C2

Jeremiah Theisen

2024-10-24

5.2.1

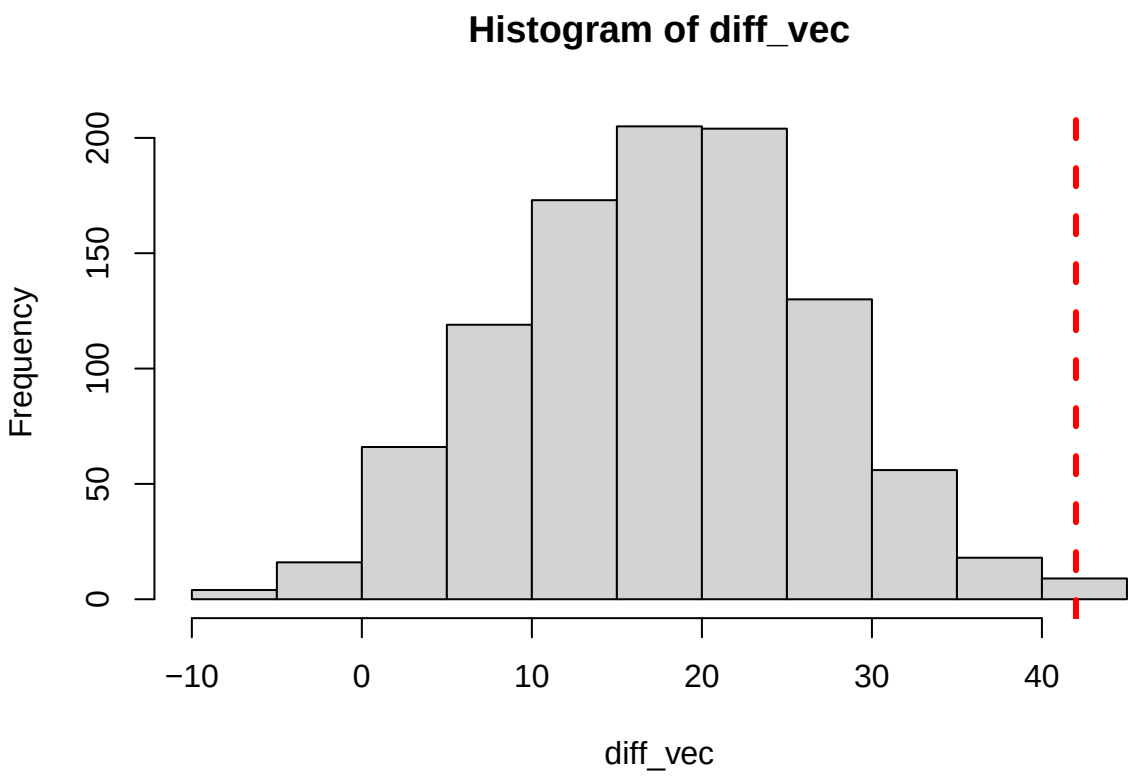
a. phat = 0.193
b. 42
c.

```
raise_rate=0.193
num_men=312
num_women=216

menraise=sample(c("RAISE", "NO"), num_men, replace=TRUE, prob=c(raise_rate, 1-raise_rate))
womenraise=sample(c("RAISE", "NO"), num_women, replace=TRUE, prob=c(raise_rate, 1-raise_rate))

d=length(which(menraise=="RAISE"))-length(which(womenraise=="RAISE"))
print(d)
```

```
## [1] 5
```

d.

```
observed_diff=42

diff_vec=rep(NA,1000)

for(i in 1:1000){
    menraise=sample(c("RAISE", "NO"), num_men, replace=TRUE, prob=c(raise_rate, 1-raise_rate))
    womenraise=sample(c("RAISE", "NO"), num_women, replace=TRUE, prob=c(raise_rate, 1-raise_rate))
    d=length(which(menraise=="RAISE"))-length(which(womenraise=="RAISE"))
    diff_vec[i]=d
}
hist(diff_vec)
abline(v = observed_diff, col="red", lwd=3, lty=2)
```




```
print(observed_diff)
```

```
## [1] 42
```

Very unlikely e.

```
length(which(diff_vec>=observed_diff))/1000
```

```
## [1] 0.004
```

f. It is very unlikely to be random chance, it falls outside of 95%.

5.2.2 SKIP

5.2.3

a.

```
icecream_n=80
icecream_p=0.6

icecream_samp=sample(c("TOPPINGS", "NO"), icecream_n, replace=TRUE, prob=c(icecream_p, 1-icecream_p))
length(which(icecream_samp=="TOPPINGS"))
```

```
## [1] 45
```

```
icecream_prop=length(which(icecream_samp=="TOPPINGS"))/icecream_n
```

b.

```
frozen_n=110
frozen_p=0.45

frozen_samp=sample(c("TOPPINGS", "NO"), frozen_n, replace=TRUE, prob=c(frozen_p, 1-frozen_p))
length(which(frozen_samp=="TOPPINGS"))
```

```
## [1] 47
```

```
frozen_prop=length(which(frozen_samp=="TOPPINGS"))/frozen_n
```

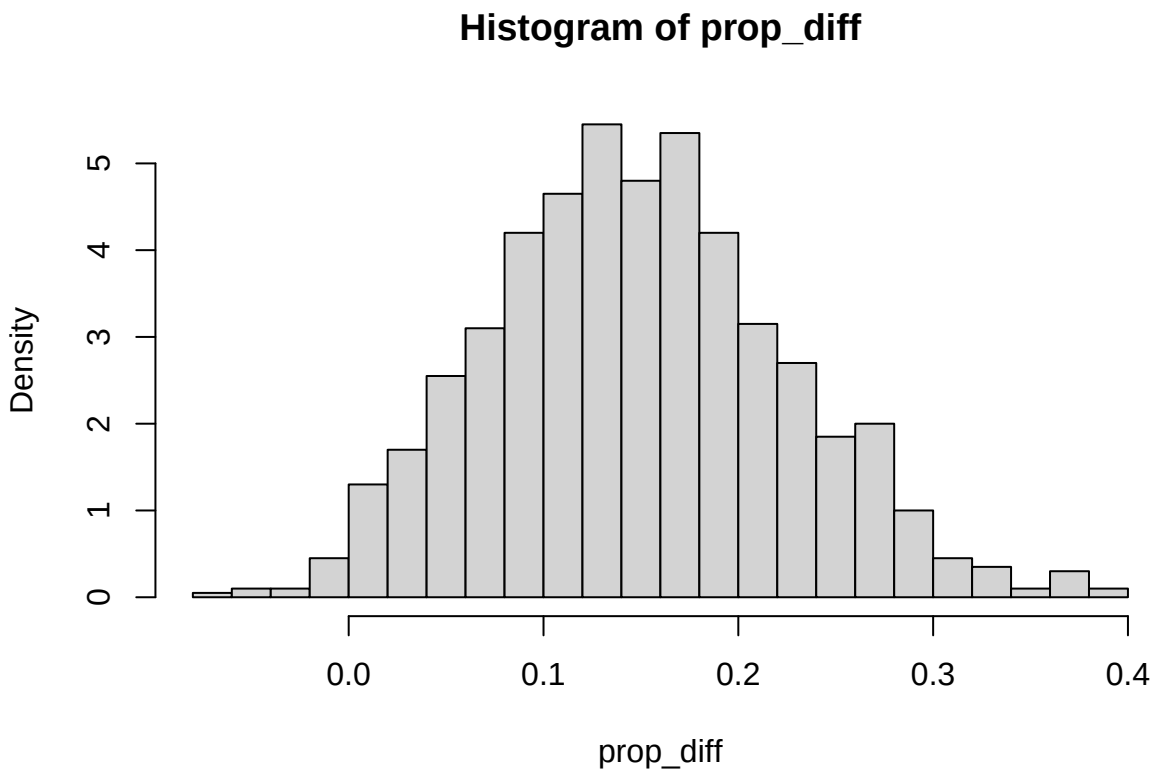
c.

```
icecream_prop-frozen_prop
```

```
## [1] 0.1352273
```

d.

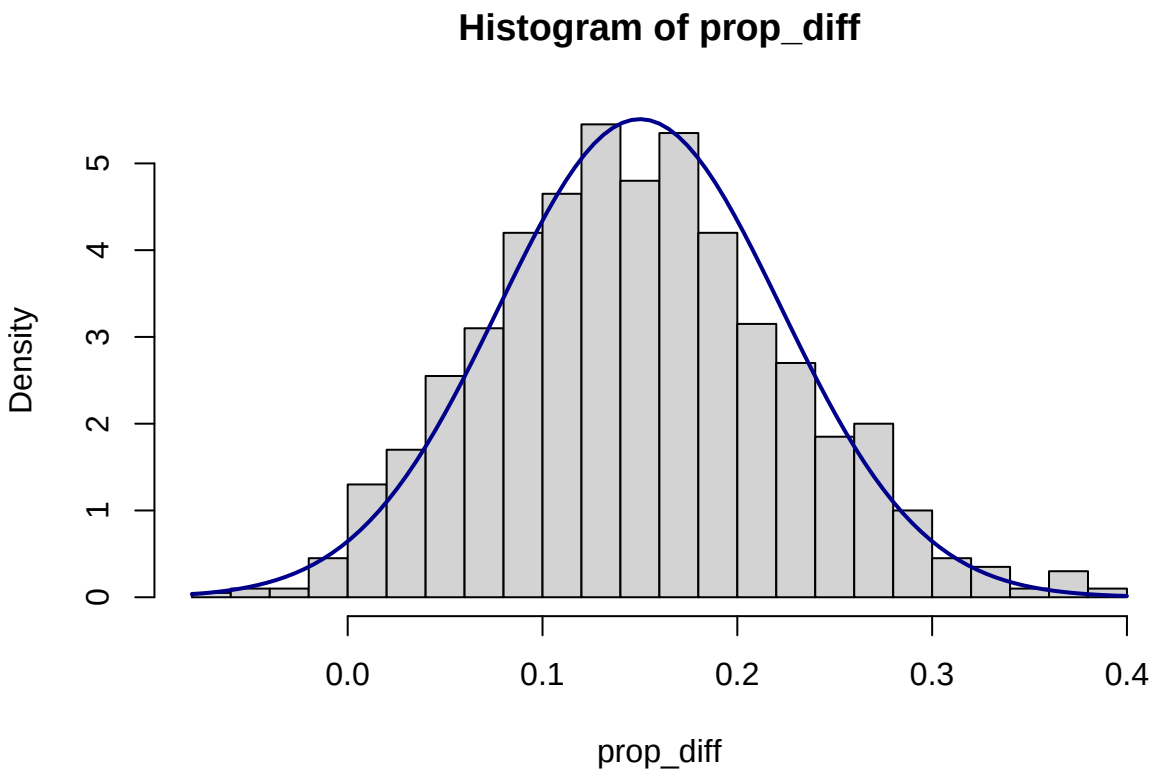
```
prop_diff=rep(NA,1000)

for(i in 1:1000){
    icecream_samp=sample(c("TOPPINGS", "NO"), icecream_n, replace=TRUE, prob=c(icecream_p, 1-icecream_p
    icecream_prop=length(which(icecream_samp=="TOPPINGS"))/icecream_n
    frozen_samp=sample(c("TOPPINGS", "NO"), frozen_n, replace=TRUE, prob=c(frozen_p, 1-frozen_p))
    frozen_prop=length(which(frozen_samp=="TOPPINGS"))/frozen_n
    prop_diff[i]=icecream_prop-frozen_prop
}
hist(prop_diff, breaks=25, prob=TRUE)
```



e. mu_D = 0.45 SE_D = 0.0724

f.

```
mu_D=0.15
SE_D=0.0724


hist(prop_diff, breaks=25, prob=TRUE)
curve(dnorm(x, mean=mu_D, sd=SE_D), col="darkblue", lwd=2, add=TRUE, yaxt="n")
```



5.2.4

a.

```
eastpoll=100
east_p=sample(c(0.3, 0.4, 0.5), 1, prob=c(0.25, 0.5, 0.25))

east_samp=sample(c("SPICYWINGS", "NO"), eastpoll, replace=TRUE, prob=c(east_p, 1-east_p))
east_x=length(which(east_samp=="SPICYWINGS"))
print(east_x)
```

```
## [1] 25
```

p1 = 0.34 b.

```
westpoll=100
west_p=sample(c(0.3, 0.4, 0.5), 1, prob=c(0.25, 0.5, 0.25))

west_samp=sample(c("SPICYWINGS", "NO"), westpoll, replace=TRUE, prob=c(west_p, 1-west_p))
west_x=length(which(west_samp=="SPICYWINGS"))
print(west_x)
```

```
## [1] 43
```

p2 = 0.63

c. d = -0.29 SED = 0.0676

d. (-0.401, -0.1787)

e. We have good reason to say that there is a negative relationship, but not enough data to say what its strength is West orders more

f.

```
east_phat=east_x/eastpoll
west_phat=west_x/westpoll

d=east_phat-west_phat

SE_D=sqrt(east_phat*(1-east_phat)/eastpoll + west_phat*(1-west_phat)/westpoll)
LB=d-1.645*SE_D
UB=d+1.645*SE_D
print(paste0("Lower Bound: ", LB , ", Upper Bound:", UB))
```

```
## [1] "Lower Bound: -0.288195490432827, Upper Bound:-0.0718045095671728"
```

g.

```
east_p-west_p
```

```
## [1] -0.1
```

This is outside the interval

5.2.5

a. p1 - p2 = 0

b. p1 - p2 > 0

c. ppool = 0.193

d. mu_D = 0 SE_D = 0.0349

e. d = 0.09

f.

```
1 - pnorm(0.09, mean=0, sd=0.0349)
```

```
## [1] 0.004957258
```

g.

h. zscore = 2.5787

i.

```r
1 - pnorm(2.5787, mean=0, sd=1)
```

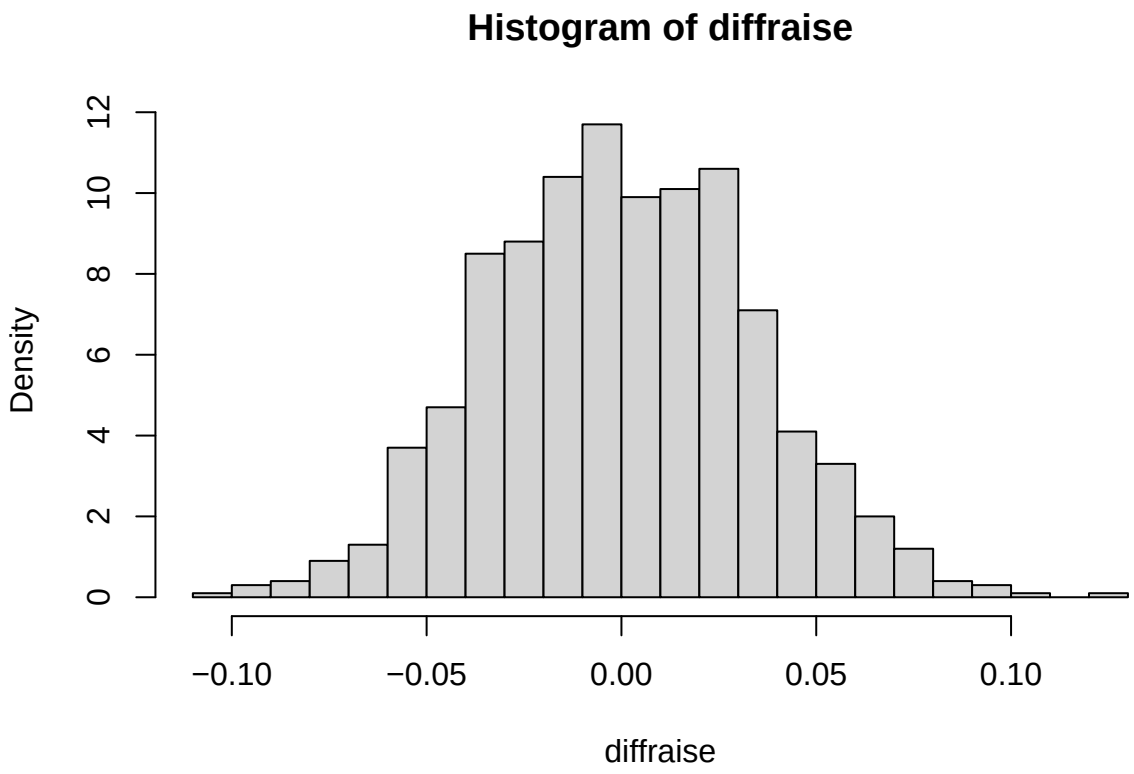
```
## [1] 0.004958644
```

```r
pnorm(2.5787, mean=0, sd=1)
```

```
## [1] 0.9950414
```

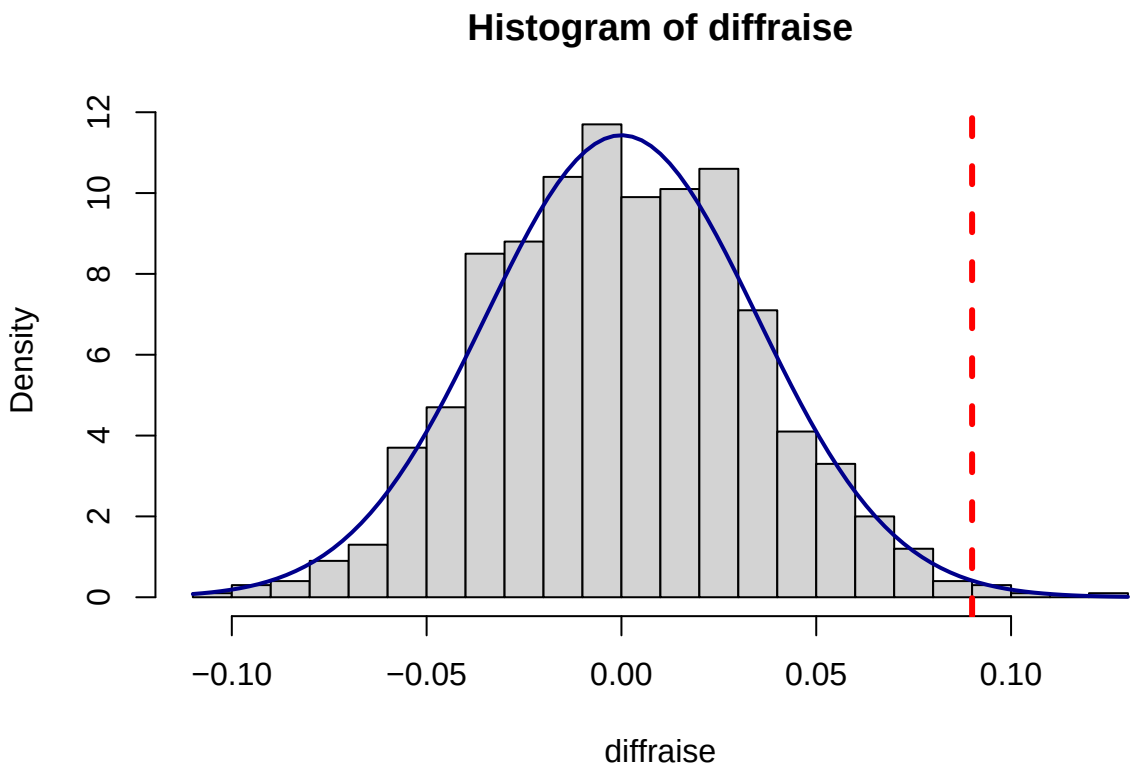
j. p-value = 0.004
k. p-value measures the probability that the observed difference, or a more extreme value, will occur given the null hypothesis.
l. The p-value is smaller than 0.05, which is less than alpha, so we reject the null hypothesis
m. Type 1, we would have rejected a true statement
n.

```r
trials=1000
num_men=312
num_women=216
m_raise=72
w_raise=30

raise_rate=(m_raise+w_raise)/(num_men+num_women)
diffraise=rep(NA, trials)

for(i in 1:trials){
    menraise=sample(c("RAISE", "NO"), num_men, replace=TRUE, prob=c(raise_rate, 1-raise_rate))
    womenraise=sample(c("RAISE", "NO"), num_women, replace=TRUE, prob=c(raise_rate, 1-raise_rate))
    diffraise[i]=length(which(menraise=="RAISE"))/num_men-length(which(womenraise=="RAISE"))/num_women
}
hist(diffraise, breaks=25, prob=TRUE)
```



o.

```
mu_D=0
SE_D=0.0349
observed_diff=0.09

hist(diffraise, breaks=25, prob=TRUE)
curve(dnorm(x, mean=mu_D, sd=SE_D), col="darkblue", lwd=2, add=TRUE, yaxt="n")
abline(v = observed_diff, col="red", lwd=3, lty=2)
```



p.

```
length(which(diffraise>=observed_diff))/trials
```

```
## [1] 0.005
```

Its a bit larger

5.2.6 SKIP

5.2.7 SKIP