

# Lecture 5: Parametric Hypothesis Testing: Comparing Means

GENOME 560

Doug Fowler, GS (dfowler@uw.edu)

# Review from last week

- What is a confidence interval?

# Review from last week

- What is a confidence interval?
  - An interval estimate of a parameter
  - Repeated sampling would produce a set of CIs that would encompass the true population parameter X% of the time

# Goals

- Introduction to hypothesis testing
- One sample hypothesis tests
  - One-sample  $t$ -test
- Two sample hypothesis tests
  - Two-sample paired and unpaired  $t$ -test
- R session
  - Doing  $t$ -tests

# Introduction to hypothesis testing

# What is hypothesis testing?

- “A statistical test examines a set of sample data and, on the basis of an expected distribution of the data, leads to a decision about whether to accept the hypothesis underlying the expected distribution or reject that hypothesis and accept an alternative one.”
- Over the next several weeks, we will come at this basic problem from many angles, but the general idea will remain the same

# What is hypothesis testing?

- “A statistical test **examines a set of sample data** and, on the basis of an expected distribution of the data, leads to a decision about whether to accept the hypothesis underlying the expected distribution or reject that hypothesis and accept an alternative one.”
- Statistical tests rely on computing a test statistic (i.e. a statistic we will compare to a parametric value or between samples)

# What is hypothesis testing?

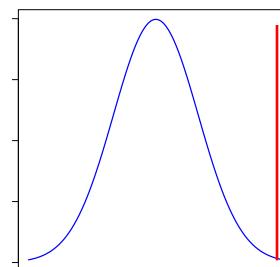
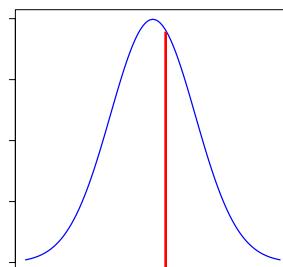
- “A statistical test examines a set of sample data and, on the basis of **an expected distribution of the data**, leads to a decision about whether to accept the hypothesis underlying the expected distribution or reject that hypothesis and accept an alternative one.”
- To generate the **expected distribution of the test statistic** (often called the null distribution) we can use parametric distributions or, through randomization, generate our own based on the data

# What is hypothesis testing?

- “A statistical test examines a set of sample data and, on the basis of an expected distribution of the data, leads to a decision about whether to accept the hypothesis underlying the expected distribution or reject that hypothesis and accept an alternative one.”
- Then, we can see what the likelihood of obtaining the test statistic value we calculated from the sample. If it is highly unlikely, we might reject the null hypothesis.

# What is hypothesis testing?

- “A statistical test examines a set of sample data and, on the basis of an expected distribution of the data, leads to a decision about whether to accept the hypothesis underlying the expected distribution or reject that hypothesis and accept an alternative one.”
- Then, we can see what the likelihood of obtaining the test statistic value we calculated from the sample. If it is highly unlikely, we might reject the null hypothesis.



# Hypothesis Testing

- Formally examine two opposing conjectures (hypotheses),  $H_0$  and  $H_A$
- These two hypotheses are mutually exclusive and exhaustive so that one is true to the exclusion of the other
- We accumulate evidence – collect and analyze sample information – for the purpose of determining which of the two hypotheses is true and which of the two hypotheses is false

# Example

- Consider a genome-wide association study (GWAS) for T2D where you measure the blood glucose level of the case/control groups
- **The null hypothesis,  $H_0$ :**
  - There is no difference between the case/control groups in the mean blood glucose levels
  - $H_0: \mu_1 - \mu_2 = 0$
- **The alternative hypothesis,  $H_A$ :**
  - The mean blood glucose levels in the case/control groups are “different”
  - $H_A: \mu_1 - \mu_2 \neq 0$

# The Null and Alternative Hypothesis

- **The null hypothesis,  $H_0$ :**
  - States the assumption (numerical, to be tested)
  - Begin with the assumption that the null hypothesis is TRUE
  - Always contains the “=” sign
  
- **The alternative hypothesis,  $H_A$ :**
  - Is the opposite of the null hypothesis
  - Challenges the status quo
  - Never contains just the “=” sign
  - Is generally the hypothesis that is believed to be true by the researcher

# One and Two Sided Tests

- Hypothesis tests can be one or two sided (tailed)
- One tailed tests are directional:

$$H_0: \mu_1 - \mu_2 \leq 0$$

$$H_A: \mu_1 - \mu_2 > 0$$

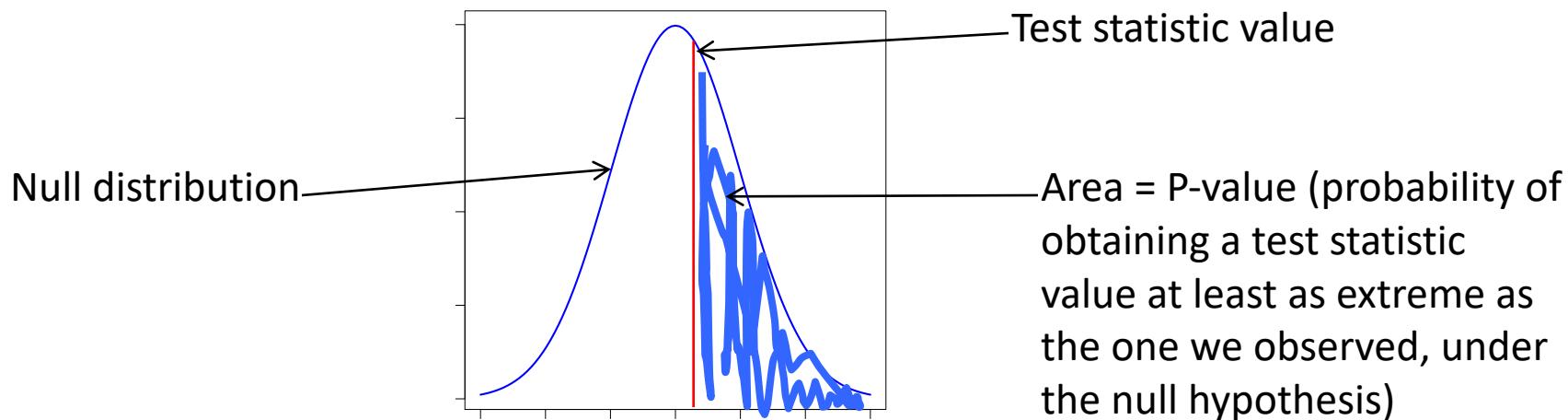
- Two tailed tests are not directional:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_A: \mu_1 - \mu_2 \neq 0$$

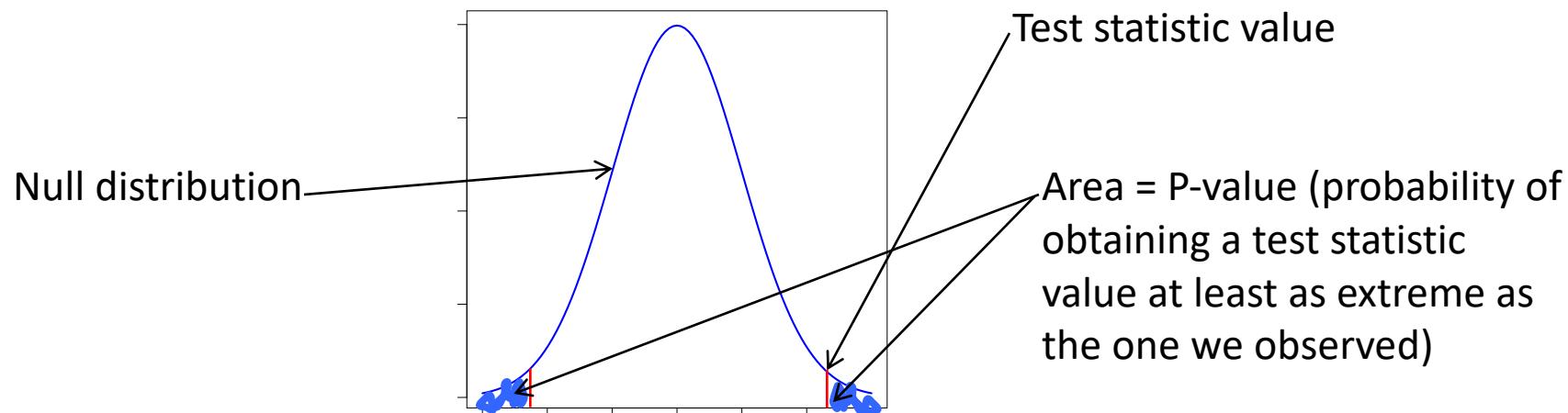
# One-sided Test P-values

- Calculate a **test statistic** from the sample data that is relevant to the hypothesis being tested
  - e.g. In our GWAS example, the test statistic can be determined based on  $\mu_1$ ,  $\mu_2$  and  $\sigma_1$ ,  $\sigma_2$  as estimated from the sample
- After calculating a test statistic we convert this to a **P-value** by comparing its value to distribution of test statistic's under the null hypothesis



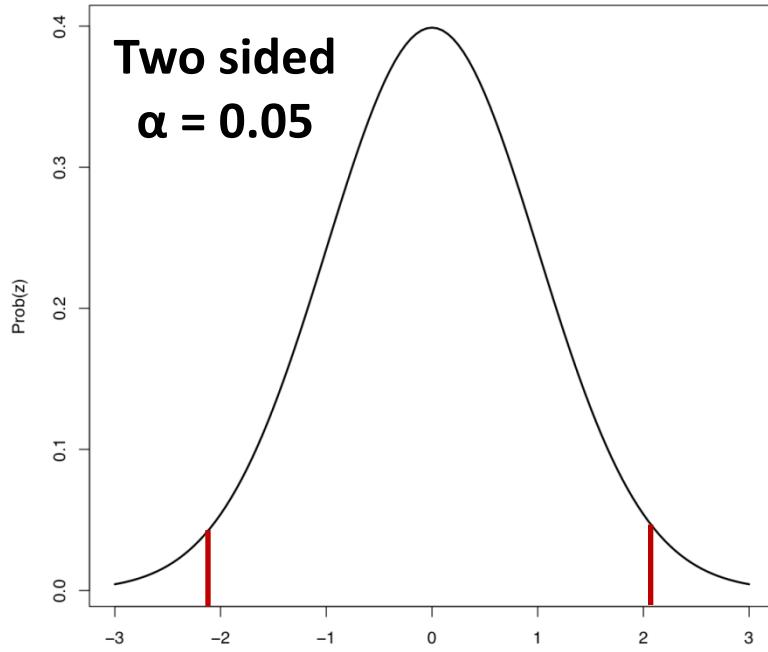
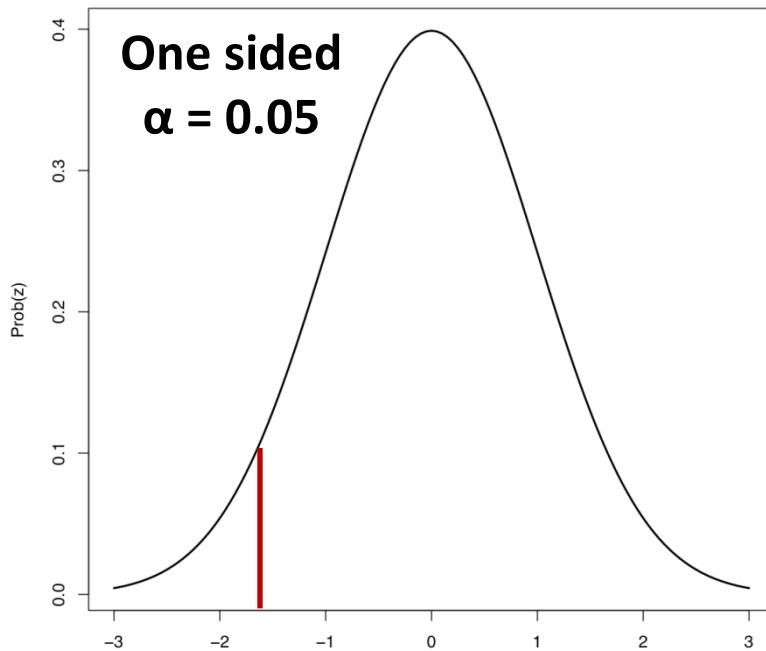
# Two-sided Test P-values

- Calculate a **test statistic** from the sample data that is relevant to the hypothesis being tested
  - e.g. In our GWAS example, the test statistic can be determined based on  $\mu_1$ ,  $\mu_2$  and  $\sigma_1$ ,  $\sigma_2$  as estimated from the sample
- After calculating a test statistic we convert this to a **P-value** by comparing its value to distribution of test statistic's under the null hypothesis



# When To Reject $H_0$

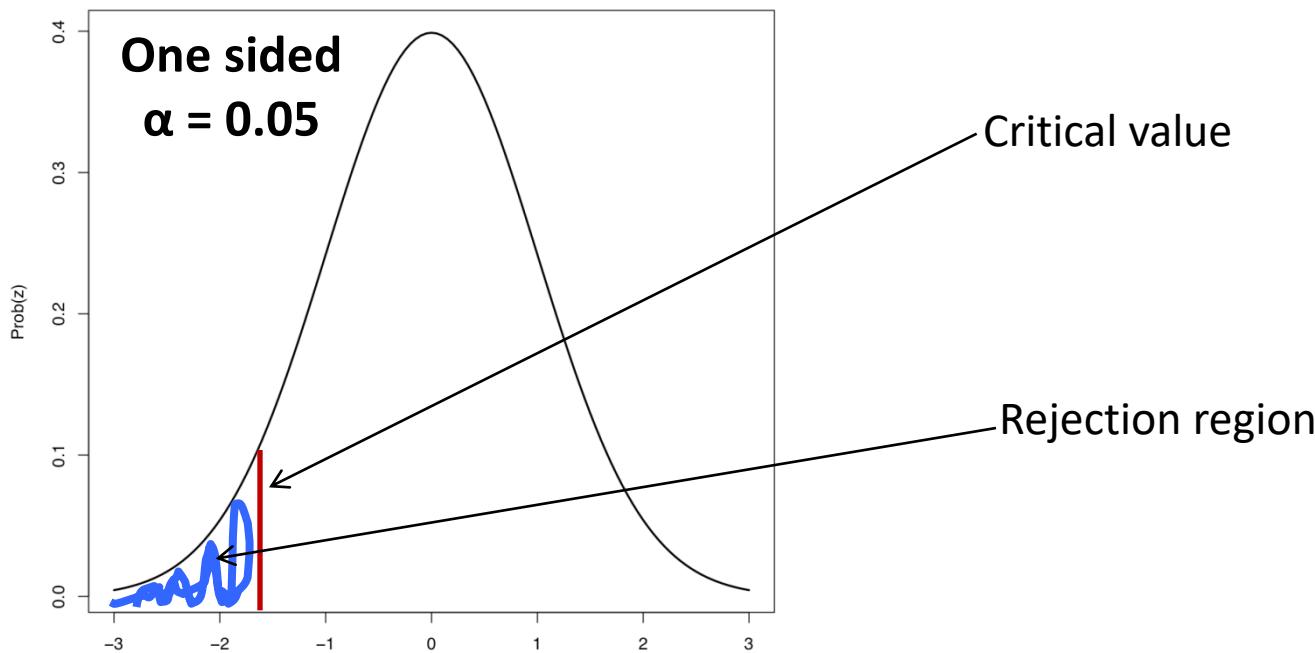
- **Level of significance,  $\alpha$ :** Specified before an experiment to define rejection region
- **Rejection region:** set of all test statistic values for which  $H_0$  will be rejected



The test statistic value required to reject is often called the critical value

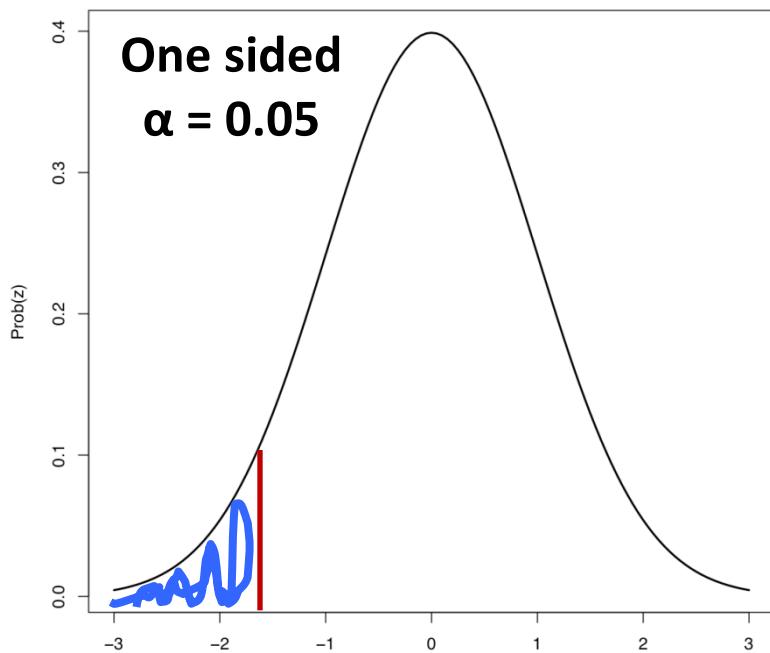
# When To Reject $H_0$

- **Level of significance,  $\alpha$ :** Specified before an experiment to define rejection region
- **Rejection region:** set of all test statistic values for which  $H_0$  will be rejected

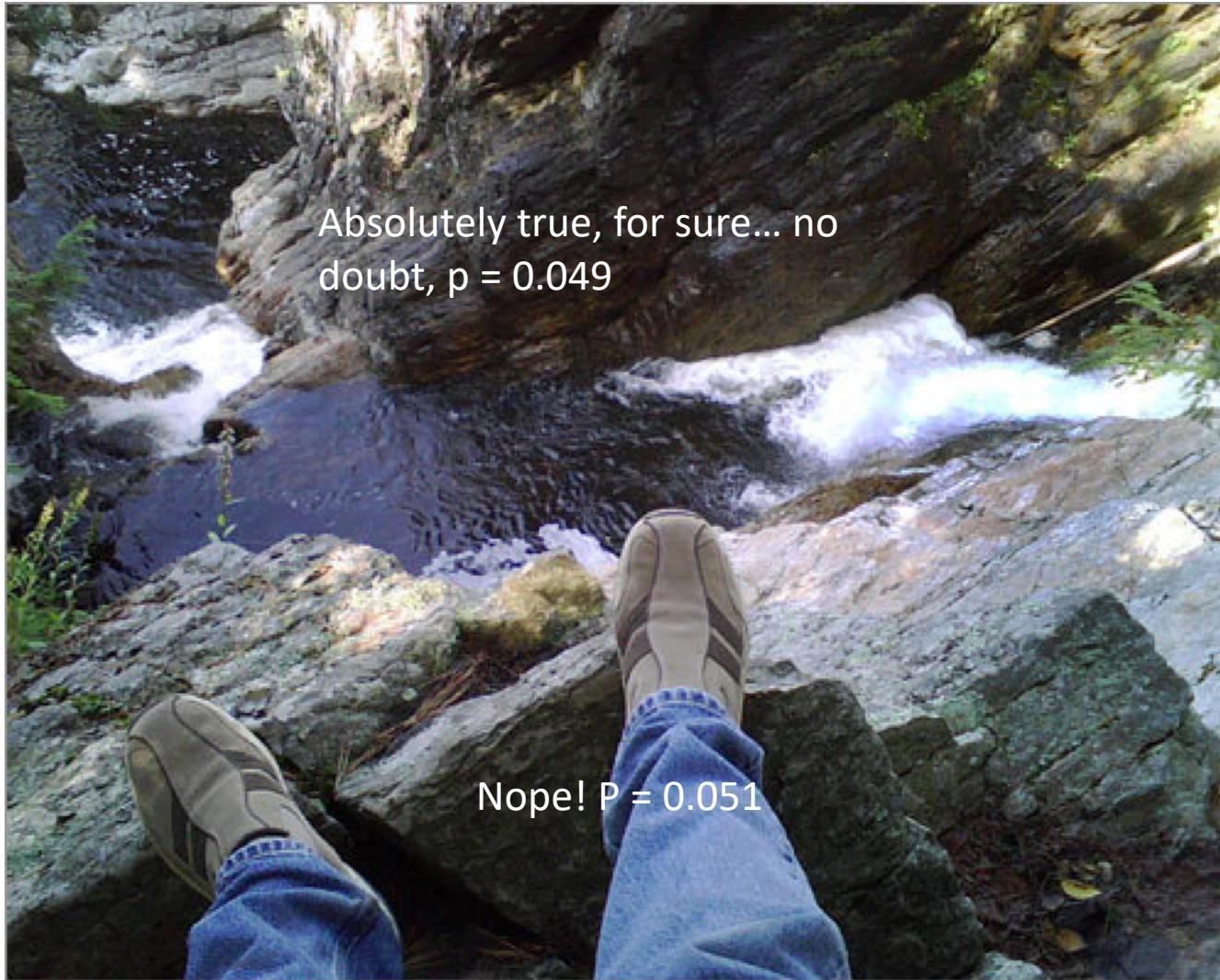


# When To Reject $H_0$

- **Level of significance,  $\alpha$ :** Specified before an experiment to define rejection region
- **Rejection region:** set of all test statistic values for which  $H_0$  will be rejected



When we obtain a test statistic in the rejection region we can conclude that **either** the null is true and a highly improbable event has occurred, **or** the null is false



# Some Notation

- In general, critical values for an  $\alpha$  level test denoted as:

One sided test:  $X_\alpha$

Two sided test:  $X_{\alpha/2}$

where  $X$  indicates the distribution of the test statistic

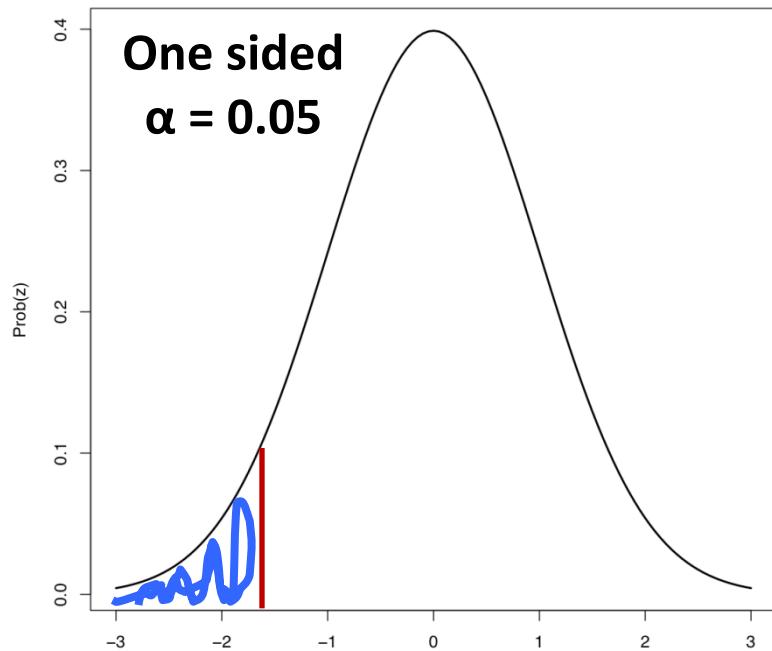
- For example, if  $X \sim N(\mu=0, \sigma=1)$ :

One sided test:  $z_\alpha$  (i.e.,  $z_{0.05} = 1.64$ )

Two sided test:  $z_{\alpha/2}$  (i.e.,  $z_{0.05/2} = z_{0.05/2} = +/-1.96$ )

# Errors in Hypothesis Testing

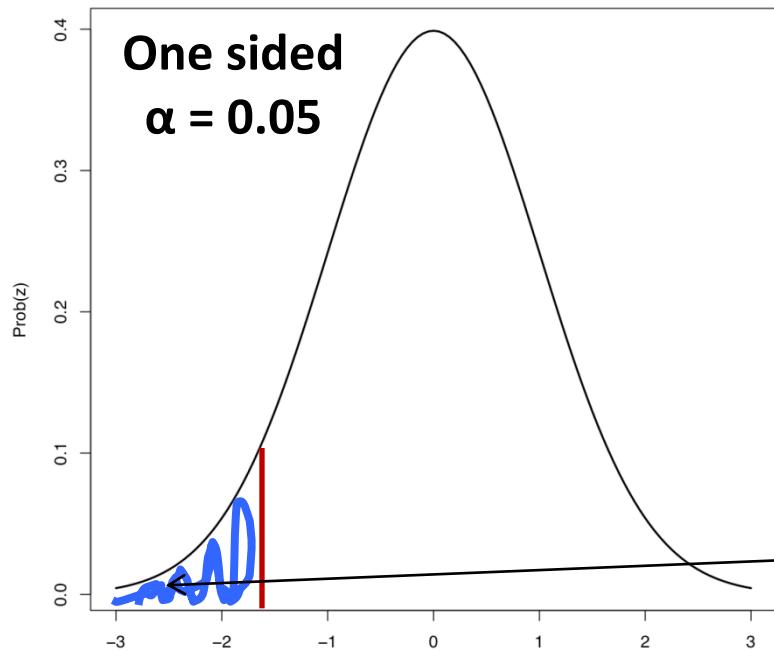
- **Level of significance,  $\alpha$ :** Specified before an experiment to define rejection region
- **Rejection region:** set of all test statistic values for which  $H_0$  will be rejected



Given  $\alpha = 0.05$ , what is the chance we will falsely reject the null hypothesis?

# Errors in Hypothesis Testing

- **Level of significance,  $\alpha$ :** Specified before an experiment to define rejection region
- **Rejection region:** set of all test statistic values for which  $H_0$  will be rejected



Given  $\alpha = 0.05$ , what is the chance we will falsely reject the null hypothesis?

5% of the time\*

# Errors in Hypothesis Testing

		Actual Situation “Truth”	
		$H_0$ True	$H_0$ False
Decision	Do Not Reject $H_0$	Correct Decision $1-\alpha$	Incorrect Decision Type II Error $\beta$
	Reject $H_0$	Incorrect Decision Type I Error $\alpha$	Correct Decision $1-\beta$

# Errors in Hypothesis Testing

		Actual Situation “Truth”	
		$H_0$ True	$H_0$ False
Decision	Do Not Reject $H_0$	Correct Decision $1-\alpha$	Incorrect Decision Type II Error $\beta$
	Reject $H_0$	Incorrect Decision Type I Error $\alpha$	Correct Decision $1-\beta$

False positive                          False negative

# Type I and II Errors

		Actual Situation “Truth”	
		$H_0$ True	$H_0$ False
Decision	Do Not Reject $H_0$	Correct Decision $1-\alpha$	Incorrect Decision Type II Error $\beta$
	Reject $H_0$	Incorrect Decision Type I Error $\alpha$	Correct Decision $1-\beta$

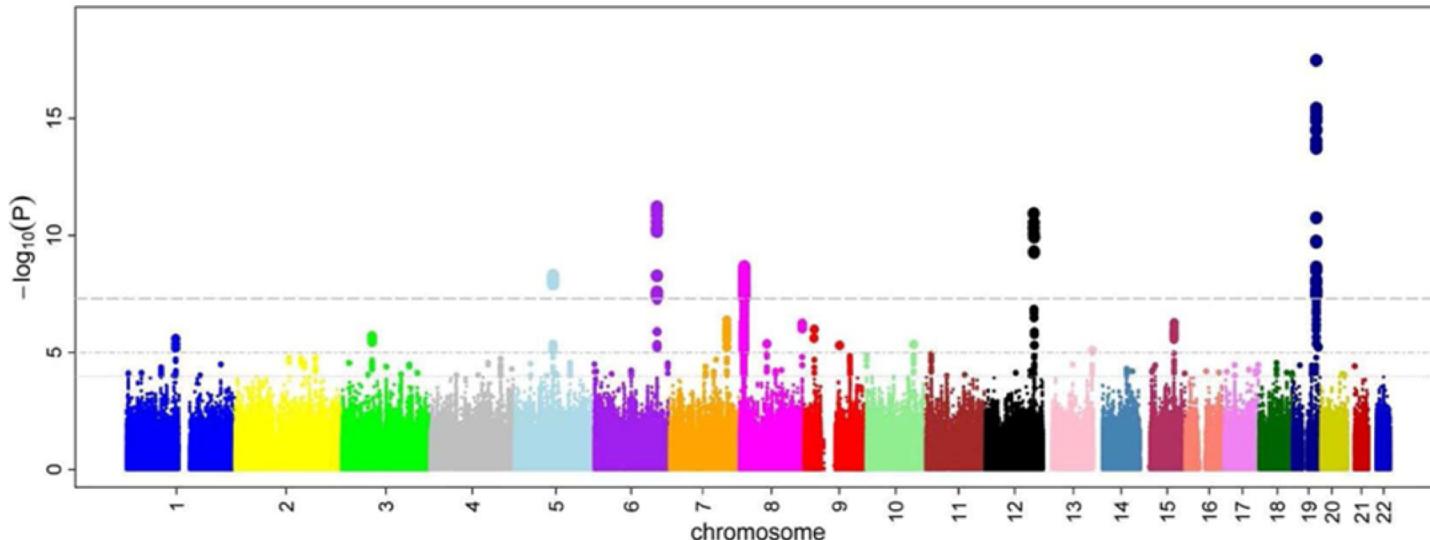
$$\alpha = P(\text{Type I Error}) \quad \beta = P(\text{Type II Error})$$

# What is a p-value

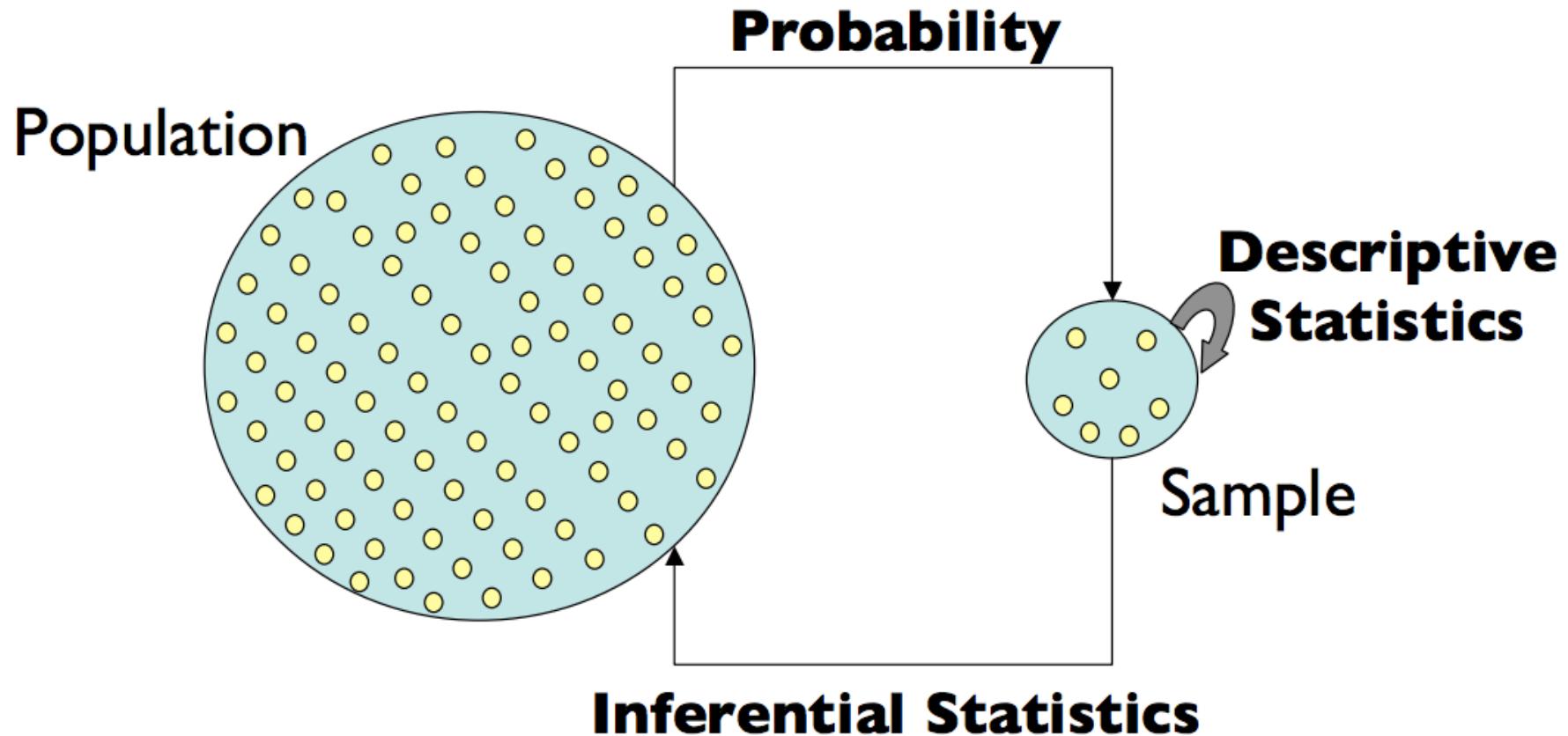
- Likelihood of obtaining a test statistic at least as extreme as the one we observed, given the assumptions we made are satisfied
  
- What is a p-value **NOT**?
  - **Not** the probability that the null hypothesis is true
  - **Not** the probability that the finding is a “fluke”
  - **Not** the probability of falsely rejecting the null
  - **Does not** indicate the size or importance of observed effects (confounded by sample size)

# No!

- What is a p-value **NOT**?
  - **Not** the probability that the null hypothesis is true
  - **Not** the probability that the finding is a “fluke”
  - **Not** the probability of falsely rejecting the null
  - **Does not** indicate the size or importance of observed effects (confounded by sample size)

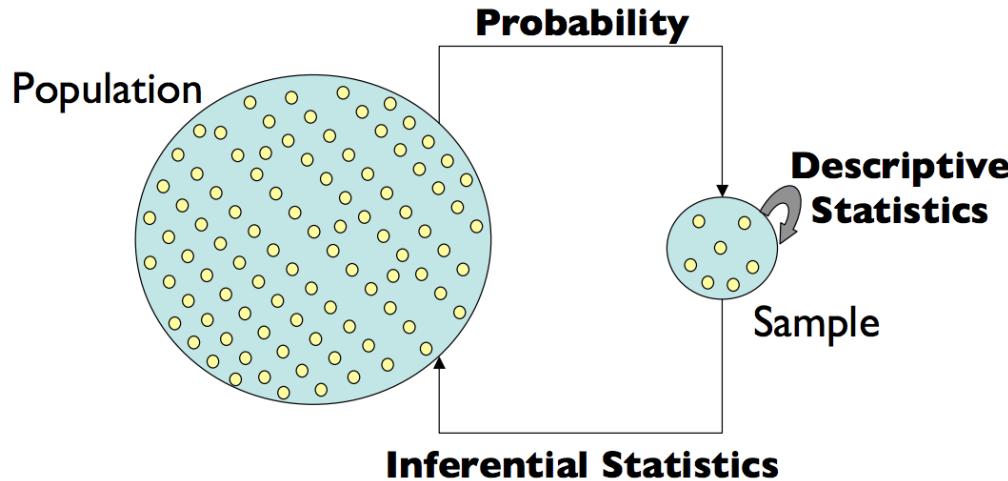


# Central Dogma of Statistics

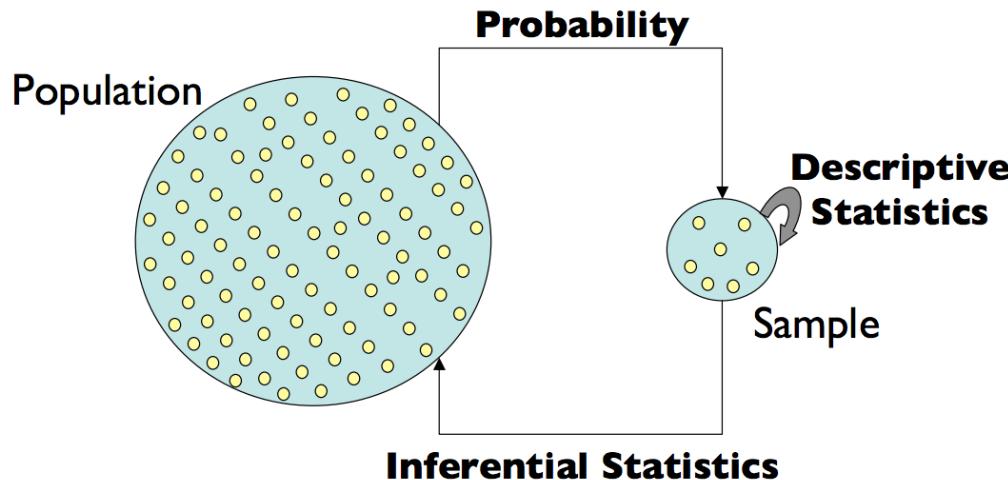


One sample: comparing a sample statistic to a known or assumed population parameter

# Central Dogma of Statistics

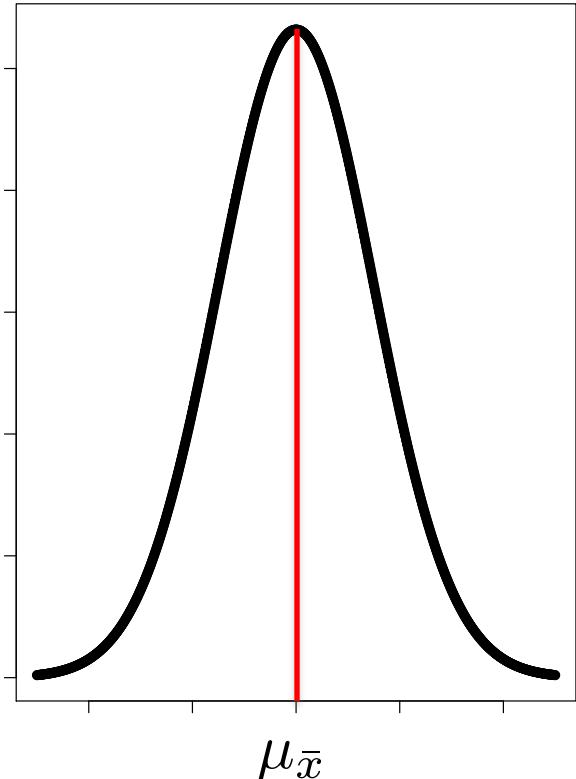


Two samples:  
comparing two  
sample statistics



# One sample hypothesis tests

- Testing the possibility of getting a result at least as extreme as  $\bar{x}$  from a sampling distribution



If  $n > 30$ , CLT says we can assume the sampling distribution is normal, in which case we can use a z-statistic

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

If  $n < 30$ , we cannot assume the sampling distribution is normal, in which case we can use a t-statistic

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

# One sample hypothesis example

- Lets say that we know a certain type of cell should have an average of 10 transcripts per cell. We measure 5 cells and find:

$$\bar{x} = 12.5$$

$$s_x^2 = 1.89$$

# One sample hypothesis example

- Lets say that we know a certain type of cell should have an average of 10 transcripts per cell. We measure 5 cells and find:

$$\bar{x} = 12.5$$

$$H_0 : \bar{x} = \mu_{\bar{x}}$$

$$s_x^2 = 1.89$$

$$H_1 : \bar{x} \neq \mu_{\bar{x}}$$

# One sample hypothesis example

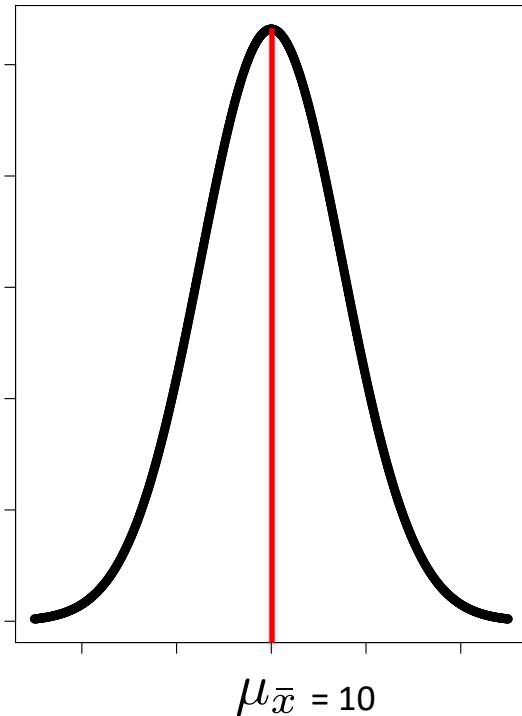
- Lets say that we know a certain type of cell should have an average of 10 transcripts per cell. We measure 5 cells and find:

$$\bar{x} = 12.5$$

$$s_x^2 = 1.89$$

$$H_0 : \bar{x} = \mu_{\bar{x}}$$

$$H_1 : \bar{x} \neq \mu_{\bar{x}}$$



# One sample hypothesis example

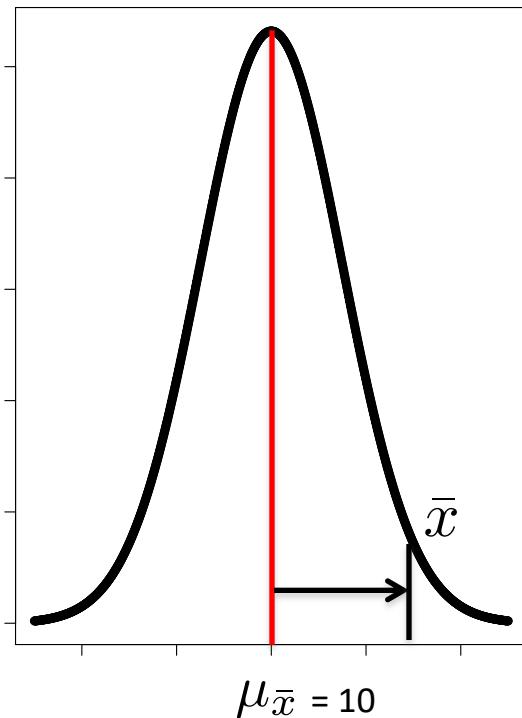
- Lets say that we know a certain type of cell should have an average of 10 transcripts per cell. We measure 5 cells and find:

$$\bar{x} = 12.5$$

$$s_x^2 = 1.89$$

$$H_0 : \bar{x} = \mu_{\bar{x}}$$

$$H_1 : \bar{x} \neq \mu_{\bar{x}}$$



# One sample hypothesis example

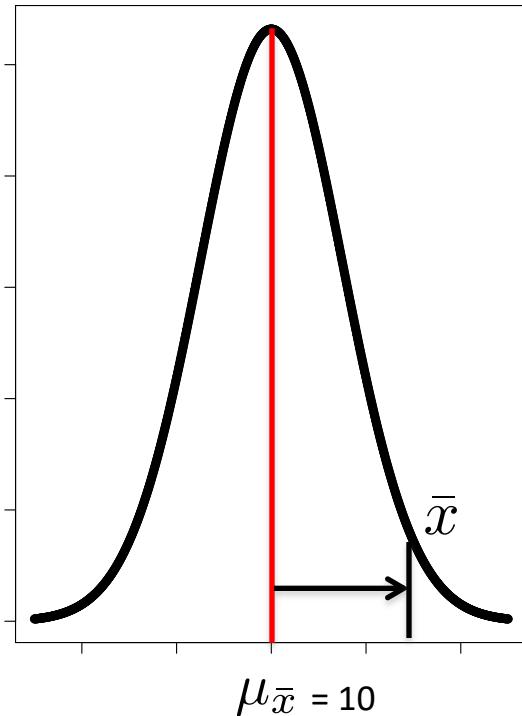
- Lets say that we know a certain type of cell should have an average of 10 transcripts per cell. We measure 5 cells and find:

$$\bar{x} = 12.5$$

$$s_x^2 = 1.89$$

$$H_0 : \bar{x} = \mu_{\bar{x}}$$

$$H_1 : \bar{x} \neq \mu_{\bar{x}}$$



$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

# One sample hypothesis example

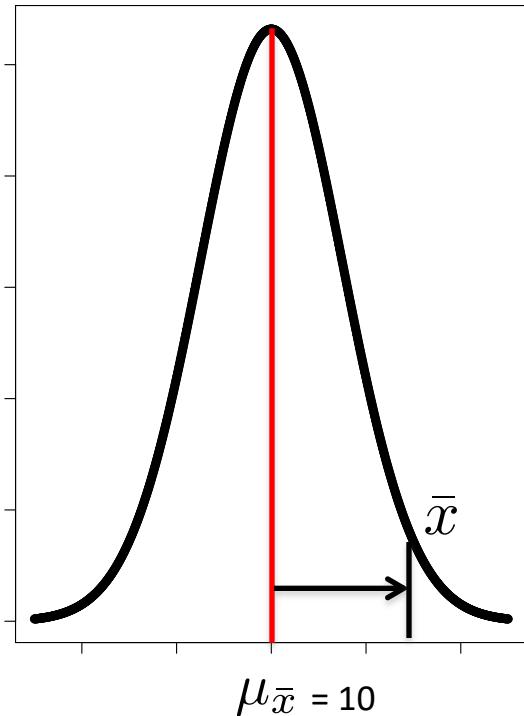
- Lets say that we know a certain type of cell should have an average of 10 transcripts per cell. We measure 5 cells and find:

$$\bar{x} = 12.5$$

$$s_x^2 = 1.89$$

$$H_0 : \bar{x} = \mu_{\bar{x}}$$

$$H_1 : \bar{x} \neq \mu_{\bar{x}}$$



$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

# One sample hypothesis example

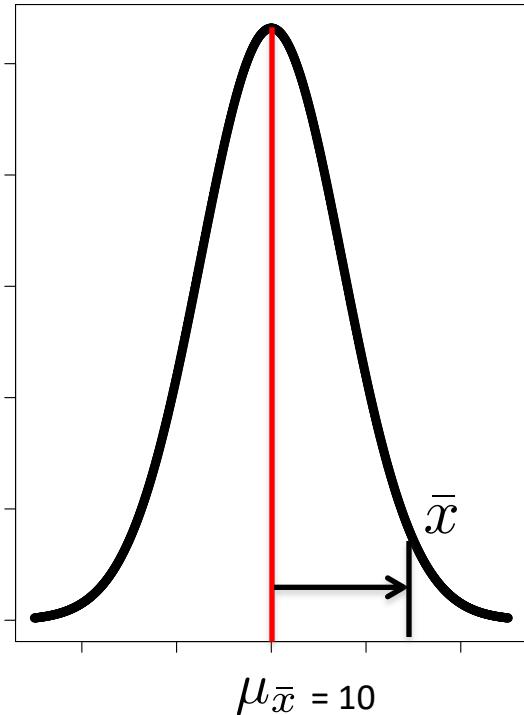
- Lets say that we know a certain type of cell should have an average of 10 transcripts per cell. We measure 5 cells and find:

$$\bar{x} = 12.5$$

$$s_x^2 = 1.89$$

$$H_0 : \bar{x} = \mu_{\bar{x}}$$

$$H_1 : \bar{x} \neq \mu_{\bar{x}}$$



$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

$$t \approx \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$$

We can approximate the standard error of the mean

# One sample hypothesis example

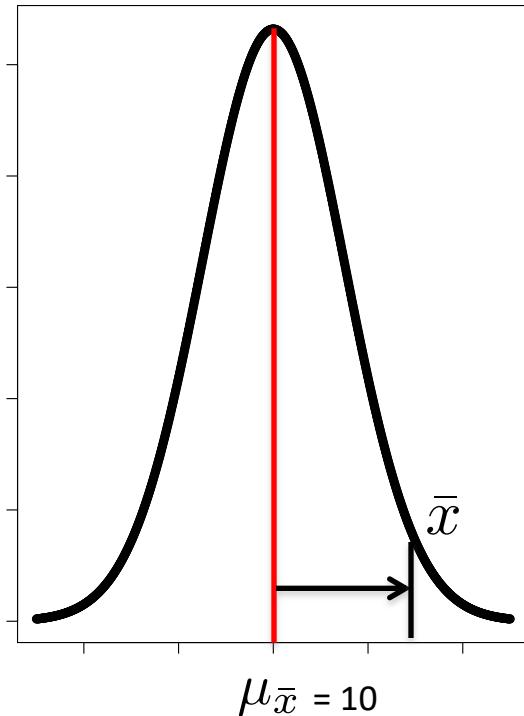
- Lets say that we know a certain type of cell should have an average of 10 transcripts per cell. We measure 5 cells and find:

$$\bar{x} = 12.5$$

$$s_x^2 = 1.89$$

$$H_0 : \bar{x} = \mu_{\bar{x}}$$

$$H_1 : \bar{x} \neq \mu_{\bar{x}}$$



$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

$$t \approx \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{12.5 - 10}{\frac{1.38}{\sqrt{5}}} = 4.03$$

# One sample hypothesis example

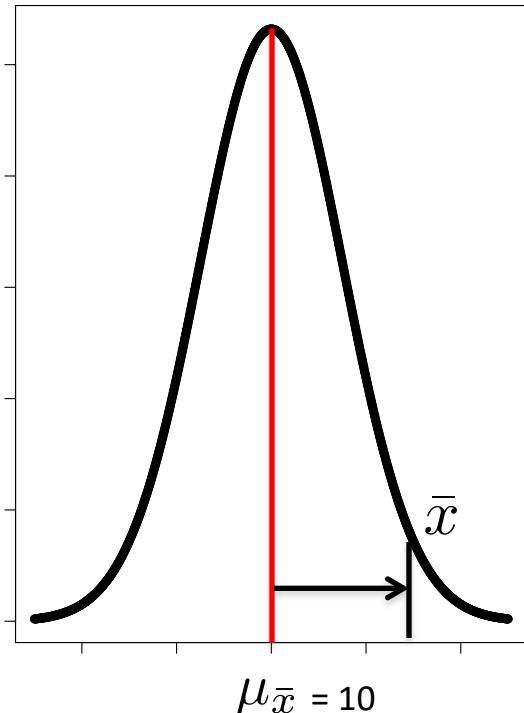
- Lets say that we know a certain type of cell should have an average of 10 transcripts per cell. We measure 5 cells and find:

$$\bar{x} = 12.5$$

$$s_x^2 = 1.89$$

$$H_0 : \bar{x} = \mu_{\bar{x}}$$

$$H_1 : \bar{x} \neq \mu_{\bar{x}}$$



$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

$$t \approx \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{12.5 - 10}{\frac{1.38}{\sqrt{5}}} = 4.03$$

Using  $pt(4.03, df = 4)$  in R, we find that 99.2% of t-values are below 4.03.

# One sample hypothesis example

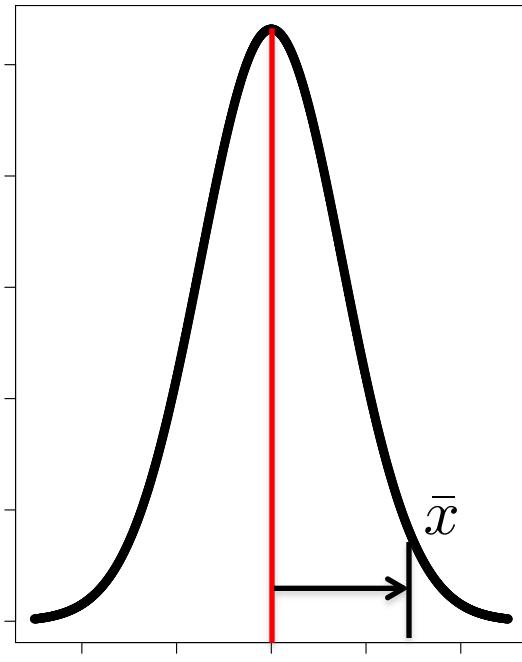
- What is the p-value?

$$\bar{x} = 12.5$$

$$s_x^2 = 1.89$$

$$H_0 : \bar{x} = \mu_{\bar{x}}$$

$$H_1 : \bar{x} \neq \mu_{\bar{x}}$$



$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

$$t \approx \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{12.5 - 10}{\frac{1.38}{\sqrt{5}}} = 4.03$$

Using `pt(4.03, df = 4)` in R, we find that 99.2% of t-values are below 4.03.

# One sample hypothesis example

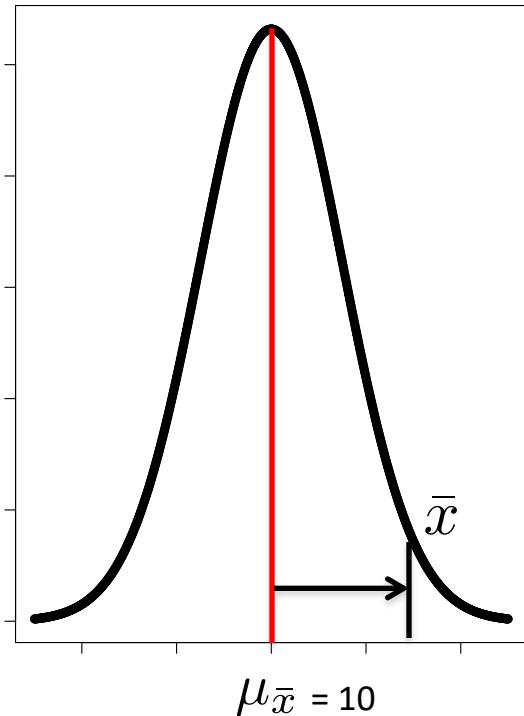
- Interpretation?

$$\bar{x} = 12.5$$

$$s_x^2 = 1.89$$

$$H_o : \bar{x} = \mu_{\bar{x}}$$

$$H_1 : \bar{x} \neq \mu_{\bar{x}}$$



$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

$$t \approx \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{12.5 - 10}{\frac{1.38}{\sqrt{5}}} = 4.03$$

Using  $pt(4.03, df = 4)$  in R, we find that 99.2% of t-values are below 4.03. Hence,  $p = 0.016$  (two-tailed).

# One sample hypothesis example

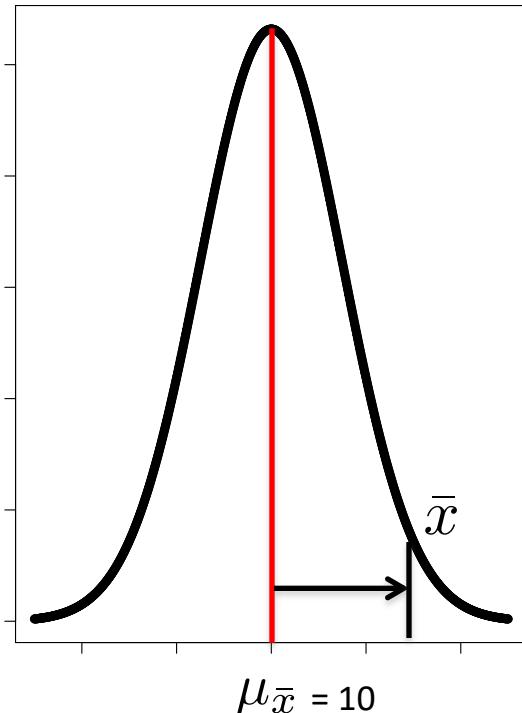
- Interpretation? We expect that 0.16% of the time we would get sample means as extreme or more extreme than 12.5.

$$\bar{x} = 12.5$$

$$s_x^2 = 1.89$$

$$H_0 : \bar{x} = \mu_{\bar{x}}$$

$$H_1 : \bar{x} \neq \mu_{\bar{x}}$$



$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

$$t \approx \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{12.5 - 10}{\frac{1.38}{\sqrt{5}}} = 4.03$$

Using  $pt(4.03, df = 4)$  in R, we find that 99.2% of t-values are below 4.03. Hence,  $p = 0.0016$ .

# One sample hypothesis example

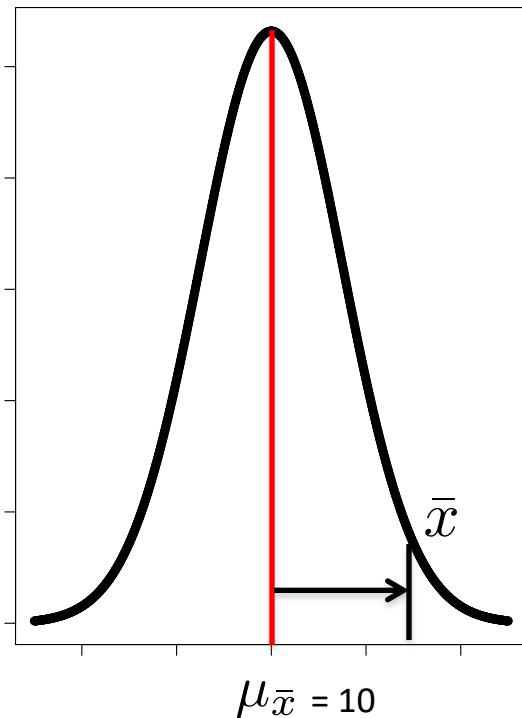
- Interpretation? Reject the null hypothesis

$$\bar{x} = 12.5$$

$$s_x^2 = 1.89$$

$$H_0 : \bar{x} = \mu_{\bar{x}}$$

$$H_1 : \bar{x} \neq \mu_{\bar{x}}$$



$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

$$t \approx \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{12.5 - 10}{\frac{1.38}{\sqrt{5}}} = 4.03$$

Using  $pt(4.03, df = 4)$  in R, we find that 99.2% of t-values are below 4.03. Hence,  $p = 0.016$ .

# T-Statistic

- **Definition:**

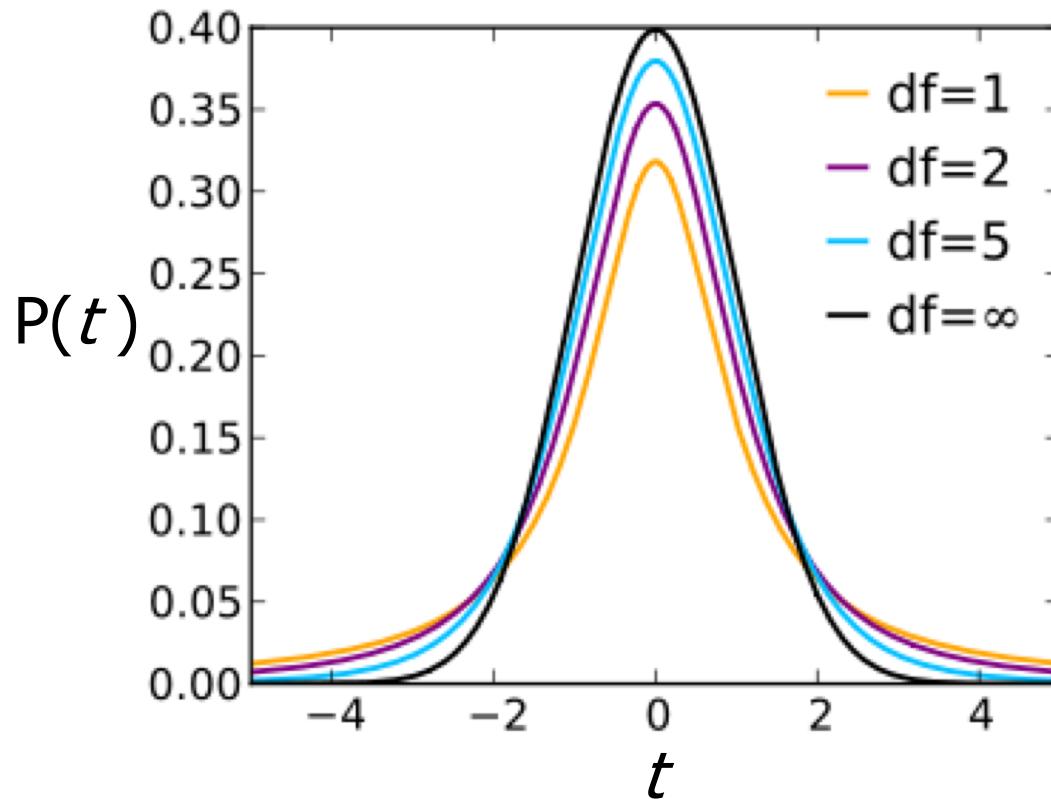
$$t = \frac{\bar{x} - \mu_{\bar{x}}}{s / \sqrt{n}}$$

- where,  $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$        $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
- **Interpretation:** The number of (estimated) standard deviations the *sample mean* is from its expected value  $\mu$
- The quantity  $(n-1)$  is called the *degrees of freedom* of the t value (we lose a degree of freedom when we use the sample mean to calculate the variance)

# Student's t-Distribution

- The t-values follow the t-Distribution

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$



- df = degrees of freedom

# Student's t-Distribution



W.S. Gosset (1876-1937) was a modest, well-liked Englishman who was a brewer and agricultural statistician for the famous Guinness brewing company in Dublin. It insisted that its employees keep their work secret, so he published under the pseudonym ‘Student’ the distribution in 1908. This was one of the first results in modern small-sample statistics.

# Two Sample t-Test

- **Paired two-sample *t*-test:**
  - There are two samples of the same size (say  $n$  numbers)
  - The corresponding numbers pair naturally
  - Examples
    - Before-and-after pairs of measurements after giving a drug
    - Expression levels of  $n$  genes in paired samples (siblings, etc)
- **Unpaired two-sample *t*-test:**
  - Two samples might even have different numbers of points (say  $n_1$  and  $n_2$ , respectively)
  - There is no natural pair

# Paired Two-Sample $t$ -Test

- Given the following data (expression levels of  $n$  genes):

$x_1, x_2, \dots, x_n$  Before drug treatment

$y_1, y_2, \dots, y_n$  After drug treatment

- Measure whether the “after” member of the pair is different from the “before” member

$d_1, d_2, \dots, d_n$  After drug treatment  
Difference  $d_i = x_i - y_i$

- Frame hypotheses
  - Null hypothesis  $H_0$ :** The mean of this sample of differences is 0
  - Alternative hypothesis  $H_A$ :** The mean is not 0
- Look familiar? Just a one-sample  $t$ -test!

# Un-Paired Two-Sample $t$ -Test

- Suppose that two samples are drawn independently

$$x_1, x_2, \dots, x_n$$

$$y_1, y_2, y_3, \dots, y_m$$

- There is no connection between point 18 from one sample, and point 18 from another

# Un-Paired Two-Sample $t$ -Test

- Suppose that two samples are drawn independently

$$x_1, x_2, \dots, x_n \quad \leftarrow \quad \text{Assume a normal sampling distribution with mean } \mu_x$$
$$y_1, y_2, y_3, \dots, y_m \quad \leftarrow \quad \text{Assume a normal sampling distribution with mean } \mu_y$$

- Are the samples drawn from distributions with the same mean or different means?

# Un-Paired Two-Sample $t$ -Test

- Supposed that two samples are drawn independently

$x_1, x_2, \dots, x_n$  ← Assume a normal sampling distribution with mean  $\mu_x$

$y_1, y_2, y_3, \dots, y_m$  ← Assume a normal sampling distribution with mean  $\mu_y$

- Are the samples drawn from distributions with the same mean or different means?
- Frame hypotheses:
  - Null hypothesis  $H_0$ :** The means of the two samples are equal  $\mu_x = \mu_y$
  - Alternative hypothesis  $H_A$ :** Not equal  $\mu_x \neq \mu_y$

# Theoretically...

- The distribution of the sample mean difference,  $\bar{x} - \bar{y}$

$$\bar{x} - \bar{y} \sim N(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}})$$

# Theoretically...

- The distribution of the sample mean difference,  $\bar{x} - \bar{y}$

$$\bar{x} - \bar{y} \sim N(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}})$$

- Let's think about how the t-value should be defined here

$$t = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

Under the null hypothesis,  $\mu_x = \mu_y$

# Theoretically...

- The distribution of the sample mean difference,  $\bar{x} - \bar{y}$

$$\bar{x} - \bar{y} \sim N(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}})$$

- Let's think about how the t-value should be defined here

$$t = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

Under the null hypothesis,  $\mu_x = \mu_y$

- As before, we usually have to use the *sample variance* because we don't know the true variance

# Pooled Variances

- If you assume that the variance is the same in both groups, you can pool all the data to estimate a common variance.
- This maximizes your degrees of freedom (and thus your power to detect a true effect)
- The  $t$ -statistic is then defined as:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}} = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

# Two-Sample $t$ -Test

- Hypothesis testing
  - Null hypothesis  $H_0$ : No difference  $\mu_x = \mu_y$
  - Alternative hypothesis  $H_A$ : Different!  $\mu_x \neq \mu_y$

- **$t$ -statistic:** 
$$t = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

- where 
$$s_p^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}{(n-1) + (m-1)}$$

- The *degrees of freedom* is  $(n+m-2)$

# T-tests gone wrong...

## **Incorrect Use of the Student *t* Test in Randomized Trials of Bilateral Hip and Knee Arthroplasty Patients**

Rajiv Gandhi, MD, MS, FRCSC,\* Holly N. Smith,\*  
Nizar N. Mahomed, MD, ScD, FRCSC,†‡§ Randy Rizek, MD,\* and  
Mohit Bhandari, MD, MSc, FRCSC||

- Review of 30 studies that compared knee/hip replacements

# T-tests gone wrong...

## **Incorrect Use of the Student *t* Test in Randomized Trials of Bilateral Hip and Knee Arthroplasty Patients**

Rajiv Gandhi, MD, MS, FRCSC,\* Holly N. Smith,\*  
Nizar N. Mahomed, MD, ScD, FRCSC, †‡§ Randy Rizek, MD,\* and  
Mohit Bhandari, MD, MSc, FRCSC||

- Review of 30 studies that compared knee/hip replacements
- Some were bilateral, treating one leg and using the other as a control. Others compared treated and untreated individuals.

# T-tests gone wrong...

## **Incorrect Use of the Student *t* Test in Randomized Trials of Bilateral Hip and Knee Arthroplasty Patients**

Rajiv Gandhi, MD, MS, FRCSC,\* Holly N. Smith,\*  
Nizar N. Mahomed, MD, ScD, FRCSC,†‡§ Randy Rizek, MD,\* and  
Mohit Bhandari, MD, MSc, FRCSC||

- Review of 30 studies that compared knee/hip replacements
- Some were bilateral, treating one leg and using the other as a control. Others compared treated and untreated individuals.
- What type of t-test is appropriate in each case?

# T-tests gone wrong...

## **Incorrect Use of the Student *t* Test in Randomized Trials of Bilateral Hip and Knee Arthroplasty Patients**

Rajiv Gandhi, MD, MS, FRCSC,\* Holly N. Smith,\*  
Nizar N. Mahomed, MD, ScD, FRCSC, †‡§ Randy Rizek, MD,\* and  
Mohit Bhandari, MD, MSc, FRCSC||

- Review of 30 studies that compared knee/hip replacements
- Some were bilateral, treating one leg and using the other as a control. Others compared treated and untreated individuals.
- What type of t-test is appropriate in each case?
  - Paired test for bilateral, unpaired for different individuals

# T-tests gone wrong...

## **Incorrect Use of the Student *t* Test in Randomized Trials of Bilateral Hip and Knee Arthroplasty Patients**

Rajiv Gandhi, MD, MS, FRCSC,\* Holly N. Smith,\*  
Nizar N. Mahomed, MD, ScD, FRCSC, †‡§ Randy Rizek, MD,\* and  
Mohit Bhandari, MD, MSc, FRCSC||

- Review of 30 studies that compared knee/hip replacements
- Some were bilateral, treating one leg and using the other as a control. Others compared treated and untreated individuals.
- What type of t-test is appropriate in each case?
- Any guesses as to how many studies got it wrong?

# T-tests gone wrong...

## **Incorrect Use of the Student *t* Test in Randomized Trials of Bilateral Hip and Knee Arthroplasty Patients**

Rajiv Gandhi, MD, MS, FRCSC,\* Holly N. Smith,\*  
Nizar N. Mahomed, MD, ScD, FRCSC, †‡§ Randy Rizek, MD,\* and  
Mohit Bhandari, MD, MSc, FRCSC||

- Review of 30 studies that compared knee/hip replacements
- Some were bilateral, treating one leg and using the other as a control. Others compared treated and untreated individuals.
- What type of t-test is appropriate in each case?
- Any guesses as to how many studies got it wrong? 60%

# A Practical Summary

- T-tests are for comparing one sample mean to a parametric mean or comparing two sample means.

# A Practical Summary

- T-tests are for comparing one sample mean to a parametric mean or comparing two sample means.
- If you have multiple samples, stay tuned for ANOVA. Don't do a bunch of pairwise t-tests!

# A Practical Summary

- T-tests are for comparing one sample mean to a parametric mean or comparing two sample means.
- If you have multiple samples, stay tuned for ANOVA. Don't do a bunch of pairwise t-tests!
- T-tests have two assumptions:

# A Practical Summary

- T-tests are for comparing one sample mean to a parametric mean or comparing two sample means.
- If you have multiple samples, stay tuned for ANOVA. Don't do a bunch of pairwise t-tests!
- T-tests have two assumptions:
  - The sampling distributions are assumed to be normal. Because of CLT this is generally true for  $n > 30$ . For  $n < 30$  CLT doesn't apply so the population itself should be normally distributed. You can check this assumption using the Kolmogorov-Smirnov test.
  - The two samples are assumed to have the same variance. While important, the t-test is more robust to unequal variances than weirdly distributed data. Also, a fancier version (Welch's t-test) works with unequal variances.

# ASA Statement Discussion

- Find on the course website and read the ASA statement on p-values
- We'll discuss in the next lecture

# R Session

- Try out the t-test
- Learn about Q-Q plots
- Try out the Kolmogorov-Smirnov and Bartlett tests

# Un-Pooled Variances

- Just replace the true variances with sample variances

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

# Un-Pooled Variances

- Just replace the true variances with sample variances

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

- The t-statistic has the Student's t-distribution with degrees of freedom  $v$

# Un-Pooled Variances

- Just replace the true variances with sample variances

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

- The t-statistic has the Student's t-distribution with degrees of freedom  $v$
- It is complicated to figure out  $v$  here!
- A good approximation is given as  $\approx$  harmonic mean 
$$\frac{2}{\frac{1}{n} + \frac{1}{m}}$$

# Pooled Variance

- Pooling variances:

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \Leftrightarrow (n-1)s_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$s_y^2 = \frac{\sum_{i=1}^m (y_i - \bar{y})^2}{m-1} \Leftrightarrow (m-1)s_y^2 = \sum_{i=1}^m (y_i - \bar{y})^2$$

$$\therefore s_p^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}{n+m-2}$$

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$$



Degrees of freedom