Cadreon R Test

Anna Andrianatou

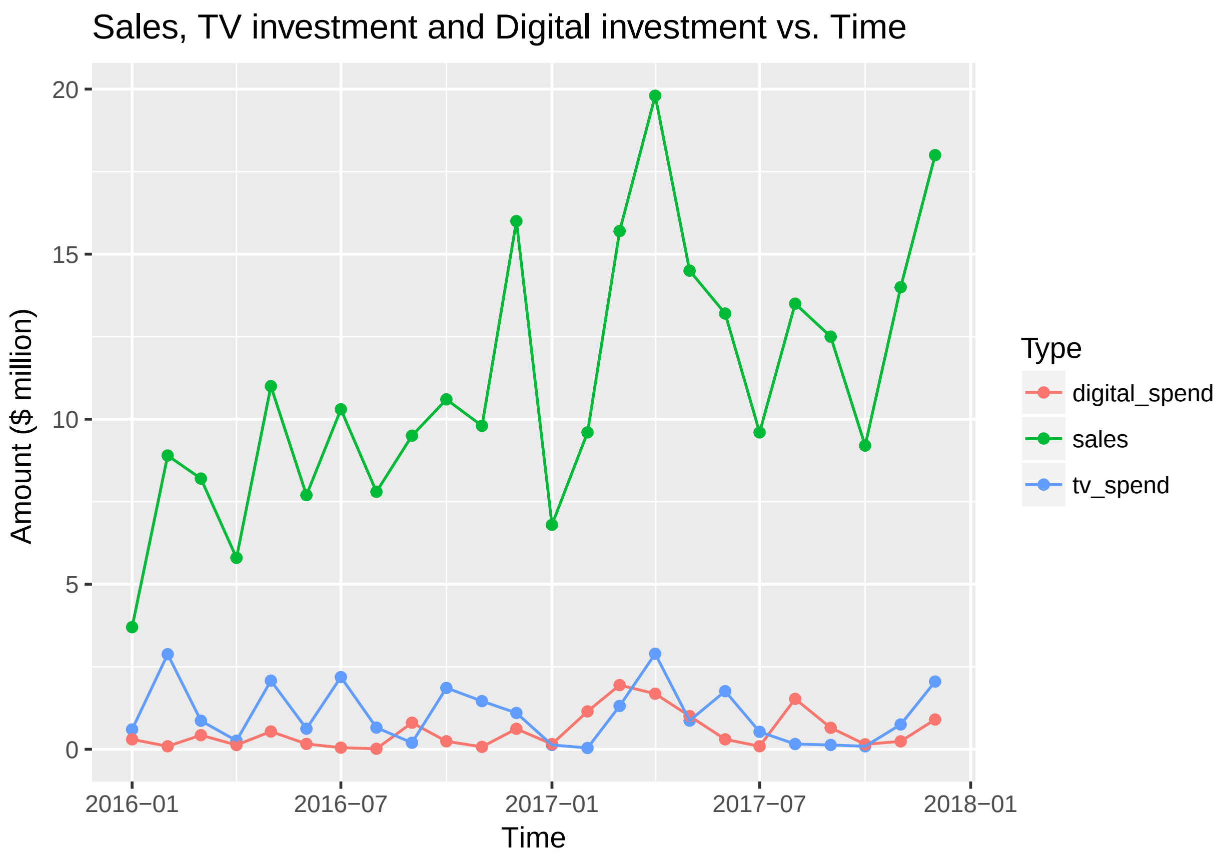
27/04/2018

The test

This is a R test given by Cadreon. It explores the sales of a toy brand across two years.

Question 2

A plot of sales, TV and digital investment is below. There does not appear to be any time trend in sales, except for a positive linear growth. The behaviour of sales does appear positively correlated with both TV investment and Digital investment.



Question 3

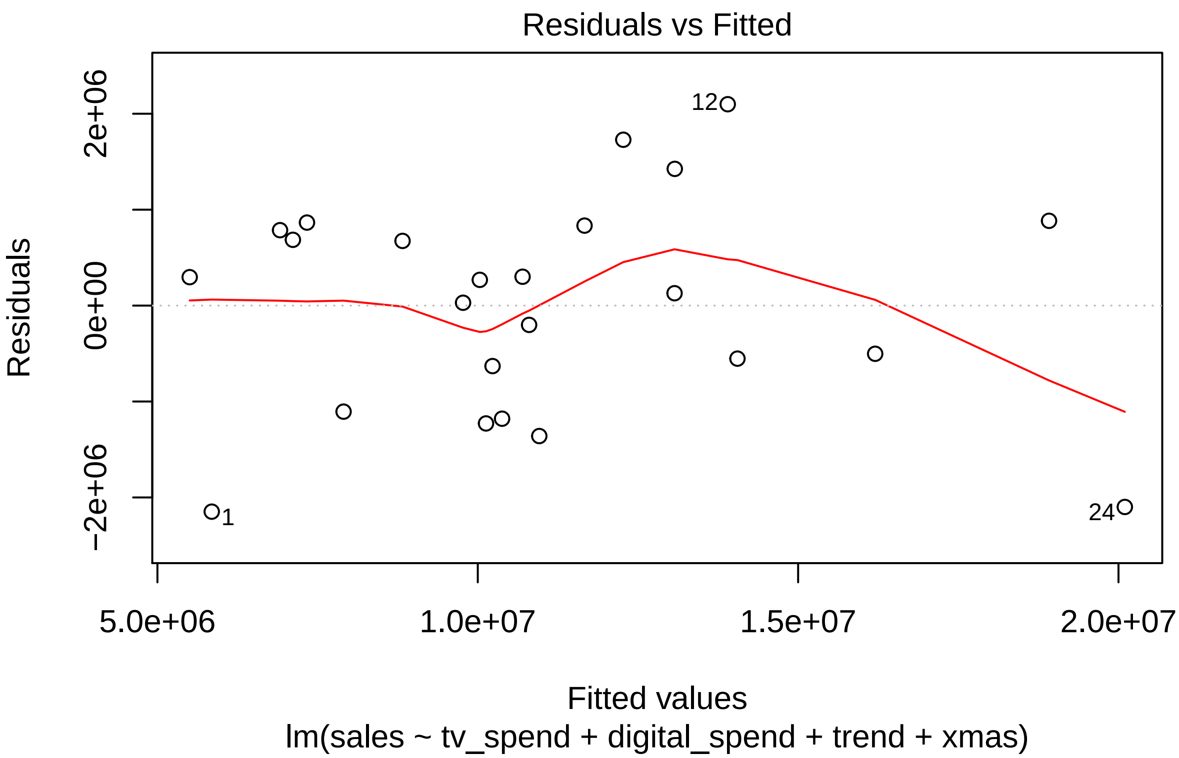
A correlation matrix between sales, TV investment and digital investment is given below.

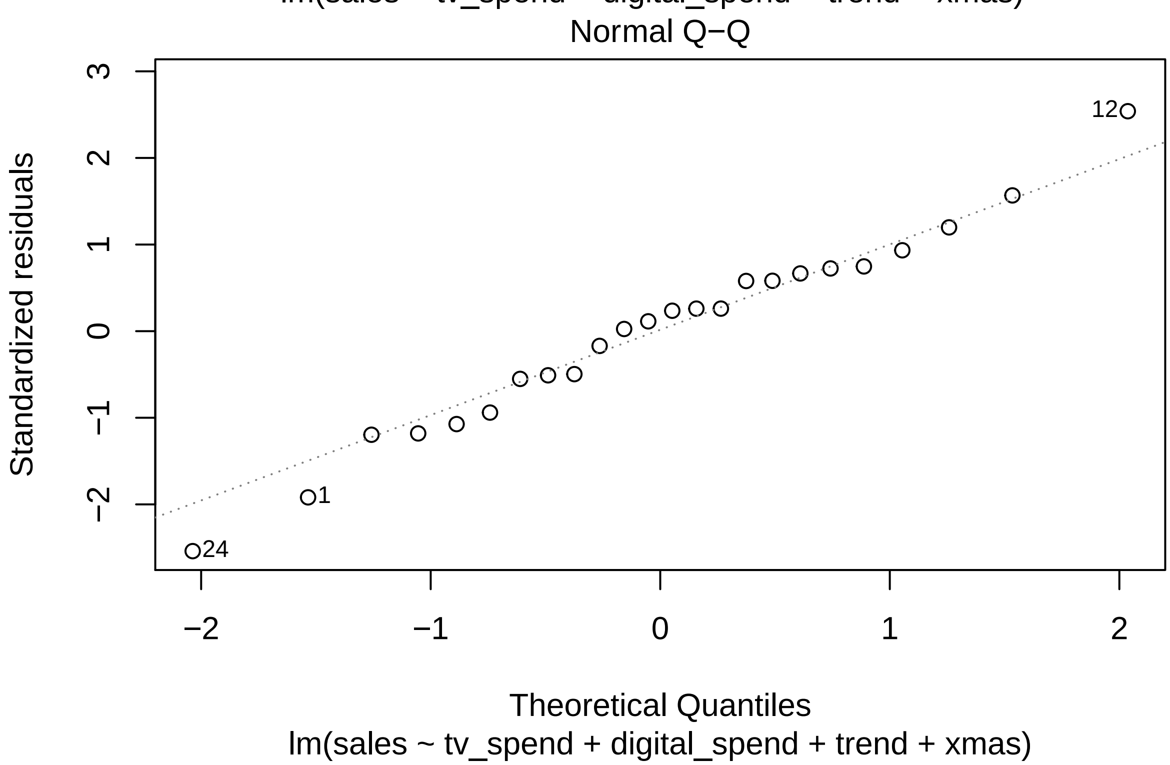
```
##                   sales  tv_spend digital_spend
## sales         1.0000000 0.4406862     0.6647654
## tv_spend      0.4406862 1.0000000     0.0720594
## digital_spend 0.6647654 0.0720594     1.0000000
```

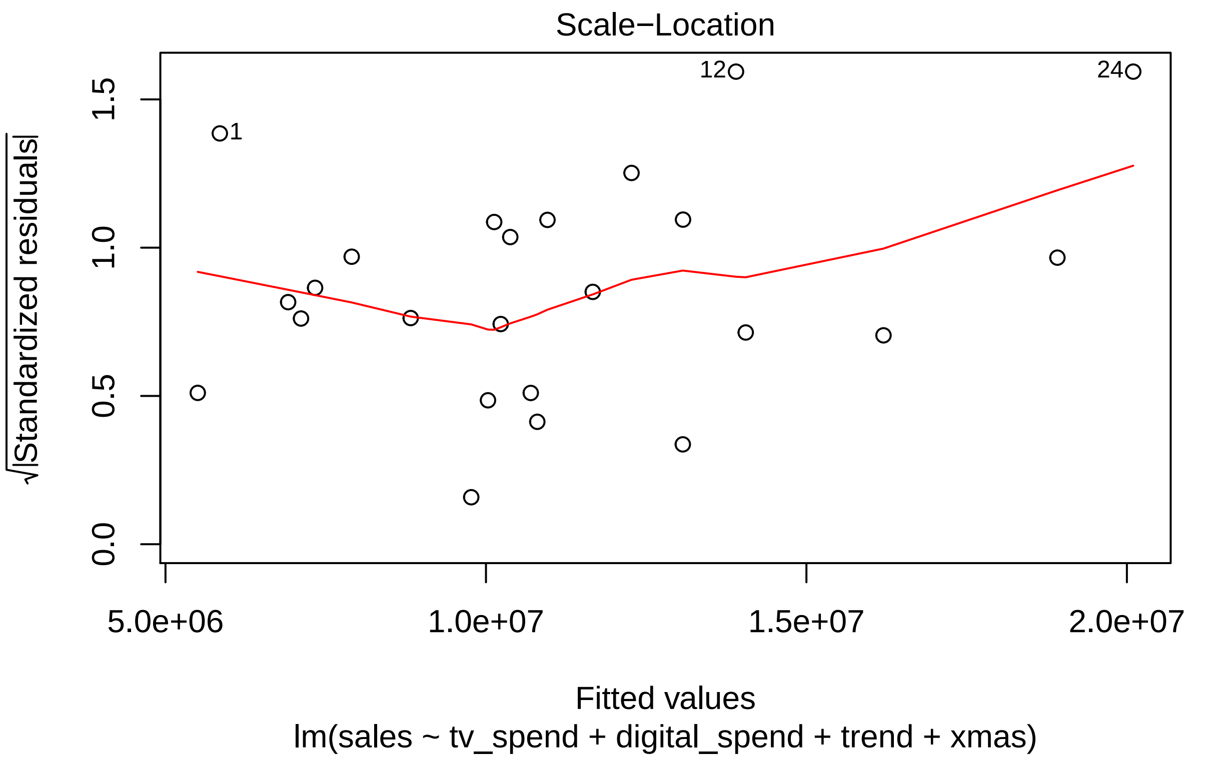
Sales is positively correlated with both types of investment. TV spending and digital spending do have some positive correlation, however is fairly small. Unlikely to cause a multicolinearity problem.

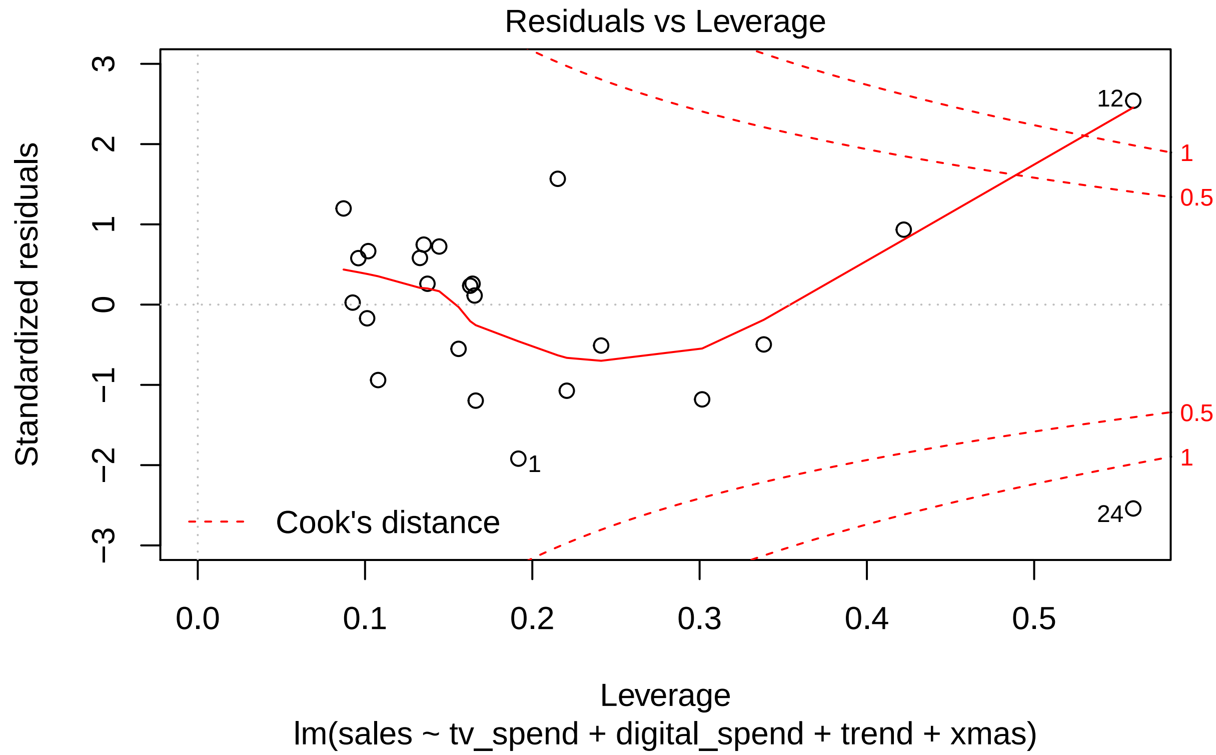
Question 4 a,b

The month variable in R converts to an integer and represents time in seconds (the spacing between each month is different due to the different number of days in various months). The trend variable is an indicator for each month. We are only really interested in the monthly trend and so the “trend” variable is included and “month” omitted.


Residuals vs Fitted
Residuals
2e+06
0e+00
−2e+06
12
1
24
5.0e+06
1.0e+07
1.5e+07
2.0e+07
Fitted values
lm(sales ~ tv_spend + digital_spend + trend + xmas)


Normal Q−Q
Standardized residuals
3
2
1
0
−1
−2
12
1
24
−2
−1
0
1
2
Theoretical Quantiles
lm(sales ~ tv_spend + digital_spend + trend + xmas)






```
##
## Call:
## lm(formula = sales ~ tv_spend + digital_spend + trend + xmas)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
```

```
## -2148071  -749281   198668   798127  2099176
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.451e+06  6.706e+05   5.146 5.74e-05 ***
## tv_spend      2.026e+00  3.071e-01   6.599 2.58e-06 ***
## digital_spend 2.983e+00  5.040e-01   5.919 1.07e-05 ***
## trend         2.863e+05  4.156e+04   6.888 1.44e-06 ***
## xmas          2.935e+06  9.730e+05   3.016  0.00711 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1245000 on 19 degrees of freedom
## Multiple R-squared:  0.9161, Adjusted R-squared:  0.8984
## F-statistic: 51.84 on 4 and 19 DF,  p-value: 5.817e-10
```

The adjusted R squared (adjusted for the number of variables in the model) is 0.8984. Therefore, approximately 90% of the variation in the sales data is explained by the model.

Assume a level of significance of 5%:

TV spend regressor has an associated p-value of 2.58e-06 (<0.05) and therefore there is statisitcally significant in the model (coefficient $\neq 0$).

Digital spend regressor has an associated p-value of 1.07e-05 and therefore there is statisitcally significant in the model (coefficient $\neq 0$).

Trend has an associated p-value of 1.44e-06 and therefore there is statisitcally significant in the model (coefficient $\neq 0$). There is some growth over time.

xmas has an associated p-value of 0.00711 and therefore there is statisitcally significant in the model (coefficient $\neq 0$). X-mas effects sales positively.

Question 5 From this regression, for every dollar in TV_spend, sales increase by $2.026

To find % change, fit a log-linear regression and interpret the coefficient of TV_spend

```
##
## Call:
## lm(formula = log(sales) ~ tv_spend + digital_spend + trend +
##     xmas)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.50464 -0.05648  0.01681  0.10188  0.25602
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.541e+01  9.255e-02 166.503  < 2e-16 ***
## tv_spend      1.957e-07  4.238e-08   4.618 0.000188 ***
## digital_spend 2.310e-07  6.955e-08   3.321 0.003590 **
## trend         3.158e-02  5.737e-03   5.506  2.6e-05 ***
## xmas          1.843e-01  1.343e-01   1.372 0.185995
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1718 on 19 degrees of freedom
## Multiple R-squared:  0.8308, Adjusted R-squared:  0.7951
## F-statistic: 23.32 on 4 and 19 DF,  p-value: 4.167e-07
```

From this log-linear regression, for every dollar in TV_spend, sales increase by 0.00001957%. ## Question 6 $ROI = (gains - investmentcost)/investmentcost$ ($2.026 for every $1 in investment)

Question 7

plug in the values to the regression

```
##        1        2        3
## 11958795 13267129 15109712
```

Question 8

The purpose of the model is to observe the return to investment of TV and digital ads and to predict future sales. Therefore, any additional data that could explain sales would be beneficial to the model. This could include any other important events other than xmas that could increase the sales of toys, prices or sales of competing toy brands or subsitution items(such as video games), the number of stores selling the toy brand in each month.

It is important to capture the impact of effects other than the advertising campaign to ensure the impact of the ads are not over estimated.