

What is Hypothesis Testing and Why Should I Care What the P-value Is?

0011 0010 1010 1101 0001 0100 1011

Mario Davidson, PhD
Biostatistician III
Vanderbilt University
Nashville, TN



Hypothesis Testing

- Means for Large Samples
- Means for Small Samples
- Proportions
- Chi Square
- Type I & II Errors

0011 0010 1010 1101 0001 0100 1011



Our Love Will Last Forever

- Before accepting my proposal, my wife asked me to prove our love was destiny.
- Dr. Davidson (1st Try): Girrrl, every since I met you my dreams have come true. The sun's brighter and the candy's sweeter.
- Dr. Davidson (2nd Try): Baby you **MUST** be the jam to my peanut butter! If our love was by chance alone, the probability of us meeting at the same university, in the same building is 0.0001. You gotta be da one!

Hypothesis Testing

1. Sometimes we want to test a claim (Ex: Present Day Love vs. Love Forever).
2. In order to test this claim, we gather “evidence (Ex: Meeting at the same university in the same building)”
3. We then test the strength of our “evidence” by asking, “What's the chance of finding that “evidence (Ex: The probability of us meeting... is 00001.)?”
4. Make a decision (Ex: Reject this notion)
 - Finally, we interpret our results (Ex: Our love is destiny!). (Thanks A. Shintani)

Hypothesis Testing (cont.)

- Stating the hypothesis test
 - The null has the sign of equality (\leq , \geq , $=$)
 - The alternative has the sign of inequality ($<$, $>$, \neq)

0011 0010 1010 1101 0001 0100 1011

$$H_0: \heartsuit \leq \text{Present Love}$$

$$H_1: \heartsuit > \text{Present Love}$$

Hypothesis Testing (cont.)

- One-sided test
 - We are interested in if the parameter is less than or greater than some value.
- Two sided
 - We are interested in if the parameter is not equal to some value.
- Alpha
 - How strong do we need our evidence to be?
- p-value

0011 0010 1010

1
2
4
5

Hypothesis Testing for the Mean: The Expense of Medical School

In 2006, the average educational debt of indebted graduates of medical school was \$135,000. Chancellor Hamilton who is interested in testing this claim conducted a study of 60 randomly chosen 2006 graduates and found the average debt to be \$126,000 with a SD of \$47,000. Do we have evidence to suggest the average is \$135,000?

Hypothesis Testing for the Mean (cont.)

- Always ask: What results do I expect?
- Claim:

H_0 : Average = 135,000 (claim) vs.

H_1 : Average \neq 135,000 (two-sided test)

- Evidence: Sample average = 126,000
Sample SD = 47,000
- Strength of “Evidence”:

Test statistic (z) = -1.48; p-value = .138

Hypothesis Testing for the Mean (cont.)

- Decision: Fail to reject the null.
- Interpretation: We have insufficient evidence to suggest that the average debt from medical school was not \$135,000 in 2006.



Hypothesis Testing for the Mean: Hours Residents Work

Dean Wickless claims that residents work 83 hours/week or less. She conducts a study of 25 randomly chosen residents and found the average to be 88 with a SD of 3.4. Test the claim that residents work more than 83 hours/week.



Hypothesis Testing for the Mean (cont.)

- Always ask: What results do I expect?

- Claim:

H_0 : Average ≤ 83 (claim) vs.

H_1 : Average > 83 (one-sided test)

- Evidence: Sample average = 88

Sample SD = 3.4

- Strength of “Evidence”

Test statistic (t) = 7.4; p-value $< .0001$

Hypothesis Testing for the Mean (cont.)

- Decision: Reject the null.
- Interpretation:

We have evidence to suggest that the average hours worked by residents is more than 83.



Test Your Knowledge (Group Discussion)

Create a problem that requires testing for the mean.

- State the problem for both a one-sided test.
- What would be the interpretations if the p-value was .04?
- What would be the interpretation if the p-value was .78?

Hypothesis Testing for Proportions: Adherence for ART

Dr. Stephen believes that less than 58% of HIV patients in Mozambique are adherent to their anit-retroviral treatment (ART). He randomly chose 68 Mozambicans and found that 54% were adherent. Test his claim.

45

Hypothesis Testing for Proportions (cont.)

- Always ask: What results do I expect?
- Claim:

H_0 : Proportion $\geq 58\%$ vs.

H_1 : Proportion $< 58\%$ (one-sided test)

- Evidence: Sample proportion = 54%
- Strength of “Evidence:”

Test statistic (z) = -.60; p-value = .27

Hypothesis Testing for Proportions (cont.)

- Decision: Fail to reject the null.
- Interpretation:

We have insufficient evidence to suggest that the proportion of Mozambicans who adhere to taking ART is less than 58%.

12
45

0011 0010 1010 1101 0001 0100 1011

Hypothesis Testing for Independence Using Chi Square

Mr. Watson is interested in if race is associated with infant mortality. He randomly chooses 107 records of women who were pregnant.

	Infant Mortality		
Race	Yes	No	Total
Black	15	27	42
White	12	53	65
Total	27	80	107



Hypothesis Testing for Independence (cont.)

- Always ask: What results do I expect?

- Claim:

H_0 : Mortality and race are independent

H_1 : Mortality and race are dependent

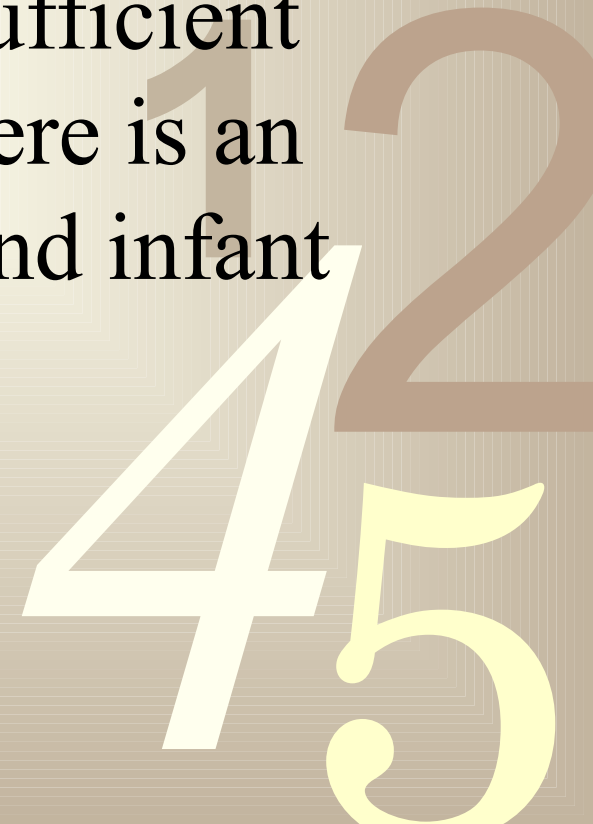
- Strength of “Evidence:”

Test statistic (χ^2) = 4.03; p-value = .045

Hypothesis Testing for Independence (cont.)

- Decision: Reject the null.
- Interpretation: We have sufficient evidence to suggest that there is an association between race and infant mortality.

0011 0010 1010



Test Your Knowledge

- Because of the importance of time in conducting intraoperative analysis on breast cancer patients, Dr. Phillips wants to study the time to conduct this procedure. She believes the average time for this procedure is 18 minutes. Dr. Phillips randomly chooses 82 surgeons and finds that the average time is 17 minutes with a SD of 2.6.

1. To test Dr. Phillips' claim, which should be conducted: a test of means, proportions or independence?
2. State the null and alternative hypothesis.
3. Is the test one-sided or two-sided?
4. If the p-value is .0003, what is your decision?
5. Interpret the results.

Test Your Knowledge (cont.)

1. To test Dr. Phillips' claim, which should be conducted: a test of means, proportions or independence? Test of means
2. State the null and alternative hypothesis:
 $H_0: \mu = 18$; $H_a: \mu \neq 18$
3. Is the test one-sided or two-sided? Two-sided
4. If the p-value is .0003, what is your decision? Reject the null hypothesis
5. Interpret the results. We have evidence to suggest that the time to conduct the surgery is not 18 minutes.

Death Row: Innocent or Not?

- 0011 0010 1010 1101 0001 0100 1011
- **Hypothesis Test:** It is estimated that more than 3% of the people on death row are innocent. A random sample of 100 death row inmates and former death row inmates was taken. Out of those 100, six (6%) had supporting evidence of their innocence.





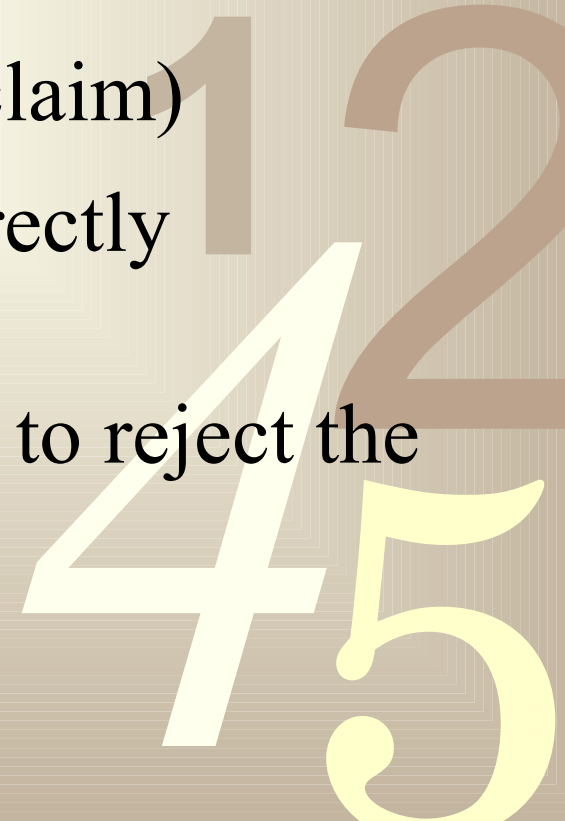
Death Row: Innocent or Not (cont.)

Group Discussion

$$H_0: p \leq 3\%$$

$$H_1: p > 3\% \text{ (claim)}$$

- What would it mean if we incorrectly rejected the null hypothesis?
- What would it mean if we failed to reject the null hypothesis incorrectly?
- Which error is worse?





Death Row: Innocent or Not (cont.)

Group Discussion

- What would it mean if we incorrectly rejected the null hypothesis? **We have evidence that the proportion of death row inmates who are innocent is more than 3%, when it is actually less than 3%.**
- What would it mean if we failed to reject the null hypothesis incorrectly? **We have insufficient evidence that the proportion of innocent death row inmates is greater than 3% when it is actually more than 3%**

Death Row: What is the Cost of an Error?

- 0011 0010 1010 1101 0001 0100 1011
- So I ask you again, which error is worse?
 - Suppose we are opponents of the death penalty.
 - Suppose we are proponents of the death penalty.



Type I & Type II Error

0011

	Truth of H_0	
Decision	H_0 is true	H_0 is false.
Do not Reject H_0	Correct decision	Type II Error
Reject H_0	Type I Error	Correct

45

Revisiting the P-value

- Using probability and under the assumption that the null hypothesis is true, the p-value is a measure of the strength of our “evidence.”
- How much “evidence” would you need for the death row example?
- The p-value is compared to your Type I error.



1
2
4
5



Death Row:

Innocent or Not (cont.)?

$$H_0: p \leq 3\%$$

$$H_1: p > 3\% \text{ (claim)}$$

$$Z = 1.76; \text{ } p\text{-value} = .039$$

Decision: (At Type I Error = .05) Reject H_0 .

Interpretation: (At Type I Error = .05) We have evidence to suggest that the proportion of death row inmates who are innocent is greater than 3%.

Conclusion

- Lecture is non-exhaustive
 - There are a host of additional hypothesis test that may be conducted.
- If planning a study, talk to a professional.
- Questions

