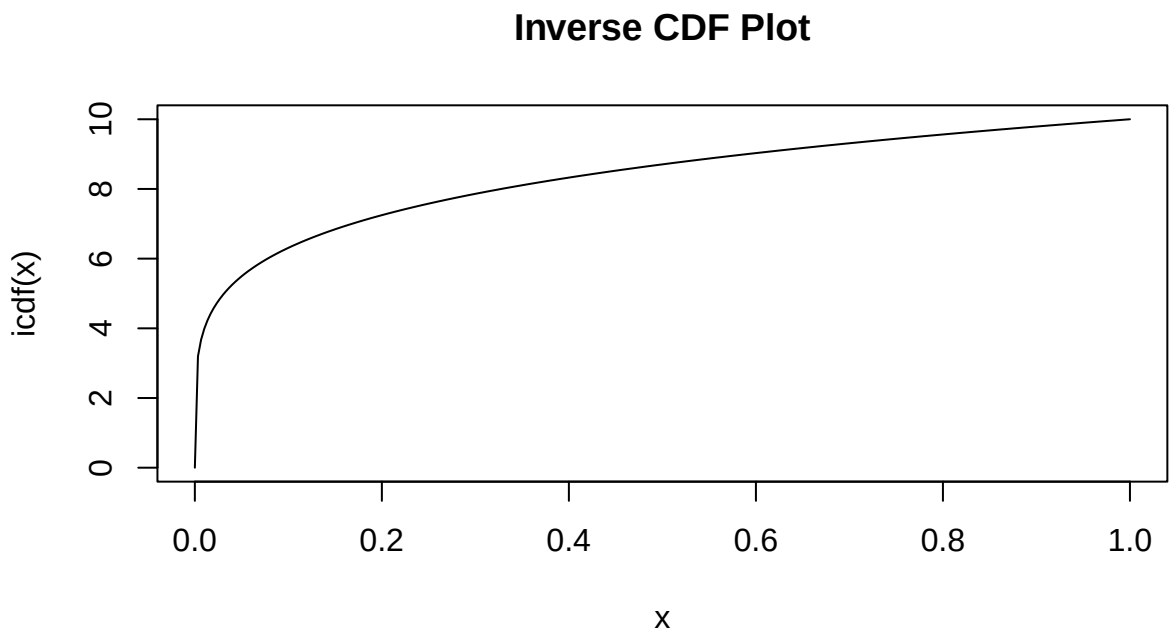
PBSR Assignment 1

Bhuvnesh Sahu

2023-10-14

```r
# answer of part 3
sample_1 <- function(a, c) {
u<-runif(1)
inv_cdf<- c*u^(1/a)
return(inv_cdf)}

sample1 <-sample_1(a=5,c=10)

#answer of part 4
 a <- 5
c <- 10
n <- 300
icdf<- function(u){
  c * u^(1/a)
}
curve(icdf, from = 0, to = 1,n=300,main = "Inverse CDF Plot")
```



```r
cat("After looking at Cdf we can say that (0,6) be the range least probability being generated")
```

```
## After looking at Cdf we can say that (0,6) be the range least probability being generated
```

```r
#answer of part 5
sample2 <- function(a,c,n) {
u<-runif(n)
inv_cdf<- c*u^(1/a)
return(inv_cdf)}
sample_data<-sample2(a,c,1000)

#answer of part 6
empirical_mean<- mean(sample_data)
theoretical_mean<- a*c/(a+1)
cat("Empirical Mean:", empirical_mean, "\n")
```

```
## Empirical Mean: 8.337815
```

```r
cat("theoretical Mean:", theoretical_mean, "\n")
```

```
## theoretical Mean: 8.333333
```

```r
cat("We can observe that the empirical and theoretical mean are approximately equal")
```

```
## We can observe that the empirical and theoretical mean are approximately equal
```

Answer of Question 2

```r
m=160
v=160
sd=sqrt(v)
norm_height=rnorm(10000,m,sd)
pois_height=rpois(10000,m)
unif_height=runif(10000,m-sqrt(3*v),m+sqrt(3*v))
#answer of part 2
smpl_norm_height=sample(size=100,norm_height)
smpl_pois_height=sample(size=100,pois_height)
smpl_unif_height=sample(size=100,unif_height)

mean_smpl_norm_height=mean(smpl_norm_height)
mean_smpl_pois_height=mean(smpl_pois_height)
mean_smpl_unif_height=mean(smpl_unif_height)

#answer of part 3
sample_means_unif=c()
sample_means_pois=c()
sample_means_norm=c()

for (i in 1:1000) {
  unif_sample <- sample(size=100,unif_height)
```

```r
  sample_means_unif[i] <- mean(unif_sample)
  pois_sample <- sample(size=100,pois_height)
  sample_means_pois[i] <- mean(pois_sample)
  norm_sample <- sample(size=100,norm_height)
  sample_means_norm[i] <- mean(norm_sample)
}
#answer of part 4

mean_norm_height=mean(sample_means_norm)
mean_pois_height=mean(sample_means_pois)
mean_unif_height=mean(sample_means_unif)

se_unif <- sd(sample_means_unif)
se_pois <- sd(sample_means_pois)
se_norm <- sd(sample_means_norm)
cat("Population Mean of Uniform Data:", mean_unif_height, "\n")
```

```
## Population Mean of Uniform Data: 159.8839
```

```r
cat("Population Mean of Poisson Data:", mean_pois_height, "\n")
```

```
## Population Mean of Poisson Data: 160.0401
```

```r
cat("Population Mean of Normal Data:", mean_norm_height, "\n")
```

```
## Population Mean of Normal Data: 160.09
```

```r
cat("Standard Error of Uniform Data:", se_unif, "\n")
```

```
## Standard Error of Uniform Data: 1.274572
```

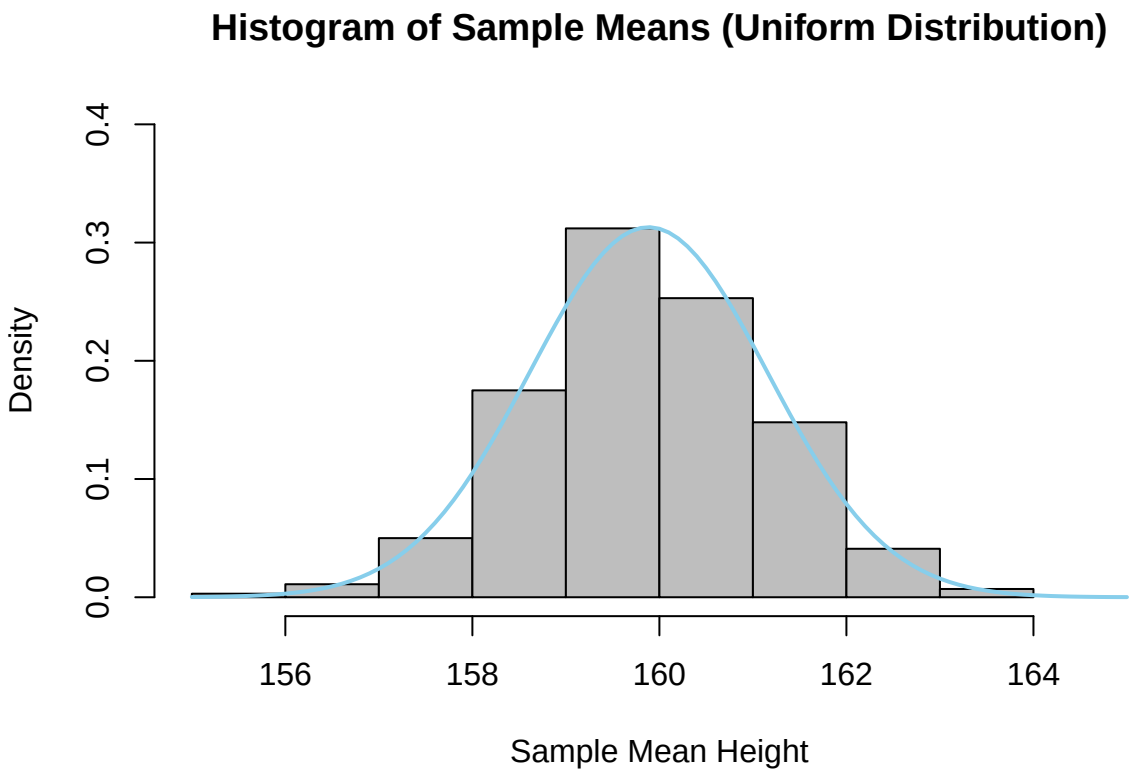
```r
cat("Standard Error of Poisson Data:", se_pois, "\n")
```

```
## Standard Error of Poisson Data: 1.262147
```

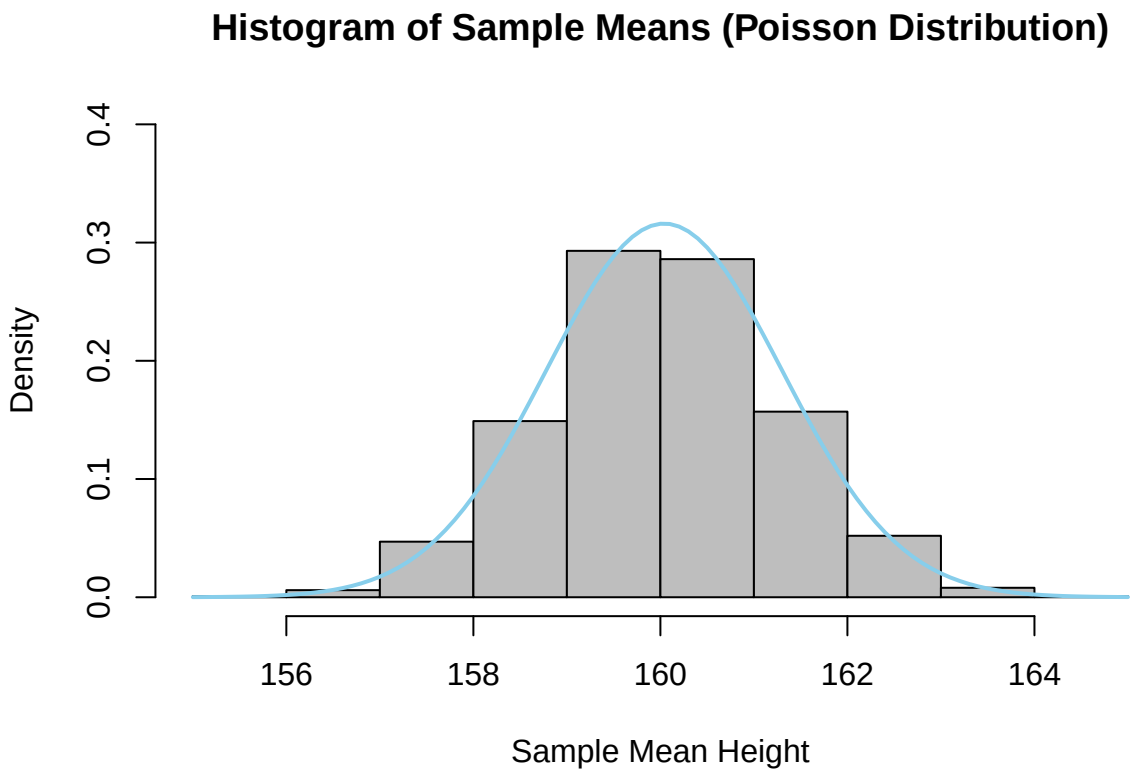
```r
cat("Standard Error of Normal Data:", se_norm, "\n")
```

```
## Standard Error of Normal Data: 1.191435
```

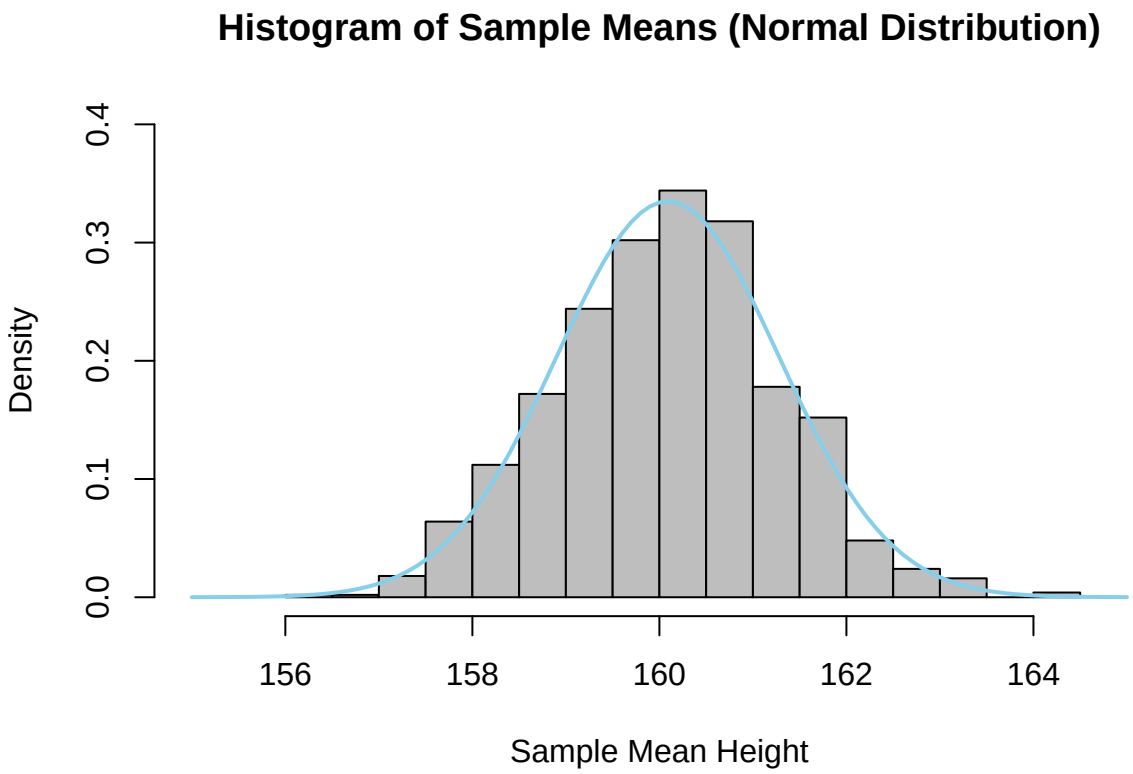
```r
#answer of part 5

hist(sample_means_unif, main = "Histogram of Sample Means (Uniform Distribution)", xlab = "Sample Mean H

curve(dnorm(x, mean =mean_unif_height, sd = se_unif), add = TRUE, col = "skyblue", lwd = 2)
```




```r
hist(sample_means_pois, main = "Histogram of Sample Means (Poisson Distribution)", xlab = "Sample Mean H

curve(dnorm(x, mean =mean_pois_height, sd = se_pois), add = TRUE, col = "skyblue", lwd = 2)
```

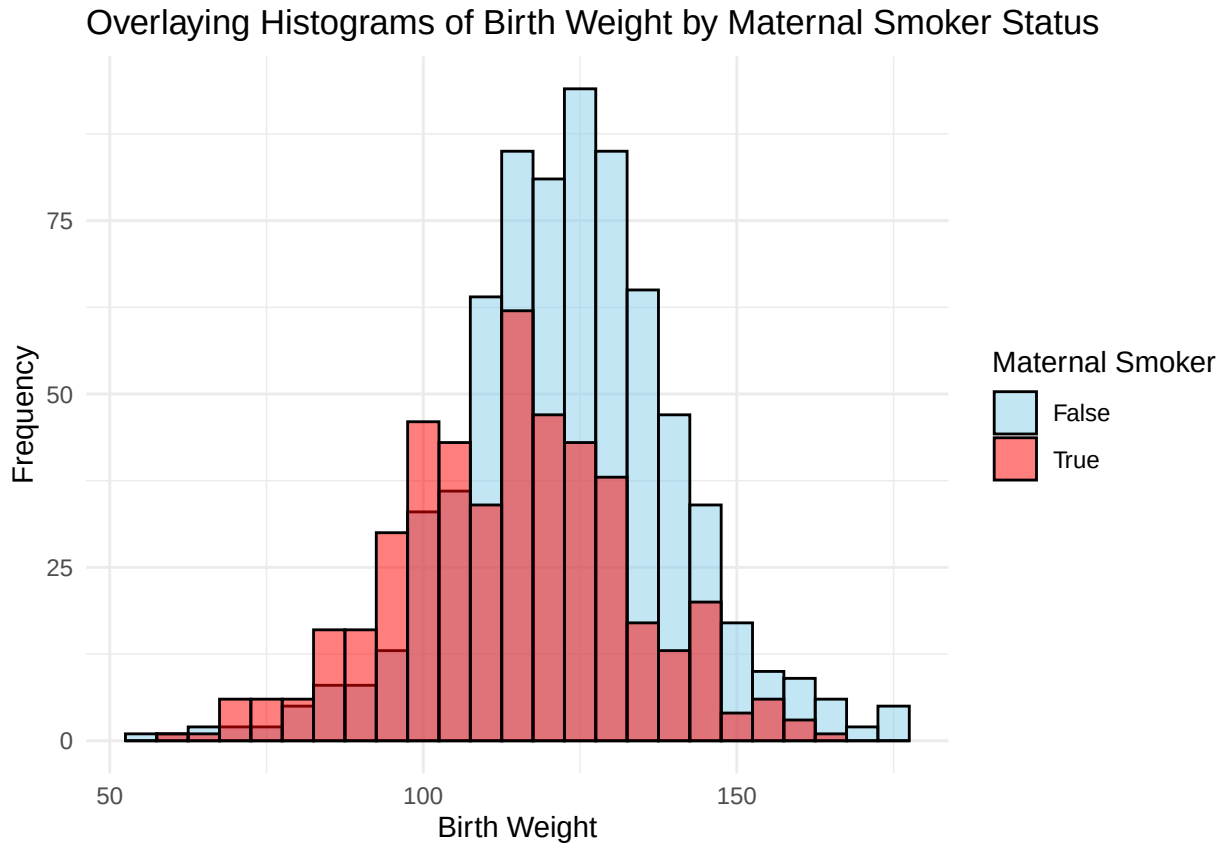



```r
hist(sample_means_norm, main = "Histogram of Sample Means (Normal Distribution)", xlab = "Sample Mean He

curve(dnorm(x, mean =mean_norm_height, sd = se_norm), add = TRUE, col = "skyblue", lwd = 2)
```


Histogram of Sample Means (Normal Distribution)
Density
0.0
0.1
0.2
0.3
0.4
156
158
160
162
164
Sample Mean Height

Answer of 3

```
#answer of part 1
library(ggplot2)
df=read.csv("C://Users//Bhuvnesh's PC//Downloads//baby.csv")
gg <- ggplot(df, aes(x = Birth.Weight, fill = Maternal.Smoker))

gg +
  geom_histogram(binwidth = 5, position = "identity", alpha = 0.5,col="black") +
  labs(
    title = "Overlaying Histograms of Birth Weight by Maternal Smoker Status",x = "Birth Weight",y = "Fr
  scale_fill_manual(values = c("False" = "skyblue", "True" = "red")) +
  theme_minimal()
```




```
result <- tapply(df$Birth.Weight, df$Maternal.Smoker, mean)
cat('Mean weight for Maternal Smoker(NO) :',result[1],'\n')
```

```
## Mean weight for Maternal Smoker(NO) : 123.0853
```

```
cat('Mean weight for Maternal Smoker(YES) :',result[2],'\n')
```

```
## Mean weight for Maternal Smoker(YES) : 113.8192
```

```
cat("It can be observed that the average weight of the child whose mother do not smoke is higher as comp
```

```
## It can be observed that the average weight of the child whose mother do not smoke is higher as compa
```

```
#Answer of part 2
result1 <- tapply(df$Birth.Weight, df$Maternal.Smoker, mean)
cat('Mean for Maternal Smoker(NO) :',result[1],'\n')
```

```
## Mean for Maternal Smoker(NO) : 123.0853
```

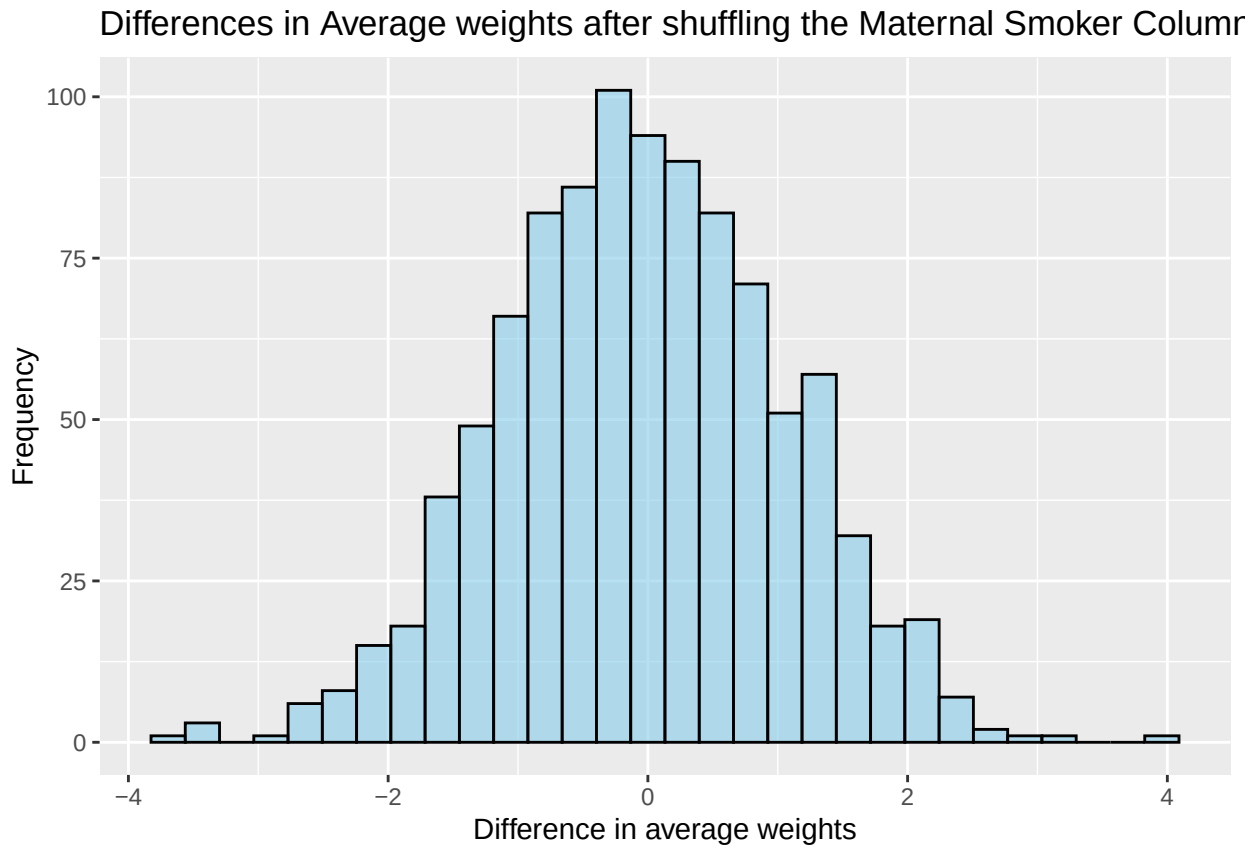
```
cat('Mean for Maternal Smoker(YES) :',result[2],'\n')
```

```
## Mean for Maternal Smoker(YES) : 113.8192
```

```
cat('The difference between the average weight of the smoking group and the average weight
of the non-smoking group :',result1[1]-result1[2])
```

```
## The difference between the average weight of the smoking group and the average weight
## of the non-smoking group : 9.266143
```

```
# Answer of part 3
df1=read.csv("C://Users//Bhuvnesh's PC//Downloads//baby.csv")
sim=1000
v=numeric(sim)
for (i in 1:1000){
  df1$Maternal.Smoker<-sample(df$Maternal.Smoker)
  result <-tapply(df1$Birth.Weight,df1$Maternal.Smoker, mean)
  v[i]=(result[1]-result[2])
}
df3=data.frame(v)
ggplot(df3,aes(x=v))+
  geom_histogram(bins = 30, fill = "skyblue", color = "black",alpha = 0.6)+
  labs(
    title = "Differences in Average weights after shuffling the Maternal Smoker Column",
    x = "Difference in average weights",
    y = "Frequency"
  )
```




```
cat('Mean of the differences:',mean(v),'\n')
```

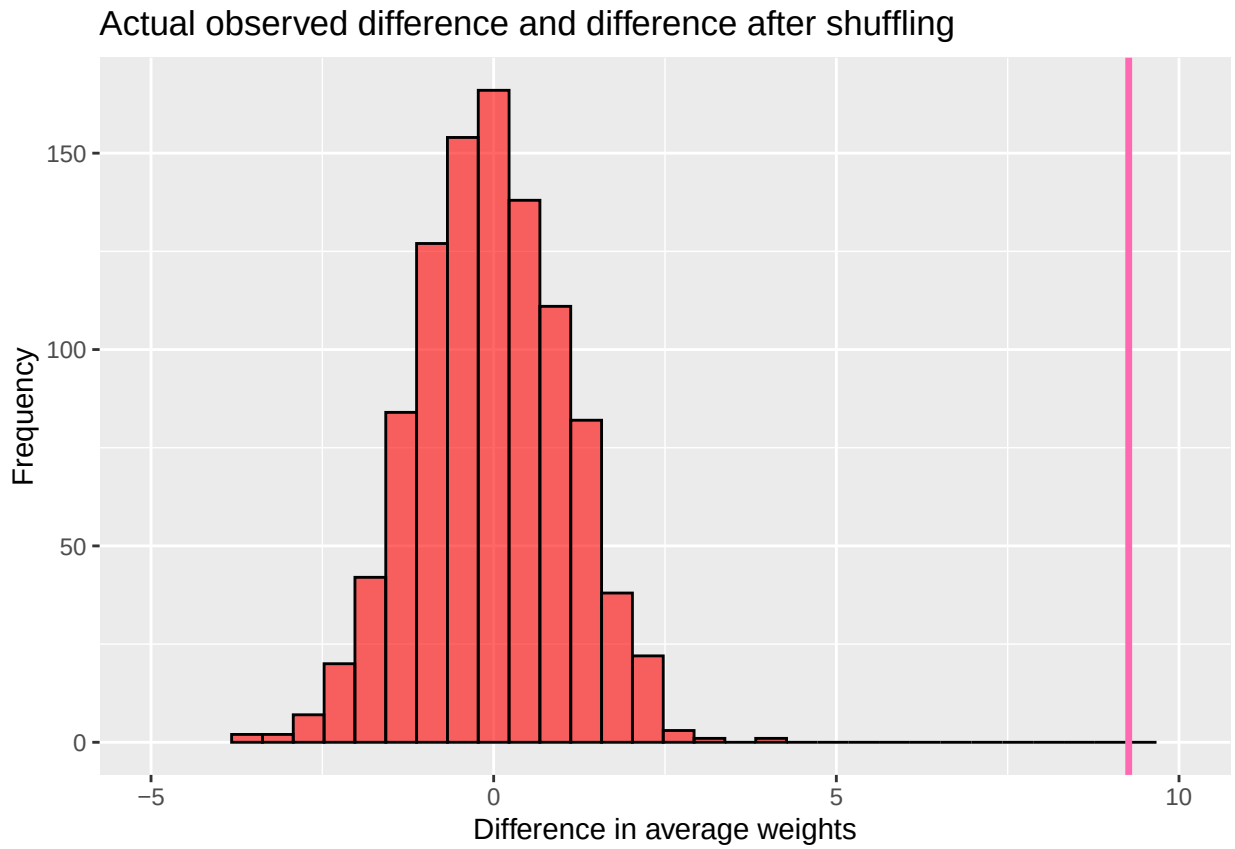
```
## Mean of the differences: -0.03883007
```

```
cat('Variance of the differences:',var(v),'\n')
```

```
## Variance of the differences: 1.145626
```

```
#Answer of part 4
ggplot(df3,aes(x=v))+
  geom_histogram(bins = 30, fill = "red", color = "black",alpha = 0.6)+
  geom_vline(xintercept=result1[1]-result1[2],col="hotpink", size = 1.2)+
  coord_cartesian(xlim=c(-5,10))+
  labs(
    title = "Actual observed difference and difference after shuffling",
    x = "Difference in average weights",
    y = "Frequency"
  )
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```




```
#Answer of part 5
obs_diff=result1[1]-result1[2]
me=mean(v)
va=var(v)
sd=sqrt(va)
cat('The actual observed difference(Before Shuffling):',obs_diff=result1[1]-result1[2],'\n')
```

```
## The actual observed difference(Before Shuffling): 9.266143
```

```
cat('Mean of the differences(After Shuffling):',mean(v),'\n')
```

```
## Mean of the differences(After Shuffling): -0.03883007
```

```
cat('Standard Deviation of the differences:',sd,'\n')
```

```
## Standard Deviation of the differences: 1.070339
```

```
cat('No of Standard Deviations  the observed difference is from the mean of the
simulated variables :',ceiling((obs_diff-me)/sd))
```

```
## No of Standard Deviations  the observed difference is from the mean of the
## simulated variables : 9
```

‘‘‘