Date: 7-11-2022

**Multivariate Regression**

**Case Name:** Experiential Retailing:Influence on Young Indian Consumer’s Response

**Objective:** To identify the impact of all the independent variables i.e (sound,light,layout,music,etc) on the customer’s retail experience.

**Justification:** Since all the variables are quantitative(numerical) in nature therefore to check the above mentioned objective,we will use a multivariate regression model.

**Data Analysis:**

**Step1:** Hypothesis for multivariate linear regression model.

Null Hypothesis(H0 ): The overall model is not statistically significant.

Alternative Hypothesis(H1): The overall model is statistically significant.

α= level of significance(0.5)

p<α => Reject H0

Here from the output **p-value=0.0001341** that is less than α(0.5) ,so we reject the null hypothesis(Ho) and accept H1 and therefore we can say that the model is statistically significant.

**Step2:** Hypothesis for β-coefficient.

H0i: All the β-coefficient. are not statistically significant

H01: At least one of the β-coefficient is statistically significant

| Coefficients: |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Estimate | Std.Error | t | value | Pr(>|t|) | Significant/Insignigicant(α=0.05) |  |
| (Intercept) | 0.818982 | 0.707732 | 1.157 | 0.250096 |  |  |  |
| shoping.when.bored | 0.009209 | 0.089835 | 0.103 | 0.918567 |  | p>α=p doesnot less than α so we do not reject Null Hypothesis | not significant |
| waste.of.time | -0.102435 | 0.110548 | -0.927 | 0.356477 |  | p>α=p doesnot less than α so we do not reject Null Hypothesis | not significant |
| wall.colour | 0.175777 | 0.099336 | 1.77 | 0.080015 | . | p>α=p doesnot less than α so we do not reject Null Hypothesis | \* significant at 10% |
| fragrance | 0.084468 | 0.110294 | 0.766 | 0.445669 |  | p>α=p doesnot less than α so we do not reject Null Hypothesis | not significant |
| emp.knowledge | -0.367028 | 0.1491 | -2.462 | 0.015635 | \* |  | significant |
| layout.flooring | -0.078535 | 0.105669 | -0.743 | 0.459183 |  |  | not significant |
| recommend | 0.044022 | 0.124292 | 0.354 | 0.723986 |  |  | not significant |
| emp.concerned | 0.176481 | 0.122625 | 1.439 | 0.153383 |  |  | not significant |
| layout.spacious | -0.151162 | 0.099258 | -1.523 | 0.131102 |  |  | not significant |
| emp.trustworthy | 0.150259 | 0.116543 | 1.289 | 0.200424 |  |  | not significant |
| layout.design.display | 0.030934 | 0.122993 | 0.252 | 0.801963 |  |  | not significant |
| entertain | -0.076202 | 0.162124 | -0.47 | 0.639416 |  |  | not significant |
| enthusiam | 0.172139 | 0.134744 | 1.278 | 0.204527 |  |  | not significant |
| moretime.spent | 0.451503 | 0.116039 | 3.891 | 0.000185 | \* |  | significant |
| buy.more | 0.107271 | 0.098752 | 1.086 | 0.28011 |  |  | not significant |
| design.good | -0.053493 | 0.160039 | -0.334 | 0.738926 |  |  | not significant |
| light.dull | 0.108277 | 0.117792 | 0.919 | 0.360307 |  |  | not significant |
| music.bothersome | -0.208864 | 0.104116 | -2.006 | 0.047693 | \* |  | significant |
| emp.not.assist | 0.12882 | 0.101218 | 1.273 | 0.20623 |  |  | not significant |

**Note:**

**\*** significant

\*\*\*

. significant at 10%

**Step3:**  Regression Model

| Coefficients: |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std.Error | t | values | Pr(>|t|) |
| (Intercept) | 1.13789 | 0.31771 | 3.582 | 0.000508 | \*\*\* |
| wall.colour | 0.1865 | 0.08737 | 2.135 | 0.034985 | \* |
| emp.knowledge | -0.22314 | 0.10351 | -2.156 | 0.033259 | \* |
| moretime.spent | 0.52414 | 0.09432 | 5.557 | 1.91E-07 | \*\*\* |

**y = α+ β1.x1 + β2.x2 + β3.x3+E**

Frequency of visit=1.1278+0.1865(wall color)-0.223(emp knowledge)+0.524(more time spent)+E

If emp knowledge and more time spent is constant and also if we will increase the wall color by 1 unit frequency of visit will increase by 18.65%

Similarly keeping wall color and more time spend constant and if we increase emp knowledge by 1 unit then frequency of visit will decrease by 22.3%

After comparing β-coefficient we conclude that the more time spent in retail store is more important and most influential factor followed by employee knowledge and wall color

**Step4:** Multicollinearity

If there is high correlation between the variables then we can say that multicollinearity is present between the variables.

| wall.colour | emp.knowledge | Moretime spent |
| --- | --- | --- |
| 1.067438 | 1.313283 | 1.298247 |

**Interpretation:**  Since the vif(Variance Inflation Factor) value for the independent variables is below 5 i.e., no multicollinearity is present between the variables

Date: 9-11-2022

Dataset: walmart\_data

Dependent Variable- Sales,Promotion Index(Quantitative Variable)

Categorical Data- Walmart - 1 represent post (walmart open) and 0 represent pre(when walmart is close)

Holiday- 1 represents holiday there 0 represents holiday not there.

**Note**:Categorical Data should be always in 0 and 1 it can't be in 1,2 or 3,etc

If n Categorical Data create dummy variable n-1

**Data Analysis:**

**Step1:** Hypothesis for multivariate linear regression model.

Null Hypothesis(H0 ): The overall model is not statistically significant.

Alternative Hypothesis(H1): The overall model is statistically significant.

α= level of significance(0.5)

p<α => Reject H0

Here from the output **p-value=2.977e-13** that is less than α(0.5) ,so we reject the null hypothesis(Ho) and accept H1 and therefore we can say that the model is statistically significant.

**Step2:** Hypothesis for β-coefficient.

H0i: All the β-coefficient. are not statistically significant

H01: At least one of the β-coefficient is statistically significant

| ANOVA |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 2 | 7.33917E+11 | 3.66959E+11 | 10.82865655 | 5.69043E-05 |
| Residual | 97 | 3.28711E+12 | 33887722216 |  |  |
| Total | 99 | 4.02103E+12 |  |  |  |

**Step3:**  Regression Model

**y = α+ β1.x1 + β2.x2 + β3.x3+E**

Sales= -332935+584857(Promotional Index)-198288(Walmart)+466048(FeatureAdv.Index)+193727(Holiday)+E

Multivariate Linear Regression With and without dummy variables:

**Objective:** To identify the impact of all the independent variables Price of egg and price of cookies on the dependent variable Sales

**Justification:** Since all the dependent variables and independent variables are numerical in nature ,therefore we will use multivariate linear regression model.

**Data Analysis:**

**Step1:** Hypothesis for multivariate linear regression model.

Null Hypothesis(H0 ): The overall model is not statistically significant.

Alternative Hypothesis(H1): The overall model is statistically significant.

α= level of significance(0.5)

If p<α => Reject H0

If p>α => We do not reject H0

Here p value =1.215e-10 which is less than α we reject the null hypothesis and accept H1 and therefore we can say that the model is statistically significant.

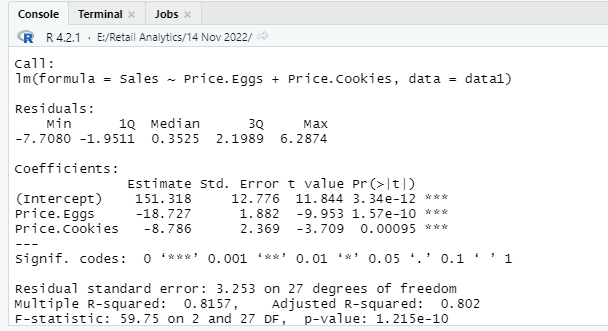
setwd("E:/Retail Analytics/14 Nov 2022")

data1 <-read.csv("supermarket.csv")

variable.names(data1)

model1 <-lm(Sales~Price.Eggs+Price.Cookies,data=data1)

summary(model1)



**Step2:** Hypothesis for β-coefficient.

H0i: All the β-coefficient. are not statistically significant

H1i: At least one of the β-coefficient is statistically significant

**y = α+ β1.x1 + β2.x2 +E**

| Coefficients: |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std.Error | t | value | Pr(>|t|) |
| (Intercept) | 151.318 | 12.776 | 11.844 | 3.34E-12 | \*\*\* |
| Price.Eggs | -18.727 | 1.882 | -9.953 | 1.57E-10 | \*\*\* |
| Price.Cookies | -8.786 | 2.369 | -3.709 | 0.00095 | \*\*\* |

**Interpretation:** For both the variables p values is 0 ie less than α.So we reject the null the hypothesis and we conclude that the β-coefficient of both the variables are statistically significant

**Step3:** Regression Model

Sales=151.318-18.727(Price of egg)-8.786(Price of cookies)+E

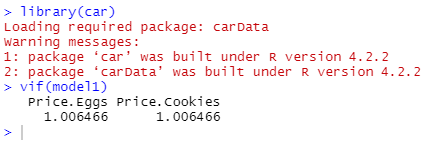
**Interpretation:** If we increase the price of egg by 1 unit sales will decrease by 18.727 unit,similarly we increase the price of cookies by 1 unit sales will decrease by 8.786.

**Step4:**  R Square means coefficient of determination(strength of model)

Here R Square value =0.8157 that means my model will predict the variation only 81.57% for the dependent variable with respect to the changes in the independent variables, and remaining 18.5% variation is due the external factors

**Step5:** Multicollinearity between independent variables.

If there is high correlation between the variables then we can say that multicollinearity is present between the variables.



| Price.Eggs | Price.Cookies |
| --- | --- |
| 1.006466 | 1.006466 |

Since the vif values is less than 5 for all independent variables there is no multicollinearity is present between the variables

**Data Analysis:**

**Step1:** Hypothesis for multivariate linear regression model.

Null Hypothesis(H0 ): The overall model is not statistically significant.

Alternative Hypothesis(H1): The overall model is statistically significant.

α= level of significance(0.5)

If p<α => Reject H0

If p>α => We do not reject H0

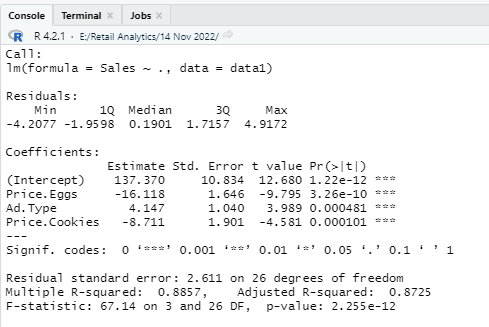
Here p value =2.255e-12 which is less than α we reject the null hypothesis and accept H1 and therefore we can say that the model is statistically significant.

**#linear regression model with the dummy variable**

model2 <-lm(Sales~Price.Eggs+Price.Cookies+Ad.Type,data=data1)

model2 <-lm(Sales~.,data=data1)

summary(model2)



**Step2:** Hypothesis for β-coefficient.

H0i: All the β-coefficient. are not statistically significant

H1i: At least one of the β-coefficient is statistically significant

**y = α+ β1.x1 + β2.x2 +E**

| Coefficients: |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std.Error | t | value | Pr(>|t|) |
| (Intercept) | 137.37 | 10.834 | 12.68 | 1.22E-12 | \*\*\* |
| Price.Eggs | -16.118 | 1.646 | -9.795 | 3.26E-10 | \*\*\* |
| Ad.Type | 4.147 | 1.04 | 3.989 | 0.000481 | \*\*\* |
| Price.Cookies | -8.711 | 1.901 | -4.581 | 0.000101 | \*\*\* |

**Interpretation:** For both the variables p values is 0 ie less than α.So we reject the null the hypothesis and we conclude that the β-coefficient of both the variables are statistically significant

**Step3:** Regression Model

Sales=137.3-16.118(Price of egg)+4.147(Ad.Type)-8.711(Price of cookies)+E

**Interpretation:** If we increase the price of egg by 1 unit sales will decrease by 16.118 unit,similarly we increase the price of cookies by 1 unit sales will decrease by 8.711

B value for base variable=137.3+4.147

**Step4:**  R Square means coefficient of determination(strength of model)

Here R Square value = 0.8857 that means my model will predict the variation only 88.57% for the dependent variable with respect to the changes in the independent variables, and remaining 11.43% variation is due the external factors

**Step5:** Multicollinearity between independent variables.

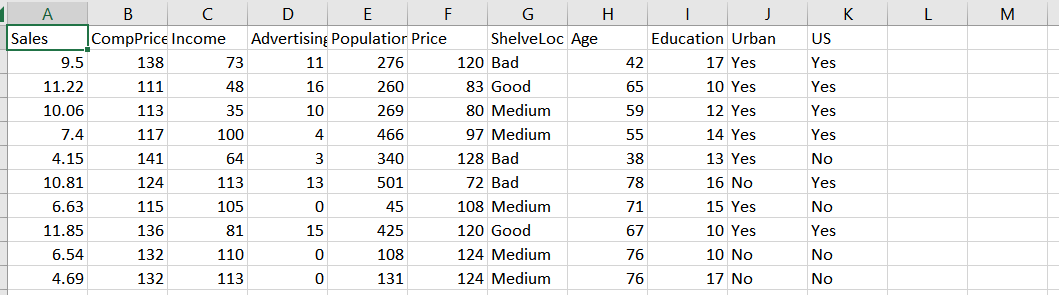
If there is high correlation between the variables then we can say that multicollinearity is present between the variables.



Since the vif values is less than 5 for all independent variables there is no multicollinearity is present between the variables

Date: 16-11-2022

Logistic Regression

****

Dependent variable: US\_yes

Independent variables:Income ,Education ,Sales

setwd("C:/Users/T2910/Desktop/Desktop/MCA 201-23/TRIM 5/Retail Analytics/16 Nov 2022")

data1 <-read.csv("carseats.csv")

variable.names(data1)

#Logistic Regression

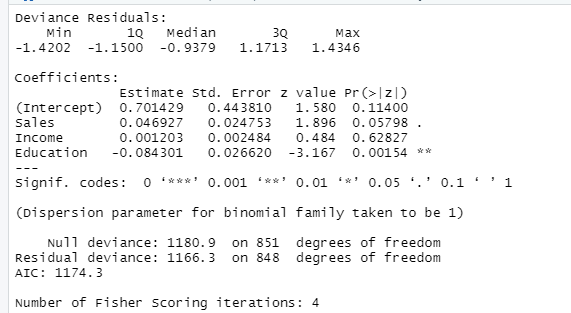
#Us\_yes->dependent variable

#creating the dummy variables for US

data1$US\_yes<-ifelse(data1$US=="Yes",1,0)

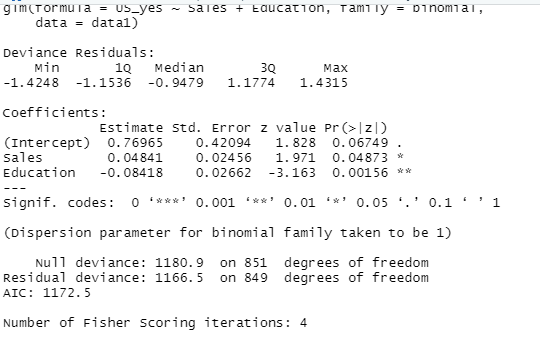
model1<-glm(US\_yes~Sales+Income+Education,data=data1,family = binomial)

summary(model1)



model2<-glm(US\_yes~Sales+Education,data=data1,family = binomial)

summary(model2)



log(P(Y=1) / 1-P(Y=1)) = Z = A + B1x1 + B2x2 = 0.7695 +0.04841 -0.08418