*(22.47689/√15) = 20.93472

Inference for Proportions Example

Pfizer, an American pharmaceutical company, recently released results from Phase 3 of an RCT trial to the determine efficacy of their mRNA-based COVID-19 vaccine candidate, BNT162b2. While they have not released the breakdown of vaccine versus placebo totals, suppose the results match the table below. Conduct a hypothesis test.

	COVID-19	no COVID-19	Total	p_hat
BNT162b2	8	22661	22669	0.00035
Placebo	162	21000	21162	0.00766

Inference for Proportions Example (Key)

1. Compute $\hat{p}$: (8 + 162) / (22669 + 21162) = 0.00388
2. Compute SE: sqrt(0.00388*(1 – 0.00388)*(1 / 22669 + 1 / 21162)) = 0.000594
3. Compute the test statistic: $z = (\hat{p}_1 - \hat{p}_2) / SE$

   = (0.00766 – 0.00035) / 0.000594 = 12.3064
4. Calculate the p-value:

```
> pnorm(q = 12.3064, lower.tail = F)*2
## [1] 8.37e-35
```

The p-value is equal to 8.37e-35. Under the null hypothesis of no difference between the proportions, there is an 8.37e-35 chance of observing the difference we saw or more extreme, which provides evidence in favor of the alternative hypothesis that these proportions are different.

Chi-Square Example

In response to a question on regular exercise, 60% of all students reported getting no regular exercise, 25% reported exercising sporadically, and 15% reported exercising regularly. The next year the school launched a health promotion campaign on campus in an attempt to increase exercise habits. The survey was then completed by 470 students and the following data were collected:

	No Regular Exercise	**Sporadic Exercise**	**Regular Exercise**	**Total**
Number of Students	255	125	90	470

1. State the null and alternative hypotheses
2. Calculate the chi-square statistic and p-value
3. Interpret the p-value

Chi-Square Example (Key 1)

In response to a question on regular exercise, 60% of all students reported getting no regular exercise, 25% reported exercising sporadically, and 15% reported exercising regularly. The next year the school launched a health promotion campaign on campus in an attempt to increase exercise habits. The survey was then completed by 470 students and the following data were collected:

First, we need to check the expected counts requirement:

- At least 80% of the cells have 5 or more observations expected ($E_i \geq 5$ for $\geq 80\%$ of the cells)
- All cells have expected counts > 1 ($E_i > 1$ for all cells)

Students	No Regular Exercise	Sporadic Exercise	Regular Exercise	Total
Observed	255	125	90	470
Expected	.6(470) = 282	.25(470)= 117.5	.15(470)= 70.5	470

Chi-Square Example (Key 2)

In response to a question on regular exercise, 60% of all students reported getting no regular exercise, 25% reported exercising sporadically, and 15% reported exercising regularly. The next year the school launched a health promotion campaign on campus in an attempt to increase exercise habits. The survey was then completed by 470 students and the following data were collected:

1. H_0: The proportions of those who exercise after the campaign are the same as the proportion of those who exercised before the campaign

   H_1: At least one of p_k is different than specified in H_0, for k

   being one of no regular exercise, sporadic exercise, or regular exercise.

2. $\chi^2 = (255-282)^2/282 + (125 - 117.5)^2/117.5 + (90 - 70.5)^2/70.5 = 8.457446$

   p-value = pchisq(q =8.457446, df = 3-1 = 2, lower.tail = F) = 0.01457099

Chi-Square Example (Key 3)

In response to a question on regular exercise, 60% of all students reported getting no regular exercise, 25% reported exercising sporadically, and 15% reported exercising regularly. The next year the school launched a health promotion campaign on campus in an attempt to increase exercise habits. The survey was then completed by 470 students and the following data were collected:

3. Since the p-value is less than our cutoff of 0.05 there is evidence against the null hypothesis in favor of the alternative hypothesis that at least one of the proportions of the exercise groups after the campaign is not the same as the proportion of the exercise group before the campaign.

Chi-Square Example #2

You are given data on the number of male and female babies and their eye colors. You want to test whether there is a difference in eye color based on the babies’ sex.

	Blue Eyes	Green Eyes	Brown Eyes
Male	20	30	25
Female	10	13	7

1. State the null and alternative hypotheses
2. Calculate the chi-square test statistic and p-value
3. Interpret the p-value

Chi-Square Example #2 (Key 1)

You are given data on the number of male and female babies and their eye colors. You want to test whether there is a difference in eye color based on the babies' sex.

1. H_0: Eye color and sex are independent

   H_0: The probability of having blue eyes conditional on being a male is equal to the probability of having blue eyes conditional on being a female.

   H_A: Eye color and sex are not independent.

   H_A: The probability of having blue eyes conditional on being a male is not equal to the probability of having blue eyes conditional on being a female.

Chi-Square Example #2 (Key 2)

Observations (Expected values)	Blue	Green	Brown	Row Total
Boys	20 (30*75/105 = 21.43)	30 (43*75/105 = 30.71)	25 (32*75/105 = 22.86)	75
Girls	10 (30*30/105 = 8.57)	13 (43*30/105 = 12.29)	7 (32*30/105 = 9.14)	30
Column Total	30	43	32	105

2. $\chi^2 = (20 - 21.43)^2 / 21.43 + (30 - 30.71)^2 / 30.71 + (25 - 22.86)^2 / 22.86$

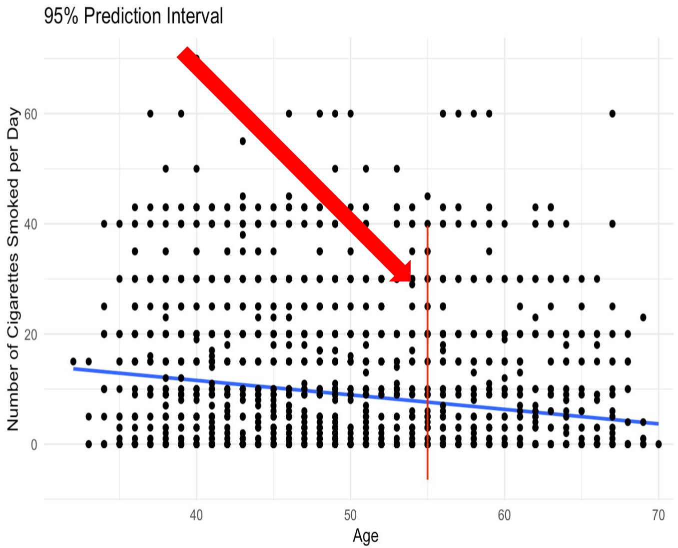
$+ (10 - 8.57)^2 / 8.57 + (13 - 12.29)^2 / 12.29 + (7 - 9.14)^2 / 9.14 = 1.092848$

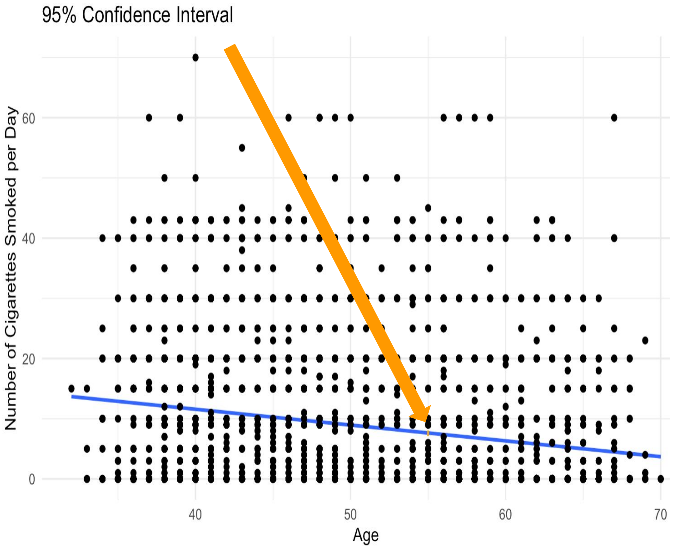
p-value = pchisq(q = 1.092848, df = 2, lower.tail = F) = 0.5790167

3. Our p-value is 0.579 which is larger than our alpha level of 0.05, so we fail to reject the null that eye color and sex of the baby are independent.

Regression Example

Why are prediction intervals wider than confidence intervals?





Regression Example (Key)

Why are prediction intervals wider than confidence intervals?

Individual values can have more variation and would fall within a wider range, which is what the prediction interval shows. On the contrary, the confidence interval shows that the variation in the mean across individual values would fall in a narrower range.

Non-Parametric Test Example

Suppose we want to test the difference in AQI scores in cities with high speed rails (group 1) and cities without high-speed rails (group 2). We randomly select six cities for each. We do not know the underlying distribution for each group.

1. Which test would you use? Why?
2. Calculate the test statistic(s) given the following AQIs for each group:

    Group 1: 70, 65, 40, 108, 70, 45

    Group 2: 50, 80, 70, 95, 105, 100

3. Calculate the p-value.
4. Interpret the p-value in terms of the null and alternative hypotheses.

Non-Parametric Test Example (Key 1)

1. Wilcoxon Rank Sum. We know nothing about the distribution and the two groups are not paired, so we would select this non-parametric test.
2. First, order the observations and rank them.
   Then, sum the ranks for each group.
   Next, assign W to the smaller rank sum.
   In this example, W = 30
   Finally, calculate the test statistic:

Group 1	Rank	Group 2	Rank
70	6	50	3
65	4	80	8
40	1	70	6
108	12	95	9
70	6	105	11
45	2	100	10
	30		47

$n_s = 6$ (# in smaller group), $n_l = 6$ (# in larger group)

$\mu_w = 6(6+6+1) / 2 = 39$

$\sigma_w = \text{sqrt}((6*6*(6+6+1))/12) = \text{sqrt}(39)$

$Z_w = (W - \mu_w) / \sigma_w = (30-39)/\text{sqrt}(39) = -1.441$

Non-Parametric Test Example (Key 2)

3. P-value: 2*pnorm(-1.441) = 0.1495847
4. Since $p > 0.05$, we fail to reject the null hypothesis that the two populations are identical.

Bootstrap Example

Suppose we are interested in estimating the 25th percentile of weights for newborns in CA. Our dataset has a SRS of 50 babies born in CA.

1. Describe the steps you would take to estimate this statistic with 10,000 samples.
2. Write the code that would construct a 95% CI for this statistic using the information on this slide.

   The 25th percentile of the sample is 6.477459

Bootstrap Example (Key)

Suppose we are interested in estimating the 25th percentile of weights for newborns in CA. Our dataset has a SRS of 50 babies born in CA.

1. Resample from our sample of size 50 with replacement from our sample. Calculate the 25th percentile of this sample.
   a. Do this 10,000 times. Each time calculate the 25th percentile and denote it.
   b. Create a histogram of all your 25th percentiles.
   c. Use this to generate your confidence interval using the quantile functions in R.
2. lower_bound <- quantile(data = estimates, p = .025)
3. upper_bound <- quantile(data = estimates, p = .975)
   a. From the simulated data, this confidence interval is (5.82522, 6.661133) which does contain our sample estimate.

Good Luck!

- The PH142 Teaching Team