Regression Analysis HW4

2020 12 6

1.

a)

```r
load("C:/Users/admin/Downloads/dataQ1.RData")
fitted<-lm(Y~x1)
fitted
```

```
##
## Call:
## lm(formula = Y ~ x1)
##
## Coefficients:
## (Intercept)           x1
##     40652.8       -515.9
```

The fitted regression equation is

$$\hat{Y} = 40652.81 - 515.93 \times x1$$

b)

```r
summary(fitted)
```

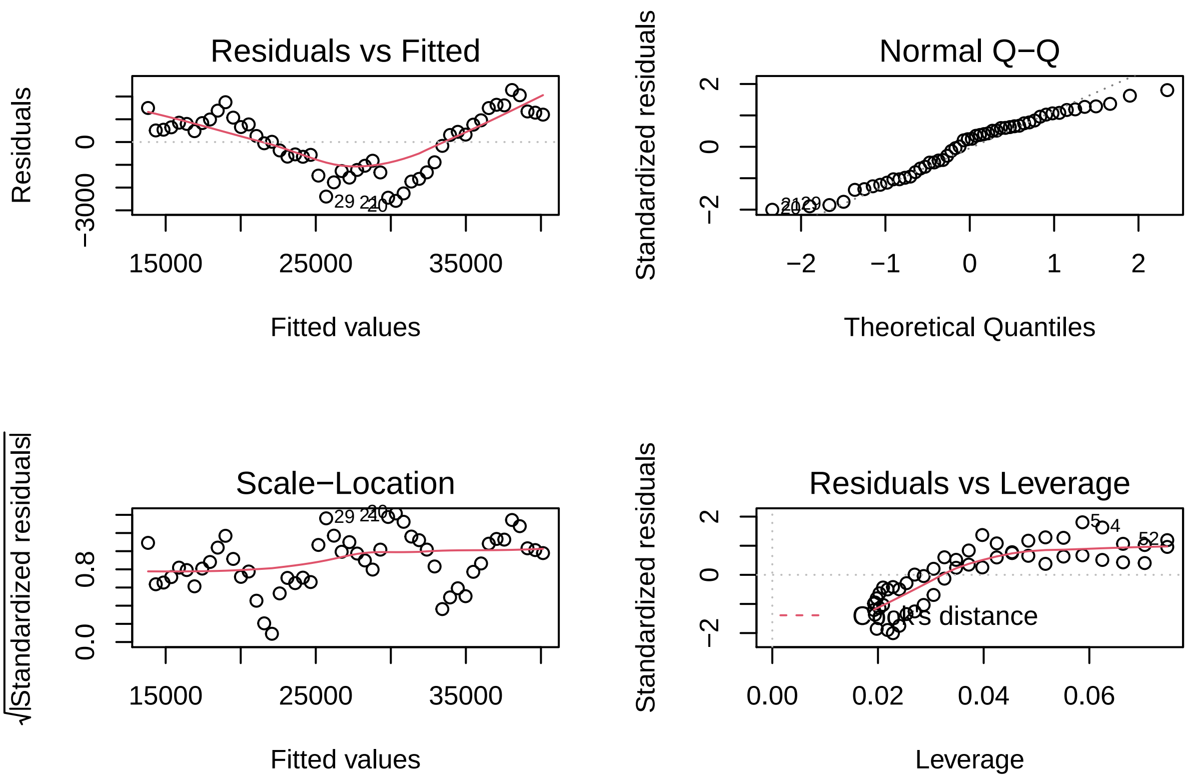
```
##
## Call:
## lm(formula = Y ~ x1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2584.2 -1088.3   317.3   962.7  2283.8
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40652.81     367.18  110.72   <2e-16 ***
## x1           -515.93      12.06  -42.79   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1305 on 50 degrees of freedom
## Multiple R-squared:  0.9734, Adjusted R-squared:  0.9729
## F-statistic:  1831 on 1 and 50 DF,  p-value: < 2.2e-16
```

In the summary above, we can find

$$R^2 = 0.9734$$

c)

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



We can see the pattern in first residual plot. It means the errors are autocorrelated.

d)

$$H_0 : \rho = 0, \quad H_1 : \rho \neq 0$$

```r
fitted$res %>% sign() %>% factor() %>% tseries::runs.test()
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

##
##  Runs Test
##
## data:  .
## Standard Normal = -6.149, p-value = 7.799e-10
## alternative hypothesis: two.sided
```

The test statistic is -6.149 and the p-value=7.799e-10 which means it is significant under $\alpha = 0.01$ Therefore we can reject H_0 and we can conclude the errors are autocorrelated.

e)

$$H_0 : \rho = 0 \quad vs \quad H_1 : \rho > 0$$

```r
lmtest::dwtest(fitted)
```

```
##
##  Durbin-Watson test
##
## data:  fitted
## DW = 0.11683, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

The test statistic is 0.11683 and the p-value<2.2e-16 and it means that it is significant under $\alpha = 0.01$ Therefore we can reject H_0 and we can conclude the errors are positively autocorrelated.

2.

a)

```r
load("C:/Users/admin/Downloads/dataQ2.RData")
fitted_1<-lm(y~n)
fitted_1
```

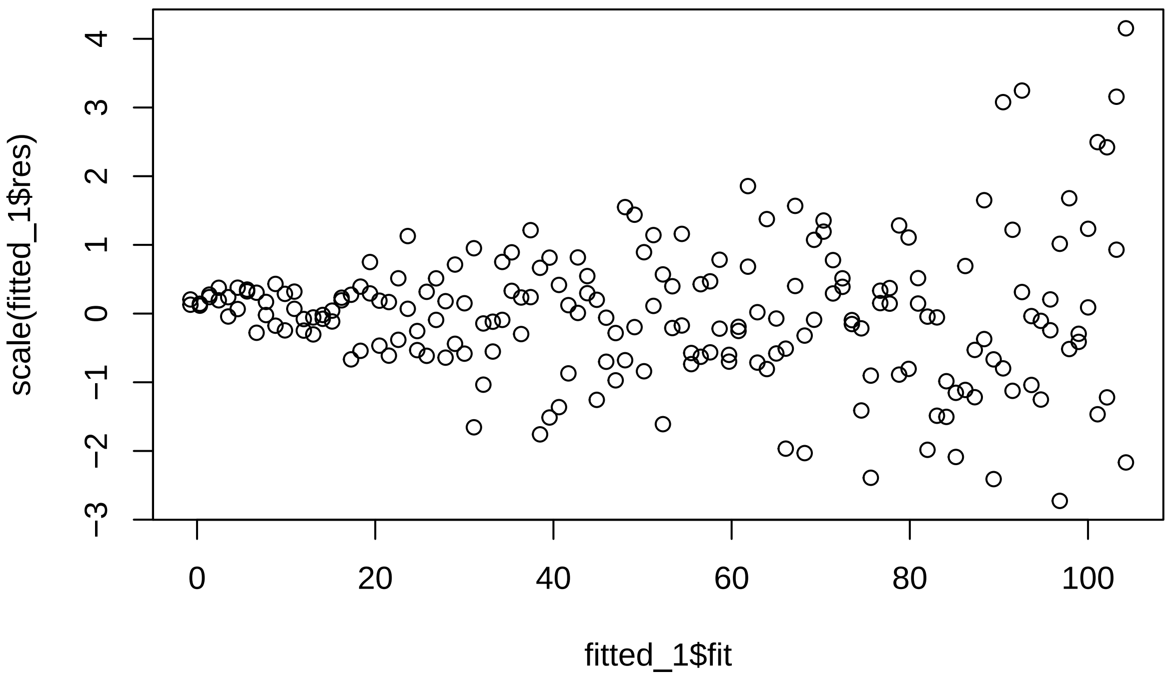
```
##
## Call:
## lm(formula = y ~ n)
##
## Coefficients:
## (Intercept)            n
##      -1.805        1.061
```

The fitted regression equation is

$$\hat{Y} = -1.805 + 1.061 \times n$$

b)

```r
plot(fitted_1$fit, scale(fitted_1$res))
```



In above plot, we can see the errors have heteroscedasticity and it violates the assumption of constancy of error variance.

c)

```
fitted_2<-lm(sqrt(y)~n)
fitted_2
```

```
##
## Call:
## lm(formula = sqrt(y) ~ n)
##
## Coefficients:
## (Intercept)            n
##     2.61275      0.08152
```

d)

The fitted regression model is

$$\hat{\sqrt{Y}} = 2.61275 + 0.08152 \times n$$

e)

```r
predict(fitted_2,newdata=data.frame(n=40))^2
```

```
##        1
## 34.49826
```

f)

```r
n_sq<-n^2
fitted_3<-lm(log(y)~n+n_sq)
summary(fitted_3)
```

```
##
## Call:
## lm(formula = log(y) ~ n + n_sq)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4744 -0.1900  0.0220  0.2386  0.7751
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.5104037  0.0768787   19.65   <2e-16 ***
## n            0.0674594  0.0035136   19.20   <2e-16 ***
## n_sq        -0.0003820  0.0000337  -11.34   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3552 on 197 degrees of freedom
## Multiple R-squared:  0.8619, Adjusted R-squared:  0.8605
## F-statistic: 614.8 on 2 and 197 DF,  p-value: < 2.2e-16
```

g) The fitted regression model in terms of the original scale is

$$\hat{Y} = exp(1.5104037 + 0.0674594 \times n - 0.0003820 \times n^2)$$