# PS4_Markdown_file

José Maria de Mattamouros Resende Fonseca de Oliveira (22-602-783)
Linda Fiorina Odermatt (17-946-310) Elena Trevisani (22-620-603)
Vasily Zhuravlev (18-502-401)

2022-12-09

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(fastDummies)
library(ggplot2)


load("~/Desktop/MEcon/II Data Analytics/Tutorial_5/drugs.RData")
```

## Exercise 1 - Calculate the share of males who consume soft drugs

```
soft_drug_male_cons <- drugs[drugs$Gender == "male" & drugs$Soft_Drug == TRUE, ]
Gender<-drugs$Gender
Soft_Drug<-drugs$Soft_Drug
sd_male_share <- round(nrow(soft_drug_male_cons)/
                         nrow(drugs[drugs$Gender == "male",])*100, 2)

print(paste("The share of males who consume soft drugs is equivalent to",
            sd_male_share,"%"))
```

```
## [1] "The share of males who consume soft drugs is equivalent to 29.18 %"
```

## Exercise 2 - Share difference between male and female hard drug consumers

```
hard_drug_male_cons <- drugs[drugs$Gender == "male" & drugs$Hard_Drug == TRUE, ]

hd_male_share <- round(nrow(hard_drug_male_cons)/
                         nrow(drugs[drugs$Gender == "male",])*100, 2)

hard_drug_female_cons <- drugs[drugs$Gender == "female" & drugs$Hard_Drug == TRUE, ]

hd_female_share <- round(nrow(hard_drug_female_cons)/
                         nrow(drugs[drugs$Gender == "female",])*100, 2)

hd_dif <- abs(hd_male_share - hd_female_share)

gender_hd_consumption <- data.frame(hd_male_share, hd_female_share, hd_dif)

colnames(gender_hd_consumption) = c("Hard Drug Male Share",
                                    "Hard Drug Female Share",
                                    "Difference")

print(gender_hd_consumption)
```

```
##   Hard Drug Male Share Hard Drug Female Share Difference
## 1                15.82                  13.09       2.73
```

```
print(paste(
  "The share difference in hard drug consumption between males and females is",
  hd_dif, "%"))
```

```
## [1] "The share difference in hard drug consumption between males and females is 2.73 %"
```

## Exercise 3 - Share of soft drug users by age group, focus on young adults

## we verify that only 3 age categories are present and store them in age_group

```
library(dplyr)


age_group=matrix(levels(factor(drugs$Age, order=TRUE)))

#Create an empty matrix
age_matrix = matrix(NA, nrow=length(age_group),
                    ncol=1)
```

```
rownames(age_matrix) <- c(age_group[1:length(age_group),])
colnames(age_matrix) <- c("Share of Soft Drug Users")
#Quick Test
nrow(drugs[drugs$Age == age_group[1,] & drugs$Soft_Drug == TRUE, ])
```

```
## [1] 3972
```

```
nrow(drugs[drugs$Age == age_group[1,] ,])
```

```
## [1] 8190
```

```
#For loop to store all the values
for (i in c(1:length(age_group))){

  # Calculate the share
  soft_drug_cons <- drugs[drugs$Age == age_group[i,] &
                            drugs$Soft_Drug == TRUE, ]

  sd_share <- round((nrow(soft_drug_cons)/
                       nrow(drugs[drugs$Age == age_group[i,] , ]))*100, 2)

  # store values in a matrix
  age_matrix[i,] <- sd_share

}

print(age_matrix)
```

```
##             Share of Soft Drug Users
## 16-17 years                     48.5
## 18-19 years                      0.0
## 20-24 years                      0.0
```

```
#Quick check
nrow(drugs[drugs$Age == age_group[1,] & drugs$Soft_Drug == TRUE, ])
```

```
## [1] 3972
```

```
nrow(drugs[drugs$Soft_Drug == TRUE, ])
```

```
## [1] 3972
```

Indeed we observe a strict abandonment of soft drug consumption as age increases. At least as young adults get over 17 years, they don't declare if they consume soft drugs. Could it be that young adults aged 18+ hide their soft drug consumption?

# Exercise 4 - Chi squared test

## Part 1 - Earnings and Soft Drug Consumption Table

```
income_group=matrix(levels(factor(drugs$Earning, order = TRUE)))
#the data is not ordered...

#Therefore we need to order manually
income_group_ordered=matrix(c("<1k USD","1-5k USD",
                              "5-10k USD", ">10k USD"))

#Create an empty matrix
income_matrix = matrix(NA, nrow=length(income_group),
                    ncol=2)
rownames(income_matrix) <- c("<1k USD","1-5k USD", "5-10k USD", ">10k USD")
colnames(income_matrix) <- c("TRUE","FALSE")
#Quick Test
nrow(drugs[drugs$Earning == income_group_ordered[1,] & drugs$Soft_Drug == TRUE, ])
```

```
## [1] 2478
```

```
nrow(drugs[drugs$Earning == income_group_ordered[1,] ,])
```

```
## [1] 6259
```

```
#For loop to store all the values
for (i in c(1:length(income_group))){

  # Calculate the share
  soft_drug_cons <- drugs[drugs$Earning == income_group_ordered[i,]
                          & drugs$Soft_Drug == TRUE, ]

  sd_share <- round((nrow(soft_drug_cons)/
                       nrow(drugs[drugs$Earning == income_group_ordered[i,]
                                  , ]))*100, 2)

  # store values in a matrix
  income_matrix[i,1] <- sd_share
  income_matrix[i,2] <- 100-sd_share
}

print(income_matrix)
```

```
##             TRUE FALSE
## <1k USD    39.59 60.41
## 1-5k USD   40.39 59.61
## 5-10k USD  38.95 61.05
## >10k USD   39.23 60.77
```

### Part 2 - Calculate Chi-Squared

```
chisq.test(x=drugs$Earning, y=drugs$Soft_Drug)
```

```
##
##  Pearson's Chi-squared test
##
## data:  drugs$Earning and drugs$Soft_Drug
## X-squared = 9.4014, df = 3, p-value = 0.0244
```

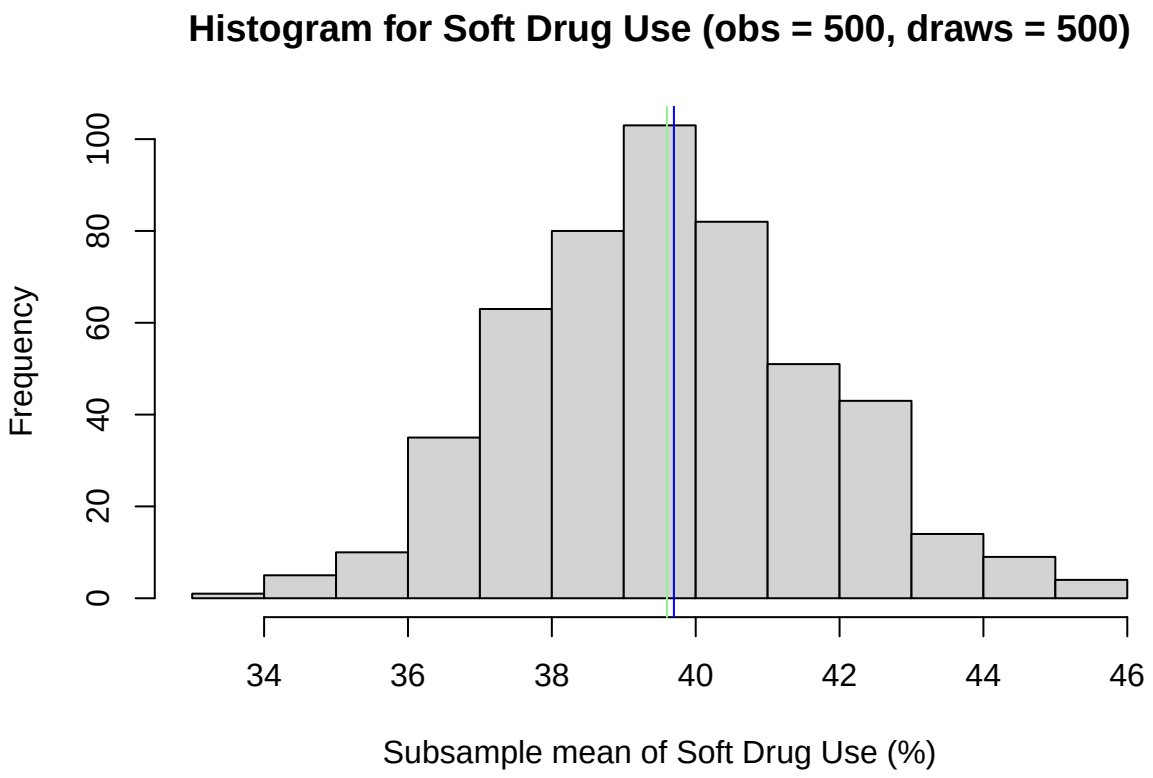
We cannot reject the independence hypothesis as the p-value is lower than the 5% threshold.

## Exercise 5 - Random sample with multiple draws

```
obs_size=500

multiple_drugs_samples <- matrix(NA, nrow=obs_size, ncol=1)

sample_numb=500

#First we convert the string variables to numeric (dummy)
boolean_sd = as.logical(drugs$Soft_Drug)
dummy_sd = as.numeric(boolean_sd)

set.seed(2122020) # for results replication

#For loop to generate  a random sample with 500 observations, 500 times
for (i in 1:sample_numb) {

  multiple_drugs_samples[i,] = mean(sample(dummy_sd,
                                           size=obs_size, replace=TRUE))*100
  }

head(multiple_drugs_samples, 5) # Test
```

```
##      [,1]
## [1,] 40.4
## [2,] 39.0
## [3,] 42.2
## [4,] 41.2
## [5,] 42.6
```

```
hist(multiple_drugs_samples, probability = FALSE,
     main = "Histogram for Soft Drug Use (obs = 500, draws = 500)",
     xlab="Subsample mean of Soft Drug Use (%)")
abline(v=mean(multiple_drugs_samples), col="blue")
abline(v=median(multiple_drugs_samples), col="lightgreen")
```



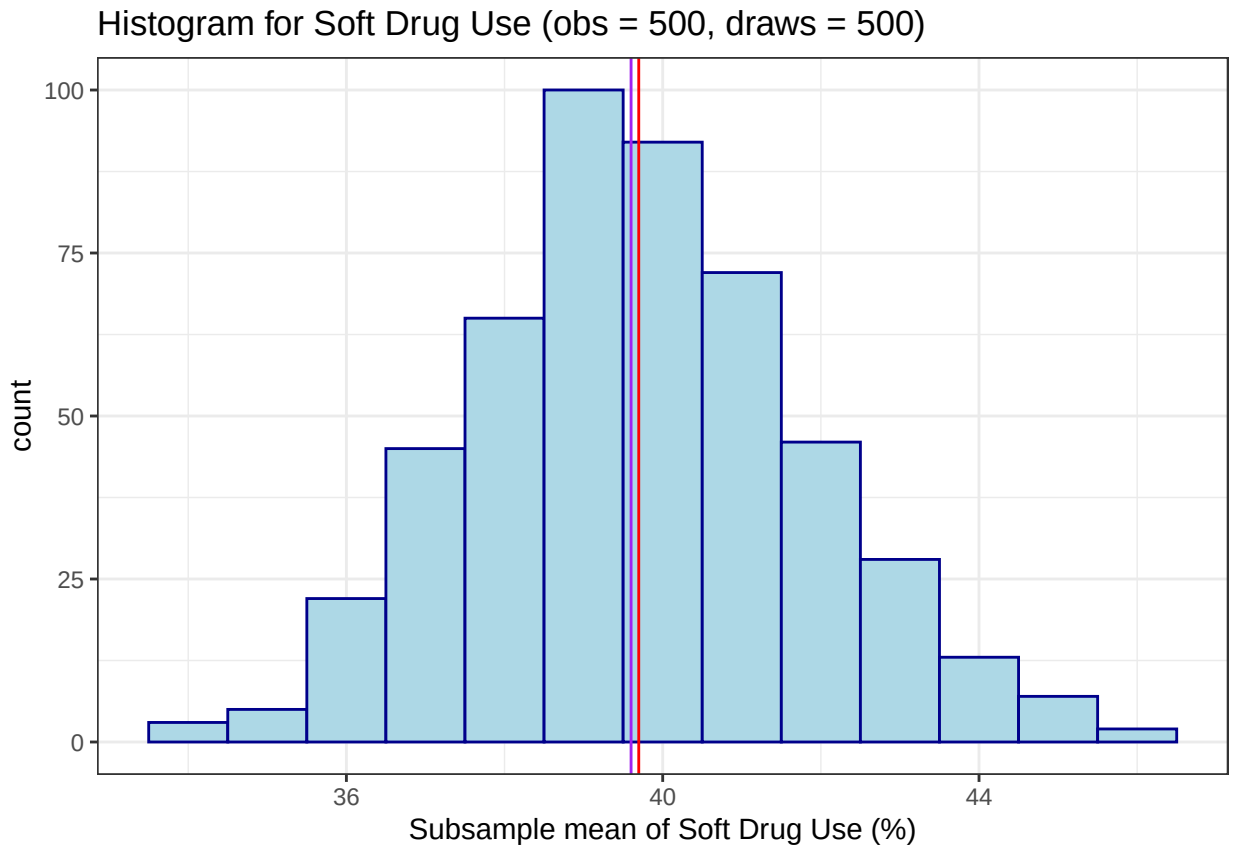

```
round(mean(multiple_drugs_samples),2) #subsample mean from 500 draws
```

```
## [1] 39.7
```

```
round(mean(dummy_sd)*100,2) #population mean
```

```
## [1] 39.72
```

```
# We observe that the recorded subsample means are very close to the average
# soft drug consumption in the full sample. This corresponds to the Central
# Limit Theory Hypothesis.

# Alternative with ggplot for aesthetics

# Convert matrix to data frame
df <- as.data.frame(multiple_drugs_samples)

#Plot
ggplot(data=df, aes(x=V1)) +
  geom_histogram(binwidth= 1 , color = "darkblue", fill = "lightblue") +
  ggtitle("Histogram for Soft Drug Use (obs = 500, draws = 500)")+
  xlab("Subsample mean of Soft Drug Use (%)") +
  geom_vline(xintercept=mean(df$V1), col="red") +
  geom_vline(xintercept=median(df$V1), col="purple") + theme_bw()
```




## Exercise 6 - Comparing random samples

## Part 1. Creating a function to generate random samples and print mean of each

## subsample

```
set.seed(2122020) # for replication of results

subsample_mean=function(data, variable, sample.size, ndraws)
{

  subsample_mean= matrix(NA, nrow=ndraws, ncol=1)

# Transform character variables into boolean
if (class(variable) == "character"){
  boolean = as.logical(variable)
  dummy = as.numeric(boolean)

  # Calculate the mean of each subsample with for loop
  for (i in 1:ndraws) {
    subsample_mean[i,] = mean(sample(dummy, sample.size, replace=TRUE))*100
```

```
  }
}
else if (class(variable) == "logical"){
  dummy = as.numeric(variable)
  for (i in 1:ndraws) {
    subsample_mean[i,] = mean(sample(dummy, sample.size, replace=TRUE))*100
  }
}
else{for (i in 1:ndraws) {
  subsample_mean[i,] = mean(sample(variable, sample.size, replace=TRUE))*100
  }}

  return(subsample_mean)
}

#Testing the function

mean(subsample_mean(drugs, drugs$Hard_Drug, 1500, 100))
```

```
## [1] 15.27467
```

```
mean(subsample_mean(drugs, drugs$Soft_Drug, 500, 1000))
```

```
## [1] 39.5684
```

```
# Part 2. Run the function for different sample sizes and number of draws

N_runs = c(100, 500, 2500)
sample_sizes=  c(100, 500, 2500)

mx_samples = matrix(data=NA, nrow=length(sample_sizes), ncol=length(N_runs))
rownames(mx_samples) = sample_sizes
colnames(mx_samples) = N_runs

set.seed(2122020) # for replication of results
for (j in 1:length(N_runs)){
  for (i in 1:length(sample_sizes)){
    mx_samples[i,j] <- round(mean(subsample_mean(drugs, drugs$Soft_Drug,
                          sample_sizes[i], N_runs[j])),2)
  }
}

#Print results
print(mx_samples)
```

```
##        100   500  2500
## 100  40.18 39.80 39.60
## 500  39.46 39.51 39.77
## 2500 39.64 39.70 39.73
```

## Part 3. Demonstrating CLT with graphs

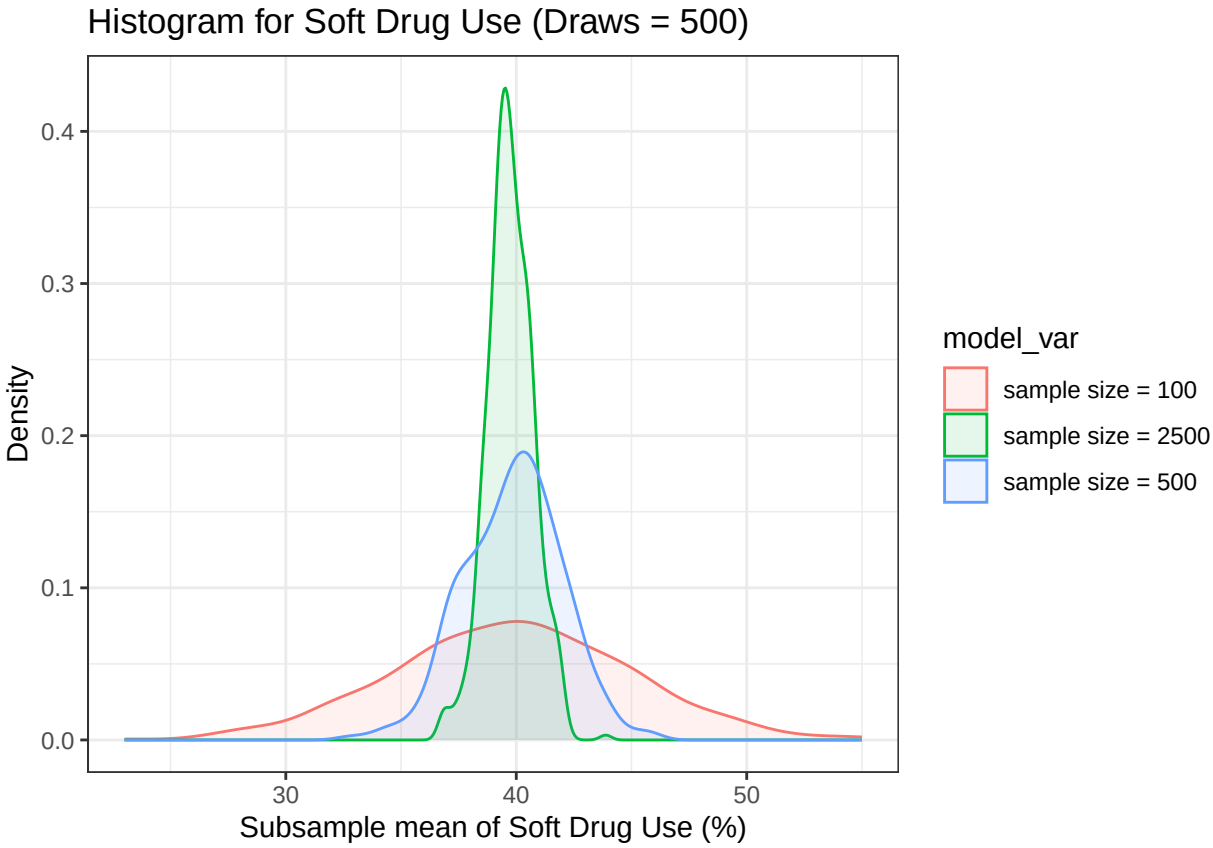
```
# Create matrix (ndraws=500, for 3 different subsample sizes )

mx_samples_500ndraws <- matrix(data=NA, nrow=500,
                               ncol=length(sample_sizes))
colnames(mx_samples_500ndraws) <-sample_sizes

for (i in 1:length(sample_sizes)){
    mx_samples_500ndraws[,i] <-subsample_mean(drugs, drugs$Soft_Drug,
                                              sample_sizes[i], 500)
  }
rep(colnames(mx_samples_500ndraws))
```

```
## [1] "100"  "500"  "2500"
```

```
length(c(mx_samples_500ndraws))
```

```
## [1] 1500
```

```
# Convert matrix to data frame for plotting

model_var <- c(rep("sample size = 100",nrow(mx_samples_500ndraws)),
               rep("sample size = 500",nrow(mx_samples_500ndraws)),
               rep("sample size = 2500",nrow(mx_samples_500ndraws)))

hist.plot <- data.frame("Share of Soft Drug Users (%)" =
                          c(mx_samples_500ndraws),
                        "models" = model_var)

# Plot Histogram
ggplot(data=hist.plot, aes(x=c(mx_samples_500ndraws), col=model_var,
                           fill=model_var)) +
  ggtitle("Histogram for Soft Drug Use (Draws = 500)")+
  xlab("Subsample mean of Soft Drug Use (%)") +
  ylab("Density")+
  geom_density(alpha=0.1) +
  theme_bw()
```




We observe that as we increase the sample size, the distribution of averages is more normal and we have get variance (i.e., the line gets narrower). This goes in line with the Central Limit Theorem for randomly distributed data. Asymptotically, the subsample averages tend to the actual mean of the full data.

```
# Create matrix (ndraws=variables, for subsample size=500 )

mx_samples_500sample <- matrix(data=NA, nrow=sum(N_runs), ncol=2)

#store N_runs values for each row
mx_samples_500sample[(1:100),2] <- "100"
mx_samples_500sample[(101:600),2] <- "500"
mx_samples_500sample[(601:3100),2] <- "2500"

x1 <- subsample_mean(drugs, drugs$Soft_Drug, 500, 100)
x2 <- subsample_mean(drugs, drugs$Soft_Drug, 500, 500)
x3 <- subsample_mean(drugs, drugs$Soft_Drug, 500, 2500)

#convert into 1 column matrix with sum(N_runs) rows

mx_samples_500sample[,1] <- as.numeric(cbind(c(x1,x2,x3)))
mx_samples_500sample[,2]
```

```
##    [1] "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"
##   [11] "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"
##   [21] "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"
##   [31] "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"
##   [41] "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"
```

```
##   [51] "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"
##   [61] "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"
##   [71] "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"
##   [81] "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"
##   [91] "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"  "100"
##  [101] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [111] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [121] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [131] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [141] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [151] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [161] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [171] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [181] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [191] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [201] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [211] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [221] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [231] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [241] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [251] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [261] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [271] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [281] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [291] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [301] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [311] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [321] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [331] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [341] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [351] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [361] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [371] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [381] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [391] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [401] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [411] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [421] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [431] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [441] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [451] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [461] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [471] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [481] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [491] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [501] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [511] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [521] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [531] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [541] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [551] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [561] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [571] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [581] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
```

```
##  [591] "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"  "500"
##  [601] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [611] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [621] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [631] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [641] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [651] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [661] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [671] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [681] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [691] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [701] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [711] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [721] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [731] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [741] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [751] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [761] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [771] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [781] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [791] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [801] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [811] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [821] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [831] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [841] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [851] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [861] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [871] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [881] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [891] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [901] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [911] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [921] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [931] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [941] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [951] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [961] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [971] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [981] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
##  [991] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1001] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1011] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1021] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1031] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1041] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1051] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1061] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1071] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1081] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1091] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1101] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1111] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1121] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
```

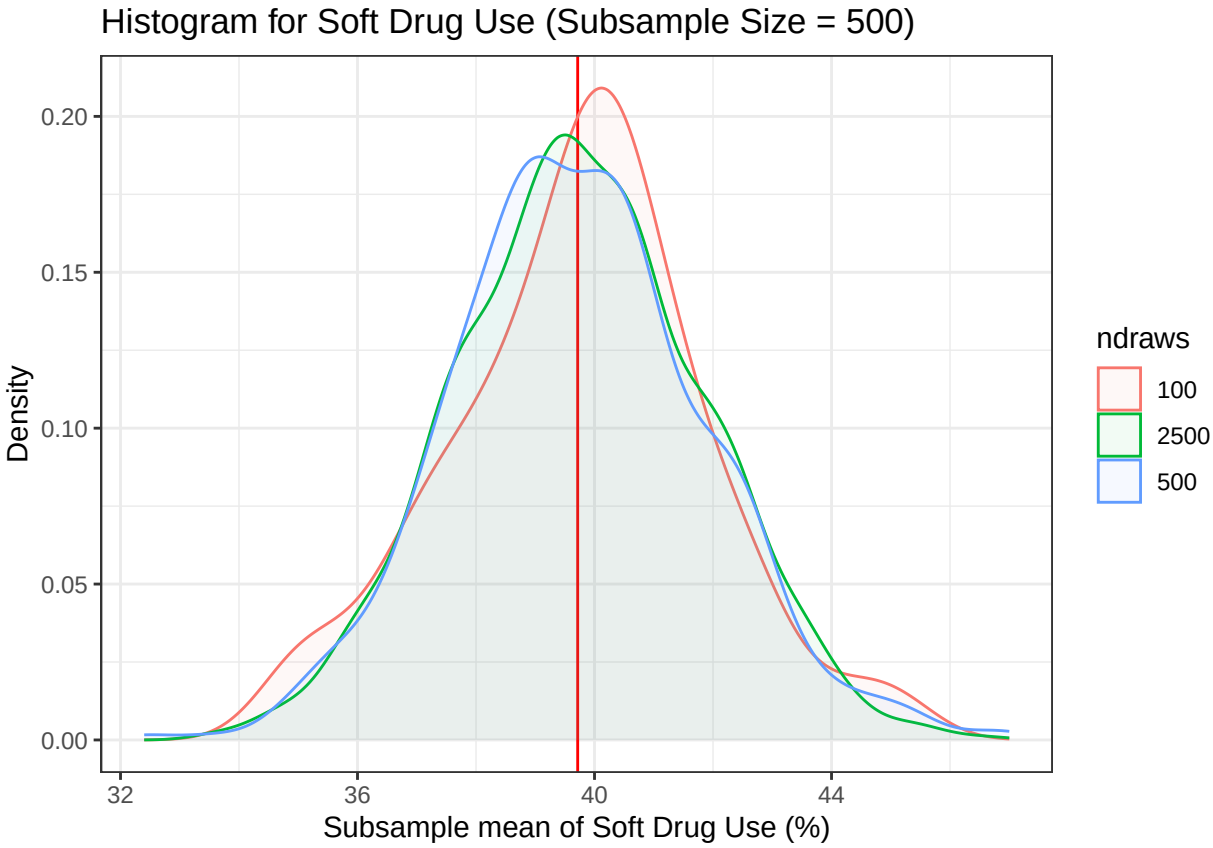
```
## [1131] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1141] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1151] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1161] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1171] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1181] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1191] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1201] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1211] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1221] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1231] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1241] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1251] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1261] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1271] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1281] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1291] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1301] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1311] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1321] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1331] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1341] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1351] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1361] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1371] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1381] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1391] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1401] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1411] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1421] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1431] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1441] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1451] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1461] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1471] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1481] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1491] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1501] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1511] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1521] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1531] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1541] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1551] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1561] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1571] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1581] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1591] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1601] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1611] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1621] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1631] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1641] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1651] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1661] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
```

```
## [1671] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1681] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1691] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1701] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1711] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1721] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1731] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1741] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1751] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1761] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1771] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1781] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1791] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1801] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1811] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1821] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1831] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1841] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1851] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1861] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1871] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1881] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1891] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1901] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1911] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1921] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1931] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1941] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1951] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1961] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1971] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1981] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [1991] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2001] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2011] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2021] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2031] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2041] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2051] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2061] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2071] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2081] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2091] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2101] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2111] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2121] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2131] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2141] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2151] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2161] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2171] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2181] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2191] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2201] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
```

```
## [2211] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2221] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2231] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2241] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2251] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2261] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2271] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2281] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2291] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2301] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2311] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2321] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2331] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2341] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2351] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2361] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2371] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2381] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2391] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2401] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2411] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2421] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2431] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2441] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2451] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2461] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2471] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2481] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2491] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2501] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2511] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2521] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2531] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2541] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2551] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2561] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2571] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2581] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2591] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2601] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2611] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2621] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2631] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2641] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2651] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2661] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2671] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2681] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2691] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2701] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2711] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2721] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2731] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2741] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
```

```
## [2751] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2761] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2771] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2781] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2791] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2801] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2811] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2821] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2831] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2841] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2851] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2861] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2871] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2881] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2891] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2901] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2911] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2921] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2931] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2941] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2951] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2961] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2971] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2981] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [2991] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [3001] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [3011] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [3021] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [3031] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [3041] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [3051] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [3061] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [3071] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [3081] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
## [3091] "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500" "2500"
```

```
# Convert to data frame for plotting
df_ndraws_500samples <- as.data.frame(x=mx_samples_500sample)
subsample_m <- as.numeric(df_ndraws_500samples$V1)
ndraws <- df_ndraws_500samples$V2

# Plot Histogram
ggplot(data=df_ndraws_500samples, aes(x=subsample_m,
                                      col=ndraws ,
                                      fill=ndraws)) +
  ggtitle("Histogram for Soft Drug Use (Subsample Size = 500)")+
  xlab("Subsample mean of Soft Drug Use (%)") +
  ylab("Density")+
  geom_vline(xintercept=(mean(dummy_sd)*100), col="red")+
  geom_density(alpha=0.05) +
  theme_bw()
```




We observe that as we increase the number of draws for a fixed sample, the distribution of subsample averages remains mostly consistent (i.e., no evident skewness). Since the draws are random, we don't observe bias in the mean (i.e., it's remains consistent for all number of draws). Asymptotically, as the number of draws tends to infinity, the mean distribution remains consistent if the estimator is unbiased. This demonstrates that for even small number of draws with random subsamples, one should be able to obtain the unbiased estimator.