```
##
## Orysya Stus
## Assignment 3
##


##
## Problem 1
##
# 1. You need to read chol.txt file for this problem.
# Use \apply" function to calculate the overall mean for each variable,
except for the sex variable.
# Use \aggregate" to calculate the sex-specific mean for each variable.
prob1 =
read.table("C:/Users/Orysya/Desktop/Introduction_to_R_Programming/Assignm
ent3_W8/chol.txt", header=T)
prob1
apply(prob1[2:12],2, mean, na.rm = T)
sum_by_sex = aggregate(prob1[2:12], by = list(prob1$sex), FUN = mean,
na.rm = T)
sum_by_sex
# 2. Create a new variable, chol2, which is based on the chol variable
from the chol data frame. If chol is greater than the mean of chol, then
chol2 = `Hi'; otherwise, chol2= `LOW'
# Use \tapply" function to calculate the standard deviation of bmi for
each chol2 category.
# Use \tapply" function to calculate the standard deviation of bmi for
each combination of sex and chol2 categories.
A = apply(prob1[3], 2, mean, na.rm=T)
A
chol2 <- ifelse(prob1$chol > A, "Hi", "Low")
chol2
prob1["chol2"] <- chol2
prob1
B = tapply(prob1$bmi, prob1$chol2, sd)
B
C = tapply(prob1$bmi, list(prob1$sex, prob1$chol2), sd)
C

##
## Problem 2
##
# Consider the simulated matrix, mat, with 10 rows and 20 columns.
set.seed(91765)
mat = matrix(rnorm(200), 10)
mat[1,1] = NA
dim(mat)
# 1. Calculate the median (use the median function) of each row by using
the for loop
medians_of_mat <- numeric()
for(i in 1:nrow(mat)) {
  medians_of_mat[i] <- median(mat[i, ])
```

```
}
medians_of_mat
# 2. Calculate the median of each row by using the apply function
apply(mat, 1, median)

##
## Problem 3
##
#You need to simulate a matrix with dimension of 1000 by 20. For example,
this matrix will look like the following:
gene <- matrix(rnorm(20000), ncol=20)
gene[1:4,] <- gene[1:4,] + 4
rownames(gene) <- rownames(gene, do.NULL = FALSE, prefix = "G")
rownames(gene)
gene
pheno <- c("case", "case", "case", "case", "case", "case", "case", "case",
"case", "case", "control", "control", "control", "control", "control",
"control", "control", "control", "control", "control")
pheno

gene_info <- gene
pheno_info <- pheno
p_values <- apply(gene_info, 1, function(gene_info) {
  t.test(gene_info ~ pheno_info)$p.value})
p_values

##
## Problem 4
##
#Calculate the two-sample t-test for each numerical variable in the chol
data frame (from the chol.txt le) by the sex variable by using the apply
function.
#The result should be a matrix that contains five columns: F.mean (mean
of the numerical variable for Female), M.mean (mean of the numerical
variable
#for Male), t (for t-statistics), df (for degrees of freedom), and p (for
p-value). The result should look like the one below
prob4 =
read.table("C:/Users/Orysya/Desktop/Introduction_to_R_Programming/Assignm
ent3_W8/chol.txt", header=T)
prob <- prob4[2:11]
aggregate(prob4[,2:11], list(prob4$sex), mean, na.rm=T)
df <- apply(prob, 1, function(prob) {
  t.test(prob ~ prob4$sex)$df})
df
p <- apply(prob, 1, function(prob) {
  t.test(prob ~ prob4$sex)$p.value})
p
```

You need to write a function for this problem.

Points: -7

In complete

Points: -8