Intro to Modern Stats

computers, not calculus
Why conduct experiments & collect data?

To discover something fundamental about a group/gene/effect/etc

We rarely care about the actual data you collected. **We CARE about the ‘truth’ behind the scenes:** the ACTUAL, complete population from which you drew your data.
True, population distributions
What is the mean salary at UCLA?

Exactly $64k?
Not $65k?

$64k
The Bootstrap
Bootstrap Resample

Resampled Means
Glossary

1. **Population Distribution**: the underlying distribution we actually care about

2. **Bootstrap**: to resample from your dataset, WITH replacement
One-Sample Bootstrap

**Lucid Chart**
The Bootstrap

Original Data

Resample

Resampled Data

Calculate Test Statistic

Test Statistic Distribution

Repeat, a lot
One-Sample Bootstrap Example

Credit: Dr. Nick Wisniewski
One-Sample Bootstrap Example code in R

** RStudio **
What can we say?

If our assumptions for applying the bootstrap are met (more on that Thursday), our claim is that we are:

There is a 95% chance the true population median lies between $42,180.00 & $59,000.05.
One-sample bootstrap code results

Histogram of resampledMedians
And then, there were two

What’s the difference in true population means?
Two-Sample Magnitude Bootstrap

**Lucid Chart**
Magnitude: How big is your effect?

**One-sample**
- Original Data
- Resample
- Resampled Data
- Calculate Test Statistic
- Test Statistic Distribution
- Calculate distribution bounds
  - left & right bounds

**Two-sample**
- Original DataA
- Original DataB
- Resample
  - Resampled DataA
  - Resampled DataB
- Calculate Difference of Means
  - Distribution: difference of actual means
  - Calculate distribution bounds
  - left & right bounds

Repeat a lot
Two-Sample Bootstrap Magnitude Example code in R

** RStudio **
What can we say?

If our assumptions for applying the bootstrap are met (more on that Thursday), our claim is that we are:

95% confident the difference between the true population means lies between $7,190.181 \ & \ $34,356.97.

This interval does not include 0, which suggests there may be a significant difference between the groups. However, this code was NOT a hypothesis test.
Is this observed difference ‘real’?

Or...

Is there no actual difference between groups; the observed difference is purely random noise?
Two-Sample Significance Bootstrap

**Lucid Chart**
Significance: Is your effect real?

1. Demeaned Original DataA
2. Demeaned Original DataB
3. Resample
4. Resampled DataA
5. Resampled DataB
6. Calculate Difference of Means
7. Distribution: difference of null means
8. Calculate % of these values that are more extreme than observed difference of means
9. p-value
Two-sample bootstrap significance example code

** RStudio **
Two-sample significance bootstrap results

Probability the observed difference is random noise
Magnitude vs. Significance

1. How big is your effect?
   a. Mean/median/mode (collectively, central tendency)
   b. Left & right confidence intervals on your

2. Is your effect ‘real’? ← Hypothesis test
   a. p-value

“Between condition A and condition B, we observed a mean effect magnitude of x [xLeftCI, xRightCI]. This effect size was significantly (or not) different from the null hypothesis (p< 0.##, two-sample percentile bootstrap).
What is the p-value?

The probability of getting your observed results OR results even more extreme, given there is no effect. Random chance.

\[ p( \text{data} | \text{null hypothesis} ) \]

**It is not:** the probability the null hypothesis is false

\[ p( \text{null hypothesis} | \text{data} ) \]

**It is not:** the chance of making a mistake by rejecting the null hypothesis
Take this home

**Magnitude**

Original DataA

Resample

Resampled DataA

Calculate Difference of Means

Distribution: difference of actual means

Calculate distribution bounds

left & right bounds

Original DataB

Resample

Resampled DataB

**Significance**

Demeaned Original DataA

Resample

Resampled DataA

Calculate Difference of Means

Distribution: difference of null means

Calculate % of these values that are more extreme than observed difference of means

Demeaned Original DataB

Resample

Resampled DataB

p-value

Repeat a lot

Repeat a lot
Intro to Modern Stats

computers, not calculus
Assessment

1) What is 95% confidence interval for the true standard deviation of sampleB from yesterday?

2) Is the standard deviation of sampleA significantly different from the standard deviation of sampleB?
True distribution of ALL the UCLA salaries
Two-sample bootstrapping, shown another way

We bootstrap to better quantify uncertainty in our inferences, not to get more accurate estimates.
Why ‘cult of normal’ stats fails

Weight Data Example

Compute likelihood of seeing a woman over 6’7” by fitting a Gaussian? $p < 1.6 \times 10^{-10}$

But there are 5 in the women’s NBA!
Paired Data
How to sample linked data

<table>
<thead>
<tr>
<th>FishID</th>
<th>Day1</th>
<th>Day7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.07</td>
<td>8.14</td>
</tr>
<tr>
<td>2</td>
<td>1.91</td>
<td>3.82</td>
</tr>
<tr>
<td>3</td>
<td>4.63</td>
<td>9.26</td>
</tr>
<tr>
<td>4</td>
<td>10.58</td>
<td>21.16</td>
</tr>
<tr>
<td>5</td>
<td>11.65</td>
<td>23.3</td>
</tr>
<tr>
<td>6</td>
<td>4.29</td>
<td>8.58</td>
</tr>
<tr>
<td>7</td>
<td>10.74</td>
<td>21.48</td>
</tr>
<tr>
<td>8</td>
<td>6.97</td>
<td>13.94</td>
</tr>
<tr>
<td>9</td>
<td>8.99</td>
<td>17.98</td>
</tr>
<tr>
<td>10</td>
<td>11.97</td>
<td>23.94</td>
</tr>
<tr>
<td>11</td>
<td>4.94</td>
<td>9.88</td>
</tr>
<tr>
<td>12</td>
<td>10.39</td>
<td>20.78</td>
</tr>
<tr>
<td>13</td>
<td>1.35</td>
<td>2.7</td>
</tr>
<tr>
<td>14</td>
<td>8.33</td>
<td>16.66</td>
</tr>
<tr>
<td>15</td>
<td>8.07</td>
<td>16.14</td>
</tr>
</tbody>
</table>
Paired Data Magnitude Bootstrap

1. Original DataA → Resample Pairs
2. Resampled DataAB → Calculate Mean of Differences
3. Distribution: mean of differences → Calculate distribution bounds
4. Repeat, a lot

left & right bounds
Paired data bootstrap magnitude example code

** RStudio **
Paired data bootstrap magnitude results
How would you do a null hypothesis test?

1) Well, what’s your actual hypothesis?

2) Null: that’s not true; you just got lucky with your data

<table>
<thead>
<tr>
<th>FishID</th>
<th>Day1</th>
<th>Day7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.07</td>
<td>8.14</td>
</tr>
<tr>
<td>2</td>
<td>1.91</td>
<td>3.82</td>
</tr>
<tr>
<td>3</td>
<td>4.63</td>
<td>9.26</td>
</tr>
<tr>
<td>4</td>
<td>10.58</td>
<td>21.16</td>
</tr>
<tr>
<td>5</td>
<td>11.65</td>
<td>23.3</td>
</tr>
<tr>
<td>6</td>
<td>4.29</td>
<td>8.58</td>
</tr>
<tr>
<td>7</td>
<td>10.74</td>
<td>21.48</td>
</tr>
<tr>
<td>8</td>
<td>6.97</td>
<td>13.94</td>
</tr>
<tr>
<td>9</td>
<td>8.99</td>
<td>17.98</td>
</tr>
<tr>
<td>10</td>
<td>11.97</td>
<td>23.94</td>
</tr>
<tr>
<td>11</td>
<td>4.94</td>
<td>9.88</td>
</tr>
<tr>
<td>12</td>
<td>10.39</td>
<td>20.78</td>
</tr>
<tr>
<td>13</td>
<td>1.35</td>
<td>2.7</td>
</tr>
<tr>
<td>14</td>
<td>8.33</td>
<td>16.66</td>
</tr>
<tr>
<td>15</td>
<td>8.07</td>
<td>16.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FishID</th>
<th>Day1</th>
<th>Day7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.07</td>
<td>8.14</td>
</tr>
<tr>
<td>2</td>
<td>3.82</td>
<td>1.91</td>
</tr>
<tr>
<td>3</td>
<td>9.26</td>
<td>4.63</td>
</tr>
<tr>
<td>4</td>
<td>10.58</td>
<td>21.16</td>
</tr>
<tr>
<td>5</td>
<td>23.3</td>
<td>11.65</td>
</tr>
<tr>
<td>6</td>
<td>4.29</td>
<td>8.58</td>
</tr>
<tr>
<td>7</td>
<td>10.74</td>
<td>21.48</td>
</tr>
<tr>
<td>8</td>
<td>13.94</td>
<td>6.97</td>
</tr>
<tr>
<td>9</td>
<td>8.99</td>
<td>17.98</td>
</tr>
<tr>
<td>10</td>
<td>23.94</td>
<td>11.97</td>
</tr>
<tr>
<td>11</td>
<td>4.94</td>
<td>9.88</td>
</tr>
<tr>
<td>12</td>
<td>20.78</td>
<td>10.39</td>
</tr>
<tr>
<td>13</td>
<td>2.7</td>
<td>1.35</td>
</tr>
<tr>
<td>14</td>
<td>8.33</td>
<td>16.66</td>
</tr>
<tr>
<td>15</td>
<td>8.07</td>
<td>16.14</td>
</tr>
</tbody>
</table>
Permutation Test

**Permutation**: a way, especially one of several possible variations, in which a set or number of things can be ordered or arranged.
Paired Data Significance Test (Permutation)

1. Original Data A, Original Data B
2. For each pair, swap the values with probability 50%
3. Permutated Data AB
4. Take the mean of the differences
5. Distribution: difference of null means
6. Calculate % of these values that are more extreme than observed difference of means.
7. p-value
Correlation / Regression
No.
Requirements for correlation coefficient

1. X is normally distributed
2. Y is normally distributed
3. Variances are uniform across the scale (homoscedasticity)
4. Not obviously nonlinear (by eye)
5. No auto-correlation (samples are independent)

Credit: Dr. Nick Wisniewski
Requirements for correlation coefficient

1. X is normally distributed
2. Y is normally distributed
3. Variances are uniform across the scale (homoscedasticity)
4. Not obviously nonlinear (by eye)
5. No auto-correlation
Intro to Modern Stats
computers, not calculus
Assessment

I sent you some new Salary data. It’s the salaries of 69 professors who have each worked at usc, ucla, and irvine in their career. Each row is a professor; each column is a university. Based on publically available data, I hypothesize that usc & irvine salaries are positively, linearly correlated.

1) Is that true, at a 5% significance level?
Assessment

I am interested in the RATIO of salaries between usc and ucla.

1) Is the ratio salaries significantly different, at a 5% significance level?

2) What is the magnitude of the true ratio of salaries between these universities, with confidence intervals?
Assessment

Let’s better characterize the salary relationship between UCLA and USC with a regression. What is the slope and intercept (w confidence intervals) of the salary of USC as a function of the salary at UCLA?

Helpful code:

bootfit = lm(d$usc ~ d$ucla)  # that’s the regression

bootfitb[i] = bootfit$coefficients[[1]]  # these are the coefficients

bootfitm[i] = bootfit$coefficients[[2]]

Bonus points for testing significance
How to quantify what we intuitively know

\[ F = \frac{\text{explained variance}}{\text{unexplained variance}}, \]

or

\[ F = \frac{\text{between-group variability}}{\text{within-group variability}}. \]
Are the means all equal?
ANOVA Example from R
<table>
<thead>
<tr>
<th></th>
<th>Hypothesis Question</th>
<th>Magnitude</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Group</td>
<td>N/A</td>
<td>bootstrap samples</td>
<td>N/A</td>
</tr>
<tr>
<td>2 Groups, different subjects</td>
<td>Is the mean different between these groups?</td>
<td>bootstrap each group's samples independently</td>
<td>shuffle group assignment across subjects</td>
</tr>
<tr>
<td>2 Groups, same subjects</td>
<td>Is the mean different between these groups?</td>
<td>bootstrap subjects</td>
<td>shuffle condition within each subject</td>
</tr>
<tr>
<td>Bivariate Data</td>
<td>Are these variables meaningfully related?</td>
<td>bootstrap subjects</td>
<td>shuffle one condition across subjects</td>
</tr>
<tr>
<td>&gt; 2 groups, diff subs</td>
<td>Do these groups all share the same mean?</td>
<td>bootstrap each group's samples independently</td>
<td>shuffle group assignment</td>
</tr>
</tbody>
</table>
Multiple Comparisons

I have a 20-sided die.

1) I have one roll; what are the odds I roll the number 1?

2) I roll 30 million times; what are the odds I roll the number 1, at least once?

This is the multiple-comparisons problem
Other

**Type I error**: incorrect rejection of a true null hypothesis (a "false positive")

**Type II error**: incorrectly retaining a false null hypothesis (a "false negative")
Rules

1. You’re visual animals; plot the raw data

2. (almost) nobody cares about your sample dataset
   a. You care about the underlying population distribution

3. Bootstrap (resample WITH replacement) to quantify uncertainty

4. Permute (resample WITHOUT replacement) for significance

5. p-value is **probability (data | null hypothesis is true)**
   a. What fraction of the time your null results are $\geq$ your effect
   b. How often would you be wrong in ‘discovering’ there’s an effect?

6. The hard work of hypothesis testing is to construct the null distribution

7. To build the null distribution, ‘break’ your effect, by shuffling, then calculate your question statistic (mean/median/ratio) 10k times.
Fin
Extra
I was wrong

In our trout-length data from yesterday, my undergrad mixed up the sampling on Day7. We don’t know if the Day7 & Day1 measurements are from the same fish or different fish. So, assume they’re 2 separate populations: are they still significantly different in length?
How to bin histogram data?