

Assignment_2

Henry Lewis 20202851

28/10/2021

Question 1:

```
# Read the txt to a variable data with first row containing column names
data <- read.table("s50_1995.txt", header = TRUE, sep = '')
```

```
# Add the factors to the alcohol variable
alcohol <- factor(data$alcohol, labels= c("Dont_drink", "Drink_bi_anually","Drink_monthly",
"Dont_weekly", "Drink_more_than_weekly"), levels= c(1,2,3,4,5), ordered = TRUE)

# Add the factors to the drugs variable
drugs <- factor(data$drugs, labels= c("Dont_take_drugs", "Drugs_tried_once","Drugs_tried_occasionally",
"Regular_drug_user"), levels= c(1,2,3,4), ordered = TRUE)

# Add the factors to the smoke variable
smoke <- factor(data$smoke, labels = c("Dont_smoke", "Smoke_occasionally","smoke_regular"), levels= c(1,2,3), ordered = TRUE)

# Add the factors to the sport variable
sport <- factor(data$sport, labels = c("Dont_Play_regularly", "Play_sport_regularly"), levels = c(1,2), ordered = TRUE)

# Create a dataframe from the factored columns
df_data <- data.frame(alcohol,drugs, smoke, sport)

#Levels(df_data$smoke)
#sapply(df_data,levels)

# spercify the dataset Structure
dataset_structure <- str(df_data)
```

```
## 'data.frame':    50 obs. of  4 variables:
## $ alcohol: Ord.factor w/ 5 levels "Dont_drink"<"Drink_bi_anually"<...: 3 2 2 2 3 4 4 4 2 4 ...
## $ drugs : Ord.factor w/ 4 levels "Dont_take_drugs"<...: 1 2 1 1 1 1 3 3 1 1 ...
## $ smoke : Ord.factor w/ 3 levels "Dont_smoke"<"Smoke_occasionally"<...: 2 3 1 1 1 1 1 3 1 1 ...
## $ sport : Ord.factor w/ 2 levels "Dont_Play_regularly"<...: 2 1 1 2 2 2 1 2 2 2 ...
```

The structure of the data shows that the object is a dataframe and states that there are 50 observations of 4 variables. Also displayed is the number of levels for each variable

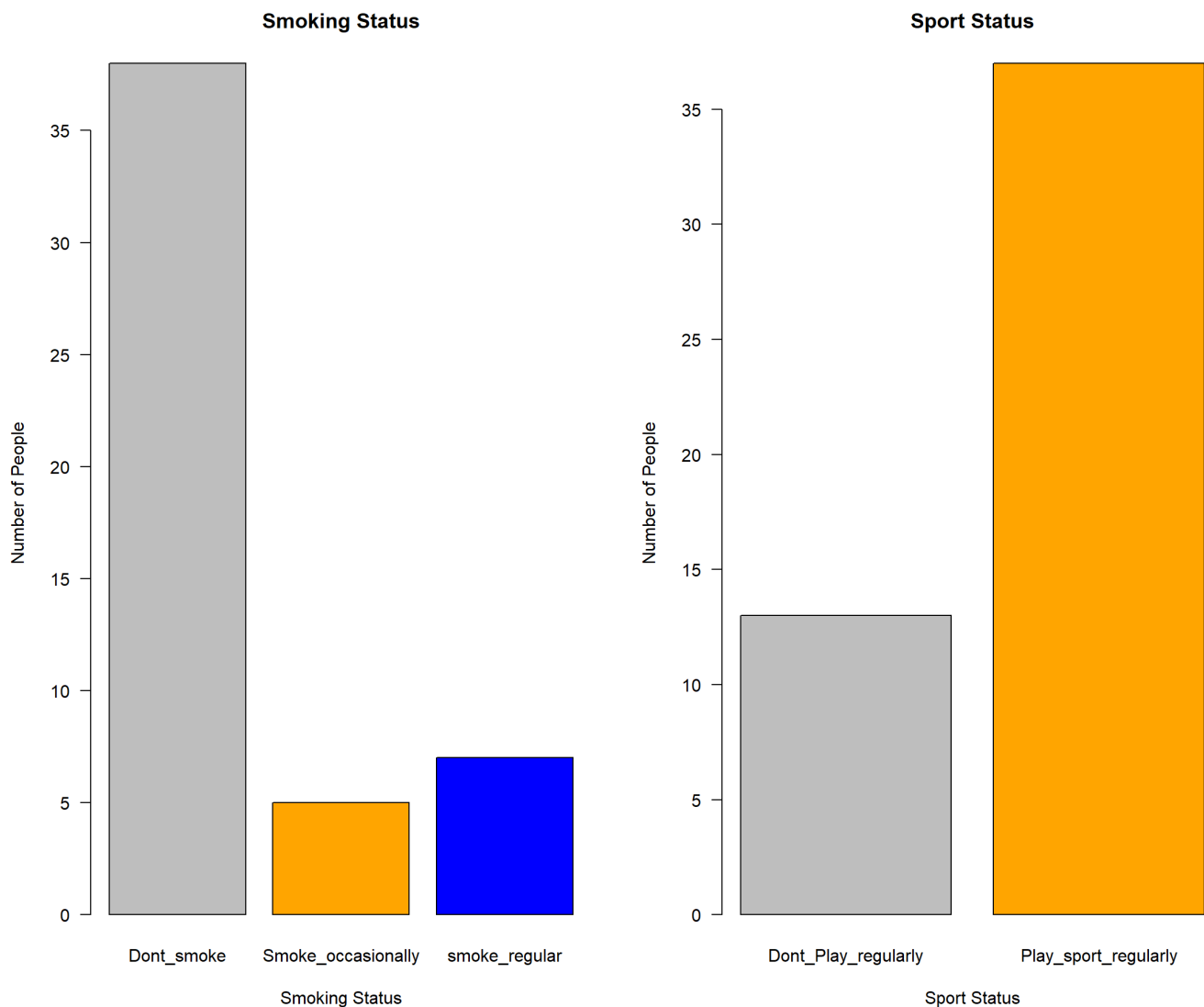
Question 2

```
# sets up a row with 2 columns for the plots
par(mfrow = c(1, 2))

# set out the colours that will be used in the barcharts
colors = c("gray", "orange", "blue")

barplot(height = table(smoke), main = "Smoking Status", xlab = "Smoking Status", ylab = "Number of People",
        col = colors, las = 1)

barplot(height = table(sport), main = "Sport Status", xlab = "Sport Status", ylab = "Number of People",
        col = colors, las = 1)
```



The two bar charts illustrate the Smoke (left) and sports (right) variables. As can be seen from the bar chart representing smoke the majority of those in the school do not smoke with less than 10 students smoking occasionally or smoking regularly. The chart representing Sport indicates that the majority of the students play sport regularly, with fewer not playing regularly.

Question 3

```
# What is the proportion of pupils who smoke at least occasionally?

v1 <- prop.table(table(df_data$smoke))
smoke_atleast_occasionally <- v1[2]+v1[3]

# What is the proportion of pupils who regularly practiced sport and smoke at least occasionally?

v2 <- prop.table(table(df_data$smoke, df_data$sport))
sport_smoke_atleast_occasionally <- v2[2,2]+v2[3,2]
```

The proportion of pupils who smoke at least occasionally is 24%, whilst the the proportion of pupils who regularly practiced sport and smoke at least occasionally is 18%

Question 4

```
# write a summary method that will show the proportion of students for every level of each variable.

list_name <- as.list(df_data)
class(list_name) <- 's50survey'

summary.s50survey <- function(input){
  cat('Alcohol','\n')
  print(prop.table(table(input$alcohol)))
  cat('Smoke','\n')
  print(prop.table(table(input$smoke)))
  cat('Drugs','\n')
  print(prop.table(table(input$smoke)))
  cat('Sport','\n')
  print(prop.table(table(input$sport)))
}
summary.s50survey(list_name)
```

```
## Alcohol
##
##          Dont_drink      Drink_bi_anually      Drink_monthly
##              0.10              0.32              0.24
##      Drink_weekly Drink_more_than_weekly
##              0.28              0.06
## Smoke
##
##      Dont_smoke Smoke_occasionally      smoke_regular
##              0.76              0.10              0.14
## Drugs
##
##      Dont_smoke Smoke_occasionally      smoke_regular
##              0.76              0.10              0.14
## Sport
##
##      Dont_Play_regularly Play_sport_regularly
##              0.26              0.74
```

The summary data expressed as a proportion for each level of each variable for the 1995 dataset is set out above

Question 5

```
# What is the proportion of pupils who did not use cannabis?  
drug_usage <- prop.table(table(df_data$drugs))  
no_cannibis <- drug_usage[1]*100
```

The percentage of pupils that don't use cannabis is 72%.

Question 6

```
# Read the txt to a variable data_1997 with first row containing column names  
data_1997 <- read.table("s50_1997.txt", header = T, sep = '')  
  
# Add the factors to the alcohol variable for 1997 updated data  
alcohol_1997 <- factor(data_1997$alcohol, labels= c("Dont_drink", "Drink_bi_anually", "Drink_m  
onthly", "Drink_weekly", "Drink_more_than_weekly"), levels= c(1,2,3,4,5), ordered = TRUE)  
  
# Add the factors to the drugs variable for the 1997 dataset  
drugs_1997 <- factor(data_1997$drugs, labels= c("Dont_take_drugs", "Drugs_tried_once", "Drugs_  
tried_occasionally", "Regular_drug_user"), levels= c(1,2,3,4), ordered = TRUE)  
  
# Add the factors to the smoke variable for the 1997 dataset  
smoke_1997 <- factor(data_1997$smoke, labels = c("Dont_smoke", "Smoke_occasionally", "smoke_re  
gular"), levels= c(1,2,3), ordered = TRUE)  
  
# Add the factors to the sport variable for the 1997 dataset  
sport_1997 <- factor(data_1997$sport, labels = c("Dont_Play_regularly", "Play_sport_regularl  
y"), levels= c(1,2), ordered = TRUE)  
  
# Create a dataframe from the factored columns for the updated 1997 dataset  
df_1997 <- data.frame(alcohol_1997, drugs_1997, smoke_1997, sport_1997)  
  
# dataset Structure  
#str(df_1997)  
  
# convert new dataset as a list  
list_name <- as.list(df_1997)  
  
# use the existing class to provide a summary of the updated dataset.  
summary.s50survey(list_name)
```

```
## Alcohol
##
##      Dont_drink      Drink_bi_anually      Drink_monthly
##           0.02           0.18           0.34
##      Drink_weekly Drink_more_than_weekly
##           0.34           0.12
## Smoke
##
##      Dont_smoke Smoke_occasionally      smoke_regular
##           0.62           0.04           0.34
## Drugs
##
##      Dont_smoke Smoke_occasionally      smoke_regular
##           0.62           0.04           0.34
## Sport
##
##      Dont_Play_regularly Play_sport_regularly
##           0.62           0.38
```

The summary data expressed as a proportion for each level of each variable for the 1997 dataset is set out above

Question 7

The proportion of students practicing sport regularly decreased from 74% in 1995 to 38% in 1997.