Lab_2

Frank Woodling

February 16, 2016

1.

$H_0 : D = 0$
$H_1 : D \neq 0$

```
(119.0 + 489.1)/2 - 147.8
```

```
## [1] 156.25
```

2.

$\frac{!(1+2)}{(!1*!2)} = 3$

There are 3 possible combinations.

Permutation 1: seeded: 147.8, 489.1; unseeded: 119.0
Permutation 2: seeded: 489.1, 119.0; unseeded: 147.8
Permutation 3: seeded: 119.0, 147.8; unseeded: 489.1

3.

Permutation 1:
(147.8 + 489.1)/2 - 119 = 199.45

Permutation 2:
(489.1 + 119.0)/2 - 147.8 = 156.25

Permutation 3:
(119.0 + 147.8)/2 - 489.1 = -355.7

4.

$P_{two-tail} = \binom{m+n}{m}$
$\frac{1}{\binom{1+2}{1}} = 0.3333$

5.

```
cloudseeding <- read.table('cloudseeding.txt', header=T)
attach(cloudseeding)

teststat.obs = mean(rainfall[treatment == "seeded"]) - mean(rainfall[treatment == "unseeded"])
teststat.obs
```

```
## [1] 697.6583
```

6.

$\binom{10}{4} = 210$

7.

```
library('gtools')
```

```
## Warning: package 'gtools' was built under R version 3.2.3
```

```
unseeded = combinations(10, 4, v = rainfall, set = F, repeats.allowed = F)

seeded = NULL
for(i in 1:210){
  seeded = rbind(seeded, setdiff(rainfall, unseeded[i,]))
}

#seeded
```

8.

```
teststat = rep(NA, 210)
for (i in 1:210){
  teststat[i] = mean(seeded[i,]) - mean(unseeded[i,])
}
```

9.

```
sum(teststat >= teststat.obs)/210
```

```
## [1] 0.03809524
```

With a p-value of 0.04285714 we can reject the null hypothesis. We can conclude that there is a difference between the means.

10.

```
teststat.obs = median(rainfall[treatment == "seeded"]) - median(rainfall[treatment == "unseeded"])
choose(10, 4)
```

```
## [1] 210
```

```
unseeded = combinations(10, 4, v = rainfall, set = F, repeats.allowed = F)

seeded = NULL
for(i in 1:210){
  seeded = rbind(seeded, setdiff(rainfall, unseeded[i,]))
}

teststat = rep(NA, 210)
for (i in 1:210){
  teststat[i] = median(seeded[i,]) - median(unseeded[i,])
}

sum(teststat >= teststat.obs)/210
```

```
## [1] 0.04285714
```

The p-value is 0.04285714. We can reject the null hypothesis and conclude that the differences of the medians are not equal to 0.

11.

```
teststat.obs = sum(rainfall[treatment == "seeded"]) - sum(rainfall[treatment == "unseeded"])

choose(10, 4)
```

```
## [1] 210
```

```
unseeded = combinations(10, 4, v = rainfall, set = F, repeats.allowed = F)

seeded = NULL
for(i in 1:210){
  seeded = rbind(seeded, setdiff(rainfall, unseeded[i,]))
}

teststat = rep(NA, 210)
for (i in 1:210){
  teststat[i] = sum(seeded[i,]) - sum(unseeded[i,])
}

sum(teststat >= teststat.obs)/210
```

```
## [1] 0.03809524
```

With a p-value of 0.03809524 we can reject the null hypothesis. We can conclude that the difference in sums of the the two treatments is not 0.

12.

```
teststat.obs = max(rainfall[treatment == "seeded"]) - max(rainfall[treatment == "unseeded"])

choose(10, 4)
```

```
## [1] 210
```

```
unseeded = combinations(10, 4, v = rainfall, set = F, repeats.allowed = F)

seeded = NULL
for(i in 1:210){
  seeded = rbind(seeded, setdiff(rainfall, unseeded[i,]))
}

teststat = rep(NA, 210)
for (i in 1:210){
  teststat[i] = max(seeded[i,]) - max(unseeded[i,])
}

sum((teststat - teststat.obs) >= 0)/210
```

```
## [1] 0.07142857
```

```
sum(teststat >= teststat.obs)/210
```

```
## [1] 0.07142857
```

The p-value is 0.07142857 so we fail to reject the null hypothesis. We cannot conclude there is a difference in maximums between the two treatments.

Lab Summary

1.

In the majority of the tests we concluded that there is a difference in rainfall between the seeded clouds and unseeded clouds. The results of the 4 different tests were similar. Three were just under 0.05 and the test using maximums was just above that figure. We can safely conclude that there is a difference between seeded and unseeded.

2.

In order to perform this test with the full data set it would take $\binom{52}{26}$ permutations. This works out to be 4.959185e+14 (there are less stars in the universe then this). Even once you get into the millions computers may start to struggle with performing calculations in a reasonable amount of time.

Code

```
### 1. ###
# H0: D = 0
# H1: D != 0

(119.0 + 489.1)/2 - 147.8
# [1] 156.25

############################
### 2. ###
factorial(1+2)/(factorial(1)*factorial(2))
# [1] 3
# There are 3 possible combinations.

# Permutation 1: seeded: 147.8, 489.1; unseeded: 119.0
# Permutation 2: seeded: 489.1, 119.0; unseeded: 147.8
# Permutation 3: seeded: 119.0, 147.8; unseeded: 489.1

############################
### 3. ###
# Permutation 1:
(147.8 + 489.1)/2 - 119
# [1] 199.45

# Permutation 2:
(489.1 + 119.0)/2 - 147.8
# [1] 156.25

# Permutation 3:
(119.0 + 147.8)/2 - 489.1
# [1] -355.7

############################
### 4. ###
# P_twotail = number of |Ds| >= |D_obs|/choose(m+n, m)    pg 27 4.
1/choose(1+2,1)
# [1] 0.3333333

############################
### 5. ###
setwd("~/Documents/STAT 3480")
cloudseeding <- read.table('cloudseeding.txt', header=T)
attach(cloudseeding)

teststat.obs = mean(rainfall[treatment == "seeded"]) - mean(rainfall[treatment == "unseeded"])
```

```
teststat.obs

############################
### 6. ###
choose(10, 4)
# [1] 210

############################
### 7. ###
# install.packages('gtools')
library('gtools')
unseeded = combinations(10, 4, v = rainfall, set = F, repeats.allowed = F)

seeded = NULL
for(i in 1:210){
  seeded = rbind(seeded, setdiff(rainfall, unseeded[i,]))
}
seeded

############################
### 8. ###
teststat = rep(NA, 210)
for (i in 1:210){
  teststat[i] = mean(seeded[i,]) - mean(unseeded[i,])
}

############################
### 9. ###
teststat >= teststat.obs
sum(teststat >= teststat.obs)
sum(teststat >= teststat.obs)/210
# [1] 0.03809524

############################
### 10. ###
teststat.obs = median(rainfall[treatment == "seeded"]) - median(rainfall[treatment == "unseeded"])
# [1] 311.5

choose(10, 4)
# [1] 210

unseeded = combinations(10, 4, v = rainfall, set = F, repeats.allowed = F)

seeded = NULL
for(i in 1:210){
  seeded = rbind(seeded, setdiff(rainfall, unseeded[i,]))
}

teststat = rep(NA, 210)
for (i in 1:210){
  teststat[i] = median(seeded[i,]) - median(unseeded[i,])
}

teststat >= teststat.obs
```

```
sum(teststat >= teststat.obs)
sum(teststat >= teststat.obs)/210
# [1] 0.04285714

############################
### 11. ###
teststat.obs = sum(rainfall[treatment == "seeded"]) - sum(rainfall[treatment == "unseeded"])
teststat.obs
# [1] 4320.9

choose(10, 4)
# [1] 210

unseeded = combinations(10, 4, v = rainfall, set = F, repeats.allowed = F)

seeded = NULL
for(i in 1:210){
  seeded = rbind(seeded, setdiff(rainfall, unseeded[i,]))
}

teststat = rep(NA, 210)
for (i in 1:210){
  teststat[i] = sum(seeded[i,]) - sum(unseeded[i,])
}

teststat >= teststat.obs
sum(teststat >= teststat.obs)
sum(teststat >= teststat.obs)/210
# [1] 0.03809524

############################
### 12. ###
teststat.obs = max(rainfall[treatment == "seeded"]) - max(rainfall[treatment == "unseeded"])
teststat.obs
# [1] 2597.8

choose(10, 4)
# [1] 210

unseeded = combinations(10, 4, v = rainfall, set = F, repeats.allowed = F)

seeded = NULL
for(i in 1:210){
  seeded = rbind(seeded, setdiff(rainfall, unseeded[i,]))
}

teststat = rep(NA, 210)
for (i in 1:210){
  teststat[i] = max(seeded[i,]) - max(unseeded[i,])
}
teststat
teststat >= teststat.obs
sum((teststat - teststat.obs) >= 0)/210
sum(teststat >= teststat.obs)/210
```

```
# [1] 0.07142857
```