|  |  |
| --- | --- |
| Lab 1 – R Language | Name:ID: |

1. Download the “yearly\_sales.csv” file from CANVAS

LOW Zhi Hao

54924670

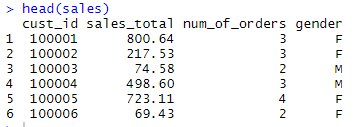
1. Open the csv file using a spreadsheet program. What is the csv file?

CSV in full form is Comma Separated Values, is heavily used in data storing as a file. Data values in cells CSV is commonly separated by `,` and data rows is separated by line breaks. Data in CSV is a table like structure, no merging of data cells is allowed.

1. Open the program “RStudio”
2. Load the previously downloaded csv file into R with the command:

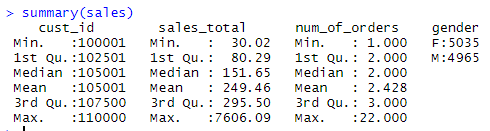
sales <- read.csv(***[your csv file path]***)

1. Type “head(sales)”. What can you observe?



First 6 rows of data is printed. This allow us to peek through the data to have an idea of the data type and headers.

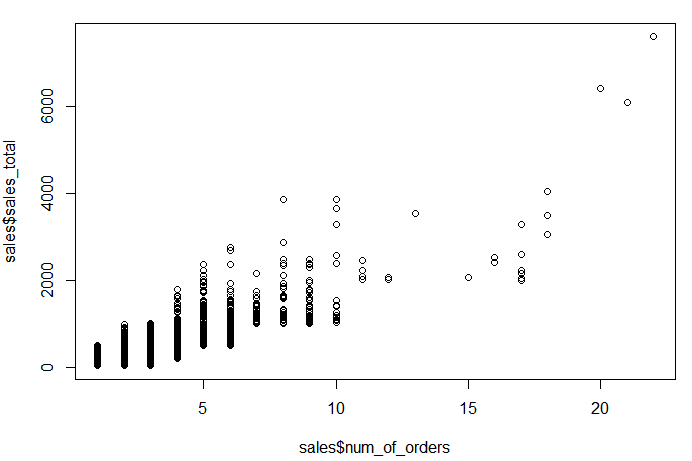
1. Type “summary(sales)”. What can you observe?



General statistical information is shown, we can see the minimum, 1st quartile, median, mean, 3rd quartile, and maximum values for numeric and ordinal data. For nominal data we only obtain the count for each class.

1. Type the following command. What can you observe?

plot(sales$num\_of\_orders,sales$sales\_total)

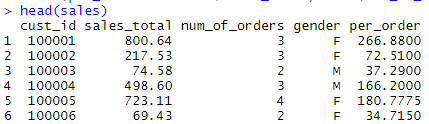


The scattered plot of number of orders to sales total is plotted. Relationship of number of orders and sales total can roughly be visualized as we are trying to gain insights from the data. The insight, when the number of orders increase the sales total increase.

1. Type the following commands. What have you done to the “sales” data?

sales$per\_order <- sales$sales\_total/sales$num\_of\_orders

head(sales)



I have added the new column for the sales data, R takes the two column’s operation and generate a new column. The `per\_order` column implies how much each order cost on average.

1. Type the following command. What have you done?

write.table(sales,"sales\_modified.txt", sep="\t", row.names=FALSE)

I have exported the data as a file named "sales\_modified.txt". This data is separated by tabs and the row’s name (index) is not exported

1. Type the following commands. What have you done?

jpeg(file="sales\_hist.jpeg")

hist(sales$num\_of\_orders)

dev.off()

We open a file “sales\_hist.jpeg” and output the histogram of number of orders to the file. The output file is an image file. File is saved when we call dev.off().

1. Type the following commands. What have you done?

x <- sales$sales\_total

y <- sales$num\_of\_orders

Selected one of the columns sales total and another number of orders and place into the variable x and variable y respectively, these data is a list like data type.

1. Type the commands in the leftmost column and fill in the following table:

|  |  |  |
| --- | --- | --- |
| R Command | Return Value | Semantic Meaning (hint: use the help panel in RStudio) |
| cor(x,y) | 0.7508015 | Measure of correlation for sales\_total and num\_of\_orders |
| cov(x,y) | 345.2111 | Measure of covarience for sales\_total and num\_of\_orders |
| IQR(x) | 215.21 | Interquartile Range of the sales\_total (3rd quartile of sales\_total – 1st quartile of sales\_total) |
| mean(x) | 249.4557 | The mean value of sales\_total or average of sales\_total |
| median(x) | 151.65 | The median value of sales\_total |
| range(x) | 30.02 7606.09 | The minimum and maximum value of sales\_total the range of the sales\_total |
| sd(x) | 319.0508 | The standard deviation of sales\_total and uses denominator *n - 1* |
| var(x) | 101793.4 | Variance of sales\_total. Same as cov(x,x) |

1. Apply the R knowledge you have learned in the previous steps to the dataset “zipIncome.csv” in CANVAS and state your data analytics insights in less than 50 words.
2. This is the end; please upload this sheet with your answers to the submission system..