Homework 11

Code

6210422036: Tanat Iempreedee

Hide

```
library(fpp2)
library(forecast)
library(ggplot2)
library(dplyr)

da <- read.table("Advert.txt", header=TRUE, colClasses=c('numeric','factor','numeric'))
str(da)
```

```
'data.frame':   22 obs. of  3 variables:
 $ advert: num  101 107 109 122 124 ...
 $ type  : Factor w/ 3 levels "Highpro","Lowpro",..: 3 1 1 3 2 2 1 1 1 2 ...
 $ sales : num  875 928 937 929 943 ...
```

EDA

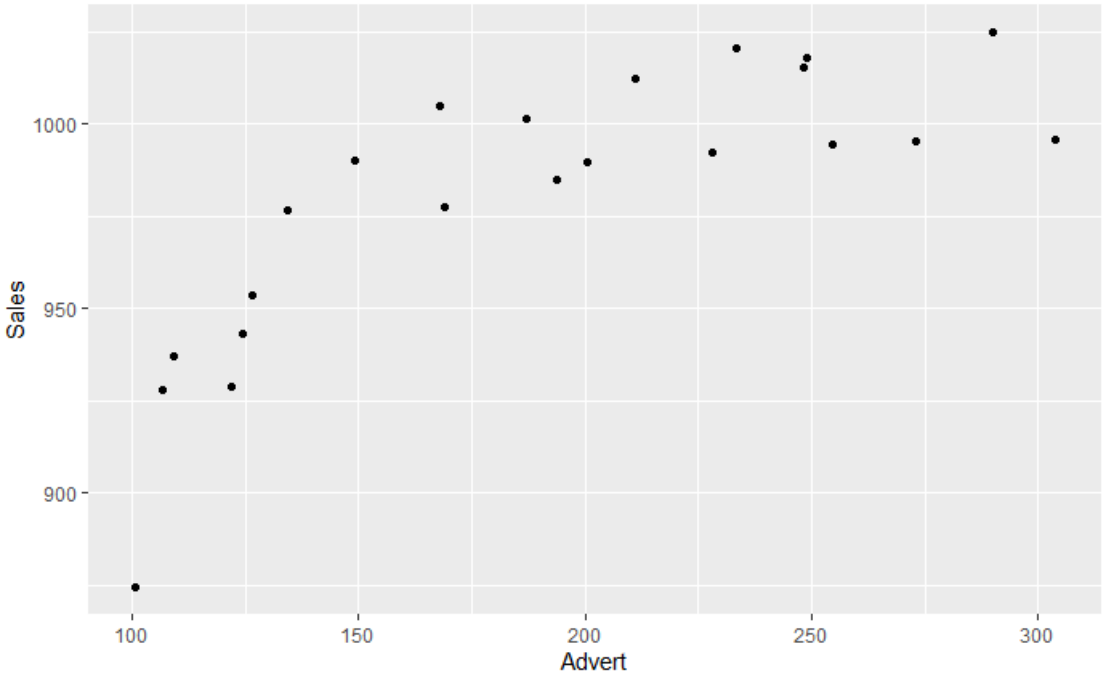
Hide

```
cor(da$advert, da$sales)
```

```
[1] 0.7908864
```

Hide

```
ggplot(data=da, aes(x=advert, y=sales)) +
  geom_point() +
  xlab('Advert') + ylab('Sales')
```



Transform

Hide

```
# Transform
da <- da %>% mutate(advert_sq=advert^2) %>%
  mutate(advert_log=log(advert)) %>%
  mutate(sales_log=log(sales))
head(da)
```

	advert <dbl>	type <fctr>	sales <dbl>	advert_sq <dbl>	advert_log <dbl>	sales_log <dbl>
1	100.57	Special	874.53	10114.32	4.610854	6.773687
2	106.80	Highpro	928.10	11406.24	4.670958	6.833139
3	109.26	Highpro	937.00	11937.75	4.693730	6.842683
4	121.78	Special	928.90	14830.37	4.802216	6.834001
5	124.46	Lowpro	942.93	15490.29	4.823984	6.848992
6	126.34	Lowpro	953.43	15961.80	4.838977	6.860066

6 rows

Hide

```
cor(da$advert_log, da$sales_log)
```

```
[1] 0.8461091
```

Model 1: Log-Log

Hide

```
# Model1: log-log
reg1 <- lm(sales_log ~ advert_log, data=da)
summary(reg1)
```

```

Call:
lm(formula = sales_log ~ advert_log, data = da)

Residuals:
      Min        1Q    Median        3Q       Max
-0.057827 -0.008113  0.001769  0.014258  0.032335

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.39213    0.06983  91.537  < 2e-16 ***
advert_log   0.09529    0.01342   7.099 7.01e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02154 on 20 degrees of freedom
Multiple R-squared:  0.7159,    Adjusted R-squared:  0.7017
F-statistic:  50.4 on 1 and 20 DF,  p-value: 7.01e-07
```

Hide

```
AIC(reg1)
```

```
[1] -102.5244
```

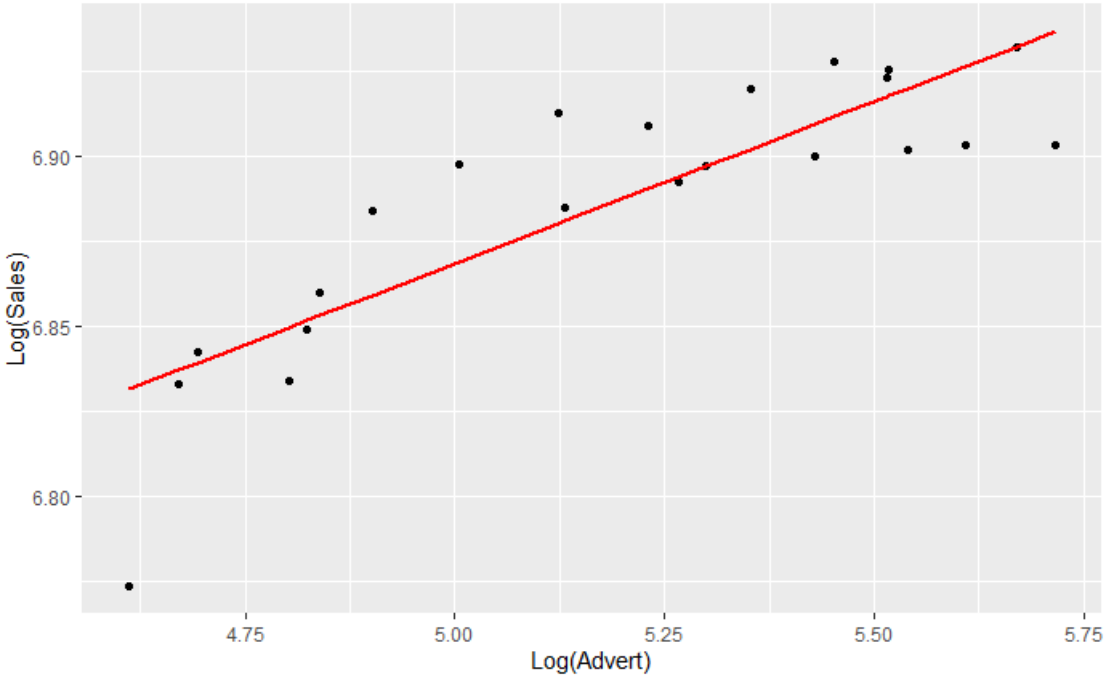
Hide

```
CV(reg1)
```

```
          CV           AIC          AICc           BIC         AdjR2
 5.489413e-04 -1.649577e+02 -1.636244e+02 -1.616846e+02  7.016957e-01
```

Hide

```
ggplot(data=da, aes(x=advert_log, y=sales_log)) +
  geom_point() +
  geom_line(aes(y=fitted(reg1)), size=1, color='red') +
  xlab("Log(Advert)") + ylab("Log(Sales)")
```



Coefficient Interpretation - Log-log

$ln(Sales) = 6.39213 + 0.09529ln(Advert)$
$\beta_1 = 0.09529$

1% increase in Advert -> 0.09% increase in Sales

Model 2: Quadratic

Hide

```
# Model2: Quadratic
reg2 <- lm(sales ~ advert + advert_sq, data=da)
summary(reg2)
```

```

Call:
lm(formula = sales ~ advert + advert_sq, data = da)

Residuals:
    Min      1Q  Median      3Q     Max
-38.289 -12.632   3.530   8.145  22.832

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 722.143426  37.826304  19.091 7.41e-14 ***
advert        2.390092   0.414941   5.760 1.50e-05 ***
advert_sq    -0.004913   0.001052  -4.673 0.000166 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.57 on 19 degrees of freedom
Multiple R-squared:  0.8257,    Adjusted R-squared:  0.8074
F-statistic: 45.02 on 2 and 19 DF,  p-value: 6.184e-08
```

Hide

```
AIC(reg2)
```

```
[1] 190.74
```

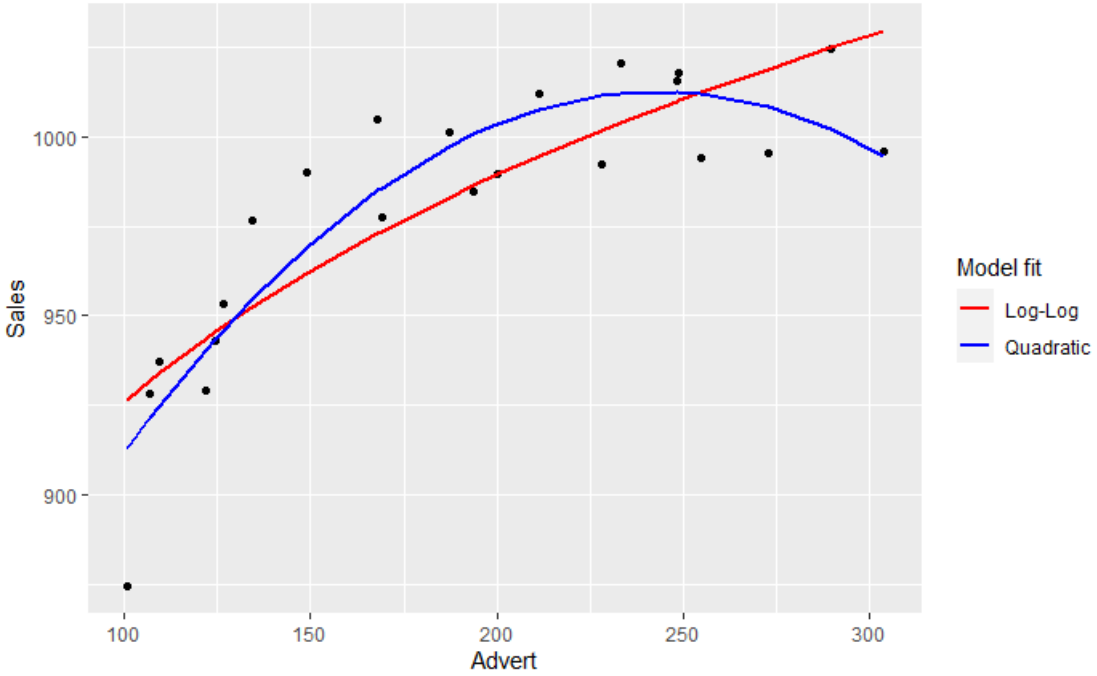
Hide

```
CV(reg2)
```

```
        CV         AIC        AICc         BIC       AdjR2
342.5593026 128.3066974 130.6596385 132.6708672   0.8074014
```

Hide

```
ggplot(data=da, aes(x=advert, y=sales)) +
  geom_point() +
  geom_line(aes(y=exp(fitted(reg1)),color='red'), size=1) +
  geom_line(aes(y=fitted(reg2), color='blue'), size=1) +
  xlab('Advert') + ylab('Sales') +
  scale_color_identity(name = "Model fit",
                       breaks = c("red", "blue"),
                       labels = c("Log-Log", "Quadratic"),
                       guide = "legend")
```



Choose between log-log and quadratic

Compare Cross Validation Result

Hide

```
rbind(LOGLOG=CV(reg1), QUADRATIC=CV(reg2))
```

```
                  CV       AIC      AICc       BIC     AdjR2
LOGLOG    5.489413e-04 -164.9577 -163.6244 -161.6846 0.7016957
QUADRATIC 3.425593e+02  128.3067  130.6596  132.6709 0.8074014
```

Interpretation

Choose **Log-log** model which gives the lower AIC, even though Quadratic gives the better Adjusted R Squared.
From the problem context, increasing advertisement expenditure should not decrease Sales. Instead, the growth of Sales could saturate at some point which aligns more with logarithmic trend.