# ARA ASSIGNMENT #6

*Rhyz C. Gomez*

*February 8, 2016*

1. A substance used in biological and medical research is shipped by airfreight to users in cartons of 1000 ampules. The data below, involving 10 shipments, were collected on the number of times the carton was transferred from one aircraft to another over the shipment route (X) and the number of ampules found to be broken upon arrival (Y).

Table 1: Number of Ampules broken (Y) Due to Number of Times the Carton is Transferred (X)

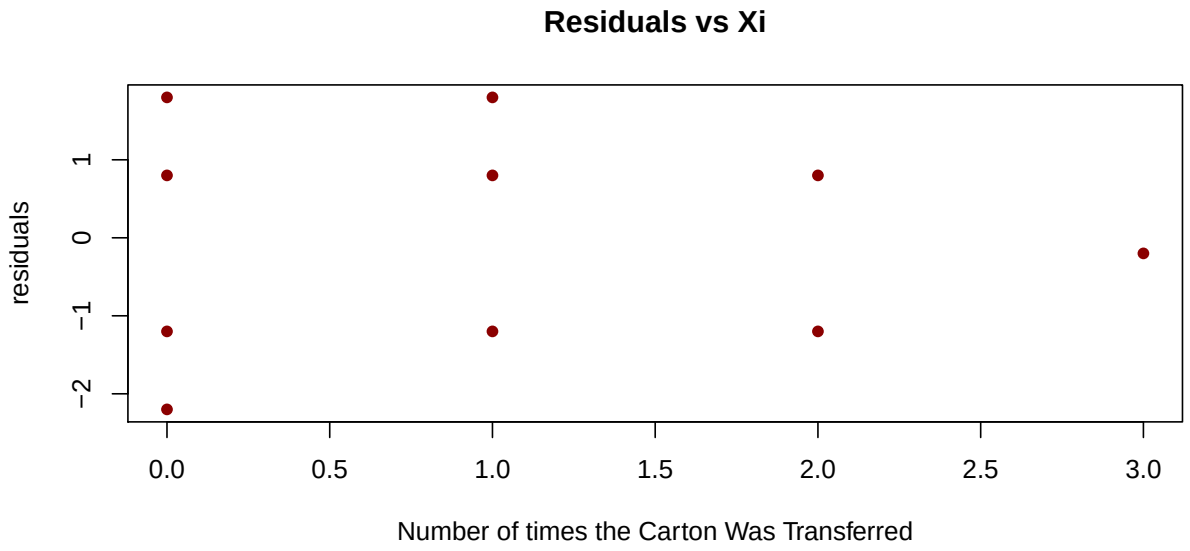
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Xi | 1 | 0 | 2 | 0 | 3 | 1 | 0 | 1 | 2 | 0 |
| Yi | 16 | 9 | 17 | 12 | 22 | 13 | 8 | 15 | 19 | 11 |

a. Obtain the residuals $e_i$ and plot them against $X_i$ to ascertain whether any departures from the assumptions are evident. What is your conclusion?

```
data1 <- data.frame(Xi=Xi,Yi=Yi)
result1 <- lm(Yi~Xi,data1)
unlist(result1$residuals)
```

```
##    1    2    3    4    5    6    7    8    9   10
##  1.8 -1.2 -1.2  1.8 -0.2 -1.2 -2.2  0.8  0.8  0.8
```

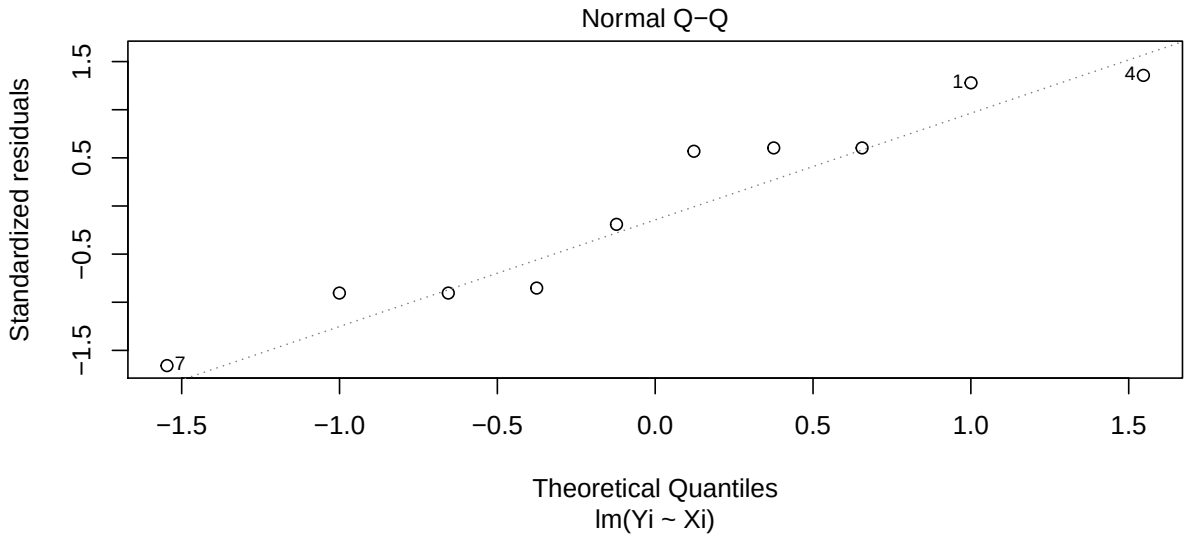
```
plot(result1$residuals~Xi,ylab="residuals",
     xlab="Number of times the Carton Was Transferred",
     main="Residuals vs Xi",col="dark red",cex=1,pch=16)
```



There is no evident pattern in the graph and hence I can conclude that there is no departure of the model based on assumption of linearity of regression function and nonconstancy of error variance.

b. Prepare a normal probability plot of the residuals. What can you conclude? Perform a formal test on the normality of the error terms and interpret the result.

```
plot(result1,which=2)
```




Based on the plot, there seems to be a violation on the assumption of normality of error terms.

$H_o$ : Error Terms are normally distributed

$H_a$ : Error Terms are not normally distributed

```
shapiro.test(result1$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  result1$residuals
## W = 0.90734, p-value = 0.2632
```

Based on Shapiro-Wilk Test, the p-value $> 0.05$. Hence, there is insufficient evidence to conclude that the error terms are not normally distributed.

2. A chemist studied the concentration of a solution (Y) over time (X). Fifteen solutions were prepared. The 15 solutions were randomly divided into five sets of three, and the five sets were measured, respectively, after 1, 3, 5, 7, and 9 hours. The results follow.

Table 2: Solution Concentration X over Time Y

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Xj | 9.00 | 9.00 | 9.00 | 7.00 | 7.00 | 7.00 | 5.00 | 5.00 | 5.00 | 3.00 | 3.00 | 3.00 | 1.00 | 1.00 | 1.0 |
| Yj | 0.07 | 0.09 | 0.08 | 0.16 | 0.17 | 0.21 | 0.49 | 0.58 | 0.53 | 1.22 | 1.15 | 1.07 | 2.84 | 2.57 | 3.1 |

a. Fit a linear regression function.

```
data2 <- data.frame(Xj=Xj,Yj=Yj)
result2 <- lm(Yj~Xj,data2)
result2
```

```
##
## Call:
## lm(formula = Yj ~ Xj, data = data2)
##
## Coefficients:
## (Intercept)           Xj
##       2.575       -0.324
```

Based on the result using R, the fitted linear regression function is

$$Y = -0.324X + 2.575$$

b. Perform an F test to determine whether or not there is a lack of fit on the linear regression function. At an $\alpha = 0.025$, interpret the result of the test.

$$H_o : E(Y) = \beta_o + \beta_1 X_i$$
$$H_a : E(Y) \neq \beta_o + \beta_1 X_i$$

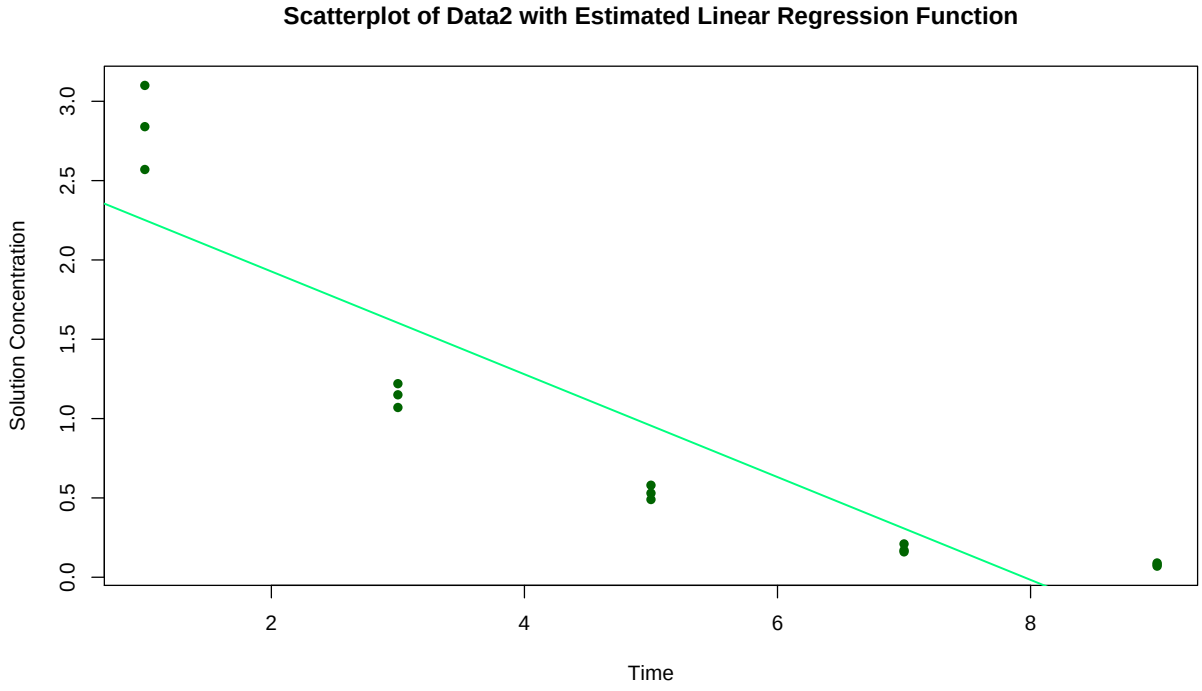
```
Model1 <- lm(Yj~Xj,data2)
Model2 <- lm(Yj~as.factor(Xj),data2)
anova(Model1,Model2)
```

```
## Analysis of Variance Table
##
## Model 1: Yj ~ Xj
## Model 2: Yj ~ as.factor(Xj)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     13 2.9247
## 2     10 0.1574  3    2.7673 58.603 1.194e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The result shows that F $< 0.025$. Hence we reject $H_o$ and conclude that the linear regression function lacks fit.

c. Prepare a scatter plot of the data. Interpret the plot.

```
plot(Yj~Xj,data2,ylab="Solution Concentration",xlab="Time",
     main="Scatterplot of Data2 with Estimated Linear Regression Function",
     pch=16,col="dark green",cex=1)
abline(Model1,lwd=1.5,col="spring green")
```




The plot shows that the estimated linear regression function is not a good fit.

d. Use the transformation $Y' = log_{10}Y$ and obtain the estimated linear regression function for the transformed data.

```
data2.1 <- data.frame(Xj=Xj,Yj=log10(Yj))
result2.1 <- lm(Yj~Xj,data2.1)
result2.1
```

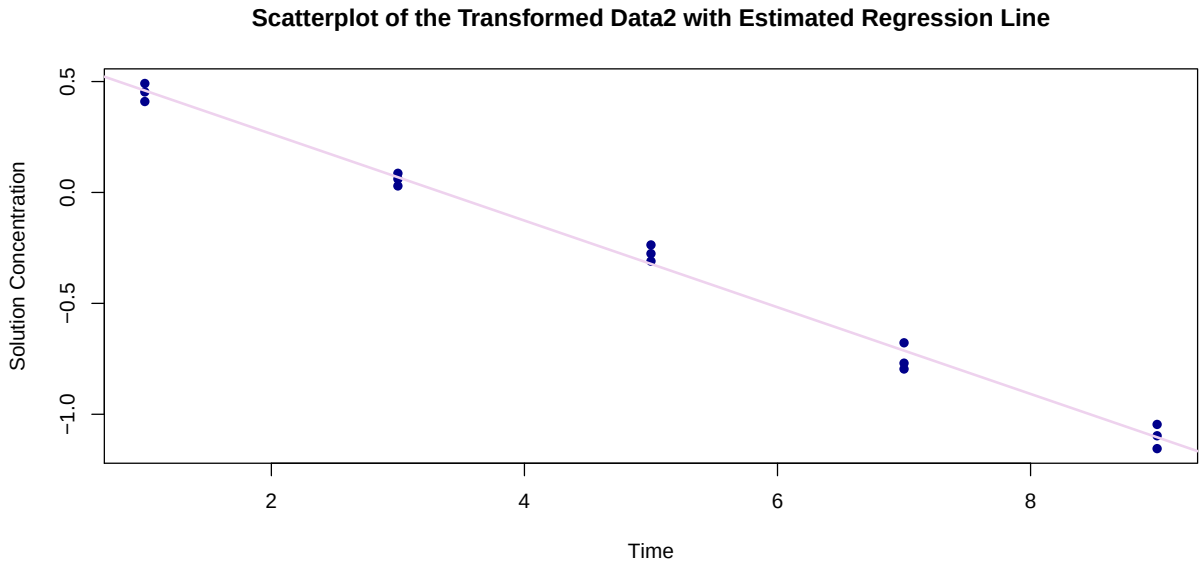
```
##
## Call:
## lm(formula = Yj ~ Xj, data = data2.1)
##
## Coefficients:
## (Intercept)           Xj
##      0.6549      -0.1954
```

The estimated linear regression function after transformation is

$$Y = -0.1954X + 0.6549$$

e. Plot the estimated regression line and the transformed data. Does the regression line appear to be a good fit to the transformed data?

```
plot(Yj~Xj,data2.1,xlab="Time",ylab="Solution Concentration",
     main="Scatterplot of the Transformed Data2 with Estimated Regression Line",
     cex=1,col="dark blue",pch=16)
abline(result2.1,lwd=2,col="thistle2")
```



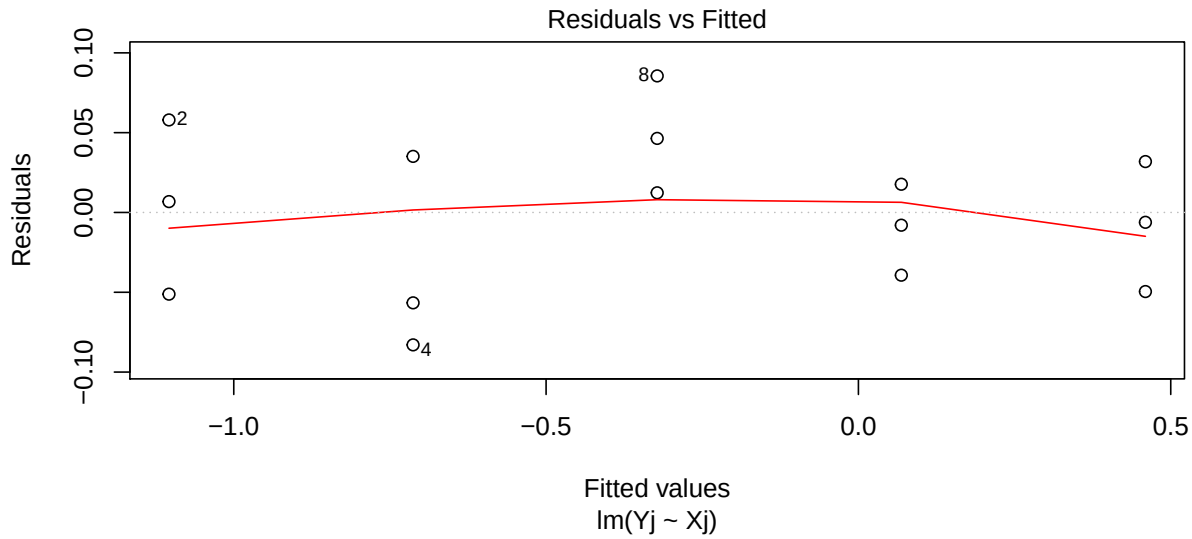

After transformation, the new estimated linear regression function is now a good fit to the transformed data based on the plot.

f. Obtain the residuals and plot them against fitted values. Also prepare a normal probability plot. What do your plots show?

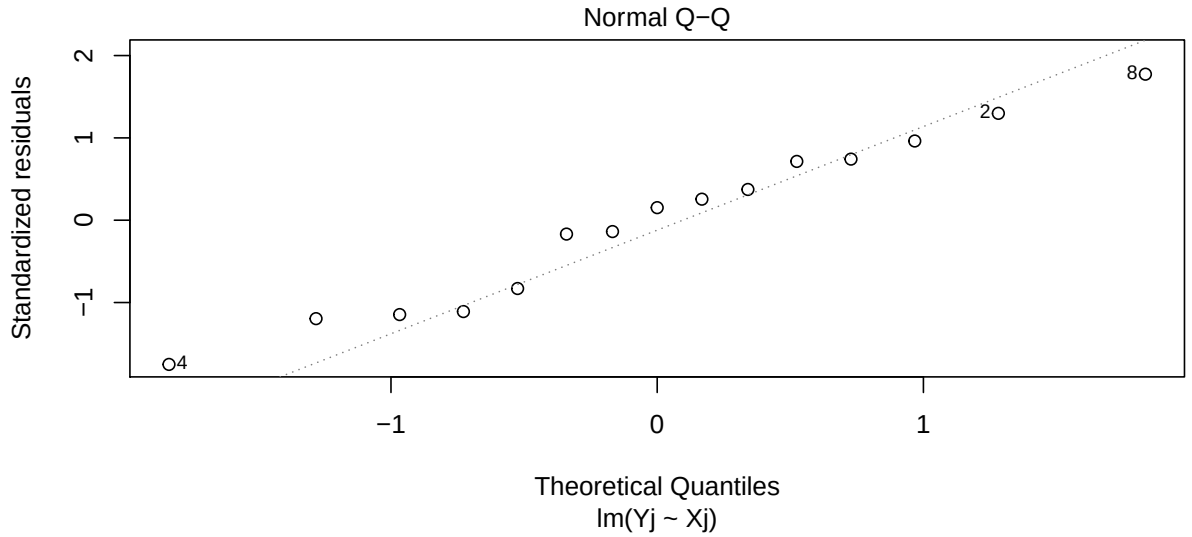
```
unlist(result2.1$residuals)
```

```
##            1            2            3            4            5
## -0.051178946  0.057965523  0.006813001 -0.082957620 -0.056628681
##            6            7            8            9           10
##  0.035141692  0.012317861  0.085549775  0.046397651  0.017680995
##           11           12           13           14           15
## -0.007980995 -0.039295058 -0.006161112 -0.049546328  0.031882242
```

```
plot(result2.1,which=1)
```




```
plot(result2.1,which=2)
```




```
shapiro.test(result2.1$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  result2.1$residuals
## W = 0.97357, p-value = 0.9069
```

Looking at the first plot, there is no evident pattern of the residuals. Thus, the assumptions of linearity of regression function and nonconstancy of error variance are not violated.

I got skeptical with the second plot as it seems like there is a certain pattern involved. However doing shapiro.test, we can say that there is not enough evidence to conclude that the error terms are not normally distributed.

g. Express the estimated regression equation in the original units.

$$antilog(Y) = antilog(-0.1954X + 0.6549)$$