ISM 4403 Homework Week 5

### Tasks:

Create a new Excel spreadsheet from the following table.

**Chart 5.1**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id | Height (inches) | gender | Hair color | Eye Color | Age |
| 1 | 67 | male | brown | brown | 25 |
| 2 | 64 | female | brown | green | 23 |
| 3 | 74 | male | blond | blue | 27 |
| 4 | 73 | Male | brown | brown | 35 |
| 5 | 60 | female | red | green | 40 |
| 6 | 61 | female | brown | green | 45 |
| 7 | 73 | female | blond | blue |  |
| 8 | 70 | female | brown | blue | 50 |
| 9 | 56 | female | blond | brown | 60 |
| 10 | 57 | male | blond | brown | 18 |
| 11 | 64 |  | brown | brown | 25 |
| 12 | 69 | male | brown | green | 23 |
| 13 | 69 | female | blond |  | 27 |
| 14 | 70 | female | brown | brown | 35 |
| 15 | 71 | female | red | green | 40 |
| 16 | 60 | female |  | green | 45 |
| 17 | 80 | male | blond | blue | 41 |
| 18 | 75 | male | brown | blue | 50 |
| 19 | 78 | male | blond | brown | 60 |
| 20 | 69 | male | blond | brown | 18 |
| 21 |  | female | brown | brown | 25 |
| 22 | 66 | male | brown | green | 23 |
| 23 | 74 | female | blond | blue | 27 |
| 24 | 72 | male | brown | brown | 35 |
| 25 | 68 | female | red | green | 40 |
| 26 | 64 | female | brown | green | 45 |
| 27 | 63 | female | blond | blue | 41 |
| 28 |  | male | brown | blue | 50 |
| 29 | 70 | male | blond | brown | 60 |
| 30 | 62 | male | blond | brown | 18 |

Import the sheet into R.

Resolve any issues with missing data as done in homework 4.

Calculate the following.

The mean height.

The sum of all heights.

The mode for gender, hair color, and eye color.

Plot a pie graph for each of the following parameters, gender (count), hair color (count), eye color (count).

Plot a line graph of height and age. In your opinion is there a relationship between height and age? Please explain why or why not.

*I don’t think there is a relationship between age and height because there are no data points of age below the 18 years old. Therefore, age and height don’t have a correlation.*

**Paste your code here**

# week 5 lab

# import dataset

library("readxl")

df = read\_excel("week5.xlsx")

head(df)

# fixing missing values using is.na

df$Height\_inches[which(is.na(df$Height\_inches))] <-mean(df$Height\_inches,na.rm = TRUE)

mean(df$Height\_inches)

df$Age[which(is.na(df$Age))] <-mean(df$Age,na.rm = TRUE)

mean(df$Age)

sum(df$Height\_inches)

head(df)

# fixing missing values

df$gender[which(is.na(df$gender))] <- 'female'

df$`Hair\_color`[which(is.na(df$`Hair\_color`))] <- 'brown'

df$`Eye\_Color`[which(is.na(df$`Eye\_Color`))] <- 'brown'

# all of the pie charts

# pie chart gender

gender\_lbls <- c("male","female")

slices <- c(13,17)

pie(slices,labels = lbls, col = rainbow(length(gender\_lbls)),main="Pie Chart of Gender")

# pie chart hair

hair\_lbls <- c("brown","blond","red")

slices <- c(15,12,3)

pie(slices, labels = lbls,col = rainbow(length(hair\_lbls)),main = "Pie Chart of Hair Color")

# pie chart eye

eye\_lbls <- c ("brown","green","blue")

slices <- c(13,9,8)

pie(slices, labels = lbls,col = rainbow(length(eye\_lbls)),main = "Pie Chart of Eye Color")

# line graoh

ggplot(df, aes(x = df$Height\_inches, y = df$Age, type = "o")) +

geom\_line()+

geom\_point()

# regressor

library(caTools)

set.seed(123)

split = sample.split(df$Height\_inches, SplitRatio = 2/3)

training\_set = subset(df, split == TRUE)

test\_set = subset(df, split == FALSE)

# Fitting Simple Linear Regression to the Training set

regressor = lm(formula = Height\_inches ~ Age,

data = training\_set)

# Predicting the Test set results

y\_pred = predict(regressor, newdata = test\_set)

# Visualizing the Training set results

library(ggplot2)

ggplot(test\_set, aes(x = test\_set$Height\_inches, y = predict(regressor, newdata = test\_set))) +

geom\_point() +

stat\_smooth()

print(y\_pred)

print(test\_set)

**End of Paste**

> # week 5 lab

>

> # import dataset

> library("readxl")

> df = read\_excel("week5.xlsx")

>

> head(df)

# A tibble: 6 x 6

id Height\_inches gender Hair\_color Eye\_Color Age

*<dbl>* *<dbl>* *<chr>* *<chr>* *<chr>* *<dbl>*

1 1 67 male brown brown 25

2 2 64 female brown green 23

3 3 74 male blond blue 27

4 4 73 Male brown brown 35

5 5 60 female red green 40

6 6 61 female brown green 45

>

> # fixing missing values using is.na

> df$Height\_inches[which(is.na(df$Height\_inches))] <-mean(df$Height\_inches,na.rm = TRUE)

> mean(df$Height\_inches)

[1] 67.82143

>

>

> df$Age[which(is.na(df$Age))] <-mean(df$Age,na.rm = TRUE)

> mean(df$Age)

[1] 36.24138

>

>

> sum(df$Height\_inches)

[1] 2034.643

>

> head(df)

# A tibble: 6 x 6

id Height\_inches gender Hair\_color Eye\_Color Age

*<dbl>* *<dbl>* *<chr>* *<chr>* *<chr>* *<dbl>*

1 1 67 male brown brown 25

2 2 64 female brown green 23

3 3 74 male blond blue 27

4 4 73 Male brown brown 35

5 5 60 female red green 40

6 6 61 female brown green 45

>

> # fixing missing values

> df$gender[which(is.na(df$gender))] <- 'female'

> df$`Hair\_color`[which(is.na(df$`Hair\_color`))] <- 'brown'

> df$`Eye\_Color`[which(is.na(df$`Eye\_Color`))] <- 'brown'

>

>

> # all of the pie charts

>

>

> # pie chart gender

> gender\_lbls <- c("male","female")

> slices <- c(13,17)

> pie(slices,labels = lbls, col = rainbow(length(gender\_lbls)),main="Pie Chart of Gender")

>

> # pie chart hair

> hair\_lbls <- c("brown","blond","red")

> slices <- c(15,12,3)

> pie(slices, labels = lbls,col = rainbow(length(hair\_lbls)),main = "Pie Chart of Hair Color")

>

> # pie chart eye

> eye\_lbls <- c ("brown","green","blue")

> slices <- c(13,9,8)

> pie(slices, labels = lbls,col = rainbow(length(eye\_lbls)),main = "Pie Chart of Eye Color")

>

>

> # line graoh

> ggplot(df, aes(x = df$Height\_inches, y = df$Age, type = "o")) +

+ geom\_line()+

+ geom\_point()

>

>

>

> # regressor

> library(caTools)

> set.seed(123)

> split = sample.split(df$Height\_inches, SplitRatio = 2/3)

> training\_set = subset(df, split == TRUE)

> test\_set = subset(df, split == FALSE)

>

> # Fitting Simple Linear Regression to the Training set

> regressor = lm(formula = Height\_inches ~ Age,

+ data = training\_set)

>

> # Predicting the Test set results

> y\_pred = predict(regressor, newdata = test\_set)

>

> # Visualizing the Training set results

> library(ggplot2)

>

> ggplot(test\_set, aes(x = test\_set$Height\_inches, y = predict(regressor, newdata = test\_set))) +

+ geom\_point() +

+ stat\_smooth()

`geom\_smooth()` using method = 'loess' and formula 'y ~ x'

>

> print(y\_pred)

1 2 3 4 5 6 7 8 9 10

67.43770 68.36283 68.74829 69.51923 67.59189 69.13376 67.05223 67.59189 68.36283 69.13376

> print(test\_set)

# A tibble: 10 x 6

id Height\_inches gender Hair\_color Eye\_Color Age

*<dbl>* *<dbl>* *<chr>* *<chr>* *<chr>* *<dbl>*

1 2 64 female brown green 23

2 4 73 Male brown brown 35

3 5 60 female red green 40

4 8 70 female brown blue 50

5 11 64 female brown brown 25

6 16 60 female brown green 45

7 20 69 male blond brown 18

8 21 67.8 female brown brown 25

9 24 72 male brown brown 35

10 26 64 female brown green 45

**Paste your results here**

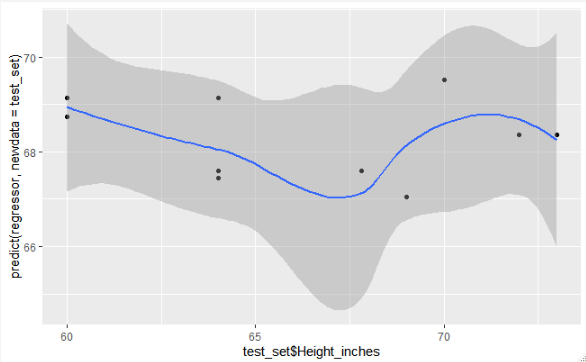
**End of Paste**

Using the example in 6.6 to create a regression model for the above data. Use this model to describe the relationship between the data in the above table.

Is there a relationship between height and age? Please explain?

There is no correlation between height and age.

**Paste your answer here**



**End of Page**

**Chart 5.2**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Customer Height (in) | Jean Length | Jean Width | Style | Price (dollars) | Profit dollars |
| 67 | 30 | 30 | baggy | 35 | 15 |
| 72 | 36 | 34 | tight | 40 | 20 |
| 60 | 24 | 30 | tight | 40 | 20 |
| 60 | 26 | 35 | baggy | 35 | 12 |
| 65 | 28 | 40 | tight | 40 | 17 |
| 66 | 28 | 38 | tight | 40 | 18 |
| 80 | 40 | 36 | baggy | 35 | 8 |
| 76 | 38 | 32 | tight | 40 | 16 |
| 80 | 40 | 30 | baggy | 35 | 9 |
| 72 | 36 | 30 | bsbby | 35 | 10 |

Given the above table (chart 5.2) plot a line graph of customer height to jean length.

If we accept that chart 5.1 represents 1% of the population who will buy jeans and we have a

monopoly. Chart a line graph that indicates how many individuals will buy each length of jean. Be sure to check the quality of the data and resolve any issues you find. The fix must be implemented in R. We will not accept solutions where you manually clean the data.

**Paste your code here**

# Part 2

# this is very unclear, I don't understand what you are asking

df2$Style[df2$Style == "bsbby"] <- "baggy"

df2 = read\_excel("week5part2.xlsx")

head(df2)

ggplot(df2, aes(x = df2$Jean\_Length, y = df2$Customer\_Height)) +

geom\_line()+

geom\_point()

**End of Paste**

**Paste your results here**

> # Part 2

> # this isn’t very unclear, I don't understand what you are asking

> df2$Style[df2$Style == "bsbby"] <- "baggy"

>

> df2 = read\_excel("week5part2.xlsx")

> head(df2)

# A tibble: 6 x 6

Customer\_Height Jean\_Length Jean\_Width Style Price\_dollars Profit\_dollars

*<dbl>* *<dbl>* *<dbl>* *<chr>* *<dbl>* *<dbl>*

1 67 30 30 baggy 35 15

2 72 36 34 tight 40 20

3 60 24 30 tight 40 20

4 60 26 35 baggy 35 12

5 65 28 40 tight 40 17

6 66 28 38 tight 40 18

> ggplot(df2, aes(x = df2$Jean\_Length, y = df2$Customer\_Height)) +

+ geom\_line()+

+ geom\_point()

**End of Paste**