

# Module 9: Hypothesis Testing

The "effect" of a particular factor on some health outcome can be described as a parameter. Statistical hypothesis testing begins with a probability model assuming there is no effect or a **null hypothesis ( $H_0$ )** and deciding whether there is sufficient evidence to reject the null hypothesis in favor of the **alternative hypothesis ( $H_1$ )**. Think of it like a legal trial; innocent until proven guilty.

One-sided Hypothesis:  $H_0: \theta = \theta_0$  vs  $H_1: \theta > \theta_0$  (or  $\theta < \theta_0$ )

Two sided Hypothesis:  $H_0: \theta = \theta_0$  vs  $H_1: \theta \neq \theta_0$

We test  $H_0$  by finding an appropriate subset  $R \subset x$ , **the rejection region**.  $R$  is defined by  $R = \{x : T(x) > c\}$ ; where  $T$  is a **test statistic** and  $c$  is a **critical value**.

- If  $X \in R \rightarrow$  reject null hypothesis
- If  $X \notin R \rightarrow$  retain the null hypothesis

## Errors

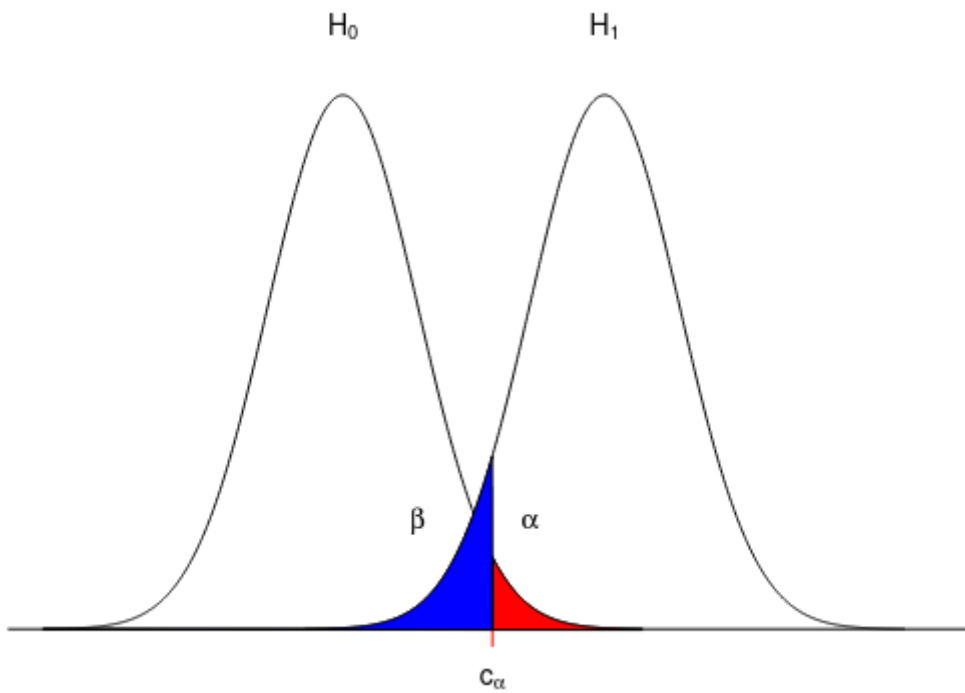
		Condition (gold standard)	
		Absent	Present
Index test	Negative	False positive rate	False negative rate
	Positive		

$$\alpha = 1 - \text{specificity}$$

$$1 - \beta = \text{sensitivity}$$

**Type I error** or **false positive** is rejecting  $H_0$  when it is true.  $\alpha$  represents the probability of this error (typically set at .05)

**Type II error** or **false negative** is failing to reject the null when it is false. Probability represented by  $\beta$ .



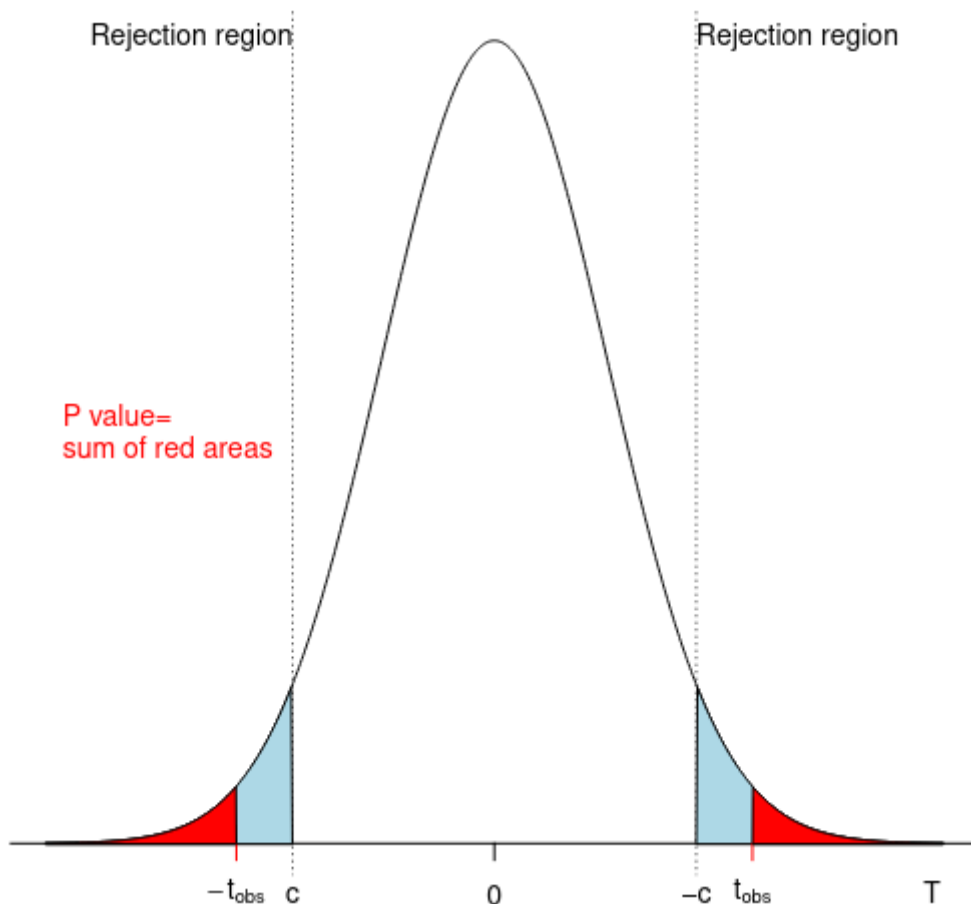
## Conducting and Interpreting the Test

1. Define null and alternative hypothesis
2. Set a desired  $\alpha$  level. We choose a critical value such that:  $P(T(x) > c \mid \theta = \theta_0) = \alpha$
3. Collect data
4. Calculate the observed test statistic value and compare to the critical value
5. Make decision

**P-value** is the smallest critical value at which the test leads to rejecting the null hypothesis. It is a probability, when assuming the null is true, of obtaining a test statistic at least as large as the one we observed.

- $p < \alpha \rightarrow$  reject null
- $p > \alpha \rightarrow$  retain null

Retaining the hypothesis does not mean the null is true, it is interpreted as a lack of evidence to accept the alternative.



We expect the data to come from the center of the distribution, there is a lower probability of pulling data from tails. If our sample statistic occurs at the tail, then this may not be a representative distribution.

- The **power function** of a test with rejection region  $R$  is:
  - $Q(\theta) = P(X \in R)$
- The **size** of a test is defined as:
  - $S(\theta) = \sup(Q(\theta))$
- A test is said to have **level  $\alpha$**  if its size is less than or equal  $\alpha$
- $1 - \beta$  is called the **power**, the probability of rejecting the null when the alternative of true
  - The alternative corresponds to a range of value, thus power will depend on whatever alternative parameter value we choose to consider
- These quantities allow identifying the adequate critical value for the test

If variance is known we can use Z-scores to compare to our critical value:

$$z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}}$$

Otherwise use a t-distribution with  $n-1$  degrees of freedom. A paired t-test is used when we are interested in the difference between two variables for the same subject when the null is that the

mean difference is 0:

$$t = \frac{\bar{d}}{\sqrt{S_d^2/n}}$$

$\bar{d}$  represents the difference in sample mean differences and the denominator is the standard error.

Likewise, we can also use t and z scores to create tests for proportion:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

## Relationship to Confidence Intervals

The null hypothesis can be tested against a two-sided alternative using a significance level  $\alpha$  by assessing if the null parameter is contained in the  $100(1 - \alpha)\%$  confidence interval for the parameter.

Also consider the comparison between two population means. If a confidence interval of the differences may not include zero while the CI of each population mean may overlap. So, overlapping intervals do not imply the samples have the same means.

---

Revision #9

Created 25 August 2022 13:57:14 by Elkip

Updated 25 August 2022 15:44:37 by Elkip