# Chapter 2 Problem 1

*Andira Putri*

For the Johnson & Johnson data, say $y_t$, shown in Figure 1.1, let $x_t = log(y_t)$. Fit the regression model: $x_t = \beta t + \alpha_1 Q_1(t) + \alpha_2 Q_2(t) + \alpha_3 Q_3(t) + \alpha_4 Q_4(t) + w_t$ where $Q_i(t) = 1$ if time t corresponds to quarter i=1,2,3,4, and zero otherwise. The $Q_i(t)$'s are called indicator variables. We will assume for now that wt is a Gaussian white noise sequence. What is the interpretation of the parameters $\beta, \alpha_1, \alpha_2, \alpha_3$, and $\alpha_4$? What happens if you include an intercept term in the model? Graph the data, $x_t$, and superimpose the fitted values on the graph. Examine the residuals, $x_t - \hat{x}_t$ and state your conclusions. Does it appear that the model fits the data well (do the residuals look white)?

```r
#import data
library(astsa)
data(jj)
#create indicator variables
Q1=rep(c(1,0,0,0),length(jj)/4)
Q2=rep(c(0,1,0,0),length(jj)/4)
Q3=rep(c(0,0,1,0),length(jj)/4)
Q4=rep(c(0,0,0,1),length(jj)/4)
#fit regression model WITHOUT intercept
fit=lm(log(jj)~time(jj)+Q1+Q2+Q3+Q4-1)
summary(fit)
```

```
##
## Call:
## lm(formula = log(jj) ~ time(jj) + Q1 + Q2 + Q3 + Q4 - 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## time(jj)  1.672e-01  2.259e-03   74.00   <2e-16 ***
## Q1       -3.283e+02  4.451e+00  -73.76   <2e-16 ***
## Q2       -3.282e+02  4.451e+00  -73.75   <2e-16 ***
## Q3       -3.282e+02  4.452e+00  -73.72   <2e-16 ***
## Q4       -3.284e+02  4.452e+00  -73.77   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9935, Adjusted R-squared:  0.9931
## F-statistic:  2407 on 5 and 79 DF,  p-value: < 2.2e-16
```

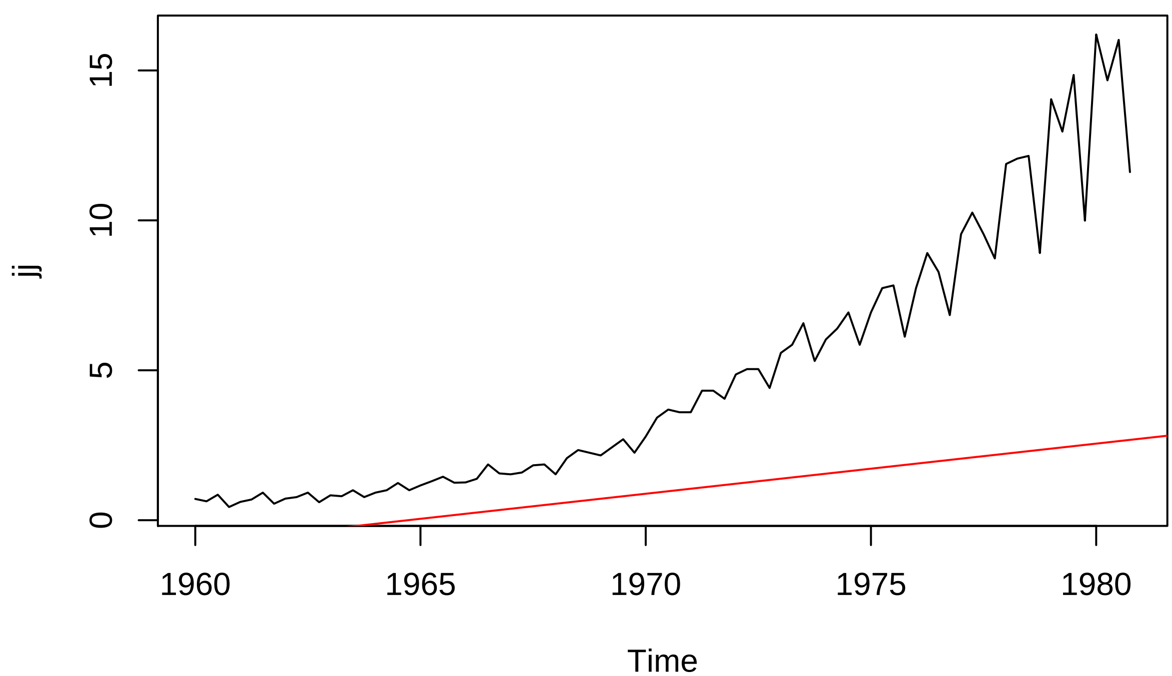
```r
#fit regression mdel WITH intercept
fit.2=lm(log(jj)~time(jj)+Q1+Q2+Q3+Q4)
summary(fit.2)
```

```
##
## Call:
## lm(formula = log(jj) ~ time(jj) + Q1 + Q2 + Q3 + Q4)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.284e+02  4.452e+00 -73.771  < 2e-16 ***
## time(jj)     1.672e-01  2.259e-03  73.999  < 2e-16 ***
## Q1           1.705e-01  3.873e-02   4.403 3.31e-05 ***
## Q2           1.986e-01  3.871e-02   5.132 2.01e-06 ***
## Q3           2.688e-01  3.870e-02   6.945 9.50e-10 ***
## Q4                  NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9859, Adjusted R-squared:  0.9852
## F-statistic:  1379 on 4 and 79 DF,  p-value: < 2.2e-16
```

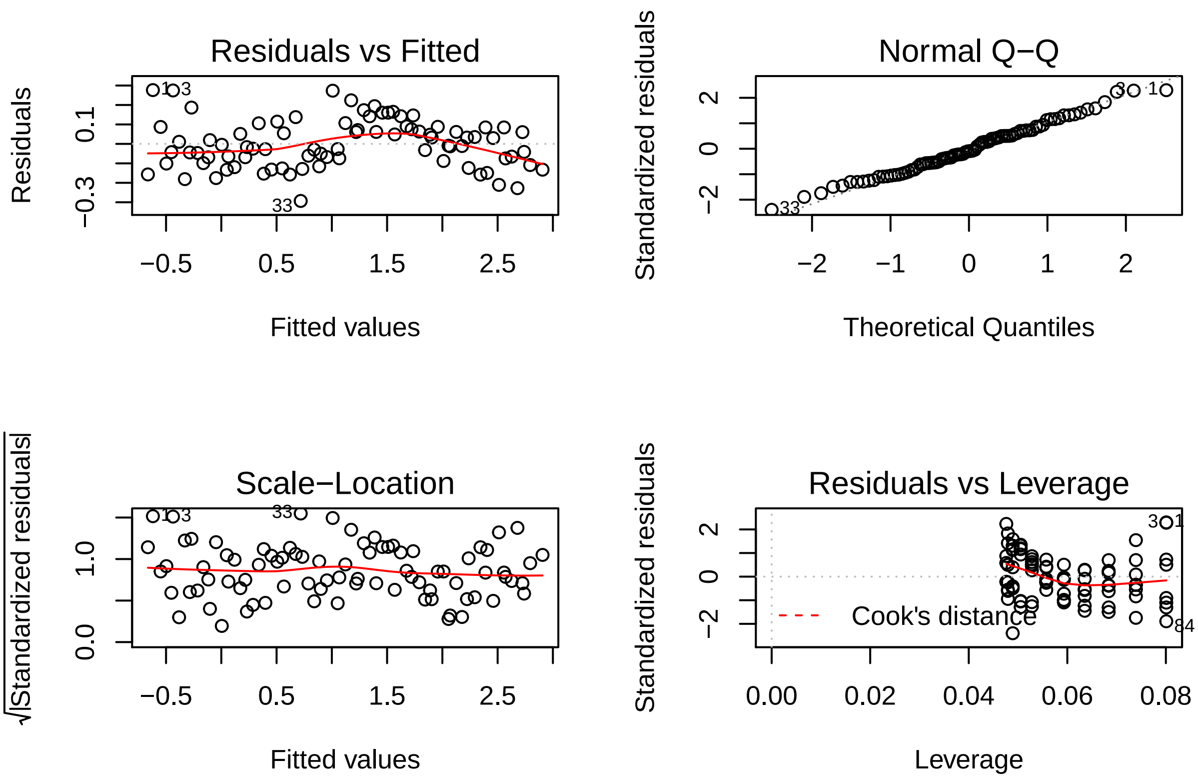
```r
plot(jj)
abline(fit.2,col="red")
```

```
## Warning in abline(fit.2, col = "red"): only using the first two of 6
## regression coefficients
```

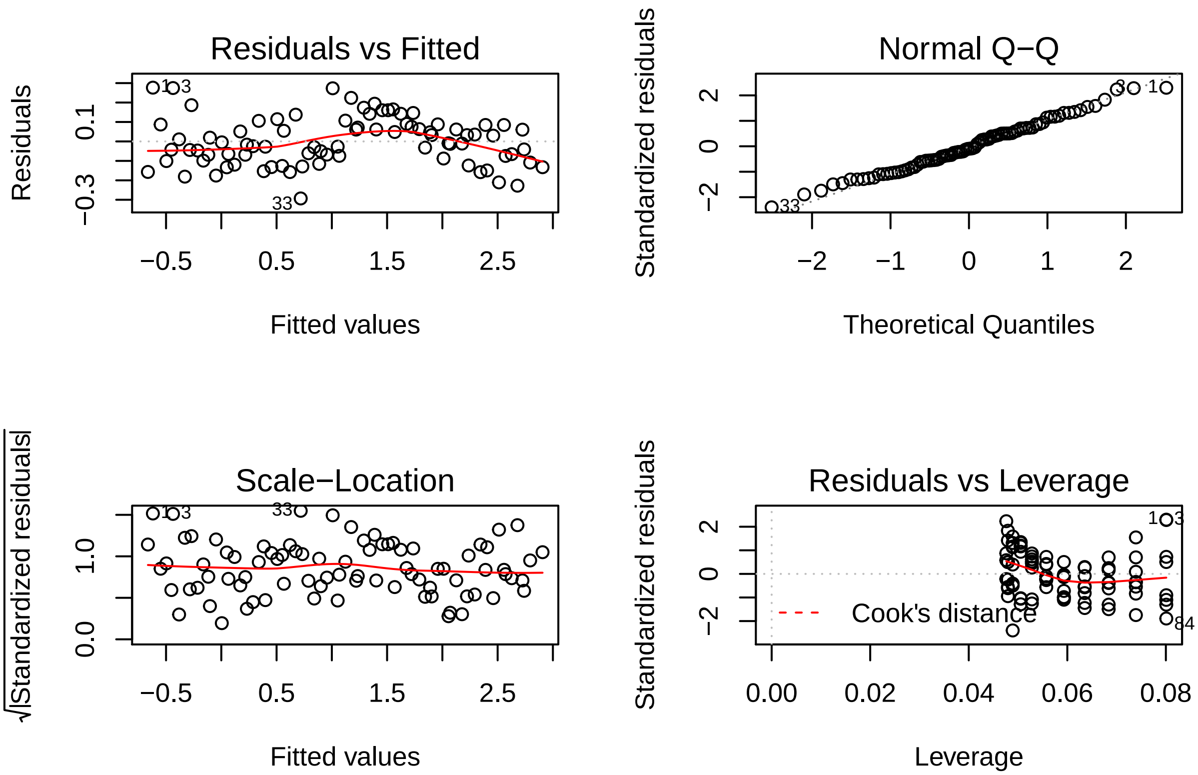



```r
#plot residuals
par(mfrow=c(2,2))
```

```r
plot(fit)
```




```r
plot(fit.2)
```



Residuals vs Fitted
Residuals
Fitted values
−0.3
0.1
−0.5
0.5
1.5
2.5
103
33
Normal Q−Q
Standardized residuals
Theoretical Quantiles
−2
0
2
−2
−1
0
1
2
103
33
Scale−Location
√|Standardized residuals|
Fitted values
0.0
1.0
−0.5
0.5
1.5
2.5
103
33
Residuals vs Leverage
Standardized residuals
Leverage
−2
0
2
0.00
0.02
0.04
0.06
0.08
Cook's distance
103
84