STATISTICAL TECHNIQUES

100594631

2023-01-23

# Loading up the Dataset into R-Markdown working environment  
df <- read.csv('C:/Users/USER/Documents/Raw\_Data/superstore\_dataset2011-2015.csv')  
head(df, 5)

## Segment Ship.Mode Region Order.Priority Sales Quantity Discount  
## 1 Consumer Standard Class Africa Medium 408.30 2 0.0  
## 2 Consumer Standard Class Oceania Medium 120.37 3 0.1  
## 3 Consumer Second Class EMEA High 66.12 4 0.0  
## 4 Corporate Second Class North High 44.86 3 0.5  
## 5 Consumer Standard Class Oceania Medium 113.67 5 0.1  
## Profit Shipping.Cost TotalSale SalesAfterDiscount  
## 1 106.14 35.46 816.600 816.60  
## 2 36.04 9.72 361.098 324.99  
## 3 29.64 8.17 264.480 264.48  
## 4 -26.06 4.82 134.595 67.30  
## 5 37.77 4.70 568.350 511.52

THE METADATA::

Ship Mode=> Shipping Mode specified by the Customer. Segment => The segment where the Customer belongs. Region => Region where the Customer belong. Sales => Sales of the Product. Quantity => Quantity of the Product. Discount => Discount provided. Shipping Cost=> The amount required to transport the product Total Sales => The product of Sales and Quantity Sales After Discount => Total sales after discount deduction

# Attaching the variables to the dataframe fro ease of calling  
attach(df)

…

# the properties of the dataset  
str(df)

## 'data.frame': 48589 obs. of 11 variables:  
## $ Segment : chr "Consumer" "Consumer" "Consumer" "Corporate" ...  
## $ Ship.Mode : chr "Standard Class" "Standard Class" "Second Class" "Second Class" ...  
## $ Region : chr "Africa" "Oceania" "EMEA" "North" ...  
## $ Order.Priority : chr "Medium" "Medium" "High" "High" ...  
## $ Sales : num 408.3 120.4 66.1 44.9 113.7 ...  
## $ Quantity : int 2 3 4 3 5 2 2 2 1 3 ...  
## $ Discount : num 0 0.1 0 0.5 0.1 0.1 0 0.15 0 0 ...  
## $ Profit : num 106.1 36 29.6 -26.1 37.8 ...  
## $ Shipping.Cost : num 35.46 9.72 8.17 4.82 4.7 ...  
## $ TotalSale : num 817 361 264 135 568 ...  
## $ SalesAfterDiscount: num 816.6 325 264.5 67.3 511.5 ...

BRIEF EXPLORATORY ANALYSIS

(QUALITATIVE)

#Viewing the frequency of the Segment Variable  
Segment.freq <- table(Segment)  
Freq\_Dist <- cbind(Segment.freq)  
Freq\_Dist

## Segment.freq  
## Consumer 25051  
## Corporate 23538

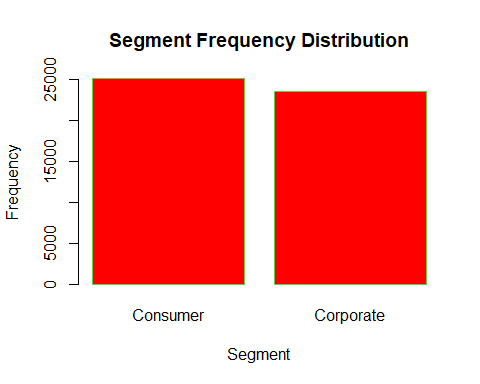
…

#the relative frequency distribution in then given by  
Segment.relfreq <- Segment.freq/nrow(df)  
Rel\_Freq\_Dist <- round(cbind(Segment.relfreq) \* 100,2)  
Rel\_Freq\_Dist

## Segment.relfreq  
## Consumer 51.56  
## Corporate 48.44

…

#Segment Frequency Distribution  
barplot((Segment.freq),  
 col="red", border=3,   
 main="Segment Frequency Distribution",   
 xlab="Segment",   
 ylab="Frequency"  
 )



QUANTITATIVE

#Relative Fequency Disricution of the the variable Quantity.  
range(Quantity)

## [1] 1 14

breaks <- seq(1, 14, by = 1)   
breaks

## [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14

…

#Classify the Quantity according to the half-unit-length sub-intervals with cut  
Quantity.cut <- cut(Quantity, breaks, right = FALSE)

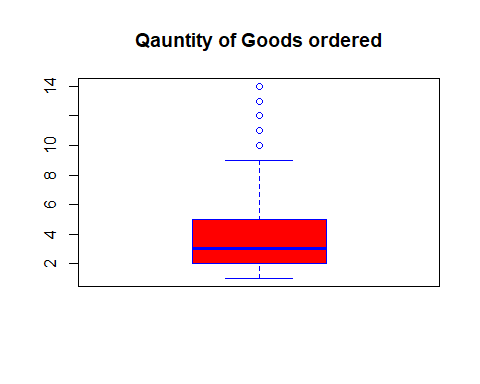
…

#frequency distribution of the quantity of goods ordered for  
Quantity.freq <- table(Quantity.cut)  
rbind(Quantity.freq)

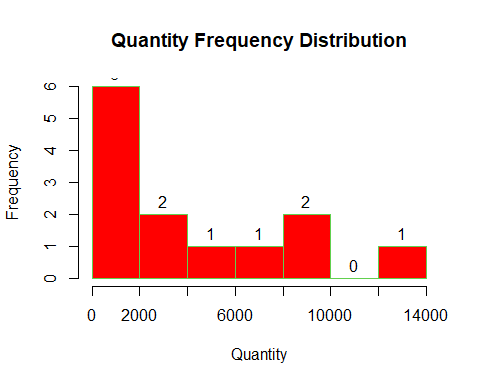
## [1,2) [2,3) [3,4) [4,5) [5,6) [6,7) [7,8) [8,9) [9,10) [10,11)  
## Quantity.freq 8470 12061 9197 6033 4641 2859 2270 1283 937 267  
## [11,12) [12,13) [13,14)  
## Quantity.freq 142 168 81

…

boxplot(Quantity, col= "Red", border= "Blue",   
 main="Qauntity of Goods ordered")

 …

# Quantity Frequency Distribution  
hist((Quantity.freq),  
 col="red", border=3,   
 main="Quantity Frequency Distribution",   
 xlab="Quantity",   
 ylab="Frequency",   
 axes=T, plot=T, labels=T, right = F)

 …

#Finding the Cumulative frequency distribution of the quantity of goods ordered for  
Quantity.cumfreq <- cumsum(Quantity.freq)  
rbind(Quantity.cumfreq)

## [1,2) [2,3) [3,4) [4,5) [5,6) [6,7) [7,8) [8,9) [9,10) [10,11)  
## Quantity.cumfreq 8470 20531 29728 35761 40402 43261 45531 46814 47751 48018  
## [11,12) [12,13) [13,14)  
## Quantity.cumfreq 48160 48328 48409

…

#Checking for the mean regional Total Sales  
Regional\_Mean\_Sales <- rbind(round(sort(tapply(TotalSale, Region, mean),  
 decreasing = TRUE),2))  
Regional\_Mean\_Sales

## North Asia Central Asia Oceania Southeast Asia North Central South  
## [1,] 1880.71 1869.95 1652.11 1396.16 1309.72 1265.28 1205.4  
## East West Caribbean Africa EMEA Canada  
## [1,] 1182.51 1122.72 905.35 683.7 652.24 627.41

…

#Checking for the mean regional Sales after discount  
Regional\_Mean\_Sales <- rbind(round(sort(tapply(SalesAfterDiscount, Region, mean),  
 decreasing = TRUE),2))  
Regional\_Mean\_Sales

## North Asia Central Asia Oceania North Central Southeast Asia South  
## [1,] 1813.27 1795.89 1444.63 1212.86 1120.99 1102.39 1064.35  
## East West Caribbean Africa Canada EMEA  
## [1,] 1019.05 982.87 803.43 642.56 627.41 590.72

…

# finding the corellation between discount and profit  
cbind(cor(Discount, Profit))

## [,1]  
## [1,] -0.3187385

’’’

plot(Discount, TotalSale,  
 col="red", border=3,   
 main="Scatterplot of Discount & Total Sales",   
 xlab="Discount",   
 ylab="Total Sales",  
 axes=T, plot=T, labels=T, right = F)

## Warning in plot.window(...): "border" is not a graphical parameter

## Warning in plot.window(...): "plot" is not a graphical parameter

## Warning in plot.window(...): "labels" is not a graphical parameter

## Warning in plot.window(...): "right" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "border" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "plot" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "labels" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "right" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "border" is not a  
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "plot" is not a  
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "right" is not a  
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "border" is not a  
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "plot" is not a  
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "right" is not a  
## graphical parameter

## Warning in box(...): "border" is not a graphical parameter

## Warning in box(...): "plot" is not a graphical parameter

## Warning in box(...): "labels" is not a graphical parameter

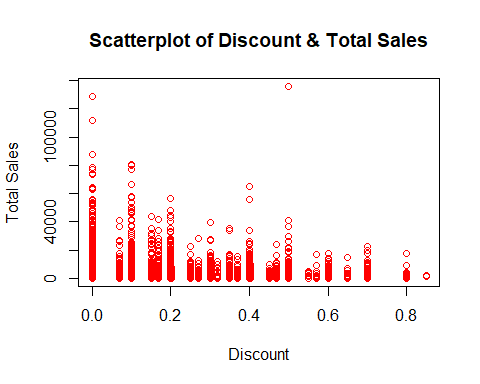
## Warning in box(...): "right" is not a graphical parameter

## Warning in title(...): "border" is not a graphical parameter

## Warning in title(...): "plot" is not a graphical parameter

## Warning in title(...): "labels" is not a graphical parameter

## Warning in title(...): "right" is not a graphical parameter



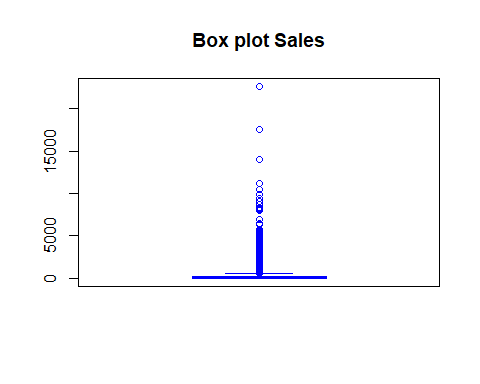
STATISTICAL TESTS

head(df)

## Segment Ship.Mode Region Order.Priority Sales Quantity Discount  
## 1 Consumer Standard Class Africa Medium 408.30 2 0.0  
## 2 Consumer Standard Class Oceania Medium 120.37 3 0.1  
## 3 Consumer Second Class EMEA High 66.12 4 0.0  
## 4 Corporate Second Class North High 44.86 3 0.5  
## 5 Consumer Standard Class Oceania Medium 113.67 5 0.1  
## 6 Consumer Standard Class Oceania Medium 55.24 2 0.1  
## Profit Shipping.Cost TotalSale SalesAfterDiscount  
## 1 106.14 35.46 816.600 816.60  
## 2 36.04 9.72 361.098 324.99  
## 3 29.64 8.17 264.480 264.48  
## 4 -26.06 4.82 134.595 67.30  
## 5 37.77 4.70 568.350 511.52  
## 6 15.34 1.80 110.484 99.44

ONE SAMPLE T-TEST

boxplot(Sales, col= "Red", border= "Blue",   
 main="Box plot Sales")



Null Hypothesis (H\_0) : Total Sale mean is 1,300 (TotalSale = 1,300)

Alternate hypothesis (H\_A) : Total Sale is not 1,300 (TotalSale != 1,300)

test\_1 <- t.test(TotalSale, mu = 1300)  
test\_1

##   
## One Sample t-test  
##   
## data: TotalSale  
## t = -5.8953, df = 48588, p-value = 3.765e-09  
## alternative hypothesis: true mean is not equal to 1300  
## 95 percent confidence interval:  
## 1176.238 1237.999  
## sample estimates:  
## mean of x   
## 1207.119

The p-Value gotten is less than the 0.05 significance level, therefore, the NULL hypothesis is rejected

TWO SAMPLE T-TEST

Null Hypothesis (H\_0) : Consumer mean is same as Corporate mean (difference = 0)

Alternate hypothesis (H\_A) : Consumer mean is not same as Corporate mean (difference != 0)

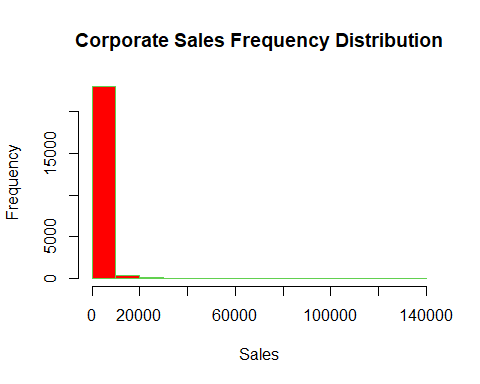
# assigning the consumer varable to the corresponding TotalSale   
seg\_consumer <- Segment == 'Consumer' # filtering the consumer from the segment column  
cons <- df[seg\_consumer, ] # creates dataset containing Consumer only  
consumer\_sales <- cons$TotalSale # selects only the corresponding TotalSale values

…

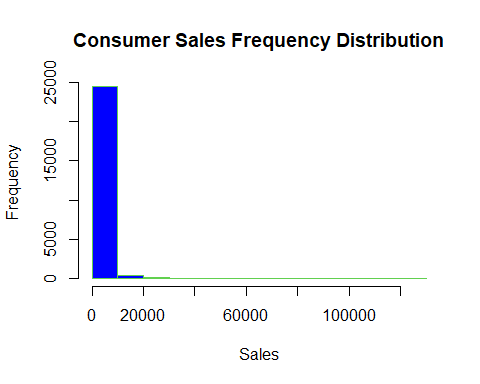
# assigningg the corporate data to a viariable   
seg\_corporate <- Segment == 'Corporate' # filtering the corporate from the segment column  
corp <- df[seg\_corporate, ] # creates dataset containing Corpotate only  
corporate\_sales <- corp$TotalSale # selects only the corresponding TotalSale values

…

hist((corporate\_sales),  
 col="red", border=3,   
 main="Corporate Sales Frequency Distribution",   
 xlab="Sales",   
 ylab="Frequency",   
 axes=T, plot=T, labels=F, right = F)



hist((consumer\_sales),  
 col="blue", border=3,   
 main="Consumer Sales Frequency Distribution",   
 xlab="Sales",   
 ylab="Frequency",   
 axes=T, plot=T, labels=F, right = F)

 …

# computing the Two sample T-Test  
test\_2 <- t.test(consumer\_sales, corporate\_sales, var.equal=TRUE)  
test\_2

##   
## Two Sample t-test  
##   
## data: consumer\_sales and corporate\_sales  
## t = -0.57179, df = 48587, p-value = 0.5675  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -79.81753 43.76482  
## sample estimates:  
## mean of x mean of y   
## 1198.386 1216.412

as seen above, P-value id greater than 0.05 which makes it not possible to reject the null hypothesis that both means are equal

CHI-SQUARE TEST

# The Ship.Mode and Region Contingency Table  
ShipRegion\_contingency <- with(df, table(Ship.Mode, Region))  
cbind(ShipRegion\_contingency)

## Africa Canada Caribbean Central Central Asia East EMEA North  
## First Class 679 70 231 1593 252 490 700 684  
## Second Class 992 91 372 2176 440 530 993 954  
## Standard Class 2659 194 1002 6780 1277 1673 3036 2901  
## North Asia Oceania South Southeast Asia West  
## First Class 385 455 958 493 515  
## Second Class 470 779 1368 523 621  
## Standard Class 1377 2048 3978 1968 1882

…

Null Hypothesis (H\_0) : Sales is independent of Discount

Alternate hypothesis (H\_A) : Consumer mean is not same as Corporate mean (difference != 0)

# implementing the chi-square test  
chisq.test(Ship.Mode, Region, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: Ship.Mode and Region  
## X-squared = 111.84, df = 24, p-value = 2.688e-13

The p-Value gotten is less than the 0.05 significance level, therefore, the NULL hypothesis is rejected …

KRUSKAL WALIS TEST

cbind(tapply(TotalSale, Order.Priority, median))

## [,1]  
## Critical 239.600  
## High 245.970  
## Low 246.546  
## Medium 247.104

…

#finding the Summary of the four variables  
Critical\_summary <- summary(TotalSale[Order.Priority=='Critical'])  
High\_summary <- summary(TotalSale[Order.Priority=='High'])  
Low\_summary <- summary(TotalSale[Order.Priority=='Low'])  
Medium\_summary <- summary(TotalSale[Order.Priority=='Medium'])  
cbind(Critical\_summary, High\_summary, Low\_summary, Medium\_summary)

## Critical\_summary High\_summary Low\_summary Medium\_summary  
## Min. 1.6020 1.344 1.6320 0.4440  
## 1st Qu. 64.2600 66.555 63.5810 68.3100  
## Median 239.6000 245.970 246.5460 247.1040  
## Mean 1218.9155 1199.713 1079.7388 1220.2185  
## 3rd Qu. 922.0635 920.655 882.2205 900.4019  
## Max. 78204.9600 80210.844 111420.1200 135830.8800

Null Hypothesis (H\_0) : same median ranks

Alternate hypothesis (H\_A) : at least one of the median rank is different

# implementing the Kruskal Walis Test  
kruskal.test(Order.Priority ~ TotalSale)

##   
## Kruskal-Wallis rank sum test  
##   
## data: Order.Priority by TotalSale  
## Kruskal-Wallis chi-squared = 26499, df = 26378, p-value = 0.2988

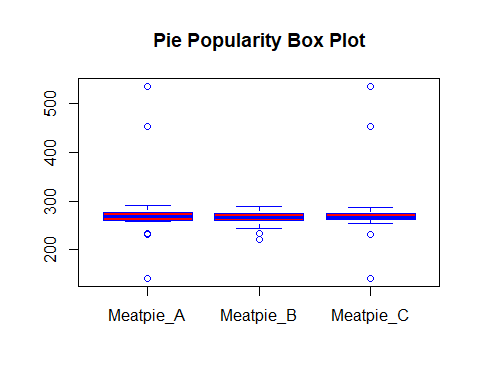
as seen above, P-value id grester than 0.05 which makes it not possible to reject the null hypothesis that both median ranks are equal

ANALYSIS OF VARIANCE(ANOVA)

df\_pie = read.csv('C:/Users/USER/Documents/Raw\_Data/meatpie\_sales.csv')  
head(df\_pie, 5)

## Meatpie\_A Meatpie\_B Meatpie\_C  
## 1 264 277 267  
## 2 261 269 279  
## 3 267 263 274  
## 4 272 266 276  
## 5 258 262 262

boxplot(df\_pie, col= "Red", border= "Blue",   
 main="Pie Popularity Box Plot")



…

# finding the number of rows  
nrow(df\_pie)

## [1] 45

…

# Concatenate the data rows in df\_pie into a single vector.  
vec <- c(t(as.matrix(df\_pie)))

…

# Assigning new variables for the treatment levels and number of control blocks.  
t\_level <- c("ï..Meatpie\_A", "Meatpie\_B", "Meatpie\_c") # treatment levels   
k <- 3 # number of treatment levels   
n <- 45

…

m\_treatment <- gl(k, 1, n\*k, factor(t\_level)) # matching treatment

…

#creating a vector of blocking factors for each element in the response data.  
block\_l <- gl(n, k, k\*n) # blocking factor

…

# implemnting the ANOVA test   
av <- aov(vec ~ m\_treatment + block\_l)  
summary(av)

## Df Sum Sq Mean Sq F value Pr(>F)  
## m\_treatment 2 2226 1113 0.529 0.591  
## block\_l 44 88819 2019 0.960 0.550  
## Residuals 88 184966 2102

WILCOXON SIGNED RANK TEST

df3 <- read.csv('C:/Users/USER/Documents/Raw\_Data/sample\_submission.csv')  
head(df3, 2)

## SalePrice  
## 1 169277.1  
## 2 187758.4

Checking the normality of the distribution

# Using the QQ-plot  
qqnorm(df3$SalePrice,  
 col="red", border=3,   
 main="QQ Plot of Sales",   
 xlab="Sample Quantile",   
 ylab="Theoretical Quantile",  
 axes=T, plot=T, labels=T, right = F)

## Warning in plot.window(...): "border" is not a graphical parameter

## Warning in plot.window(...): "labels" is not a graphical parameter

## Warning in plot.window(...): "right" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "border" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "labels" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "right" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "border" is not a  
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "right" is not a  
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "border" is not a  
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "right" is not a  
## graphical parameter

## Warning in box(...): "border" is not a graphical parameter

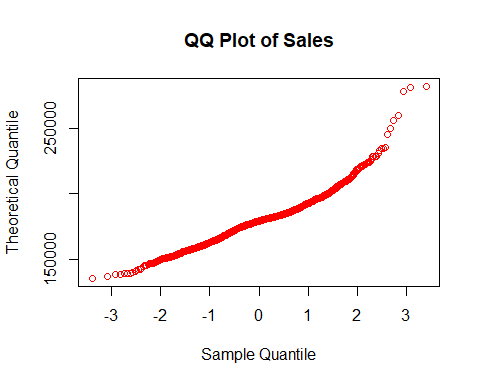
## Warning in box(...): "labels" is not a graphical parameter

## Warning in box(...): "right" is not a graphical parameter

## Warning in title(...): "border" is not a graphical parameter

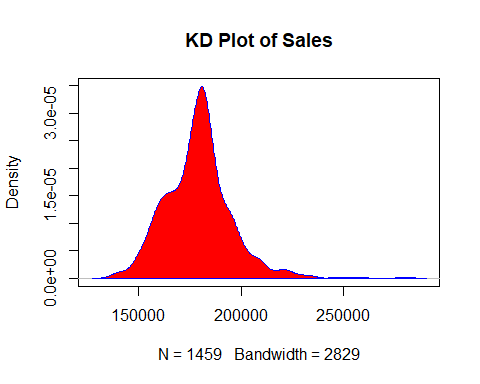
## Warning in title(...): "labels" is not a graphical parameter

## Warning in title(...): "right" is not a graphical parameter



The bent line indicates abnormality of the distribution

# using KDE plot  
d <- density(df3$SalePrice)  
plot(d, main="KD Plot of Sales")  
polygon(d, col="red", border="blue")



The KDE plot also indicates left-skewness.

Using Shapiro-Wilt Test to further check if the data value is normally distributed

Ho = The distribution is normally distributed

Ha = The distribution in not normally distributed

# implementing the shapiro wilk test  
shapiro.test(df3$SalePrice)

##   
## Shapiro-Wilk normality test  
##   
## data: df3$SalePrice  
## W = 0.95287, p-value < 2.2e-16

Since a p-value extremely lower than 0.05 is gotten, then we can reject the NULL hypothesis that the data value is normally distributed and go ahead with the Wilcoxon signed rank test

Ho = The median(mean) is 180000

Ha = The median(mean) is not 180000

wilcox.test(df3$SalePrice, mu = 180000)

##   
## Wilcoxon signed rank test with continuity correction  
##   
## data: df3$SalePrice  
## V = 479160, p-value = 0.0009131  
## alternative hypothesis: true location is not equal to 180000

LINEAR REGRESSION

#perform label encoding on variable cegorical variables  
df$Segment <- as.numeric(factor(df$Segment))  
df$Ship.Mode <- as.numeric(factor(df$Ship.Mode))  
df$Region <- as.numeric(factor(df$Region))  
df$Order.Priority <- as.numeric(factor(df$Order.Priority))

head(df, 3)

## Segment Ship.Mode Region Order.Priority Sales Quantity Discount Profit  
## 1 1 3 1 4 408.30 2 0.0 106.14  
## 2 1 3 10 4 120.37 3 0.1 36.04  
## 3 1 2 7 2 66.12 4 0.0 29.64  
## Shipping.Cost TotalSale SalesAfterDiscount  
## 1 35.46 816.600 816.60  
## 2 9.72 361.098 324.99  
## 3 8.17 264.480 264.48

…

# the general correlation  
cor(df)

## Segment Ship.Mode Region Order.Priority  
## Segment 1.000000e+00 0.0021656341 -0.0062785699 0.0171324961  
## Ship.Mode 2.165634e-03 1.0000000000 0.0002833827 0.3999254413  
## Region -6.278570e-03 0.0002833827 1.0000000000 -0.0033776228  
## Order.Priority 1.713250e-02 0.3999254413 -0.0033776228 1.0000000000  
## Sales 5.371098e-04 0.0007061361 0.0358444932 -0.0001722673  
## Quantity 2.707446e-05 0.0036710969 0.1105517929 0.0047451822  
## Discount -2.394395e-03 -0.0087209407 0.0353443127 0.0059758526  
## Profit 2.988576e-03 0.0025837369 -0.0007597047 0.0014790284  
## Shipping.Cost -9.860328e-04 -0.1428019658 0.0321320444 -0.1827469727  
## TotalSale 2.594053e-03 0.0021146674 0.0352344306 0.0022781203  
## SalesAfterDiscount 2.291732e-03 0.0022098848 0.0281236262 0.0019970528  
## Sales Quantity Discount Profit  
## Segment 0.0005371098 2.707446e-05 -0.002394395 0.0029885756  
## Ship.Mode 0.0007061361 3.671097e-03 -0.008720941 0.0025837369  
## Region 0.0358444932 1.105518e-01 0.035344313 -0.0007597047  
## Order.Priority -0.0001722673 4.745182e-03 0.005975853 0.0014790284  
## Sales 1.0000000000 3.135900e-01 -0.087911347 0.5038083069  
## Quantity 0.3135899910 1.000000e+00 -0.021977616 0.1098541962  
## Discount -0.0879113473 -2.197762e-02 1.000000000 -0.3187384969  
## Profit 0.5038083069 1.098542e-01 -0.318738497 1.0000000000  
## Shipping.Cost 0.7740772442 2.757009e-01 -0.080111654 0.3703900756  
## TotalSale 0.9020813556 4.404530e-01 -0.060877184 0.4443254207  
## SalesAfterDiscount 0.8847158799 4.291082e-01 -0.106150618 0.5376436275  
## Shipping.Cost TotalSale SalesAfterDiscount  
## Segment -0.0009860328 0.002594053 0.002291732  
## Ship.Mode -0.1428019658 0.002114667 0.002209885  
## Region 0.0321320444 0.035234431 0.028123626  
## Order.Priority -0.1827469727 0.002278120 0.001997053  
## Sales 0.7740772442 0.902081356 0.884715880  
## Quantity 0.2757008795 0.440452987 0.429108230  
## Discount -0.0801116537 -0.060877184 -0.106150618  
## Profit 0.3703900756 0.444325421 0.537643627  
## Shipping.Cost 1.0000000000 0.700676529 0.696777455  
## TotalSale 0.7006765290 1.000000000 0.985091081  
## SalesAfterDiscount 0.6967774551 0.985091081 1.000000000

Ho = the predictors have no influence on the target variables

Ha = the predictors have no influence on the target variables

lmr <- lm(Sales ~ Segment + Ship.Mode + Region +Order.Priority + Quantity + Discount)  
summary(lmr)

##   
## Call:  
## lm(formula = Sales ~ Segment + Ship.Mode + Region + Order.Priority +   
## Quantity + Discount)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -918.9 -185.0 -89.4 38.5 22314.7   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 53.5096 11.4955 4.655 3.25e-06 \*\*\*  
## SegmentCorporate 0.4511 4.1841 0.108 0.914145   
## Ship.ModeSecond Class 3.4676 7.0654 0.491 0.623581   
## Ship.ModeStandard Class 1.3903 6.5521 0.212 0.831954   
## RegionCanada -10.7822 25.4850 -0.423 0.672237   
## RegionCaribbean -78.6790 13.5350 -5.813 6.17e-09 \*\*\*  
## RegionCentral -17.9445 8.4323 -2.128 0.033336 \*   
## RegionCentral Asia 82.0436 12.6317 6.495 8.38e-11 \*\*\*  
## RegionEast -32.7779 11.3893 -2.878 0.004004 \*\*   
## RegionEMEA -3.4373 9.6992 -0.354 0.723046   
## RegionNorth -19.6776 9.9037 -1.987 0.046941 \*   
## RegionNorth Asia 74.3300 12.1365 6.124 9.17e-10 \*\*\*  
## RegionOceania 53.2756 10.7442 4.959 7.13e-07 \*\*\*  
## RegionSouth -31.4985 9.2011 -3.423 0.000619 \*\*\*  
## RegionSoutheast Asia 34.6609 11.1172 3.118 0.001823 \*\*   
## RegionWest -56.5017 11.0320 -5.122 3.04e-07 \*\*\*  
## Order.PriorityHigh -11.6606 9.3633 -1.245 0.213005   
## Order.PriorityLow -16.5175 13.3937 -1.233 0.217495   
## Order.PriorityMedium -9.7691 9.4272 -1.036 0.300085   
## Quantity 67.0369 0.9478 70.728 < 2e-16 \*\*\*  
## Discount -185.1281 10.1227 -18.288 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 460.7 on 48568 degrees of freedom  
## Multiple R-squared: 0.1111, Adjusted R-squared: 0.1107   
## F-statistic: 303.4 on 20 and 48568 DF, p-value: < 2.2e-16

…

plot(lmr, col = 'red')

