Exercise7

Mohammad Imtiaz Nur

3/27/2021

ID: 1878074

```
data("marketing", package = "datarium")
head(marketing, 5)
```

```
##   youtube facebook newspaper sales
## 1  276.12    45.36     83.04 26.52
## 2   53.40    47.16     54.12 12.48
## 3   20.64    55.08     83.16 11.16
## 4  181.80    49.56     70.20 22.20
## 5  216.96    12.96     70.08 15.48
```

Task 0: Multiple Linear Regression Model

Multiple linear regression model as follows to estimate sales based on the advertising budget for youtube, facebook and newspaper:

$$sales = b0 + b1 * youtube + b2 * facebook + b3 * newspaper$$

```
model <- lm(sales ~ youtube + facebook + newspaper, data = marketing)
model_summary <- summary(model)
model_summary
```

```
##
## Call:
## lm(formula = sales ~ youtube + facebook + newspaper, data = marketing)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5932  -1.0690   0.2902   1.4272   3.3951
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.526667   0.374290   9.422   <2e-16 ***
## youtube      0.045765   0.001395  32.809   <2e-16 ***
## facebook     0.188530   0.008611  21.893   <2e-16 ***
## newspaper   -0.001037   0.005871  -0.177     0.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1

```
## 
## Residual standard error: 2.023 on 196 degrees of freedom
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8956
## F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```
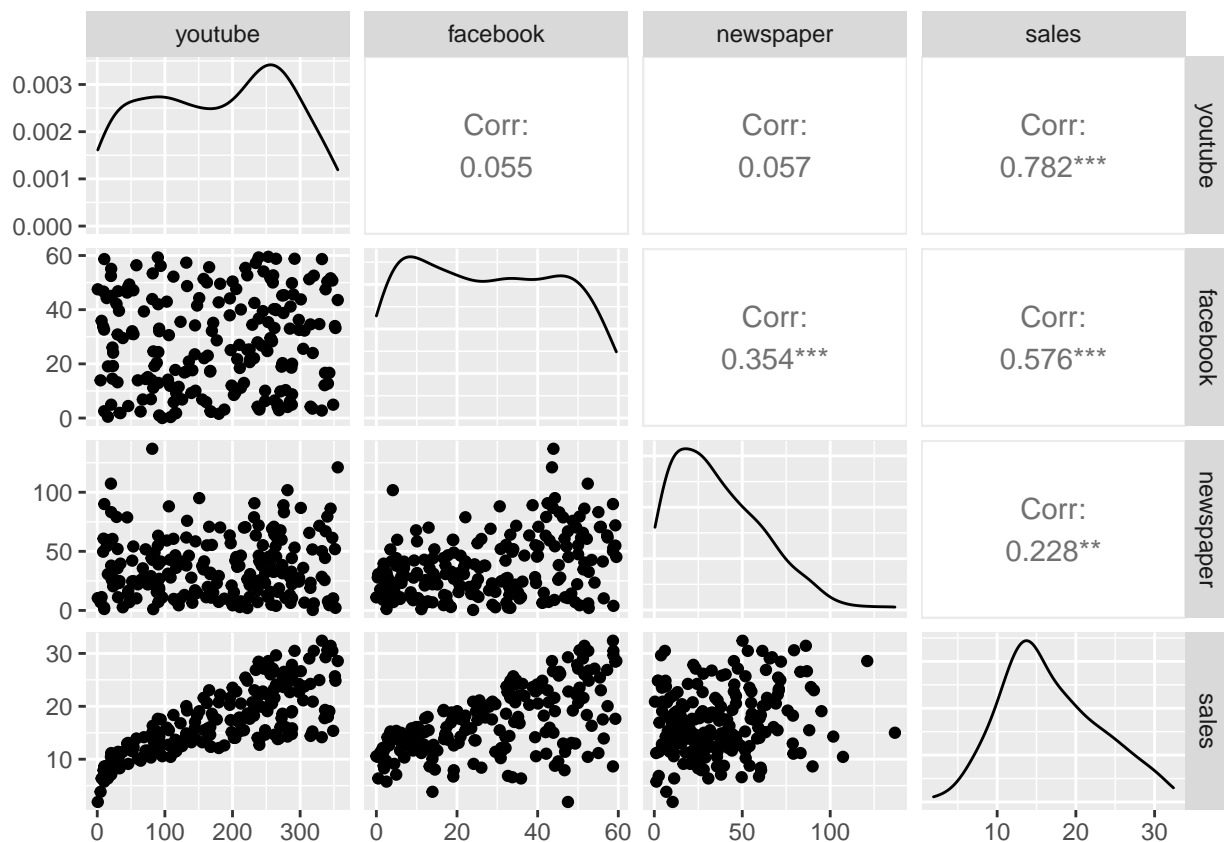
Task 1: Plots
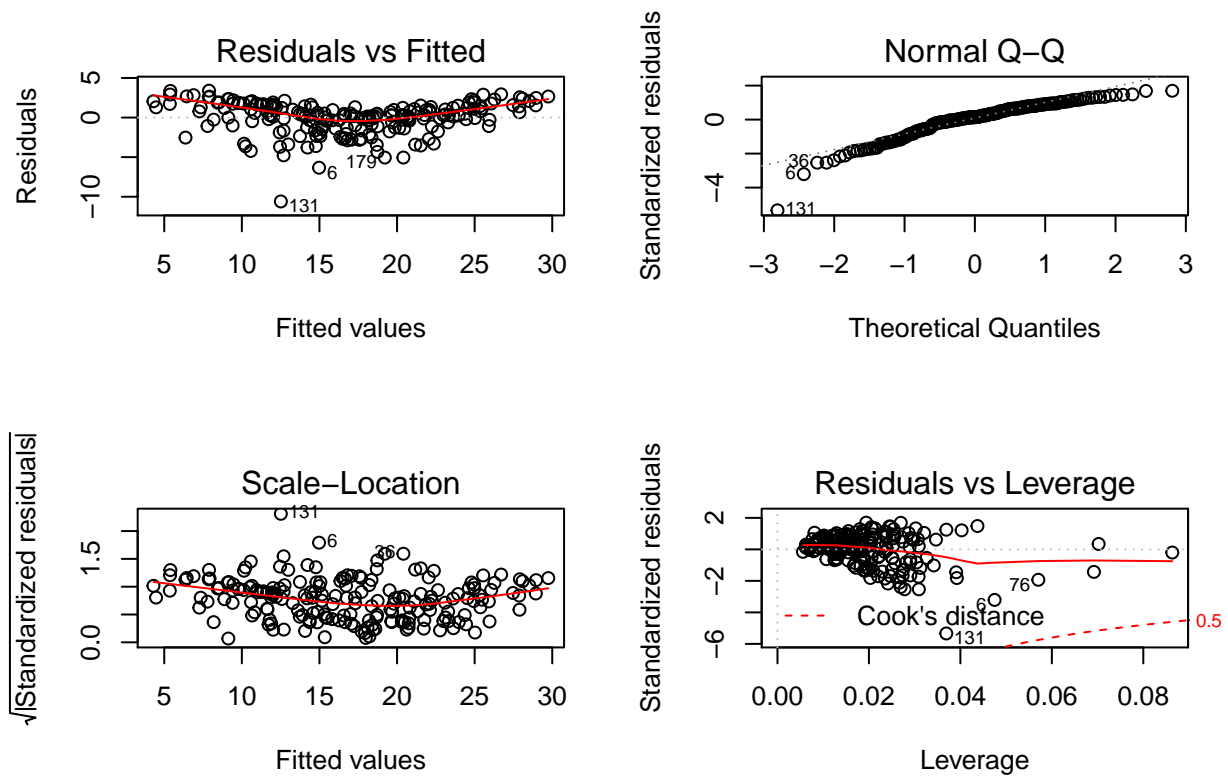
```r
library(GGally)
```

```
## Loading required package: ggplot2
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
ggpairs(marketing)
```



```r
par(mfrow=c(2, 2))
plot(model)
```

Residuals vs Fitted

Residuals

Fitted values

6 179 131

Normal Q–Q

Standardized residuals

Theoretical Quantiles

36 60 131

Scale–Location

$\sqrt{|\text{Standardized residuals}|}$

Fitted values

131 6

Residuals vs Leverage

Standardized residuals

Leverage

76 6 131

Cook's distance

0.5

Task 2: Examining F-statistic

Let,

H_0 = No predictor variables are significantly related to the outcome variable.

H_1 = At least one predictor variables is significantly related to the outcome variable.

From the above model, p-value of the F-statistic is $< 2.2\text{e-}16$, which is highly significant. Hence, we can reject the null hypothesis as at least one of the predictor variables is significantly related to the outcome variable.

Task 3: Examining Predictor Variable

We can examine the coefficients table, which shows the estimate of regression beta coefficients and the associated t-statistic p-values.

```
summary(model)$coef
```

```
##                 Estimate  Std. Error    t value     Pr(>|t|)
## (Intercept)  3.526667243 0.374289884  9.4222884 1.267295e-17
## youtube      0.045764645 0.001394897 32.8086244 1.509960e-81
## facebook     0.188530017 0.008611234 21.8934961 1.505339e-54
## newspaper   -0.001037493 0.005871010 -0.1767146 8.599151e-01
```

For a given predictor, the t-statistic evaluates whether or not there is significant association between the predictor and the outcome variable, i.e. whether the beta coefficient of the predictor is significantly different from zero.

From the t-values it can be seen that, changing in youtube and facebook advertising budget are significantly associated to changes in sales while changes in newspaper budget is not significantly associated with sales.

So the significant predictor variables are Youtube and Facebook.

Task 4: Regression Equation

$sales = 3.53 + 0.046 * youtube + 0.188 * facebook + (-0.001) * newspaper$

We can remove the newspaper variable from the model as it is not significant.

```
updated_model <- lm(sales ~ youtube + facebook, data = marketing)
summary(updated_model)
```

```
##
## Call:
## lm(formula = sales ~ youtube + facebook, data = marketing)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5572  -1.0502   0.2906   1.4049   3.3994
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.50532    0.35339   9.919   <2e-16 ***
## youtube      0.04575    0.00139  32.909   <2e-16 ***
## facebook     0.18799    0.00804  23.382   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.018 on 197 degrees of freedom
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8962
## F-statistic: 859.6 on 2 and 197 DF,  p-value: < 2.2e-16
```

So, the multiple linear regression model equation can be written as follow:

$sales = 3.50 + 0.046 * youtube + 0.188 * facebook$

Task 5 : Recommendations

For the budget planning based on the reviewed data, as both Youtube and Facebook have a very strong positive correlation with sales, we certainly can say that increasing the advertisement budget in these two sector will increase the sales for the company. And an increase/cut in the advertising budget for newspaper will not significantly affect sales.