

HYPOTHESIS TESTING

The basics

Hypothesis testing

$$y_1, y_2, \dots, y_n$$
$$\bar{Y}, S^2$$

- Recall: Statistical inference involves using sample data to make conclusions about a population.
- With estimation (point or interval) the goal is to estimate some parameter that summarizes some aspect of the population.
- With hypothesis testing, the goal is to use sample data to decide between two competing statements about the population.
- Typically, these statements ("hypotheses") are about the value(s) of some population parameter.

$$\mu, \sigma^2, \tau, p$$

Some examples

- Example 1: We gather some data trying to determine whether the rate of failure of some medical device is more than the claimed 3%.
- Example 2: We gather some data trying to decide whether one medication is better at reducing blood pressure than another medication.

μ_1 = "true" average reduction
with med 1

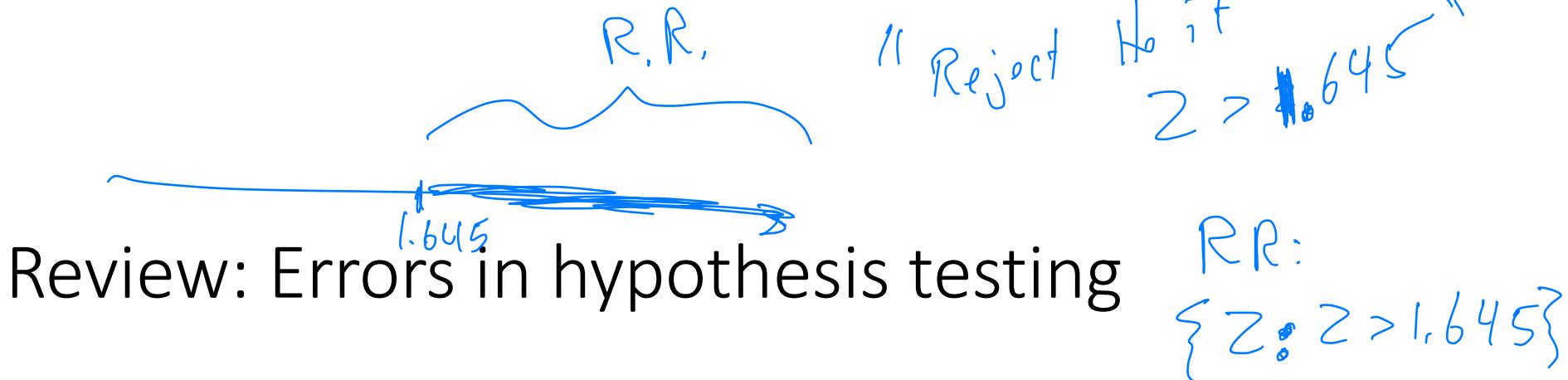
μ_2

Elements of a hypothesis test

- Definition 1: A STATISTICAL HYPOTHESIS is a statement about the value of one or more parameters.
- Two kinds:
 - Alternative hypothesis (H_a) – this is often the “research hypothesis”, something we might like to show is true.
 - Null hypothesis (H_0) – this is the hypothesis counter to H_a , complement of H_a , “previously held belief”, “status quo”, “no effect”, etc.
- Example 1 again: Parameter is p. *unknown*
- Example 2 again: Parameters are μ_1 and μ_2 .

Basics of statistical hypothesis testing

- Example 1 again: If we gathered 100 devices, tested them all, and 14 of them failed, what would we conclude?
- If the “true” failure rate is really 3% (i.e., $H_0: p = 0.03$), is it impossible to have 14 out of 100 fail? No
- If the “true” failure rate is really 3% (i.e., $H_0: p = 0.03$), is it unlikely to have 14 out of 100 fail? Yes!
- If our observed data are “extreme”, that would suggest that H_0 is false (equivalently, that H_a is true). True
- How extreme is “extreme enough”?
- We choose a test statistic to measure this.



Review: Errors in hypothesis testing

- Type I error – reject H_0 when in fact, H_0 is true.
- Type II error – fail to reject H_0 when in fact, H_0 is false (equivalently, when $, H_a$ is true).
- Statistical hypothesis testing involves defining a rejection region. If the observed test statistic falls into the rejection region, then we will reject H_0 . (If the observed test statistic does not fall into the rejection region, we fail to reject H_0 .)
- The basic idea: given a test statistic, determine a rejection region such that the probability of a Type I error is small, say α .
- We will use β to refer to the probability of a Type II error.

Example 1 revisited

- For the failure rate problem, the hypotheses are $H_0: p = 0.03$ and $H_a: p > 0.03$.
- Test statistic: Y , the number of failures (out of 100) in the sample.
- One option: set $RR = \{y: y \geq 7\}$. (A)
- Another option: set $RR = \{y: y \geq 12\}$. (B)
- Exercise 1: Calculate the α that corresponds to each RR. (Use normal approximation to the binomial.)
- Exercise 2: Calculate the β that corresponds to each RR if the true failure rate is 0.08.
- There is an α / β tradeoff!

$$\text{P}(Y \geq 7) \text{ if } p = 0.08$$

$$(A) \dots = \beta = 0.303$$

$$(B) \dots = 0.897$$

$$\begin{aligned} \text{(A)} \quad \alpha &= P(Y \geq 7) \text{ if } p = 0.03 \\ &= 1 - \text{pbinom}(6, p = 0.03, \text{size} = 100) \\ &= 0.0312 \end{aligned}$$

$$\begin{aligned} \text{(B)} \quad \beta &= 0.0000481 \end{aligned}$$