Inferential STATS

Chapter 10: Hypothesis Tests Regarding a Parameter

Section 10.1: Basics of Hypothesis Testing

GOAL: Make a decision about p or μ based on $\hat{p}$ or $\bar{x}$ using probability theory. (sampling dist)

CI	
Chapter 9: Find a sample then **estimate** whether the population fits within a certain interval.	**Chapter 10**: Given a *past* claim of the parameter, we will test whether or not the claim has *changed*.

STRUCTURE OF A HYPOTHESIS TEST

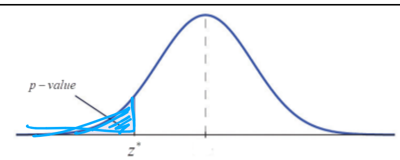
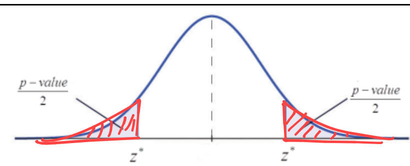
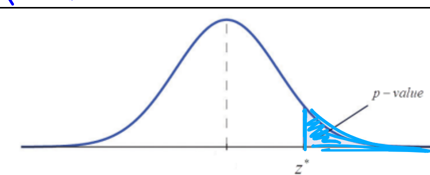
1) Make an assumption about reality	• make a hypothesis about a population parameter. (p or μ)
2) Look at a sample evidence	• find a point estimate to test claim about pop. parameter.
3) Determine whether it contradicts our assumption.	• make a decision if pop. parameter has changed.

We won't be 100% certain, we will just be able to tell if sample data supports a statement or not.

HYPOTHESES STATEMENTS

NULL HYPOTHESIS (H_0)	ALTERNATIVE HYPOTHESIS (H_A) also: H_1
A statement of no change, no effect, no difference and is assumed true until *statistical* evidence indicates otherwise.	A statement that we are trying to find evidence to support instead of the null hypothesis.

THREE TYPES OF HYPOTHESIS TESTS → how phrase alternative hyp.

LEFT-TAILED	TWO-TAILED	RIGHT-TAILED
H_0: *parameter* = # "past" H_A: *parameter* < #	H_0: *parameter* = # H_A: *parameter* ≠ #	H_0: *parameter* = # H_A: *parameter* > #
		
ie got smaller	ie it changed!	ie got bigger

GTQ

Ex1: **1)** What's the parameter? **2)** What do "they say"? (H_0) **3)** What do we think? (H_A) **4)** What type of test?

The packaging on a light bulb says it should last 500 hours. Consumer Reports wants to know if the mean lifetime is actually less than that.	The standard deviation of the rate of return for some mutual funds is 0.08%. A manager believes the standard deviation might be higher than that.	According to a Gallup poll in 2008, 80% of Americans felt satisfied with the way things are going in their lives. A researcher wonders if the percentage is different now.
1) parameter: mean life bulb μ 2) and 3) $H_0: \mu = 500$ hrs/bulb. $H_A: \mu < 500$ 4) Left-tailed test	1) parameter: standard dev. σ 2) and 3) $H_0: \sigma = 0.0008$ "past" $H_A: \sigma > 0.0008$ 4) RIGHT-tailed test	1) parameter: proportion p 2) and 3) $H_0: p = 0.8$ "past" $H_A: p \neq 0.8$ 4) Two-tailed Test

TWO POSSIBLY CORRECT CONCLUSIONS:

1) We decide there **is** evidence to *support* H_A	2) We decide there **is NOT** enough evidence to *support* H_A
"Reject the null hypothesis" Note we're not going to say we "proved" the alternative hyp.	"Fail to Reject the null hypothesis" Note we keep the null hyp but never "proved it"

- The language we choose is like in a court system. We don't say a defendant is proven innocent or that we accept that they are innocent, we say "not guilty" when there isn't enough evidence to conclude guilt.
- The role of the "alternative hypothesis": this is what we are trying to find evidence for. Namely, this is the claim we are interested in determining whether there's enough statistical evidence to support it (or not).

Ex2: Historically, Jimbo's pizza had a mean delivery time of 48 minutes. After getting a new pizza oven, he takes a sample of 50 orders and finds that the mean delivery time is now 45 minutes, which makes Jimbo think that the mean delivery time has been reduced.

parameter: mean μ = average delivery time for pizza.

State Jimbo's hypotheses in statistical notation:

$$\begin{cases} H_0: \mu = 48 \text{ min} \\ H_A: \mu < 48 \text{ min} \end{cases}$$

State the conclusion if the **null is rejected:** → support H_A

"There is enough statistical evidence to support that the true (population) mean pizza delivery time has decreased."

State the conclusion if the **null is *not* rejected:** → keep H_0 don't support H_A

"There is not enough statistical evidence to support that the true (population) mean pizza delivery time has decreased."

FOUR POSSIBLE OUTCOMES (2 ERRORS)

Ex3: Is there poison in the well?

H_0: **No** poison in the well

H_1: There **IS** poison in the well

		Truth about the Population (**Reality**)	
		H_0 is true	H_0 is false
Decision Based On Sample (**Our Conclusion**)	Fail to Reject H_0 keep H_0	Conclude there is no poison in well when no poison in well Good	Conclude there is NO poison in well when there is poison in well BAD! II
	Reject H_0 (accept H_A)	Conclude there IS poison in well when no poison in well BAD! I	Conclude there is poison in well when there is poison in well Good

NOTE: We never *PROVE* that there's no poison in the well!!!

conclusions	H_0 T	H_0 F
FTR H_0	☺	II
R H_0	I	☺

TYPE I AND TYPE II ERRORS

Type I error: The mistake of rejecting the null hypothesis when it is actually true.
The symbol α (alpha) is used to represent the probability of such an error.

$$\alpha = P(\,Reject\ H_0 \mid H_0\ is\ True\,)$$

Type II error: The mistake of failing to reject the null hypothesis when it is actually false.
The symbol β (beta) is used to represent the probability of such an error

$$\beta = P(\,Fail\ to\ Reject\ H_0 \mid H_0\ is\ False\,)$$

$$P(\text{conclusion} \mid \text{reality})$$

Ex4: In the poison in the well example, which type of error is worse? Type II (people get sick!)

Ex5: On average, it used to take 30 minutes to find parking, but we think we have sufficient evidence to say that the time has decreased. But, in fact, the true parking time is still 30 minutes. What kind of error did we make?

past μ now

$H_0: \mu = 30$ min

$H_A: \mu < 30$ min

conclude: $H_A \to$ reject H_0

reality H_0 is still true.

Error $P(\text{conclusion} \mid \text{reality})$

$= P(R\ H_0 \mid H_0\ T) = \alpha$ $\boxed{\text{Type I Error}}$

	H_0 T	H_0 F
FTR H_0		II
R H_0	I	

Hypothesis Tests about Majorities

The **majority** is a number or percentage equaling more than 50% of a total.

To test a claim about a majority, your null hypothesis will be $p =$ 0.5.

Ex6: A Gallup survey reports that 57% of 504 randomly selected gun owners support stricter gun laws. Test the claim that a majority of gun owners favor stricter gun laws. Write out the hypotheses for this example. What would a Type II error be in this scenario?

sample past Hypothesis Test

parameter: proportion

$H_0: p = 0.5$

$H_A: p > 0.5$ (majority)

Type II Error: $\beta = P(\underset{\text{conclude}}{\text{FTR } H_0} \mid \underset{\text{reality}}{H_0\ F})$

FTR H_0: We conclude that proportion of people who favor stricter gun laws is 50%.

H_0 F: when in reality the proportion of people who favor stricter gun laws is greater.

Ex7: Your company markets a computerized device to test a patient's mean resting heart rate. Based on the sample results, the device determines whether there is significant evidence that the patient's mean resting heart rate is greater than 100 beats per minute. If so, your company recommends that the person seeks medical attention.

μ

a. State appropriate null and alternative hypotheses in this setting.

b. Which error is worse for your company?

• parameter: mean μ
↳ mean beats per minute of heart rate of patient

H_0: $\mu = 100$ bpm

H_1: $\mu > 100$ bpm

reality

conclusions

	H_0 is true	H_0 is false
Fail to Reject H_0	conclude: $\mu = 100$ bpm reality: $\mu = 100$ bpm	**Type II Error** conclude: $\mu = 100$ bpm reality: $\mu > 100$ bpm
Reject H_0	**Type I Error** conclude: $\mu > 100$ bpm reality: $\mu = 100$ bpm	conclude: $\mu > 100$ bpm reality: $\mu > 100$ bpm

Seek Med Attention? YES NO
Did They Need It? YES NO

↳ Terrible outcome!

Seek Med Attention? YES NO
Did They Need It? YES NO

↳ "false alarm"
Bad, but not as bad as Type II Error!

We will NOT know 100% if our conclusion of our Hypothesis Test is correct, but we can assign probabilities to making Type I and Type II Errors when we complete a hypothesis test.

Level of Significance (Type I Error)	The probability of making a Type I Error. In other words, we take a sample that makes H_0 look WRONG when it's actually TRUE. $\alpha = P(R\ H_0 \mid H_0\ T)$

FACT: As you decrease the probability of one type of error,
THEN the probability of the other type increase.

Choosing a Significance Level

Typically the significance level, α is given to be between 0.01 (1%) and 0.1 (10%).

When a Type I error is…