Section 4.2

Measuring Evidence with p-values
Paul the Octopus

http://www.youtube.com/watch?v=3ESGpRUMj9E
Paul the Octopus

• In 2008, Paul the Octopus predicted 8 World Cup games, and predicted them all correctly

• Is this evidence that Paul’s chance of guessing correctly, $p$, is really greater than 50%?

• What are the null and alternative hypotheses?

$$H_0: p = 0.5$$

$$H_a: p > 0.5$$
How unusual is it to see a sample statistic as extreme as that observed, if $H_0$ is true?

- If it is very unusual, we have statistically significant evidence against the null hypothesis.

- **Today’s Question:** How do we measure how unusual a sample statistic is, if $H_0$ is true?
Measuring Evidence against $H_0$

To see if a statistic provides evidence against $H_0$, we need to see what kind of sample statistics we would observe, just by random chance, if $H_0$ were true.
Paul the Octopus

- We need to know what kinds of statistics we would observe just by random chance, if the null hypothesis were true
- How could we figure this out???

*Simulate many samples of size $n = 8$ with $p = 0.5$*
Simulate!

• We can simulate this with a coin!
• Each coin flip = a guess between two teams (Heads = correct, Tails = incorrect)
• Flip a coin 8 times, count the number of heads, and calculate the sample proportion of heads
• Come to the board to add your sample proportion to a class dotplot
• How extreme is Paul’s sample proportion of 1?
• We just created our first randomization distribution!
Randomization Distribution

A *randomization distribution* is a collection of statistics from samples simulated assuming the null hypothesis is true.

- The randomization distribution shows what types of statistics would be observed, just by random chance, if the null hypothesis were true.
Lots of simulations!

- To estimate the p-value more accurately, we need many more simulations!

www.lock5stat.com/statkey
Randomization Distribution

Randomization Dotplot of Proportion  Null hypothesis: $p = 0.5$

- Left Tail
- Two-Tail
- Right Tail

- samples = 200
- mean = 0.494
- st. dev. = 0.184
Key Question

How unusual is it to see a sample statistic as extreme as that observed, if $H_0$ is true?

- A randomization distribution tells us what kinds of statistics we would see just by random chance, if the null hypothesis is true.
- This makes it straightforward to assess how extreme the observed statistic is!
p-value

The $p$-value is the chance of obtaining a sample statistic as extreme (or more extreme) than the observed sample statistic, if the null hypothesis is true.

- The $p$-value can be calculated as the proportion of statistics in a randomization distribution that are as extreme (or more extreme) than the observed sample statistic.
p-value

- Paul the Octopus: the \textit{p-value} is the chance of getting all 8 out of 8 guesses correct, if $p = 0.5$
- What proportion of statistics in the randomization distribution are as extreme as $\hat{p} = 1$?
If Paul is just guessing, the chance of him getting all 8 correct is 0.004.
Calculating a p-value

1. What kinds of statistics would we get, just by random chance, if the null hypothesis were true? *(randomization distribution)*

2. What proportion of these statistics are as extreme as our original sample statistic? *(p-value)*
Death Penalty

- A random sample of people were asked “Are you in favor of the death penalty for a person convicted of murder?”

<table>
<thead>
<tr>
<th>Year</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>663</td>
<td>342</td>
</tr>
<tr>
<td>2010</td>
<td>640</td>
<td>360</td>
</tr>
</tbody>
</table>

- Did the proportion of Americans who favor the death penalty decrease from 1980 to 2010?

Death Penalty

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>663</td>
<td>342</td>
</tr>
<tr>
<td>2010</td>
<td>640</td>
<td>360</td>
</tr>
</tbody>
</table>


$H_0$: $p_{1980} = p_{2010}$  \hspace{1cm}  $\hat{p}_{1980} = 0.66$  \hspace{1cm}  $\hat{p}_{2010} = 0.64$

$H_a$: $p_{1980} > p_{2010}$  \hspace{1cm}  So the sample statistic is:

$\hat{p}_{1980} - \hat{p}_{2010} = 0.66 - 0.64 = 0.02$

How extreme is 0.02, if $p_{1980} = p_{2010}$?  

*StatKey*
Death Penalty

If proportion supporting the death penalty has not changed from 1980 to 2010, we would see differences this extreme about 16% of the time.
p-value

Using the randomization distribution below to test

\[ H_0 : \rho = 0 \quad vs \quad H_a : \rho > 0 \]

Match the sample statistics: \( r = 0.1, \ r = 0.3, \text{ and } \ r = 0.5 \)

With the p-values: \( 0.005, \ 0.15, \ \text{ and } \ 0.35 \)

Which sample statistic goes with which p-value?
Extrasensory Perception

• Recall our ESP experiment from last class
• Let’s use that data to create a randomization distribution and find a p-value!

StatKey
Alternative Hypothesis

- A **one-sided** alternative contains either > or <
- A **two-sided** alternative contains ≠

- The p-value is the proportion in the tail in the direction specified by $H_a$

- For a two-sided alternative, the p-value is twice the proportion in the smallest tail
p-value and $H_a$

**Upper-tail (Right Tail)**
- $H_0: \mu = 0$
- $H_a: \mu > 0$
- $\bar{x} = 2$

**Lower-tail (Left Tail)**
- $H_0: \mu = 0$
- $H_a: \mu < 0$
- $\bar{x} = -1$

**Two-tailed**
- $H_0: \mu = 0$
- $H_a: \mu \neq 0$
- $\bar{x} = 2$
Sleep versus Caffeine

• Recall the sleep versus caffeine experiment from last class

• $\mu_s$ and $\mu_c$ are the mean number of words recalled after sleeping and after caffeine.

• $H_0: \mu_s = \mu_c$
  $H_a: \mu_s \neq \mu_c$

• Let’s find the p-value!

• [www.lock5stat.com/statkey](http://www.lock5stat.com/statkey)
Sleep or Caffeine for Memory?

Randomization Dotplot of $\overline{x}_1 - \overline{x}_2$; Null hypothesis $\mu_1 = \mu_2$

- $\text{Left Tail}$ ✓ $\text{Two-Tail}$ □ $\text{Right Tail}$
- $\# \text{samples} = 1000$
- $\text{mean} = -0.021$
- $\text{st. dev.} = 1.499$

$p$-value $= 2 \times 0.022$

$= 0.044$

$\overline{X}_S - \overline{X}_C = 3$

$\overline{X}_S - \overline{X}_C$ when $H_0$ true
Using the randomization distribution below to test

$$H_0 : \rho = 0 \quad vs \quad H_a : \rho > 0$$

Which sample statistic shows the most evidence for the alternative hypothesis?  
$r = 0.1$,  $r = 0.3$, or  $r = 0.5$

Therefore, which $p$-value shows the most evidence for the alternative hypothesis?  
$0.35$,  $0.15$, or  $0.005$
p-value and $H_0$

• If the p-value is small, then a statistic as extreme as that observed would be unlikely if the null hypothesis were true, providing significant evidence against $H_0$

• The smaller the p-value, the stronger the evidence against the null hypothesis and in favor of the alternative
The **smaller** the p-value, the **stronger** the evidence against $H_0$. 

**p-value and $H_0$**
Summary

• The randomization distribution shows what types of statistics would be observed, just by random chance, if the null hypothesis were true.

• A p-value is the chance of getting a statistic as extreme as that observed, if $H_0$ is true.

• A p-value can be calculated as the proportion of statistics in the randomization distribution as extreme as the observed sample statistic.

• The smaller the p-value, the greater the evidence against $H_0$. 