# #1 The R code include in another file.

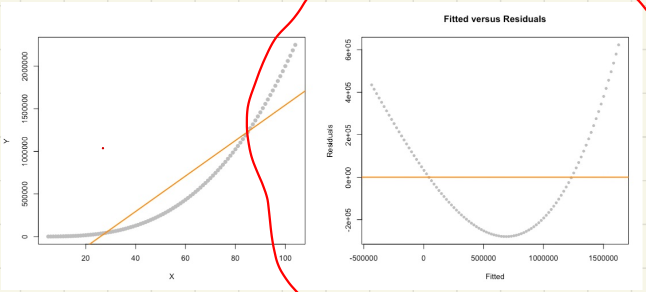
Base :

```
Call:
lm(formula = Y ~ X, data = Dataset_1)

Residuals:
    Min      1Q  Median      3Q     Max
-279166 -224109  -64146  183868  622869

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -538500.6    53216.9  -10.12   <2e-16 ***
X             20820.8      862.9   24.13   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 249100 on 98 degrees of freedom
Multiple R-squared:  0.8559,	Adjusted R-squared:  0.8545
F-statistic: 582.2 on 1 and 98 DF,  p-value: < 2.2e-16
```




Since the residuals summary and plot are bad, we start transform below.

```
	studentized Breusch-Pagan test

data:  Dataset_1_fit
BP = 5.4522, df = 1, p-value = 0.01954
```

because $p < 0.05$, reject the $H_0$ . So the error has non-constant variance about the true model.

```
	Shapiro-Wilk normality test

data:  resid(Dataset_1_fit)
W = 0.9059, p-value = 2.707e-06
```

because $P < 0.05$. reject the $H_0$. The data is not consistent with have been generated according to a normal distribution.

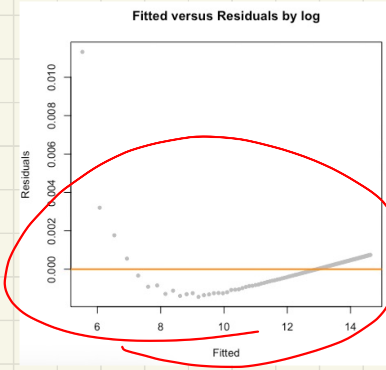
Log:

```
Call:
lm(formula = log(Y) ~ log(X), data = Dataset_1)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0014421 -0.0006845 -0.0000171  0.0004487  0.0113220

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.6986105  0.0007272   960.7   <2e-16 ***
log(X)      2.9986635  0.0001884 15919.7   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.001377 on 98 degrees of freedom
Multiple R-squared:      1,     Adjusted R-squared:      1
F-statistic: 2.534e+08 on 1 and 98 DF,  p-value: < 2.2e-16
```




The residuals and $R^2$ are good, but plot is terriable.

```
        studentized Breusch-Pagan test

data:  Dataset_1_log_fit
BP = 12.181, df = 1, p-value = 0.0004828
```

because $p < 0.05$, reject the $H_0$. So the error has non-constant variance about the true model.

```
        Shapiro-Wilk normality test

data:  resid(Dataset_1_log_fit)
W = 0.53812, p-value = 3.308e-16
```

because $p < 0.05$. reject the $H_0$. The data is not consistent with have been generated according to a normal distribution.

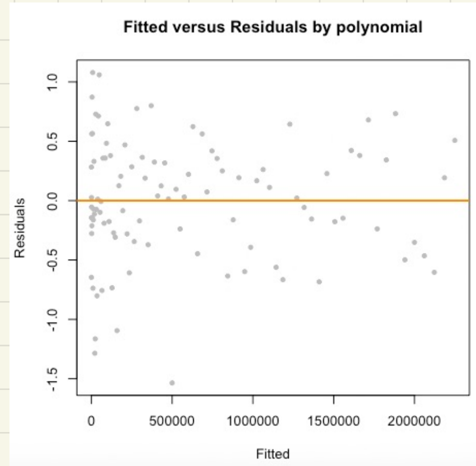
Poly :

```
Call:
lm(formula = Y ~ X + I(X^2) + I(X^3), data = Dataset_1)

Residuals:
     Min       1Q   Median       3Q      Max
-1.53555 -0.27802  0.01789  0.35573  1.07893

Coefficients:
              Estimate Std. Error    t value Pr(>|t|)
(Intercept)  3.628e+00  2.856e-01     12.701   <2e-16 ***
X           -4.989e-02  2.163e-02     -2.307   0.0232 *
I(X^2)       1.010e-03  4.509e-04      2.240   0.0274 *
I(X^3)       2.000e+00  2.725e-06 733855.023   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5147 on 96 degrees of freedom
Multiple R-squared:      1,	Adjusted R-squared:      1
F-statistic: 5.311e+13 on 3 and 96 DF,  p-value: < 2.2e-16
```




The Residual report is good and $R^2$. p all good. this model is significaitan.
And plot is good, the poly transformation is good for this model.

```
	studentized Breusch-Pagan test

data:  Dataset_1_quad_fit
BP = 3.7037, df = 3, p-value = 0.2953
```

because $p > 0.05$, accept the $H_0$. So the error has constant variance about the true model.

```
	Shapiro-Wilk normality test

data:  resid(Dataset_1_quad_fit)
W = 0.98633, p-value = 0.3939
```

because $p > 0.05$. accept the $H_0$. The data is consistent with have been generated according to a normal distribution.

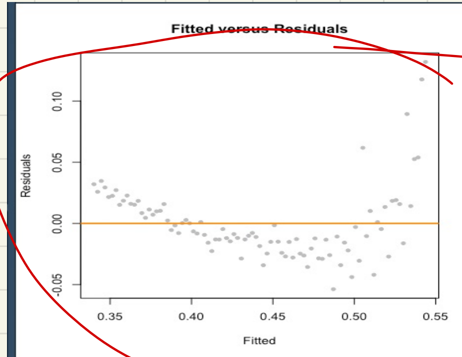
Boxcox :

```
Call:
lm(formula = z ~ x)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.05375 -0.01604 -0.00783  0.01468  0.13196 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.5663797  0.0074264   76.27   <2e-16 ***
x           -0.0022630  0.0001218  -18.57   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03053 on 89 degrees of freedom
Multiple R-squared:  0.7949,	Adjusted R-squared:  0.7926 
F-statistic:   345 on 1 and 89 DF,  p-value: < 2.2e-16
```




The $R^2$ kind of small by comparing with other transformation. and plot is not even at all.

```
	studentized Breusch-Pagan test

data:  m
BP = 12.132, df = 1, p-value = 0.0004958

> shapiro.test(resid(m))

	Shapiro-Wilk normality test

data:  resid(m)
W = 0.85577, p-value = 5.988e-08
```

because $p < 0.05$, reject the $H_0$ . so the error has non-constant variance about the true model.

because $P < 0.05$. reject the $H_0$. The data is not consistent with have been generated according to a normal distribution.

Result: By comparing those above transformations. we can get the poly is the best models for this dataset.

since it satisfy two test, and have good residuals (small, symmetic)

high $R^2$, $R_a^2$. small P which imply very significant, and the plot are even and nice.

#2

Assume $d=2$, then we have $G: \mathbb{R}^2 \to \mathbb{R}$

Since $G$ has a local minimum at point $x_0$. then we have $x_0 = (u_0, v_0)$

$$\frac{\partial G(u,v)}{\partial u} = \lim_{u \to u_0} \frac{G(u,v_0) - G(u_0,v_0)}{u-u_0} = \partial_u G(u_0,v_0)$$

limit from left: $\lim_{u \uparrow u_0} \frac{G(u,v_0) - G(u_0,v_0)}{u-u_0} \begin{matrix} + \\ - \end{matrix} \leq 0$

Since $(u_0, v_0)$ is local minimum, so $G(u,v_0) \geq G(u_0,v_0)$

limit from right: $\lim_{u \downarrow u_0} \frac{G(u,v_0) - G(u_0,v_0)}{u-u_0} \begin{matrix} + \\ + \end{matrix} \geq 0$.

Then we can concluded that $\partial_u G(u_0, v_0) = 0$.

$$\frac{\partial G(u,v)}{\partial v} = \lim_{v \to v_0} \frac{G(u_0,v) - G(u_0-v_0)}{v-v_0} = \partial_v G(u_0,v_0)$$

limit from left: $$\lim_{v \uparrow v_0} \frac{G(u_0, v) - G(u_0, v_0)}{v - v_0} \frac{+}{-} \leq 0$$

Since $(u_0, v_0)$ is local minimum, so $G(u_0, v) \geq G(u_0, v_0)$

limit from right: $$\lim_{v \downarrow v_0} \frac{G(u_0, v) - G(u_0, v_0)}{v - v_0} \frac{+}{+} \geq 0.$$

Then we can concluded that $\partial_v G(u_0, v_0) = 0$.

So since we have $\partial_u G(u_0, v_0) = 0$ & $\partial_v G(u_0, v_0) = 0$ then we can concluded that $\nabla G(x_0) = 0$.

#3

(a) Assume we have $G_1$ and $G_2$, then we can have:

$$G_1[\lambda u + (1-\lambda)v] + G_2[\lambda u + (1-\lambda)v] \leq \lambda G_1(u) + (1-\lambda)G_1(v) + \lambda G_2(u) + (1-\lambda)G_2(v)$$

$$(G_1+G_2)[\lambda u + (1-\lambda)v] \leq \lambda G_1(u) + \lambda G_2(u) + (1-\lambda)G_1(v) + (1-\lambda)G_2(v)$$

$$\underbrace{(G_1+G_2)}_{G}[\lambda u + (1-\lambda)v] \leq \lambda [\underbrace{G_1(u) + G_2(u)}_{G(u)}] + (1-\lambda)[\underbrace{G_1(v) + G_2(v)}_{G(v)}]$$

So we can know the sum of convex function is convex.

(b) $$H(\alpha_0, \alpha_1, \dots \alpha_m) = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_m x_m - y$$

$$H[\lambda u + (1-\lambda)v] = H[\lambda u_0 + (1-\lambda)v_0 + \lambda u_1 + (1-\lambda)v_1 + \cdots + \lambda u_m + (1-\lambda)v_m]$$

$$= \lambda u_0 + (1-\lambda)v_0 + [\lambda u_1 + (1-\lambda)v_1]x_1 + [\lambda u_2 + (1-\lambda)v_2]x_2 \cdots [\lambda u_m + (1-\lambda)v_m]x_m - y$$

$$= \lambda(u_0 + u_1x_1 + u_2x_2 \cdots u_m x_m - y) + (1-\lambda)(v_0 + v_1x_1 + v_2x_2 + \cdots + v_m x_m - y)$$

$$= \lambda H(u) + (1-\lambda)H(v)$$

(c) $f(t) = t^2$ Satisfies: $G(\lambda u + (1-\lambda)v) - \lambda G(u) - (1-\lambda)G(u) \leq 0$

Let $t = \lambda u + (1-\lambda)u$ then $t^2 = [\lambda u + (1-\lambda)v]^2$

$$= \lambda^2 u^2 + 2\lambda(1-\lambda)u \cdot v + (1-\lambda)^2 v^2$$

$$f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda) f(v)$$

$$\lambda^2 u^2 + 2\lambda(1-\lambda)u \cdot v + (1-\lambda)^2 v^2 \leq \lambda u^2 + (1-\lambda)v^2$$

$$\lambda^2 u^2 + 2\lambda(1-\lambda)u \cdot v + (1-\lambda)^2 v^2 - \lambda u^2 - (1-\lambda)v^2 \leq 0$$

$$(\lambda^2 - \lambda)u^2 - 2(\lambda^2 - \lambda)u \cdot v + (\lambda^2 - \lambda)v^2 \leq 0$$

$$(\lambda^2 - \lambda)(u - v)^2 \leq 0$$

Since $\lambda \in [0,1]$, so this function is satisfies for all real number. So is convex.

(d) We can start by part(3) to get $H(\lambda u + (1-\lambda)v) = \lambda H(u) + (1-\lambda)H(v)$. So we can get $F(\dots) = \sum_{n=1}^{N} H_n^2(\dots)$. Then by part(4), we can get $F(\dots) = H_n^2(\dots)$ is convex. Then by using part(1) we get the sum $\sum_{n=1}^{N} H_n^2(\dots)$ is convex.

#4

Since $x_0$ is local mininum, so for any $y$ in convet set, we can choose a small enough $\lambda > 0$. then we have:

$$G(x_0) \le G[x_0 + \lambda(y - x_0)]$$

$$= G[\lambda y + (1-\lambda)x_0]$$

$$\le \lambda G(y) + (1-\lambda)G(x_0)$$

$$\le G(y)$$

Since $y$ is an arbitrary point in convet set, then we can prove $x_0$ is a global minimum.

#5

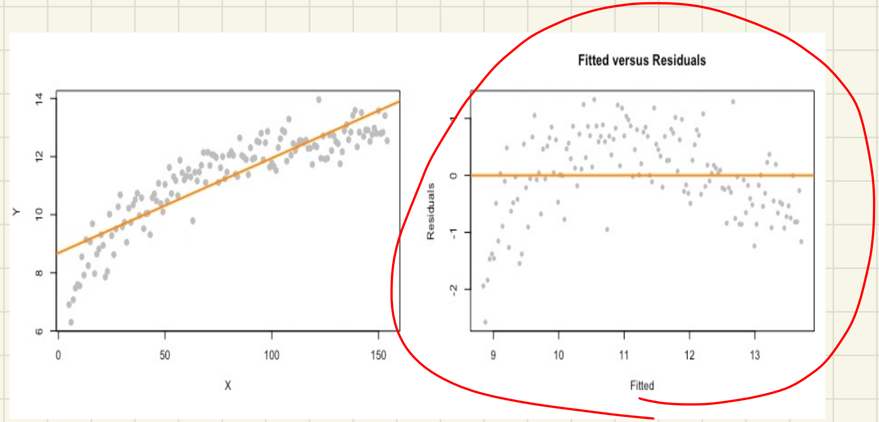
Base :

```
Call:
lm(formula = Y ~ X, data = Dataset_2)

Residuals:
     Min       1Q   Median       3Q      Max
-2.57773 -0.47843  0.02257  0.61561  1.33017

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 8.683026   0.126456   68.66   <2e-16 ***
X           0.032645   0.001397   23.37   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7408 on 148 degrees of freedom
Multiple R-squared:  0.7868,    Adjusted R-squared:  0.7853
F-statistic: 546.1 on 1 and 148 DF,  p-value: < 2.2e-16
```




The residual, $R^2$, p-value are good. but the plot is not even. so try transform.

```
studentized Breusch-Pagan test

data:  Dataset_2_fit
BP = 13.756, df = 1, p-value = 0.0002082
```

because $p < 0.05$, reject the $H_0$. So the error has non-constant variance about the true model.

```
Shapiro-Wilk normality test

data:  resid(Dataset_2_fit)
W = 0.97478, p-value = 0.00733
```

because $p < 0.05$. reject the $H_0$. The data is not consistent with have been generated according to a normal distribution.

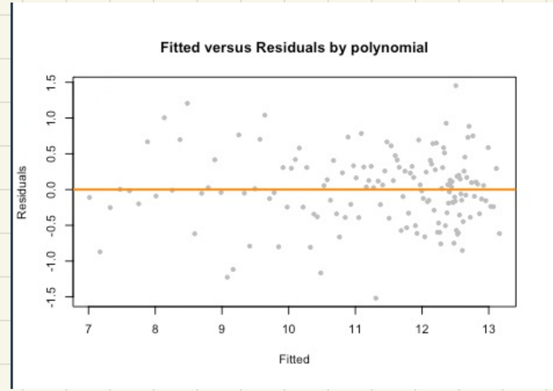
poly :

```
Call:
lm(formula = Y ~ X + I(X^2) + I(X^3) + I(X^4) + I(X^5), data = Dataset_2)

Residuals:
     Min       1Q   Median       3Q      Max
-1.51761 -0.31603  0.00755  0.30747  1.45321

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.136e+00  4.113e-01  14.919  < 2e-16 ***
X            1.916e-01  4.690e-02   4.084 7.32e-05 ***
I(X^2)      -3.358e-03  1.711e-03  -1.963   0.0516 .
I(X^3)       3.797e-05  2.646e-05   1.435   0.1535
I(X^4)      -2.289e-07  1.815e-07  -1.261   0.2093
I(X^5)       5.451e-10  4.546e-10   1.199   0.2324
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5052 on 144 degrees of freedom
Multiple R-squared:  0.9035,    Adjusted R-squared:  0.9002
F-statistic: 269.7 on 5 and 144 DF,  p-value: < 2.2e-16
```




The Residual is good. $R^2$ is high and P-value is low, and plot is even.

```
        studentized Breusch-Pagan test

data:  Dataset_2_quad_fit
BP = 6.2775, df = 5, p-value = 0.2801
```

because $P > 0.05$, accept the $H_0$. So the error has constant variance about the true model.

```
        Shapiro-Wilk normality test

data:  resid(Dataset_2_quad_fit)
W = 0.99505, p-value = 0.8948
```

because $P > 0.05$. accept the $H_0$. The data is consistent with have been generated according to a normal distribution.

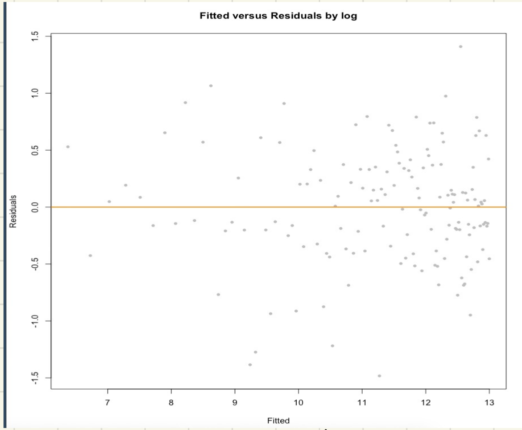
log:

```
Call:
lm(formula = Y ~ log(X), data = Dataset_2)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.48239 -0.31379  0.03454  0.33745  1.40947 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   3.2640     0.2207   14.79   <2e-16 ***
log(X)        1.9329     0.0523   36.95   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5017 on 148 degrees of freedom
Multiple R-squared:  0.9022,	Adjusted R-squared:  0.9016 
F-statistic:  1366 on 1 and 148 DF,  p-value: < 2.2e-16
```




The Residual is good. $R^2$ is high and P-value is low, and plot is ok (not too even)

```
        studentized Breusch-Pagan test

data:  Dataset_2_log_fit
BP = 2.1306, df = 1, p-value = 0.1444
```

because $P > 0.05$, accept the $H_0$. So the error has constant variance about the true model.

```
        Shapiro-Wilk normality test

data:  resid(Dataset_2_log_fit)
W = 0.99136, p-value = 0.4953
```

because $P > 0.05$. accept the $H_0$. The data is consistent with have been generated according to a normal distribution.

Result : Base on poly and log tranform both is work on this dataset ( both have high $R^2$, low p-value, and small, symetric residuals ). And from BP and Shapiro shows it has constant variance and consistent So both transform is work. But I perfer poly one since it has a more even plot.