

Hypothesis tests for two means

Outline for today

Better know a player David Winfield

Hypothesis tests for two means

- Experimental design and testing the effectiveness of a pill
- Review of whether baseball games have gotten longer

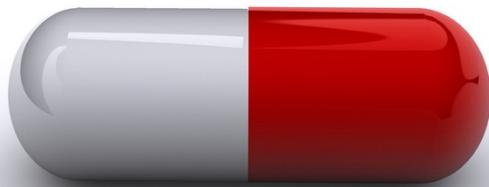
Parametric tests 2 means (t-test)

Better know a player: David
Winfield

Review: Steps for doing a hypothesis test

1. State the null and alternative hypothesis
2. Calculate the observed statistic
3. Create a null distribution
 - Typical statistics you would expect to get if the null hypothesis was true
4. Create a p-value
 - calculate the probability of getting a statistics as great or greater than the observed from the null distribution

Hypothesis tests for two means



Question: Is this pill effective?

Question: Does this steroid work?

Question: What does it mean for a pill to work?

Answer: On average the people who take the pill will be better than the people who do not

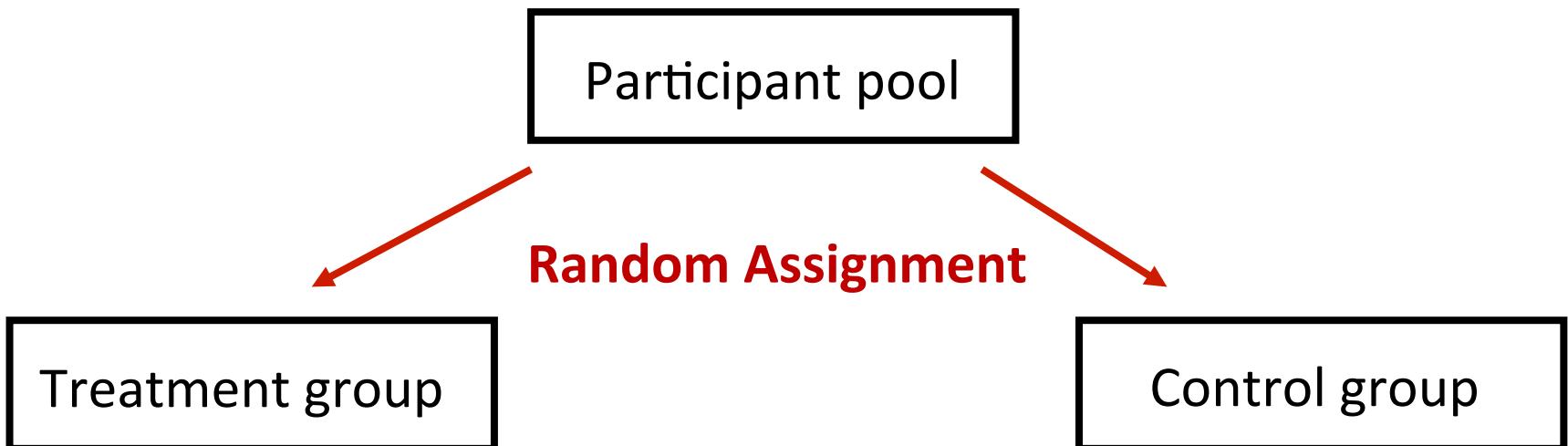
- $\mu_{\text{pill}} > \mu_{\text{no-pill}}$

Experimental design

Question: How can we design a study to test whether a pill *causes* an improvement in health?

Take a group of participant and *randomly assign*:

- Half to a *treatment group* where they get the pill
- Half in a *control group* where they get a fake pill (placebo)
- See if there is more improvement in the treatment group compared to the control group



Experimental design

Question: If the pill did not work, would you expect better outcomes in the treatment or control groups?

If we find that the treatment group performs better, then this could be due to:

- A systematic improvement due to the pill
- The treatment group is better due to the random assignment of people

If we can show that random assignment of people is not likely to account for the differences between the groups, then the pill must have caused the improvement in the treatment group

Assessing whether the random assignment could account for our results

Question: How can we assess if the random assignment of people could account for the observed effect?

- **Answer:** Hypothesis tests!

Question: what is the first we do when testing hypotheses?

- State the null and alternative hypotheses in symbols and words
 - $H_0: \mu_T = \mu_C$ or $\mu_T - \mu_C = 0$
 - $H_A: \mu_T > \mu_C$ or $\mu_T - \mu_C > 0$

What do we do next?

- Let's look at a real experiment!

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:

- Blood samples were taken from all participants and exposed to an antigen
- The production of interferon gamma (which is related to an immune response) was measured

The data showed the following interferon gamma levels:

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

Implementing this permutation test in R

```
# the tea and the coffee data
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)

# the observed statistic
obs.diff <- mean(tea) - mean(coffee)

# combine data from both seasons together
combined.data <- c(tea, coffee)
```

Implementing this permutation test in R

```
null.dist <- NULL

for (i in 1:10000) {

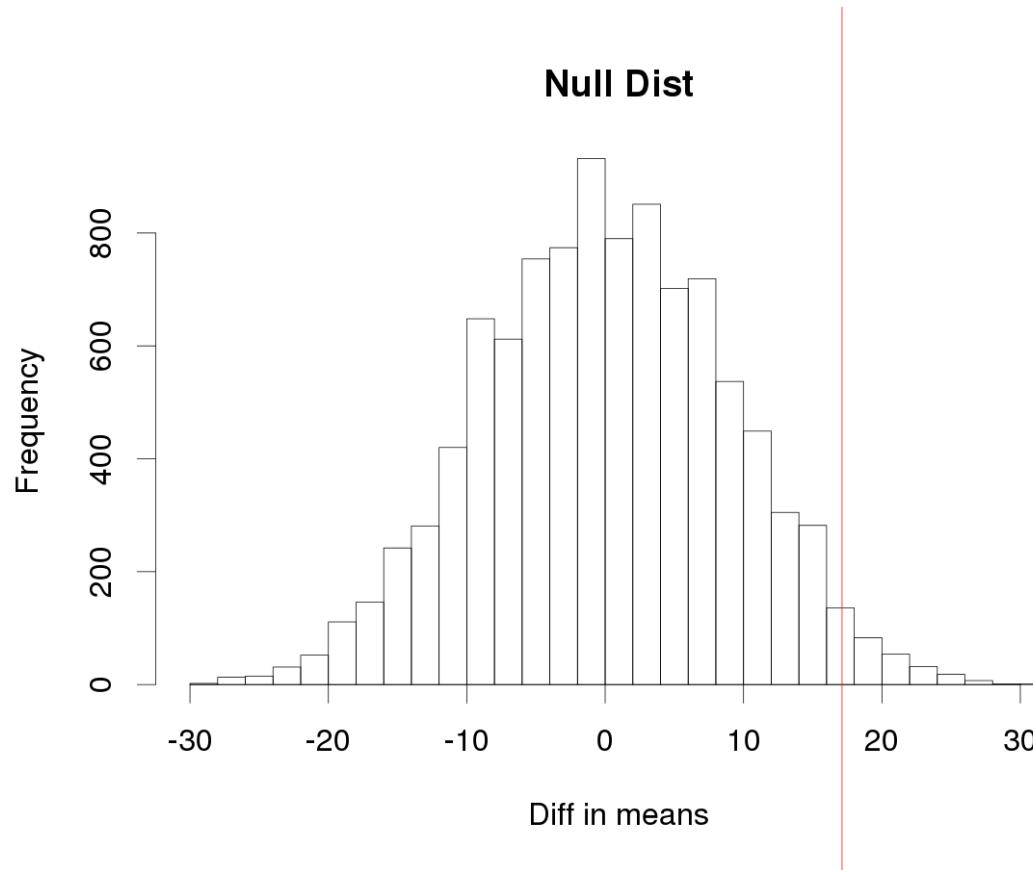
  # shuffle the combined data
  shuffled.durations <- sample(combined.data)

  # get the 'tea' and 'coffee' data sets
  shuff.tea <- shuffled.data[1:11]
  shuff.coffee <- shuffled.data[12:21]

  # calculate the observed statistic under the null hypothesis
  null.dist[i] <- mean(shuff.tea) - mean(shuff.coffee)
}

hist(null.dist, n = 100, main = 'Null Dist', xlab = 'Mean diff')      # display the null distribution
p.value <- sum(null.dist >= obs.diff)/10000
```

3. Creating a null distribution



p-value = .0255
Conclusions?

Review: Comparing two means

Have baseball games gotten longer in the past 50 years?

We will address this question by comparing mean lengths of games in 1964 to those in 2014

After plotting the data (descriptive statistics) what is the first step of hypothesis testing?

1. Null and Alternative Hypotheses

1a. State the null and alternative hypotheses in words

- **Null hypothesis:** Baseball games are the same length in 1964 as they are in 2014
- **Alternative hypothesis:** Baseball games are longer in 2014 than in 1964

1b. State 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 $\mu_{2014} - \mu_{1964} > 0$

What do we do next?

- 2. Compute the statistic of interest

What is the statistic of interest?

2. Compute the statistic of interest

Average game length 1964 is: $\bar{x}_{1964} = 154.21$ minutes

- (based on $n = 684$ games with 54 outs)

Average game length 2014 is: $\bar{x}_{2014} = 187.64$ minutes

- (based on $n = 1021$ games with 54 outs)

`observed.stat <- 187.64 - 154.21 = 33.42 minutes`

What do we do next?

- 3. Create a null distribution

Hypothesis tests for two means

3. Calculate the null distribution

- How can we create a null distribution???

One way: under the null hypothesis all games lengths from 1964 and 2014 are equally likely

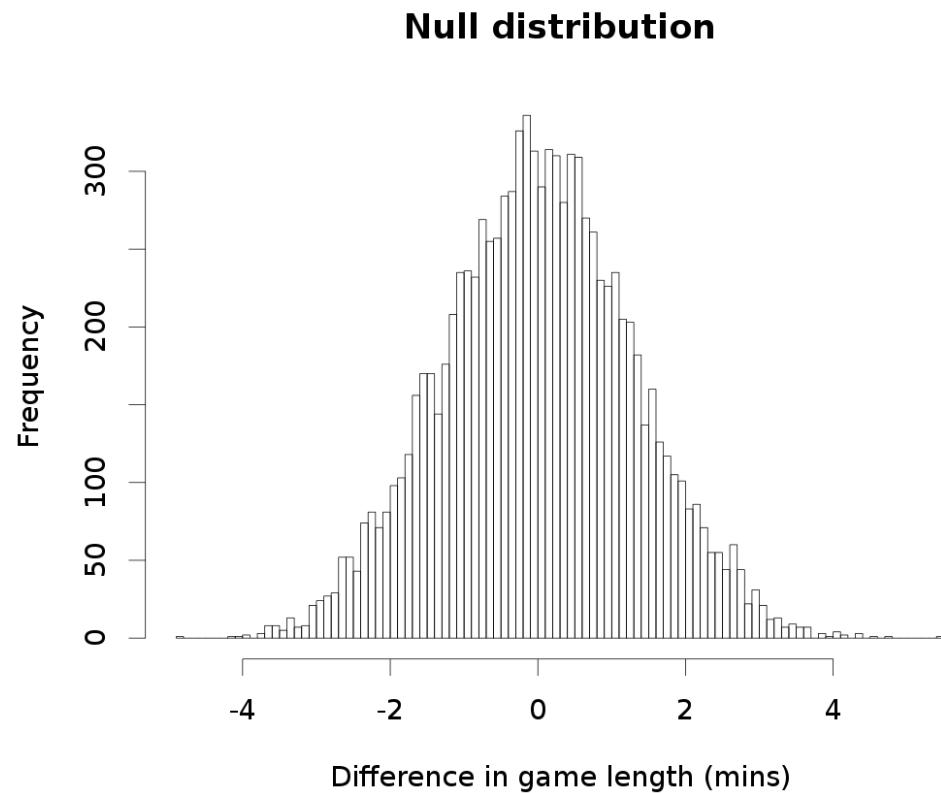
Thus combine all the games lengths from the 1964 and 2014 seasons into one vector

We can then randomly select **684** games to simulate the 1964 season and take the remaining **1021** to simulate the 2014 season

The difference in these means of these 684 and 1021 games gives us one point in the null distribution

If we repeat this 10,000 times we will get a full null distribution

Hypothesis tests for two means



Do the results seem statistically significant?

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

Implementing this permutation test in R

Getting the data (see Moodle)

```
load("/home/shared/all.game.logs.Rda")  
  
game.logs$year <- substr(game.logs$Date, 1, 4)  
  
game.logs54 <- filter(game.logs, LengthInOuts == 54)  
  
game.logs.54.out.1964 <- filter(game.logs54, year == 1964)  
  
game.logs.54.out.2014 <- filter(game.logs54, year == 2014)
```

Implementing this permutation test in R

```
# game durations for 1964 and 2014
```

```
game.duration.1964 <- game.logs.54.outs.1964$Duration  
game.duration.2014 <- game.logs.54.outs.2014$Duration
```

```
# number of games in 1964 and 2014
```

```
num.games.1964 <- length(game.duration.1964)  
num.games.2014 <- length(game.duration.2014)
```

```
# the observed statistic
```

```
obs.diff <- mean(game.duration.2014) - mean(game.duration.1964)
```

```
# combine data from both seasons together
```

```
combined.durations <- c(game.duration.1964, game.duration.2014)
```

Implementing this permutation test in R

```
null.dist <- NULL

for (i in 1:10000) {

  # shuffle the combined game durations
  shuffled.durations <- sample(combined.durations)

  # get the random durations for 1964 and 2014
  shuff.1964 <- shuffled.durations[1:num.games.1964]
  shuff.2014 <- shuffled.durations[(num.games.1964 +1) :length(shuffled.durations)]

  # calculate the observed statistic under the null hypothesis
  null.dist[i] <- mean(shuff.2014) - mean(shuff.1964)
}

hist(null.dist, n = 100, main = 'Null Dist', xlab = 'Mean diff')      # display the null distribution
p.value <- sum(null.dist >= obs.diff)/10000
```

Worksheet 9, problem 2

Please resubmit worksheet 9 with question 2
answered by midnight on Tuesday