4211 Homework 9

Matthew DeSantis

2023-04-20

1

(a)

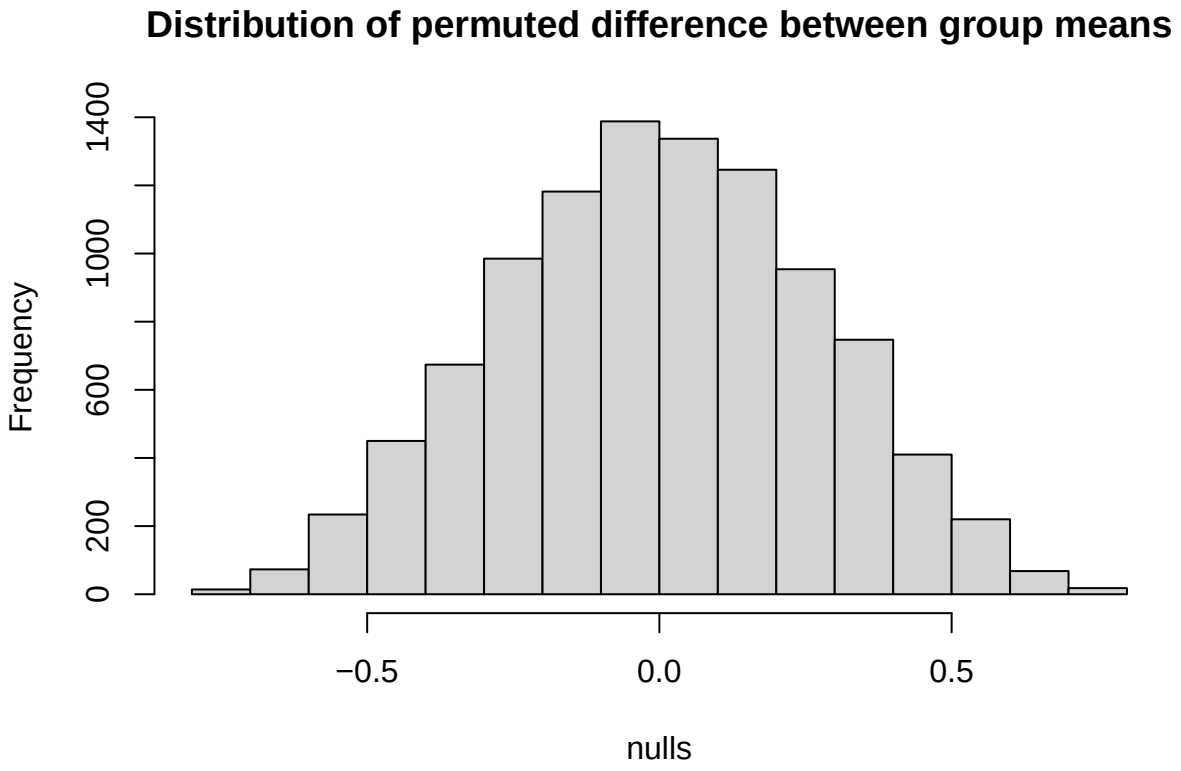
```
x = data.frame(
  msce = c(8.27,8.20,8.25,8.14,9.00,8.10,7.20,8.32,7.70,8.50,9.48,8.65,8.16,8.83,7.76,8.63),
  eth = c(rep(0,9), rep(1,7))
)


permtest = function(x,y){
  xstar = sample(x)
  return(mean(y[xstar==1])-mean(y[xstar==0]))
}

set.seed(488103)
observeddiff = mean(x$msce[x$eth==1]-x$msce[x$eth==0])
```

```
## Warning in x$msce[x$eth == 1] - x$msce[x$eth == 0]: longer object length is not
## a multiple of shorter object length
```

```
nulls = replicate(10000, permtest(x$eth, x$msce))
hist(nulls, main = "Distribution of permuted difference between group means")
```



(b)

```
mean(abs(nulls)>abs(observeddiff))
```

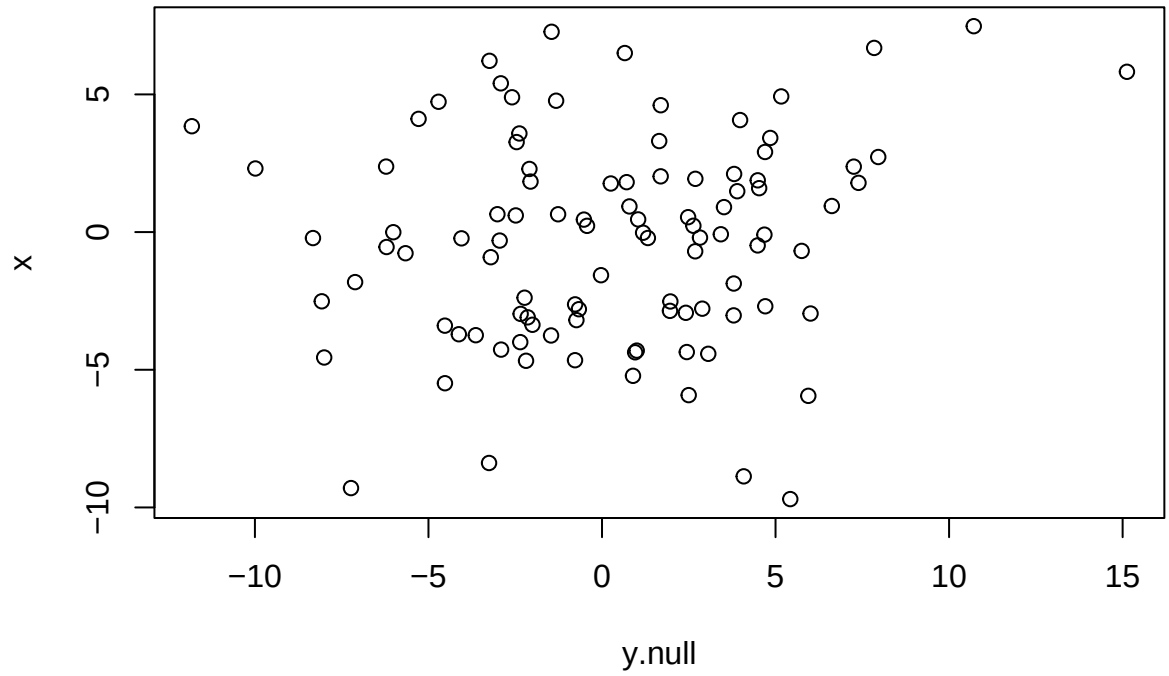
```
## [1] 0.0442
```

The difference is significant at the level alpha = 0.05.

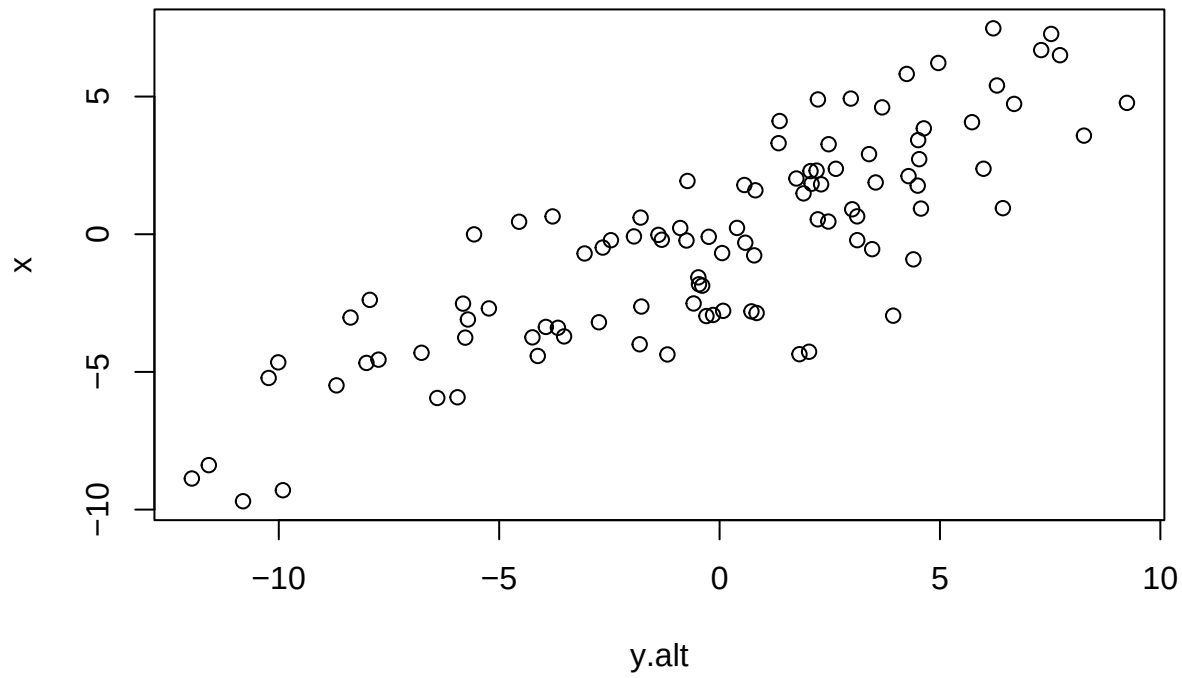
2

(a)

```
set.seed(4211)
x=rnorm(100,0,4)
y.null=rnorm(100,0,5) # null hypothesis
y.alt=x+rnorm(100,0,3) # alternative hypothesis
## from a location model with normal errors
plot(y.null, x)
```




```
plot(y.alt, x)
```




```
cor.test(x, y.null)
```

```
##
##  Pearson's product-moment correlation
##
## data:  x and y.null
## t = 1.5235, df = 98, p-value = 0.1309
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.04567892  0.33841165
## sample estimates:
##       cor
## 0.1521039
```

```
cor.test(x, y.alt)
```

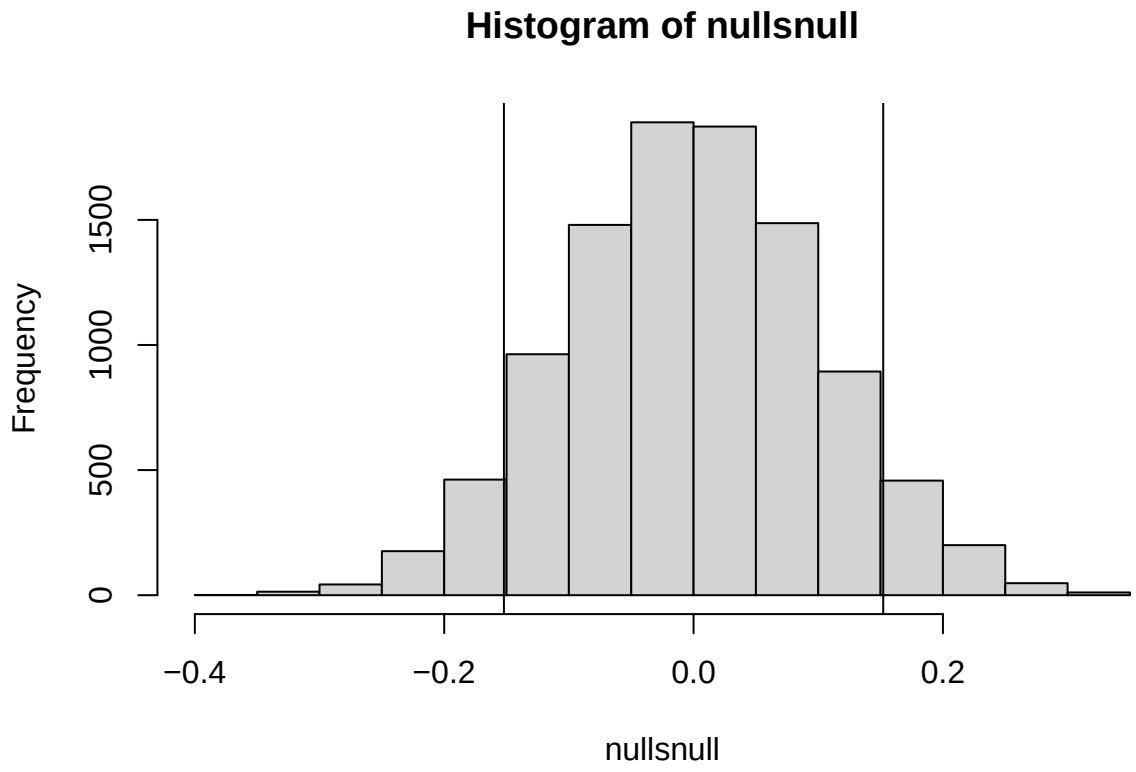
```
##
##  Pearson's product-moment correlation
##
## data:  x and y.alt
## t = 14.757, df = 98, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7576060 0.8828581
```

```
## sample estimates:
##       cor
## 0.8304473
```

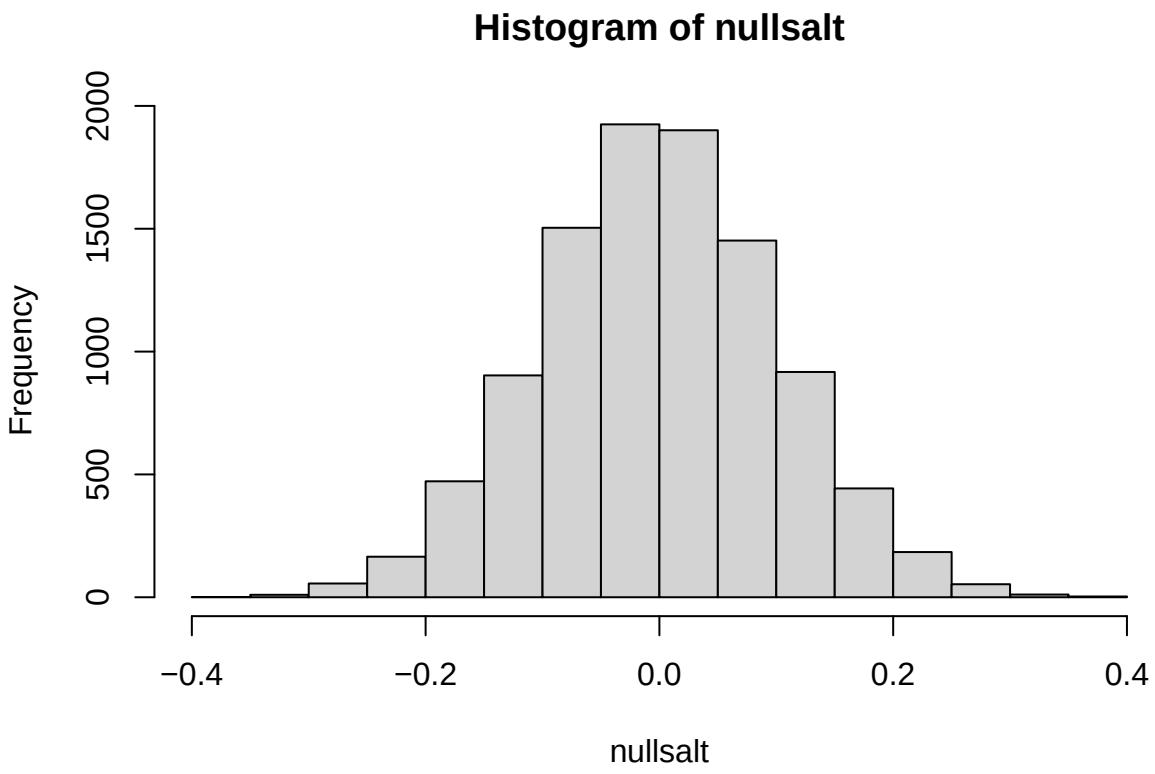
(b)

```
permtest = function(x,y){
  xstar = sample(x)
  return(cor(xstar, y))
}

set.seed(488103)
observedcornull = cor(x, y.null)
observedcoralt = cor(x, y.alt)
nullsnull = replicate(10000, permtest(x, y.null))
nullsalt = replicate(10000, permtest(x, y.alt))
hist(nullsnull)
abline(v=observedcornull)
abline(v=-observedcornull)
```




```
hist(nullsalt)
```




```
mean(abs(nullsnull)>abs(observedcornull))
```

```
## [1] 0.1362
```

```
mean(abs(nullsalt)>abs(observedcoralt))
```

```
## [1] 0
```

3

(a)

```
bootstrap = function(x,y){
  index = sample(1:length(x), replace = TRUE)
  xstar = x[index]
  ystar = y[index]
  return(cor(xstar, ystar))
}

set.seed(488103)
bootnull = replicate(10000, bootstrap(x, y.null))
bootalt = replicate(10000, bootstrap(x, y.alt))
var(bootnull)
```

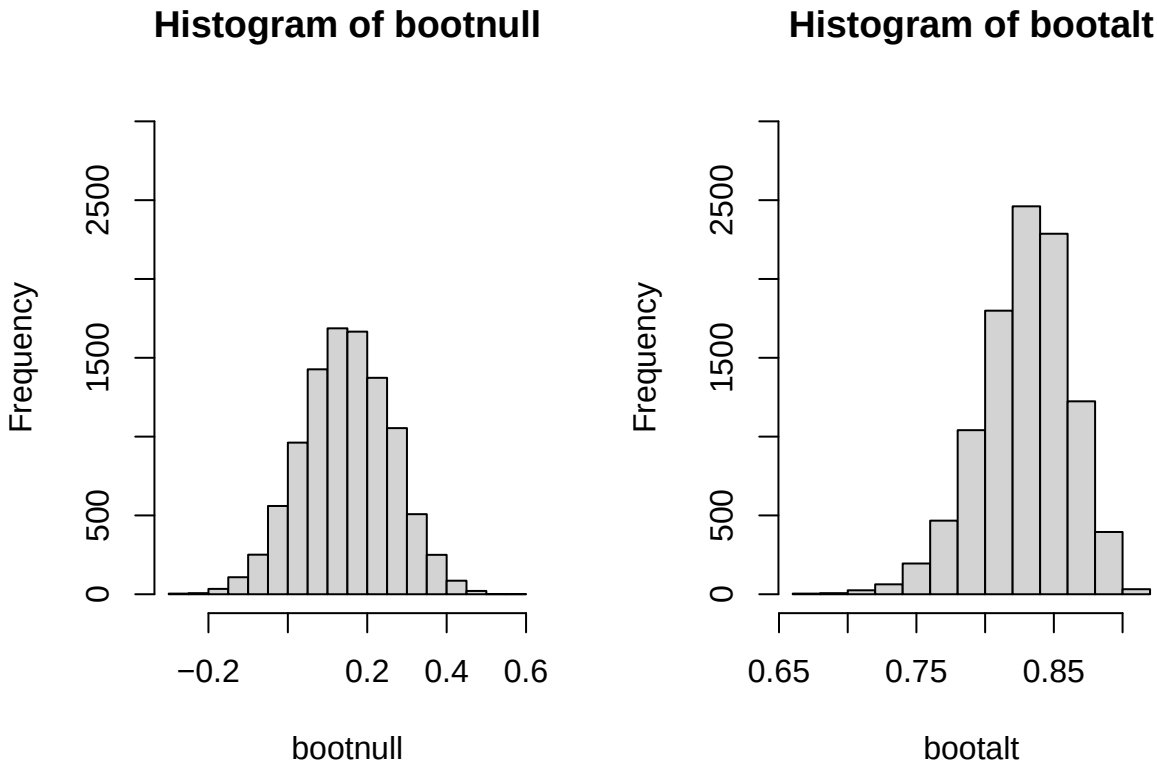
```
## [1] 0.01283102
```

```
var(bootalt)
```

```
## [1] 0.001072049
```

(b)

```
par(mfrow=c(1,2))
hist(bootnull, ylim=c(0,3000))
hist(bootalt, ylim=c(0,3000))
```



The distribution of correlation under the null hypothesis appears to be normally distributed about 1.5, while the distribution of correlation under the alternative hypothesis is skewed left, with a peak around 0.83.