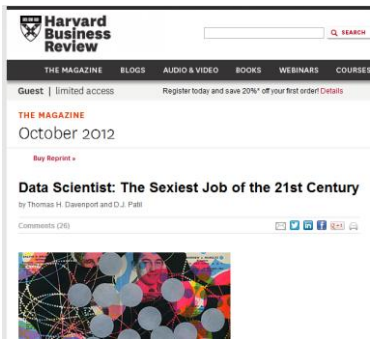
STAT 101
Dr. Kari Lock Morgan
9/25/12

Hypothesis Testing: p-value

SECTION 4.2

- Randomization distribution
- p-value



"If "sexy" means having rare qualities that are much in demand, data scientists are already there. They are difficult and expensive to hire and, given the very competitive market for their services, difficult to retain."

http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/

A Note on Random Samples

- Why do we take random samples?
- If we have access to data from the entire population, why would we take a random sample?
- For Project 1, if you have access to data on the entire population, USE IT!
- The methods of inference of no longer needed (mention this in your paper and explain why), but do a CI and test anyway just to prove you can…

Paul the Octopus

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

Key Question

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?

Paul the Octopus

- Based on your simulation results, for a sample size of $n = 8$, do you think $\hat{p} = 1$ is *statistically significant?*

a) Yes
b) No

Randomization Distribution

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

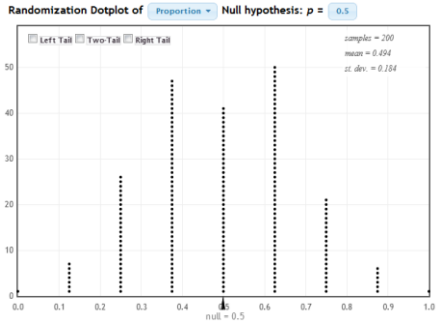
- The randomization distribution shows what types of statistics would be observed, just by random chance, if the null hypothesis were true

Lots of simulations!

- For a better randomization distribution, we need many more simulations!

www.lock5stat.com/statkey

Randomization Distribution



Paul the Octopus

- Based on StatKey's simulation results, for a sample size of $n = 8$, do you think $\hat{p} = 1$ is *statistically significant?*

a) Yes
b) No

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!

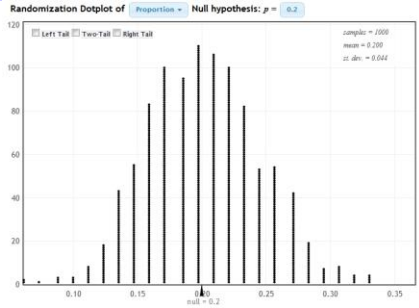
What about ESP?

- How could we simulate what would happen, just by random chance, if the null hypotheses were true for the ESP experiment?
- Roll a die.
  - 1 = "correct letter"
  - 2-5 = "wrong letter"
  - 6 = roll again
- Did you get the correct letter?

(a) Yes
(b) No

ESP Randomization Distribution

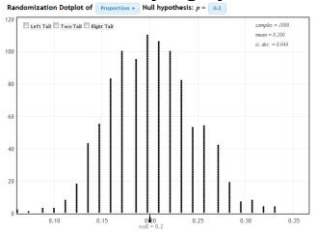
- www.lock5stat.com/StatKey



ESP

Based on the randomization distribution, do you think the results of our class ESP experiment (24/85 = 0.29) are *statistically significant?*

a) Yes
b) No



ESP

- What does this imply about ESP?

a) Evidence that ESP exists
b) Evidence that ESP does not exist
c) Impossible to tell

Quantifying Evidence

- We need a way to **quantify** evidence against the null…

p-value

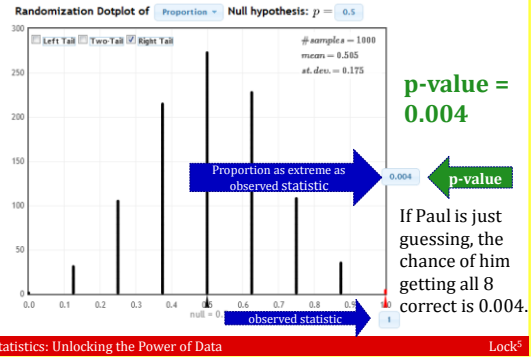
The ***p-value*** is the chance of obtaining a sample statistic as extreme (or more extreme) than the observed sample statistic, if the null hypothesis is true

- 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 ***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 $\hat{p} = 1$?

1000 Simulations



p-value = 0.004

If Paul is just guessing, the chance of him getting all 8 correct is 0.004.

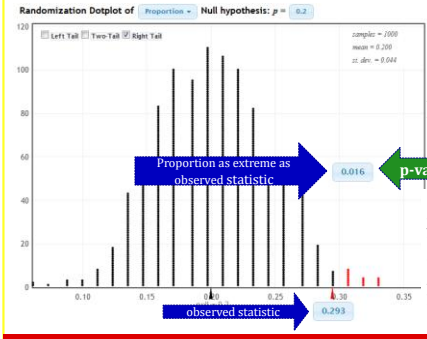
Calculating a p-value

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)*

p-value

- ESP: the ***p-value*** is the chance of getting $\hat{p} \geq 0.29$, if $p = 0.2$, with $n = 85$.
- What proportion of statistics in the randomization distribution are as extreme as $\hat{p} = 0.29$?
- www.lock5stat.com/statkey

ESP p-value



p-value = 0.016

If you were all just guessing randomly, the chance of us getting a sample proportion of 0.293 is 0.016.

Death Penalty

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

	Yes	No
1980	663	342
2010	640	360

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

"Death Penalty," Gallup, www.gallup.com

Death Penalty

	Yes	No
1980	663	342
2010	640	360

p_{1980}, p_{2010}: proportion of Americans who favor the death penalty in 1980, 2010

$H_0: p_{1980} = p_{2010}$
$H_a: p_{1980} > p_{2010}$

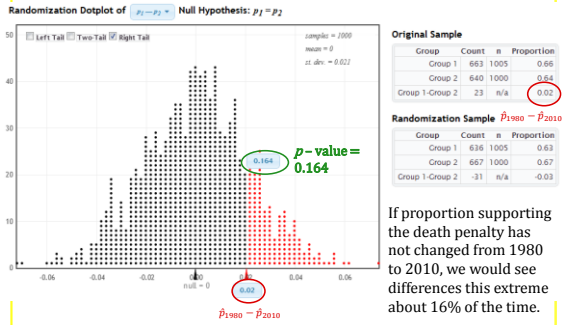
$\hat{p}_{1980} = 0.66 \quad \hat{p}_{2010} = 0.64$

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

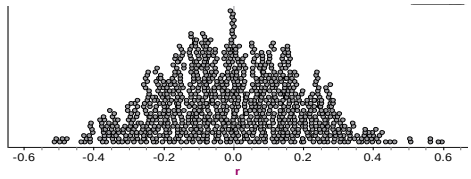
Using the randomization distribution below to test

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

Match the sample statistics: r = 0.1, r = 0.3, and r = 0.5

With the p-values: 0.005, 0.15, and 0.35

Which sample statistic goes with which p-value?



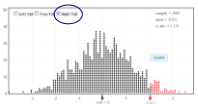
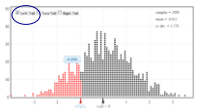
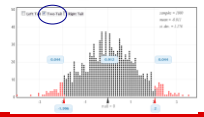
Alternative Hypothesis

- 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

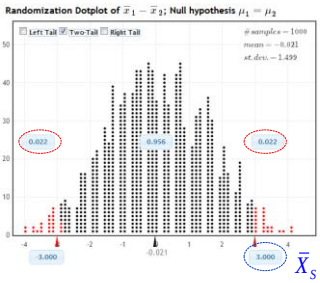
p-value and H_a

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
- μ_s and μ_c are the mean number of words recalled after sleeping and after caffeine.
- $H_0: \mu_s = \mu_c$
  $H_a: \mu_s \neq \mu_c$ ← Two-tailed alternative
- Let's find the p-value!
- www.lock5stat.com/statkey

Sleep or Caffeine for Memory?



p-value
= 2 × 0.022
= 0.044

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

$\bar{X}_S - \bar{X}_C$ when H_0 true

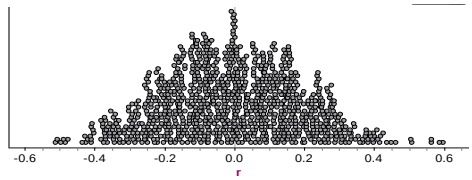
p-value

Using the randomization distribution below to test

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

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

p-value and H_0

The smaller the p-value, the stronger the evidence against H_0.

p-value and H_0

Which of the following p-values gives the strongest evidence *against* H_0?

a) 0.005
b) 0.1
c) 0.32
d) 0.56
e) 0.94

The lower the p-value, the stronger the evidence against the null hypothesis.

p-value and H_0

Which of the following p-values gives the strongest evidence *against* H_0?

a) 0.22
b) 0.45
c) 0.03
d) 0.8
e) 0.71

The lower the p-value, the stronger the evidence against the null hypothesis.

p-value and H_0

Two different studies obtain two different p-values. Study A obtained a p-value of 0.002 and Study B obtained a p-value of 0.2. Which study obtained stronger evidence *against* the null hypothesis?

a) Study A
b) Study B

The lower the p-value, the stronger the evidence against the null hypothesis.

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

To Do

- Read Section 4.2
- Idea and data for Project 1 (proposal due 9/27)
- Do Homework 4 (due Thursday, 10/4)