Hypothesis tests for two or more means

Outline for today

Review of permutations tests for two means

Parametric tests for two means (t-test)

Tests for more than two means

Announcement: Final projects

All final project presentations are due at 11:59pm on Sunday April 30th

- A week from this Sunday

5 minute presentations: 3 slides

Final project written reports are due the Sunday after the last day of class (May 7th)

Big Data Baseball – Q&R

"Tampa Bay is believed to be one of the first teams to employ PITCHf/x data to monitor its pitchers' health and predict and preempt injury… Tampa Bay had only one pitcher in the entire organization undergo Tommy John surgery from late 2005 to mid-2009" (pg 189).

Question: Do we believe that Tampa's program is contributing to fewer pitchers having Tommy John surgery?

From the MLB website, on average, 30 pitchers have Tommy John surgery per year.

Considering that there are around 360 pitchers in the MLB (assuming that there are 12 pitchers per team), this means that about 8% of pitchers get Tommy John surgery per year.

To account for the 4 years, we can assume that there were 48 pitchers on Tampa Bay who could have had to undergo Tommy John surgery

Big Data Baseball – Q&R

Step 1:

- H_0: $\pi = .08$ H_A: $\pi = .08$

Step 2:

- 1/48 = .021 (or k = 1)

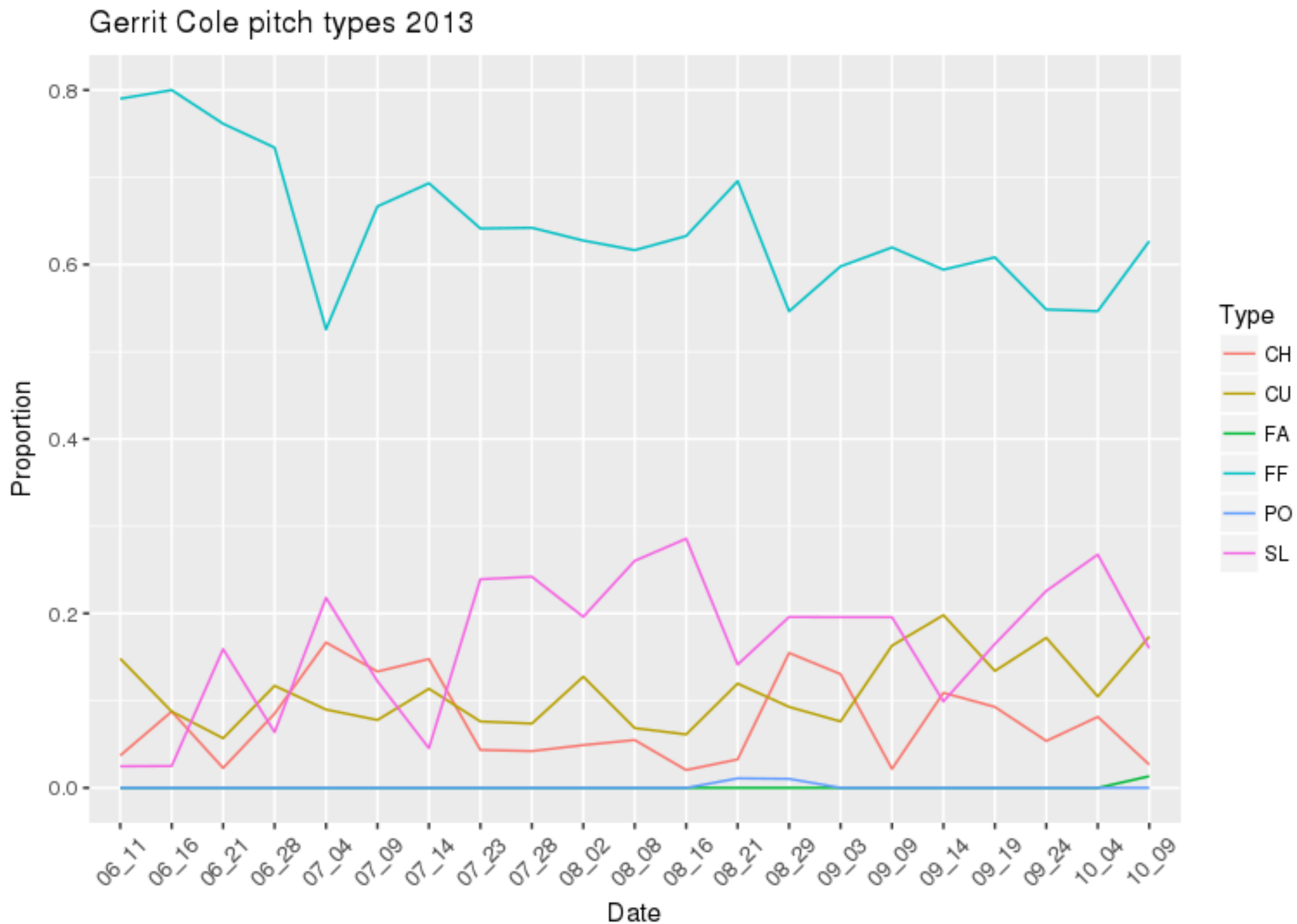
Step 3:

- Binomial distribution n = 48, $\pi = .08$

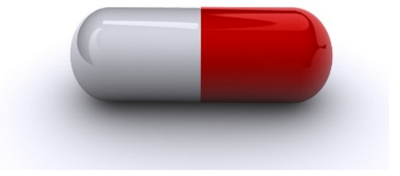
Step 4:

- sum(dbinom (0:1, 48, .08)) = .0945

What can we conclude?


Gerrit Cole pitch types 2013
Proportion
0.8
0.6
0.4
0.2
0.0
Date
06_11
06_16
06_21
06_28
07_04
07_09
07_14
07_23
07_28
08_02
08_08
08_16
08_21
08_29
09_03
09_09
09_14
09_19
09_24
10_04
10_09
Type
CH
CU
FA
FF
PO
SL

Hypothesis tests for two means



Question: How can design an experiment to test if this pill is effective?

Participant pool

Random Assignment

Treatment group

Control group

Question: What hypotheses are we testing?

- Write down null and alternative hypotheses using symbols
- H_0: $\mu_T = \mu_C$ or $\mu_T - \mu_C = 0$
- H_A: $\mu_T > \mu_C$ or $\mu_T - \mu_C > 0$

2. Compute the statistic of interest

Kamath, et al, 2003, suspected that drinking tea might help boost one's immune system

To test this hypothesis recruited 21 healthy volunteers:

- 11 volunteers were assigned to drink 5 to 6 cups of tea a day
- The remaining 10 were assigned to drink that much coffee

After two weeks they measured immune system response (interferon gamma levels) which showed:

Tea	5	11	13	18	20	47	48	52	55	56	58
Coffee	0	0	3	11	15	16	21	21	38	52	

State the null and alternative hypotheses for this experiment and compute the statistic of interest

3. Creating a null distribution

Question: how can we create a null distribution here?

Under that null hypothesis there is no difference between the tea and coffee drinks, so we can generate a null distribution by:

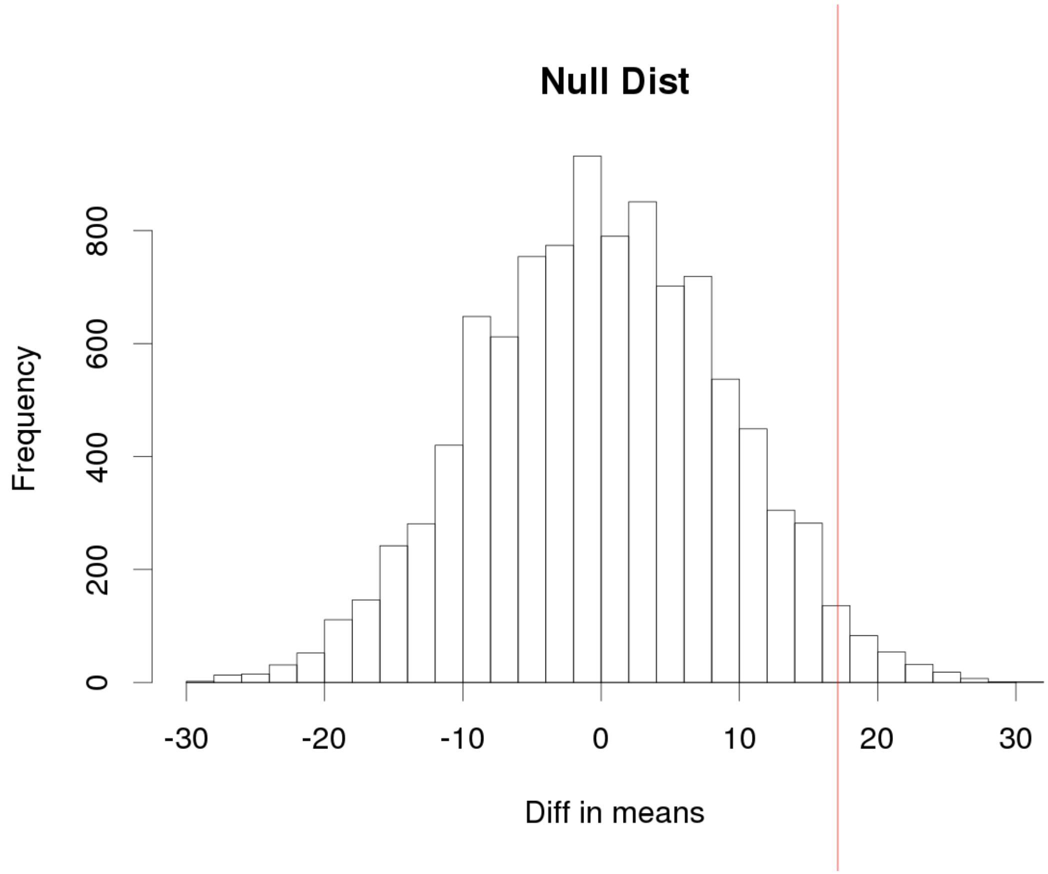
1. Combining the data from both groups together

2. Randomly selecting **11** subjects to simulate the tea drinkers and the remaining **10** subjects to simulate the coffee drinkers

3. Calculating the difference between the means of these two shuffled groups

4. repeating this process 10,000 times to get a full null distribution

3. Creating a null distribution



p-value = .0255

Conclusions?

Review: Comparing two means

Have baseball games gotten longer in the past 50 years?

We can use a hypothesis tests to assess whether the mean length of games in 2014 is longer than those in 2014

1. The null and alternative hypotheses using symbols
   - H_0: $\mu_{2014} = \mu_{1964}$ or $\mu_{2014} - \mu_{1964} = 0$
   - H_A: $\mu_{2014} > \mu_{1964}$ or $\mu_{2014} - \mu_{1964} > 0$

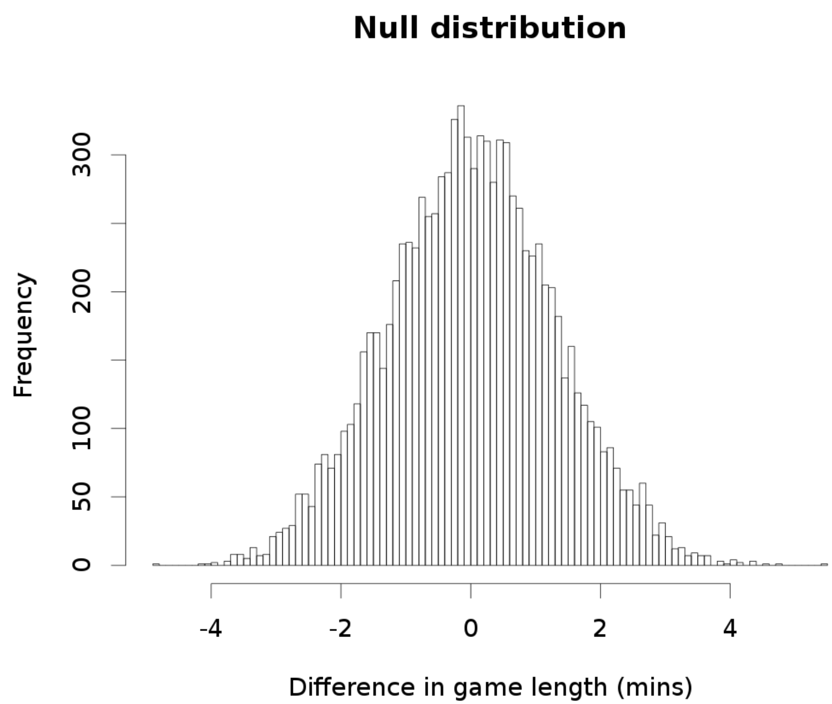
2. The observed statistic:
   - $\bar{x}_{2014} = 187.64$ $\bar{x}_{1964} = 154.21$ $\bar{x}_{2014} - \bar{x}_{1964} =$ 33.42 minutes

Hypothesis tests for two means

How can we create a null distribution?

Hypothesis tests for two means

How can we create a null distribution?

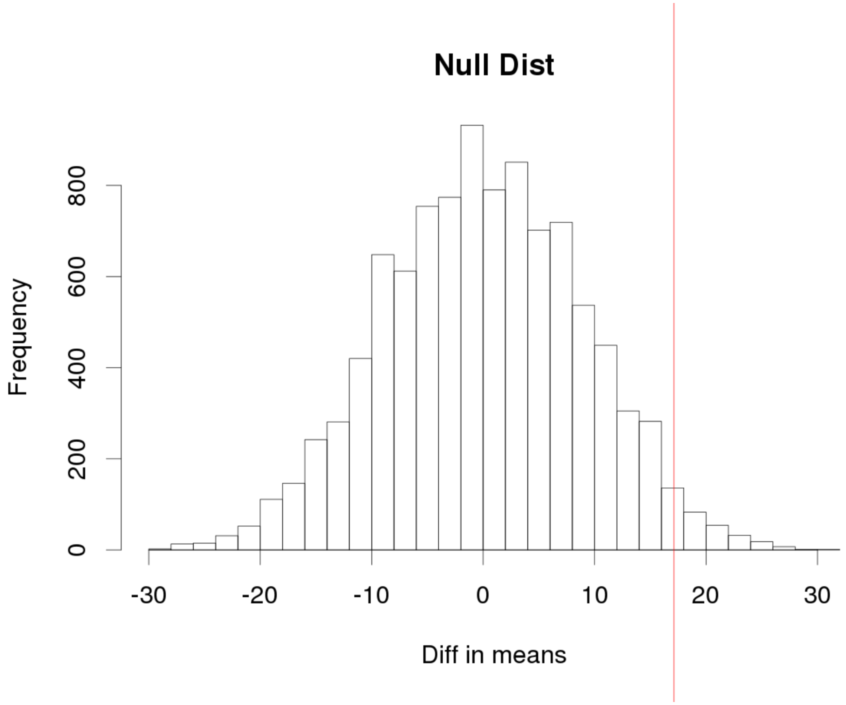


Do the results seem statistically significant?

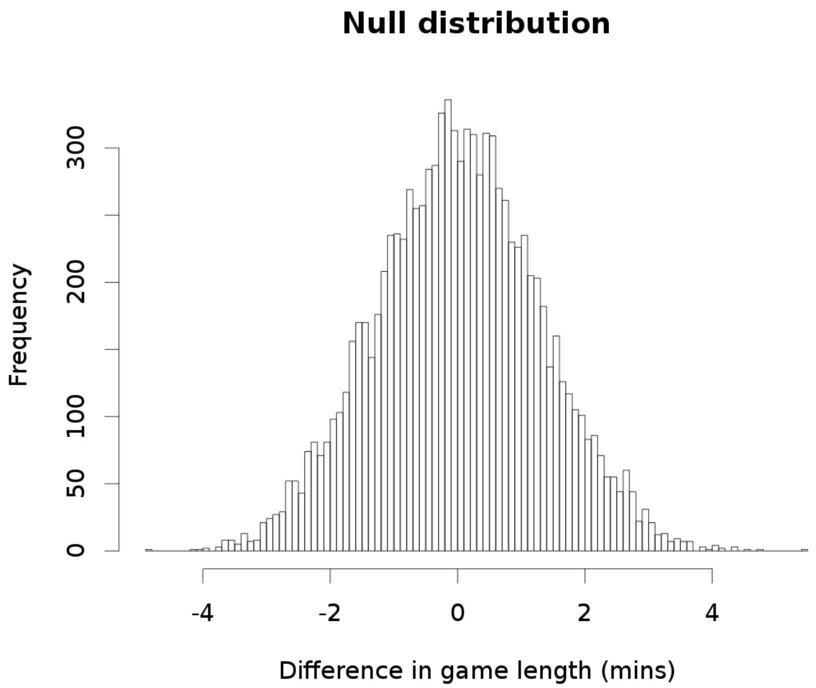
- Observed difference of 33 minutes is not even close to being on this figure
- Conclusions?

Parametric tests for differences in means

Tea and coffee example



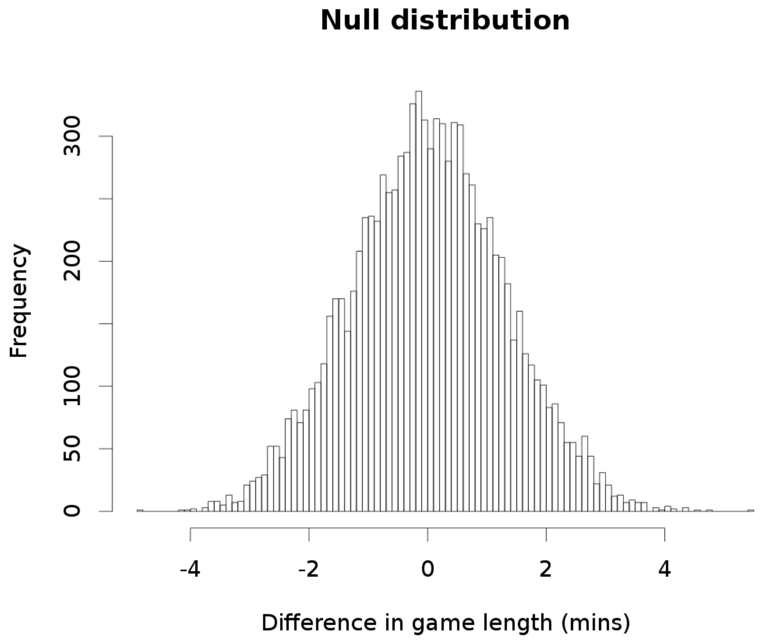
Game duration example



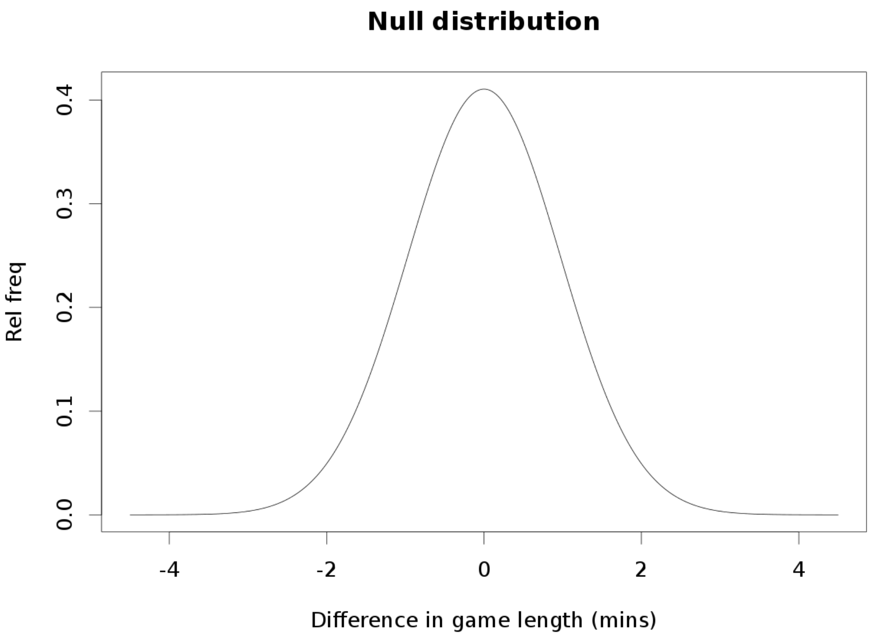
Notice both these distributions appear close to Normal Distributions

Game Duration example

<u>Randomization distribution</u>



<u>Normal distribution: N(0, SE)</u>



Where is 33 on this plot?

Parametric tests for differences in means

Let's look at the tea and coffee example again

- tea <- c(5, 11, 13, 18, 20, 47, 48, 52, 55, 56, 58)
- coffee <- c(0, 0, 3, 11, 15, 16, 21, 21, 38, 52)

Let's use the following notation:

- $\bar{x}_T$ is the mean of the tea data i.e., mean(tea)
- $\bar{x}_C$ is the mean of the coffee data i.e., mean(coffee)
- s_T is the standard deviation of the tea data i.e., sd(tea)
- s_C is the standard deviation of the coffee data i.e., sd(coffee)
- n_T is the number of points in the tea data i.e., length(tea)
- n_C is the number of points in the coffee data i.e., length(coffee)

We can then define a 't-statistic' as:

$$t = \frac{\bar{x}_T - \bar{x}_C}{\sqrt{\frac{s_T^2}{n_T} + \frac{s_C^2}{n_C}}}$$

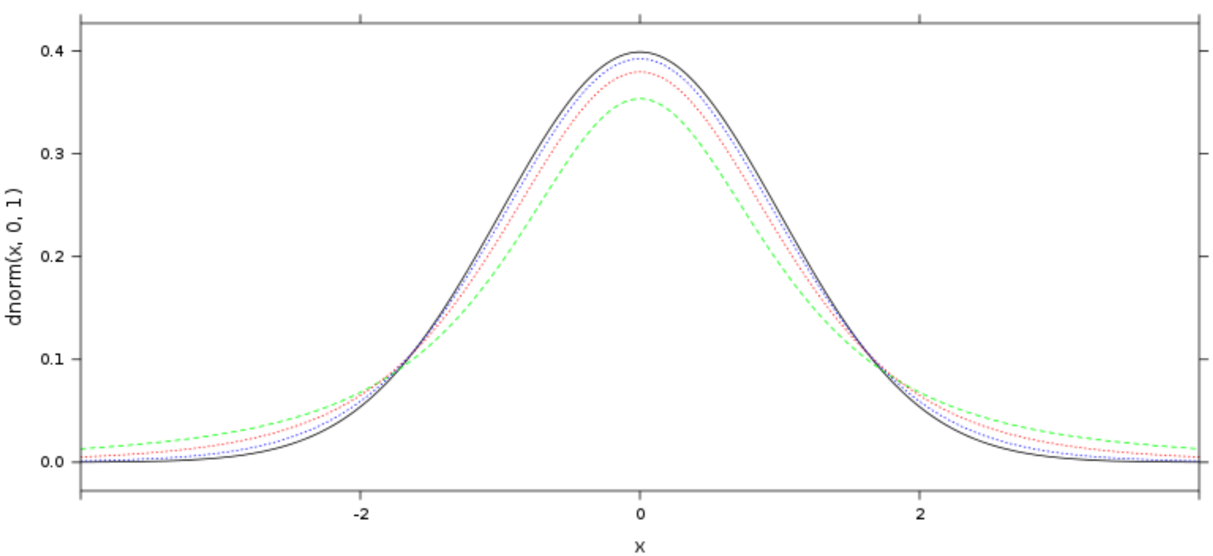
Compute the t-statistic for this data!

Parametric tests for differences in means

Suppose the null hypothesis H_0: $\mu_T = \mu_C = 0$ is true

Then our t-statistic will come from a t-distribution

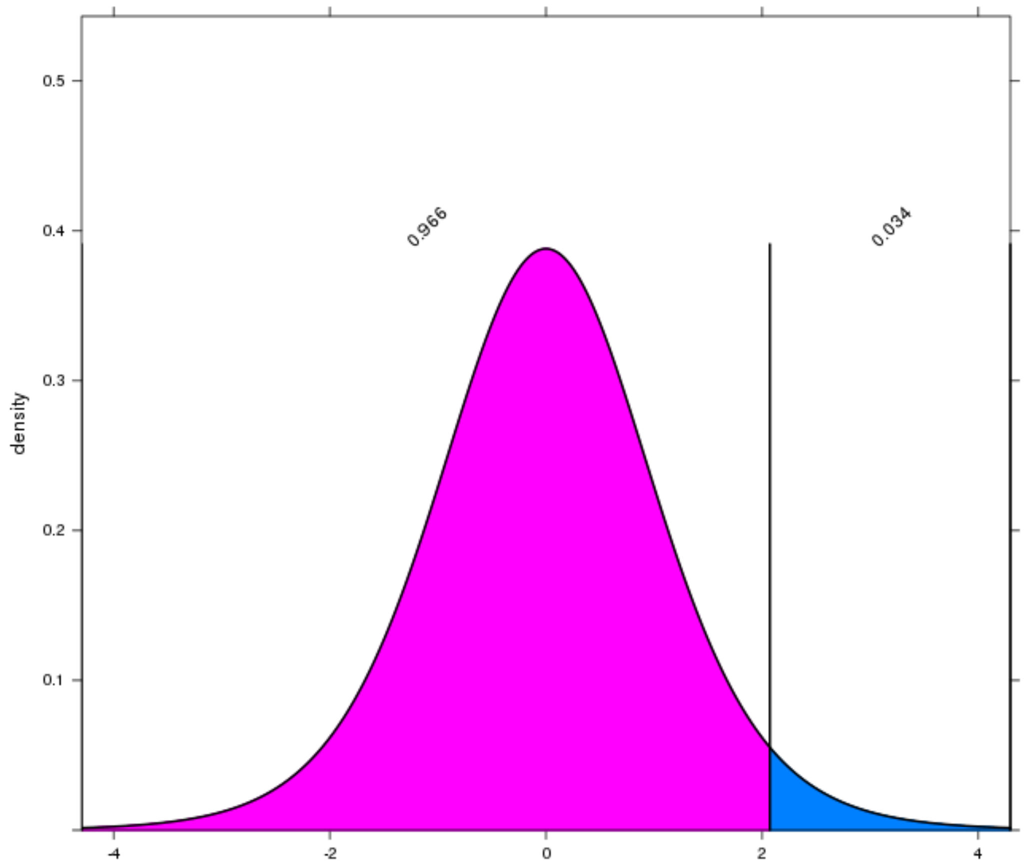
- with degrees of freedom approximately equal to the smaller of $n_1 - 1$ and $n_2 - 1$
- i.e., our null distribution is a t-distribution!



N(0, 1), df = 2, df = 5, df = 15

t-test for differences in means

Tea example



t-statistic = 2.071

t-test in R

```
# difference in the means
mean.diff <- mean(tea) - mean(coffee)

# pooled standard deviation
pool.sd <- sqrt(var(tea)/11 + var(coffee)/10)

# t-statistic
t.stat <- mean.diff/pool.sd

# p-value
p.value <- pt(t.stat, df = 9, lower.tail = FALSE)
```

$$t = \frac{\bar{x}_T - \bar{x}_C}{\sqrt{\frac{s_T^2}{n_T} + \frac{s_C^2}{n_C}}}$$

Does taking an examine increase your pulse rate?

Students pulse rates were measured during lecture and while taking an examine to see if pulse rate goes up while taking an examine

- What are the null and alternative hypotheses?

The data for the 10 students are:

Student	1	2	3	4	5	6	7	8	9	10	Mean	Std. Dev.
Quiz	75	52	52	80	56	90	76	71	70	66	68.8	12.5
Lecture	73	53	47	88	55	70	61	75	61	78	66.1	12.8

Does pulse rate increase while taking an exam?

load('/home/shared/baseball_stats_2017/QuizPulse10.rda')

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

pt(t.stat, df = (n – 1), lower.tail = FALSE)

Answer in R

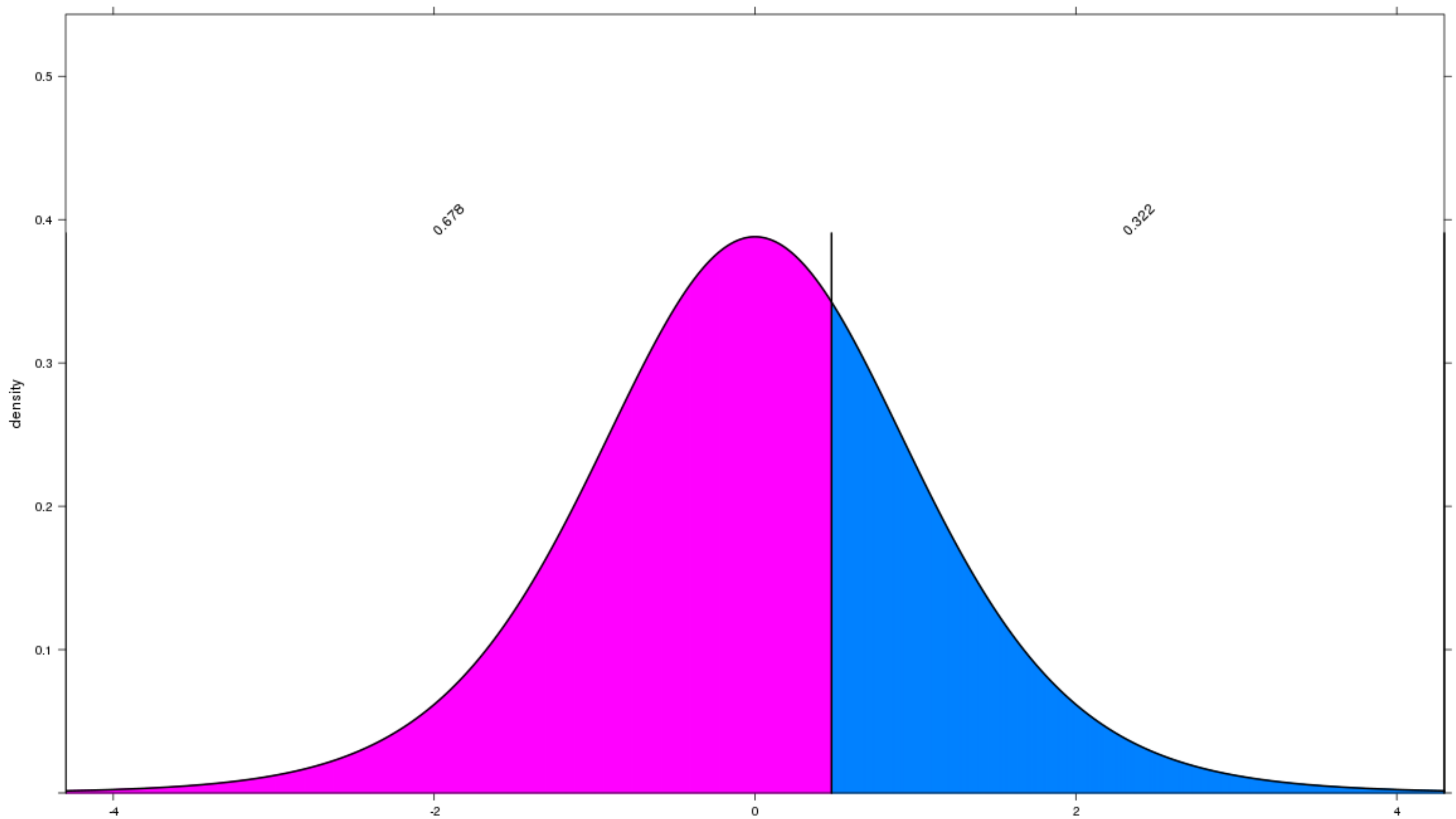
```
quiz <- QuizPulse10$Quiz
lecture <- QuizPulse10$Lecture

mean.diff <- mean(quiz) - mean(lecture)
pool.sd <- sqrt(var(quiz)/10 + var(lecture)/10)

t.stat <- mean.diff/pool.sd

pt(t.stat, df = 9, lower.tail = FALSE)
```

Quiz vs. lecture null distribution



p-value = .322