

MH3510 Assignment 1

Neo Shun Xian Nicholas

24 September 2020

Goal: To explore the relationship between the amount of beta-erythoidine in an aqueous solution and the colorimeter reading of the turbidity

X: Concentration (mg/mL)

Y: Colorimeter Reading

Declaration of data

Storing the data into vector X and Y for concentration and colorimeter reading respectively

```
X <- c(40,50,60,70,80,90,40,60,80,50)
Y <- c(69,175,272,335,490,415,72,265,492,180)
```

a: Fit a simple regression to the data

Fit linear model

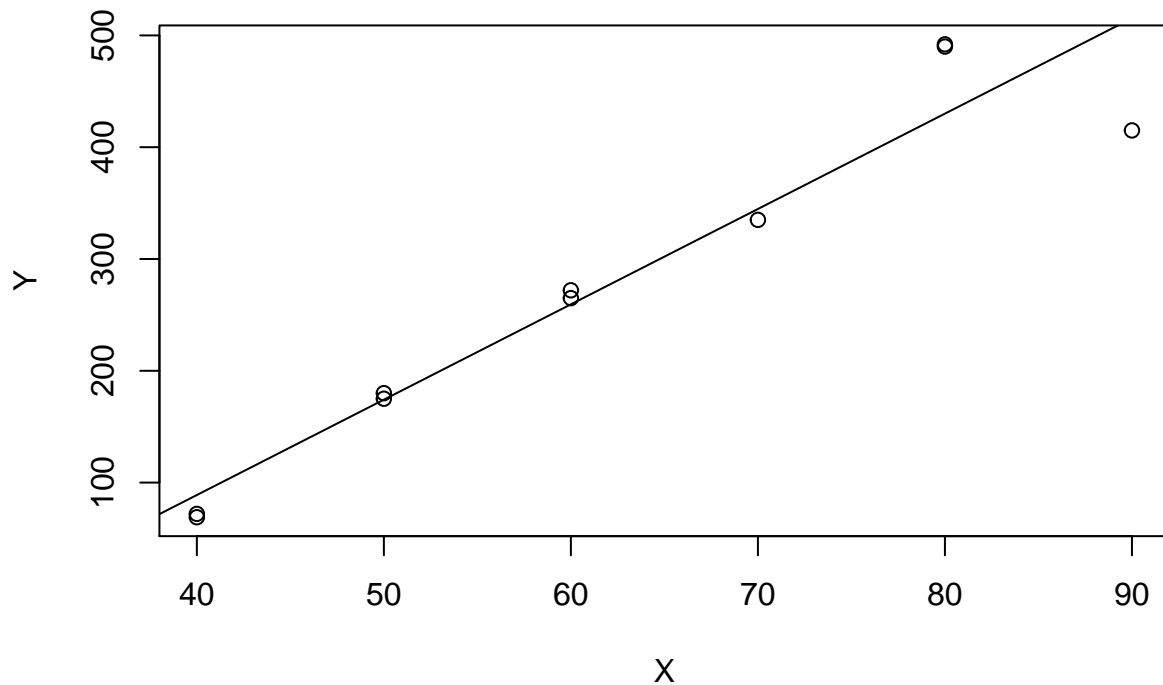
```
slr <- lm(Y~X)
# Display basic information of this model
slr
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Coefficients:
## (Intercept)          X
##    -252.297         8.529
```

From the plot above, we can see that for the best fit line of the Colorimeter Reading against Concentration, the Y-intercept is -252.297 and the gradient is 8.529. i.e $Y = -252.297 + 8.529 \cdot X$ is the best fit line.

Plot data and the best fit line

```
# Plot data
plot(X,Y)
# Plot best fit line
abline(slr)
```



As shown in the above plot, it seems that the linear model of Y against X fits fairly well to the data, except with larger deviations of actual value of Y and predicted value of Y when X equals to 80 and 90.

b: Obtain the residuals and examine them

Obtain and display the residual values

```
# Obtain residuals
residual <- slr$res
# display the residuals value
residual
```

```
##          1          2          3          4          5
## -19.8623188  0.8478261 12.5579710 -9.7318841 59.9782609
##          6          7          8          9         10
## -100.3115942 -16.8623188  5.5579710 61.9782609  5.8478261
```

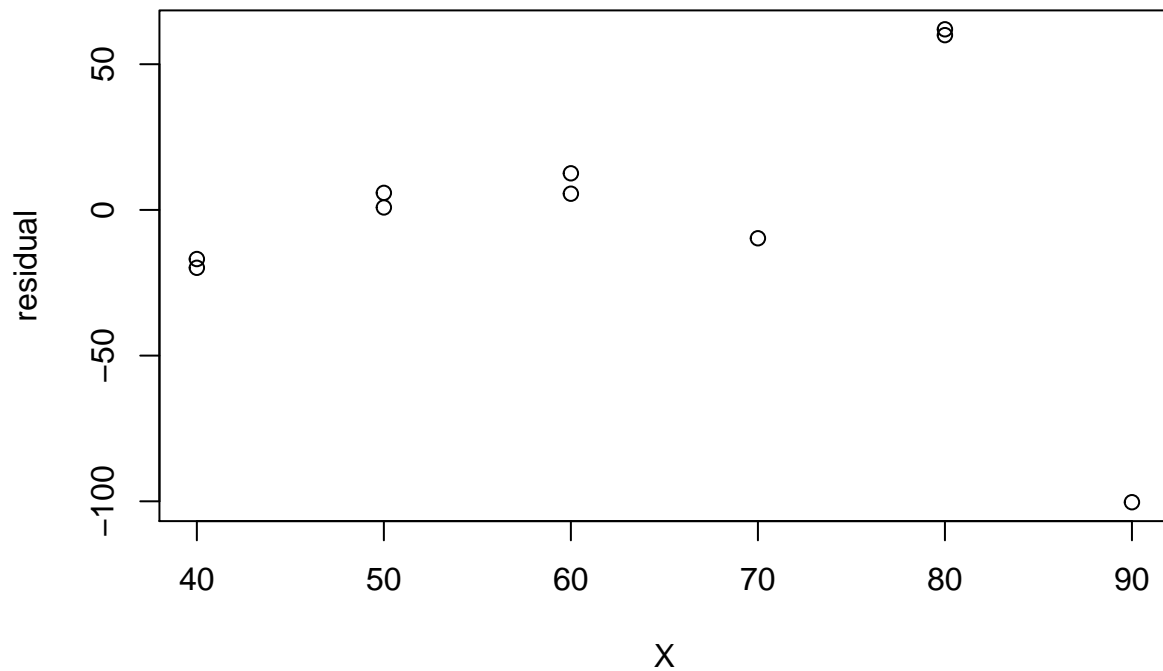
Examining the residual values based on the output above

From the residual values, there are 3 points of predicted Y values that has significant deviations from the actual Y values with residual values -100.31, 59.978 and 61.978

c: Comment on the adequacy of the model

Examine the residuals by plotting

```
# Plot residual against X plot  
plot(X,residual)
```



Comment

From the above plot, we can see that the plot between the residual and Concentration(X) is **not linear**. This implies that there is **no linear relationship** between the response and predicted variable (Y_actual vs Y_predicted). Hence, the model is **not adequate** to show the relationship between the Concentration(X) and Colorimeter Reading(Y). Moreover, as shown in the plot above, it seems to imply that a **higher order polynomial term such as x^2 should be fitted** in order to show the adequacy of the model, even without the abnormal behaviour of Y when X is 80 & 90.