STAT 101
Dr. Kari Lock Morgan

Hypothesis Testing: Hypotheses

SECTION 4.1

- Statistical test
- Null and alternative hypotheses
- Statistical significance

Review of Last Class

- The standard error of a statistic is the standard deviation of the sample statistic, which can be estimated from a bootstrap distribution
- Confidence intervals can be created using the standard error or the percentiles of a bootstrap distribution
- Confidence intervals can be created this way for any parameter, as long as the bootstrap distribution is approximately symmetric and continuous

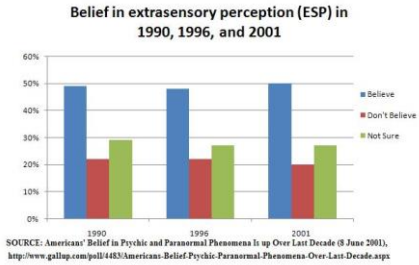
Extrasensory Perception

Is there such a thing as extrasensory perception (ESP) or a "sixth sense"?

Do you believe in ESP?

a) Yes
b) No

Extrasensory Perception



Extrasensory Perception

- One way to test for ESP is with Zener cards:

- Subjects draw a card at random and telepathically communicate this to someone who then guesses the symbol

Extrasensory Perception

- There are five cards with five different symbols
- If there is no such thing as ESP, what proportion of guesses should be correct?

a) $p = 0$
b) $p = 1/4$
c) $p = 1/5$
d) $p = 1/2$

Because there are 5 cards, each person has a 1/5 chance of guessing correctly each time, if ESP does not exist.

Extrasensory Perception

- As we've learned, statistics vary from sample to sample
- Even if the population proportion is 1/5, not every sample proportion will be exactly 1/5
- ***How do we determine when a sample proportion is far enough above 1/5 to provide evidence of ESP?***

Statistical Test

A ***statistical test*** uses data from a sample to assess a claim about a population

- In the ESP experiment, we want to use sample data to determine whether the population proportion of correct guesses is really higher than 1/5

Statistical Evidence

- Let $\hat{p}$ denote the sample proportion of correct guesses in an ESP experiment
- Which of these sample statistics would give the strongest evidence for ESP?

a) $\hat{p} = 0$
b) $\hat{p} = 1/5$
c) $\hat{p} = 1/2$
d) $\hat{p} = 3/4$

3/4 is the highest, so provides the strongest evidence of ESP.

Extrasensory Perception

- Let's create our own sample proportion!
- Randomly choose a letter from A B C D E, and write it down (don't show anyone!)
- Find a partner, telepathically communicate your letter (no auditory or visual clues!), and have them guess your letter. Switch roles.
- Did you guess correctly?
  a) Yes
  b) No

Extrasensory Perception

- What is the sample proportion for our class?
- This provides
  a) Strong evidence for ESP
  b) Weak evidence for ESP
  c) No evidence for ESP
  d) Not sure
- Next class, we'll learn how to quantify this evidence!

Statistical Hypotheses

- Statistical tests are framed formally in terms of two competing hypotheses:

Null Hypothesis (H_0): Claim that there is no effect or difference.

Alternative Hypothesis (H_a): Claim for which we seek evidence.

Statistical Hypotheses

H_o: Null hypothesis

H_a: Alternative hypothesis

Competing claims about a population

- The alternative hypothesis is established by observing evidence (data) that contradicts the null hypothesis and supports the alternative hypothesis
- Hypotheses are always about population parameters

Statistical Hypotheses

ALL POSSIBILITIES

Usually the null is a very specific statement

Can we reject the null hypothesis?

ESP Hypotheses

For the ESP experiment:

H_o: $p = 1/5$ ← No "effect" or no "difference"

H_a: $p > 1/5$ ← Claim we seek "evidence" for

- Helpful hints:
  - H_0 usually includes =
  - H_a usually includes >, <, or ≠
  - The inequality in H_a depends on the question

Sleep versus Caffeine

- Students were given words to memorize, then randomly assigned to take either a 90 min nap, or a caffeine pill. 2 ½ hours later, they were tested on their recall ability.
- Explanatory variable: sleep or caffeine
- Response variable: number of words recalled
- Is sleep or caffeine better for memory?

Mednick, Cai, Kanady, and Drummond (2008). "Comparing the benefits of caffeine, naps and placebo on verbal, motor and perceptual memory," *Behavioral Brain Research*, 193, 79-86.

Sleep versus Caffeine

What is the parameter of interest in the sleep versus caffeine experiment?

a) Proportion
b) Difference in proportions
c) Mean
d) Difference in means
e) Correlation

The response variable (number of words recalled) is quantitative and the explanatory variable (sleep or caffeine) is categorical, so we are interested in a difference in means.

Sleep versus Caffeine

- Let μ_s and μ_c be the mean number of words recalled after sleeping and after caffeine.
- Is there a difference in average word recall between sleep and caffeine?
- What are the null and alternative hypotheses?

a) H_0: $\mu_s \neq \mu_c$, H_a: $\mu_s = \mu_c$
b) H_0: $\mu_s = \mu_c$, H_a: $\mu_s \neq \mu_c$
c) H_0: $\mu_s \neq \mu_c$, H_a: $\mu_s > \mu_c$
d) H_0: $\mu_s = \mu_c$, H_a: $\mu_s > \mu_c$
e) H_0: $\mu_s = \mu_c$, H_a: $\mu_s < \mu_c$

The null hypotheses is "no difference," or that the means are equal. The alternative hypothesis is that there is a difference.

Difference in Hypotheses

- Note: the following two sets of hypotheses are equivalent, and can be used interchangeably:

$H_0: \mu_1 = \mu_2$ $\quad$ $H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 \neq \mu_2$ $\quad$ $H_a: \mu_1 - \mu_2 \neq 0$

Hypotheses

- Take a minute to write down the hypotheses for each of the following situations:
  - Does the proportion of people who support gun control differ between males and females?

p_f: proportion of females who support gun control
p_m: proportion of males who support gun control

$H_0: p_f = p_m$
$H_a: p_f \neq p_m$

  - Is the average hours of sleep per night for college students less than 7?

μ: average hours of sleep per night for college students

$H_0: \mu = 7$
$H_a: \mu < 7$

Your Own Hypotheses

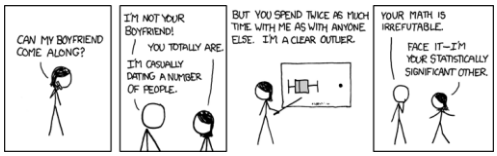
- Come up with a situation where you want to establish a claim based on data
- What parameter(s) are you interested in?
- What would the null and alternative hypotheses be?
- What type of data would lead you to believe the null hypothesis is probably not true?

Statistical Significance

When results as extreme as the observed sample statistic are *unlikely* to occur by random chance alone (assuming the null hypothesis is true), we say the sample results are ***statistically significant***

- If our sample is statistically significant, we have convincing evidence against H_0, in favor of H_a
- If our sample is not statistically significant, our test is inconclusive

Statistical Significance



www.xkcd.com

Note on Statistical Significance

- Statistical significance is a difficult concept, but also one of the most fundamental concepts of the course
- We return to this concept almost every class for the rest of the semester, so
  - it will get easier!
  - it's worth thinking deeply about!

Sleep versus Caffeine

μ_s and μ_c: mean number of words recalled after sleeping and after caffeine

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

The sample difference in means is $\bar{x}_s - \bar{x}_c = 3$, and this is statistically significant. We can conclude…

a) there is a difference between sleep and caffeine for memory (and data show sleep is better)

b) there is not a difference between sleep and caffeine for memory

c) nothing

Extrasensory Perception

p = Proportion of correct guesses

H_0: $p = 1/5$

H_a: $p > 1/5$

- If results are statistically significant…
  - the sample proportion of correct guesses is higher than is likely just by random chance (if ESP does not exist and $p = 1/5$)
  - we have evidence that the true proportion of correct guesses really is higher than 1/5, and thus have evidence of ESP

Extrasensory Perception

p = Proportion of correct guesses

H_0: $p = 1/5$

H_a: $p > 1/5$

- If results are NOT statistically significant…
  - the sample proportion of correct guesses could easily happen just by random chance (if ESP does not exist and $p = 1/5$)
  - we do not have enough evidence to conclude that $p > 1/5$, or that ESP exists

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
- How do we measure how unusual a sample statistic is, if H_0 is true?

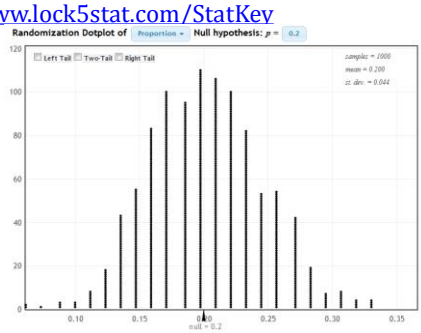
SIMULATE what would happen if H_0 were true!

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 – Random Chance

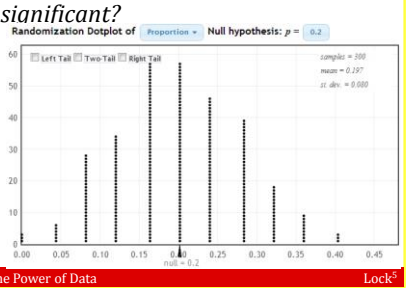
- www.lock5stat.com/StatKey



ESP

Based on the distribution below, do you think the results of our class ESP experiment 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

Summary

- Statistical tests use data from a sample to assess a claim about a population
- Statistical tests are usually formalized with competing hypotheses:
  - Null hypothesis (H_0): *no effect or no difference*
  - Alternative hypothesis (H_a): *what we seek evidence for*
- If it would be unusual to get results as extreme as that observed, just by random chance, if the null were true, then the data is statistically significant
- If data are statistically significant, we have convincing evidence against the null hypothesis, and in favor of the alternative

Project 1

- Pose a question that you would like to investigate. If possible, choose something related to your major!
- Find or collect data that will help you answer this question (you may need to edit your question based on available data)
- You can choose either a single variable or a relationship between two variables
- See Finding Data for help finding data

Project 1

- The result will be a five page paper including
  - Description of the data collection method, and the implications this has for statistical inference
  - Descriptive statistics (summary stats, visualization)
  - Confidence interval
  - Hypothesis test
- Proposal due next Monday, 2/17
  - If using existing data: include link to data, relevant summary statistic
  - If collecting your own data, proposal should include a detailed data collection plan

Sample or Population?

- If your data represents a sample from a population, inference makes sense
- If you have data on the entire population, you will be asked to take a random sample and do inference

To Do

- Read Section 4.1
- Project 1 Proposal (due Monday 2/17)