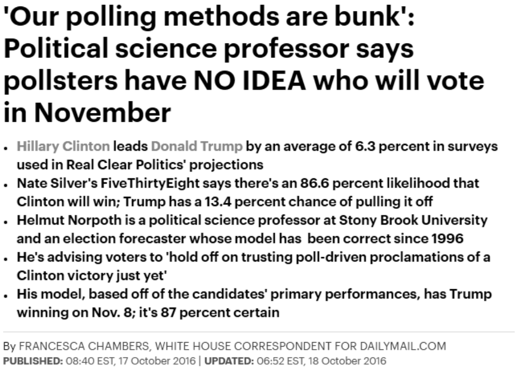
Stat 104: Quantitative Methods

Lecture 23: The Language of Hypothesis Testing

Introduction to Hypothesis Testing, Part I



'Our polling methods are bunk': Political science professor says pollsters have NO IDEA who will vote in November

- **Hillary Clinton leads Donald Trump by an average of 6.3 percent in surveys used in Real Clear Politics' projections**
- **Nate Silver's FiveThirtyEight says there's an 86.6 percent likelihood that Clinton will win; Trump has a 13.4 percent chance of pulling it off**
- **Helmut Norpoth is a political science professor at Stony Brook University and an election forecaster whose model has been correct since 1996**
- **He's advising voters to 'hold off on trusting poll-driven proclamations of a Clinton victory just yet'**
- **His model, based off of the candidates' primary performances, has Trump winning on Nov. 8; it's 87 percent certain**

By FRANCESCA CHAMBERS, WHITE HOUSE CORRESPONDENT FOR DAILYMAIL.COM
PUBLISHED: 08:40 EST, 17 October 2016 | **UPDATED:** 06:52 EST, 18 October 2016

Los Angeles Times

Op-Ed Sick of political polls? Try prediction markets

A hybrid between sports betting and derivatives markets, these allow traders to buy and sell shares that will pay out if a certain political

Who will win the 2016 U.S. presidential election?
Top Predictions — 840570 Comments
Clinton 1624 — 82¢ ↑ 1¢
Trump 3182 — 20¢ ↑ 1¢

Which party will win the White House?
Top Predictions — 3844 Comments
Democratic 673 — 82¢ ↓ 1¢
Republican 1088 — 21¢ NC

Will the next elected President be a woman?
Current Market — 1099 Comments
Yes — 81¢ ↓ 1¢
No — 19¢ ↑ 1¢

https://www.predictit.org/

Hypothesis Testing (intro)

We have discussed two methods of making inference on parameters in a population based on a random sample (in English, how do we figure out the true mean or proportion)

- *Point estimates* give us a single "guess"
- *Confidence intervals* give us a region.

Alternatively, we might be interested in using the information in a sample to *test hypothesis* about parameters in the population.

What is a Hypothesis?

- ❑ A **hypothesis** is a statement regarding a characteristic of one or more populations.
- ❑ In this section, we look at hypotheses regarding a single population parameter.

Hypothesis Examples, Single Population

- ■ In 2008, 62% of American adults regularly volunteered their time for charity work. A researcher believes that this percentage is different today.
- ■ According to a study published in March, 2006 the mean length of a phone call on a cellular telephone was 3.25 minutes. A researcher believes that the mean length of a call has increased since then.
- ■ Using an old manufacturing process, the standard deviation of the amount of wine put in a bottle was 0.23 ounces. With new equipment, the quality control manager believes the standard deviation has decreased.

Caution

- We test these types of statements using sample data because it is usually impossible or impractical to gain access to the entire population.
- So there's always a chance of making a mistake ☹
- If population data are available, there is no need for inferential statistics

Recap : Confidence Intervals

- Allow us to use sample data to **estimate** a population value, like the true mean or the true proportion, and then put bounds on our estimate.
- *Example:* Give a 95% confidence interval for the true average amount students spend weekly on alcohol.

Recap : Hypothesis Testing

- Allows us to use sample data to **test a claim** about a population, such as testing whether a population proportion or population mean equals some number.
- *Example:* Is the true average amount that students spent weekly on alcohol $20?

Parameter Identification

- Hypothesis tests can be carried out on all the **population parameters** (the Greeks), such as the population mean, median, proportion or variance.
- But in this section, we will conduct tests of hypothesis only regarding the population mean μ or the population proportion p.

Hypothesis Testing Steps

1. Make a statement regarding the nature of the population.
2. Collect evidence (sample data) to test the statement.
3. Analyze the data to assess the plausibility of the statement.

The Null Hypothesis

- The **null hypothesis**, denoted H_0, is a statement to be tested.
- The null hypothesis is a statement of no change, no effect or no difference and **is assumed true until evidence indicates otherwise.**

The Alternative Hypothesis

- The **alternative hypothesis**, denoted H_a, is a statement that we are trying to find evidence to support.
- Since we only have sample data, we can really only disprove a theory, not prove it, since we haven't seen all the data.

H_0: all cats have four legs

Setting up Ho and Ha

1. Equal versus not equal hypothesis (**two-tailed test**)
   H_0: parameter = some value
   H_a: parameter ≠ some value
2. Equal versus less than (**left-tailed test**)
   H_0: parameter = some value
   H_a: parameter < some value
3. Equal versus greater than (**right-tailed test**)
   H_0: parameter = some value
   H_a: parameter > some value

Status Quo or "No Effect"

- The null hypothesis is a statement of *status quo* or *no difference* and always contains a statement of equality.
- The null hypothesis is assumed to be true until we have evidence to the contrary.
- We seek evidence that supports the statement in the alternative hypothesis.

Forming Hypotheses

- Set up Ho and Ha for the following situation
- In 2008, 62% of American adults regularly volunteered their time for charity work. A researcher believes that this percentage is different today.

Forming Hypotheses

- Set up Ho and Ha for the following situation
- According to a study published in March, 2006 the mean length of a phone call on a cellular telephone was 3.25 minutes. A researcher believes that the mean length of a call has increased since then.

Forming Hypotheses

- Set up Ho and Ha for the following situation
- Using an old manufacturing process, the standard deviation of the amount of wine put in a bottle was 0.23 ounces. With new equipment, the quality control manager believes the standard deviation has decreased.

Four Outcomes from Hyp Testing

1. Reject the null hypothesis when the alternative hypothesis is true. This decision would be correct.
2. Do not reject the null hypothesis when the null hypothesis is true. This decision would be correct.
3. Reject the null hypothesis when the null hypothesis is true. This decision would be incorrect. This type of error is called a **Type I error.**
4. Do not reject the null hypothesis when the alternative hypothesis is true. This decision would be incorrect. This type of error is called a **Type II error.**

Recap-What can go wrong ?

Because the conclusion we will be making is based on sample data, the possibility of making an error always exists.

and we Claim that

IF	H_o is true	H_o is false
H_o is true	Correct Decision (no error)	**Type I error**
H_o is false	**Type II error**	Correct Decision (no error)

Example: Type I and Type II Errors

- In 2008, 62% of American adults regularly volunteered their time for charity work. A researcher believes that this percentage is different today.
- A **Type I** error is made if the researcher concludes that $p \neq 0.62$ when the true proportion of Americans 18 years or older who participated in some form of charity work is currently 62%.
- A **Type II** error is made if the sample evidence leads the researcher to believe that the current percentage of Americans 18 years or older who participated in some form of charity work is still 62% when, in fact, this percentage differs from 62%.

Example: Type I and Type II Errors

- According to a study published in March, 2006 the mean length of a phone call on a cellular telephone was 3.25 minutes. A researcher believes that the mean length of a call has increased since then.
- A **Type I** error occurs if the sample evidence leads the researcher to conclude that $\mu > 3.25$ when, in fact, the actual mean call length on a cellular phone is still 3.25 minutes.
- A **Type II** error occurs if the researcher fails to reject the hypothesis that the mean length of a phone call on a cellular phone is 3.25 minutes when, in fact, it is longer than 3.25 minutes.

Notation for Type I and Type II Errors

$$\alpha = P(\text{Type I Error}) = P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true})$$

$$\beta = P(\text{Type II Error}) = P(\text{not rejecting } H_0 \text{ when } H_1 \text{ is true})$$

Which error is worse?

- The construction of the hypotheses are such that the Type I error is considered the worse of the two.
- The probability of making a Type I error, α, is chosen by the researcher *before* the sample data is collected.
- The **level of significance**, α, is the probability of making a Type I error

Example

H_o: Defendant is not guilty.

H_a: Defendant is guilty.

What is the Type I Error ?

What is the Type II Error ?

Which error is more important ?

The milemaster tire company has decided that their new tire must last more than 45,000 miles or they wont market it.

H_o: tire lasts 45,000 (or less)

H_a: tire lasts more than 45,000.

What is the type I error ? The type II error ?

What is the cost of a type I error here ?

What is the cost of a type II error ?

The Significance Level

We define

$$\alpha = \text{Prob(Type I error)} = \text{P(reject } H_o | H_o \text{ true)}$$

We normally use α=.05 : called the **significance level**

This is a pretty standard value to use, but not set in stone. Usually the greater the cost of a type I error, the smaller this number is.

Caution in Stating Your Conclusion

- We never “accept” the null hypothesis, because, without having access to the entire population, we don’t know the exact value of the parameter stated in the null.
- Rather, we say that we do not reject the null hypothesis. This is just like the court system. We never declare a defendant “innocent”, but rather say the defendant is “not guilty”.

Example: Stating Your Conclusion

According to a study published in March, 2006 the mean length of a phone call on a cellular telephone was 3.25 minutes. A researcher believes that the mean length of a call has increased since then.

a) Suppose the sample evidence indicates that the null hypothesis should be rejected. State the wording of the conclusion.

b) Suppose the sample evidence indicates that the null hypothesis should not be rejected. State the wording of the conclusion.

Reject the Null

Suppose the sample evidence indicates that the null hypothesis should be rejected. State the wording of the conclusion.

The statement in the alternative hypothesis is that the mean call length is greater than 3.25 minutes. Since the null hypothesis (μ = 3.25) is rejected, there is sufficient evidence to conclude that the mean length of a phone call on a cell phone is greater than 3.25 minutes.

Fail to Reject the Null

Suppose the sample evidence indicates that the null hypothesis should not be rejected. State the wording of the conclusion.

Since the null hypothesis (μ = 3.25) is not rejected, there is insufficient evidence to conclude that the mean length of a phone call on a cell phone is greater than 3.25 minutes. In other words, the sample evidence is consistent with the mean call length equaling 3.25 minutes.

Actually Testing Hypotheses

- As you can see, there is a lot of terminology we had to go through before we get to test any hypotheses
- There are 3 methods for actually doing the hypothesis test:
  - Confidence Interval Method (classical)
  - Test statistic method (classical)
  - P-values (using the computer)

Two Tailed Test Using a Confidence Interval

- The level of confidence, (1 - α) x 100%, in a confidence interval represents the percentage of intervals that will contain the unknown parameter if repeated samples are obtained.
- When testing $H_o:\theta=\theta_o$ vs $H_a:\theta\neq\theta_o$, if a (1 - α) x 100% confidence interval contains θ_o, we do not reject the null hypothesis. If the confidence interval does not contain θ_o, we conclude that $\theta\neq\theta_o$ at the level of significance α.
- For now we are assuming *n* is large.

Testing a Two-Sided Hypothesis

- ❑ Step 1: Define Hypothesis
  - $H_0 : \theta = \theta_o$
  - $H_a : \theta \neq \theta_o$
- ❑ Step 2: Construct confidence interval
- ❑ Step 3: Accept or Reject
  - If θ_o falls within this interval, then we fail to reject the null, otherwise we reject it.

Example

A machine being used for packaging seedless golden raisins has been set so that on the average 15 ounces of raisins will be packaged per box. The quality control engineer wishes to test the machine setting and selects a sample of 30 raisin boxes. He finds $\bar{x}$=15.11 and *s*=0.4058

He wants to know if the mean weight per box is different from 15 ounces.

$$H_0: \mu = \underset{\underset{\mu_0}{\uparrow}}{15} \quad vs. \quad H_a: \mu \neq 15$$

Solution

- The 95% Confidence Interval is

```
> tsum.test(mean.x=15.11,s.x=.4058,n.x=30)
        One-sample t-Test
data:  Summarized x
t = 203.94, df = 29, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 14.95847 15.26153
```

- What can we conclude about

$$H_0: \mu = 15 \quad vs. \quad H_a: \mu \neq 15$$

Logic

- Do we know the true value of μ?
- However, we are pretty confident that the true value of μ is somewhere in the confidence interval!
- So when we ask if μ=15, all we need to do is see if 15 is in the confidence interval. If it is, its plausible that μ=15. But if 15 is not in the confidence interval, we would reject the null hypothesis since then 15 is not a plausible value for μ.

What's the chance of making an error?

- How much confidence do we have in our 95% confidence interval?
- 95%!
- Or better stated (1-α)x100%
- We call α our <u>level of significance</u>; it is also referred to as a Type I error

Example

- A manufacturer of sports equipment has developed a new synthetic fishing line that the company claims has a mean breaking strength of 8 kilograms.
- Test the hypothesis that μ=8 kilograms against the alternative that μ ≠ 8 kilograms if a random sample of 50 lines is tested and found to have a mean breaking strength of 7.8 kilograms with a standard deviation of 0.5 kilograms.

Fishing Line Example

- The 95% confidence interval is

```
> tsum.test(mean.x=7.8,s.x=.5,n.x=50)
        One-sample t-Test
data:  Summarized x
t = 110.31, df = 49, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 7.657902 7.942098
```

- Conclusion?

Example

- When a robot welder is in adjustment, its mean time to perform its task is 1.3250 minutes.
- An incorrect mean operating time can disrupt the efficiency of other activities along the production line.
- For a recent random sample of 80 jobs, the mean cycle time for the welder was 1.3229 minutes with a std deviation of 0.0396.
- Does the machine appear to be in need of adjustment?

Example (cont)

- We want to test

$$H_0 : \mu = 1.325 \quad vs. \quad H_a : \mu \neq 1.325$$

- The 95% CI is

$$1.325 \pm 1.96\left(\frac{.0396}{\sqrt{80}}\right) = (1.316, 1.334)$$

- Conclusion; fail to reject Ho-the robot is not in need of adjustment.

Example with Proportions

- The manufacturer of the Bic Extended Lighter claims that it lights on the first time 75% of the time. Test this claim.
- We want to test

Two sided Proportion Test

- Suppose we make 300 attempts and the lighter lights on the first try 214 times.
- The resultant confidence interval is then

$$\frac{214}{300} \pm 1.96\sqrt{\frac{\frac{214}{300}\left(1-\frac{214}{300}\right)}{300}} = (0.662, 0.764)$$

Example with Proportions

- A new blog by politico.com says that 60% of Americans trust the president. You want to test this claim so you randomly survey 1000 adults and find that 637 trust the president. <u>What can you conclude?</u>
- You want to test the claim

$$H_o : p = 60\% \qquad H_a : p \neq 60\%$$

Testing the claim

- We will be lazy and use R for the calculations

```
> prop.test(637,1000)

        1-sample proportions test with continuity correction
data:  637 out of 1000, null probability 0.5
X-squared = 74.529, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.6062176 0.6667163
```

- What can we conclude?

One Sided Tests

- It is very easy and straightforward to test a two sided hypothesis test using a confidence interval.
- However, we would also like to be able to test one-sided hypotheses.
- These can be done using one-sided confidence intervals, or as we will see next time, using what is called a <u>test statistic</u>.

<u>Things you should know</u>

- ☐ Null and alternative hypothesis
- ☐ Type I and II error
- ☐ Using a confidence interval to test a large sample, two sided hypothesis for the mean and proportion.