Gabriella Weis
Mr. Lewis
H Applied Statistics
13 November 2023

#14 Assessment: Multiple Regression, F-Test, R-Studio

Part A:

a).

1. SSResid = $\sum(y - \hat{y})^2$

SSResid = 390.4347

2. SSTo = $\sum(y - \bar{y})^2$

$\bar{y} = \frac{\sum(y)}{14} = 21.10714286$

SSTo = 1618.209286

3. SSReg = $\sum(\hat{y} - \bar{y})^2$

SSReg = 1227.116129

b).

$R^2 = 1 - \frac{SSResid}{SSTotal}$
$R^2 = 1 - \frac{390.4347}{1618.209286}$
R^2 = 0.7587242247

Adj. $R^2 = 1 - [\frac{(n-1)}{n-(k+1)}]\frac{SSResid}{SSTo}$
Adj. $R^2 = 1 - \frac{(13)(390.4347)}{(8)(1618.209286)}$
Adj. R^2 = 0.60793

I would recommend using the adjusted R^2 because it accounts for the number of predictors in the model and penalizes for excessive variables, creating a more accurate measure of the model's goodness of fit.

c).

Model Utility F-test

$\hat{y}$ = predicted shear strength of sandy soil

x_1= depth

x_2= water content

$x_3 = (x_1)^2$

$x_4 = (x_2)^2$

$x_5 = (x_1)(x_2)$

β_1= estimated increase in predicted shear strength of sandy soil when depth increases by one unit.

β_2= estimated increase in predicted shear strength of sandy soil when water content increases by one unit.

β_3= estimated increase in predicted shear strength of sandy soil when x_3 increases by one unit.

β_4= estimated increase in predicted shear strength of sandy soil when x_4 increases by one unit.

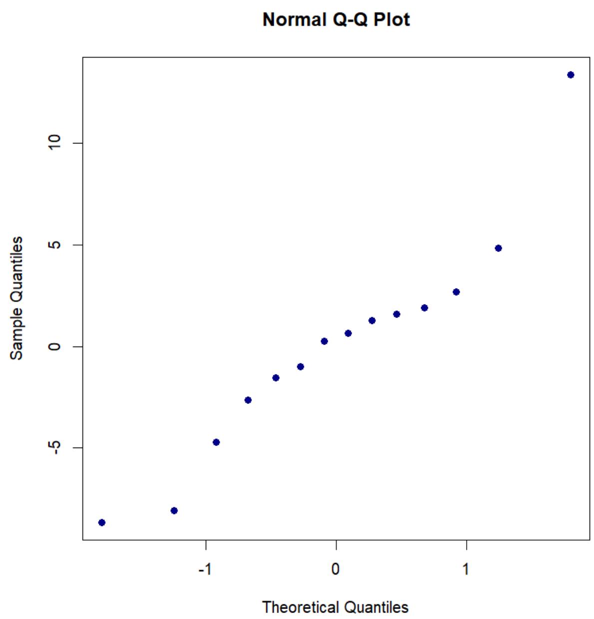
β_5= estimated increase in predicted shear strength of sandy soil when x_5 increases by one unit.

H_o: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_a: At least one slope $\neq 0$

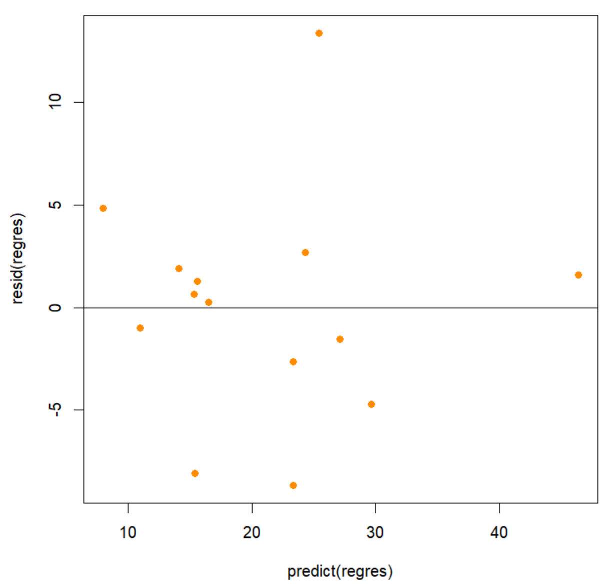
$\alpha = 0.05$

Checks:
Normality:



Probability plot appears curved. Proceed with caution.

Equal Variance:



Residual plot appears approximately normally distributed with a few major outliers and one closely clumped group of four points above the y=0 line. Proceed with caution.

Calculations:

$$F = \frac{R^2/k}{(1-R^2)/(n-(k+1))} \quad \text{w/df} = \frac{k}{n-(k+1)}$$

$$F = \frac{0.7587242247/5}{(1-0.7587242247)/(8)} \quad \text{w/df} = \frac{5}{8} = 0.625$$

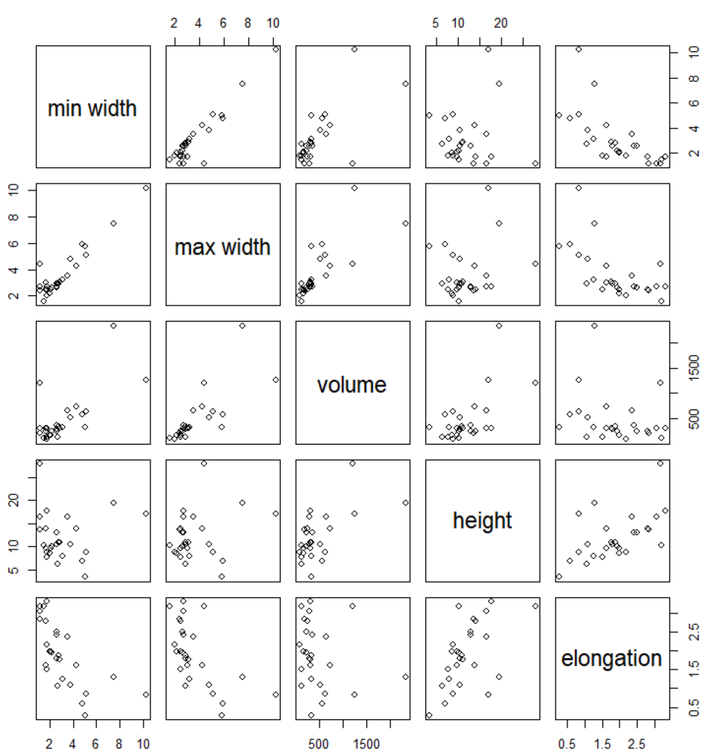
F = 5.03142 w/df = 0.625

p-value = p(F > 5.03142) = 0.0222

Since the p-value (0.0222) < α (0.05), we reject the null hypothesis. Therefore, there is convincing evidence to suggest that at least one of the slopes is not equal to zero.

Part B:

a).



Model	Equation	R^2	R^2(Adj.)	S_e	F-Stat.
Max Width(x_1)	$\hat{y}$ = -231.02+190.65x_1	0.5843	0.5677	312	35.14
Max Width(x_1)+Min Width(x_2)	$\hat{y}$ = -263.72+248.07x_1-56.99x_2	0.5916	0.5575	315.7	17.38
Max Width(x_1)+Min Width(x_2)+Height(x_3)	$\hat{y}$ = -660.46+55.89x_1+104.08x_2+49.95x_3	0.7951	0.7683	228.4	29.74
Max Width(x_1)+Min Width(x_2)+Height(x_3)+elongation(x_4)	$\hat{y}$ = -75.94-86.02x_1+124.44x_2+95.59x_3-357.49x_4	0.8437	0.8153	203.9	29.7
Min Width(x_1)+Height(x_2)+Elongation(x_3)	$\hat{y}$ = -216.39-73.21x_1+81.64x_2+-276.82x_3	0.8371	0.8158	203.7	39.39
Height(x_1)+Elongation(x_2)	$\hat{y}$ = 130.16+100.54x_1-459.53x_2	0.8087	0.7928	216	50.73

I chose the model: Predicted Volume = Min Width(x_1)+Height(x_2)+Elongation(x_3), or $\hat{y}$ = -216.39-73.21x_1+81.64x_2+-276.82x_3, because it has high R^2 and adjusted R^2 values, a low standard error, and a high F-statistic.

b).

We should consider adjusted R^2 instead of R^2 when attempting to determine the quality of the fit of the data to the model because it accounts for all of the x terms.

c).

Model Utility F-test

$\hat{y}$ = predicted volume

x_1 = min width

x_2 = height

x_3 = elongation

β_1 = estimated increase in predicted volume when min width increases by one unit.

β_2 = estimated increase in predicted volume when height increases by one unit.

β_3 = estimated increase in predicted volume when elongation increases by one unit.

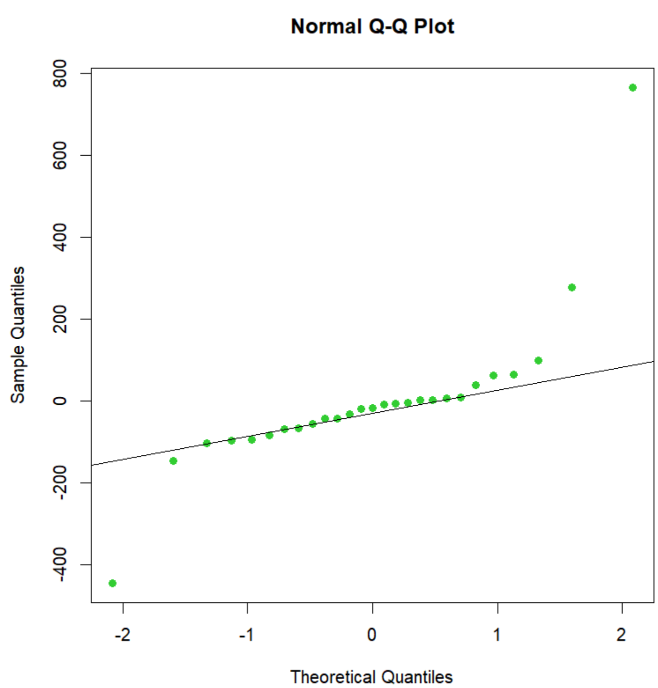
H_o: $\beta_1 = \beta_2 = \beta_3 = 0$

H_a: At least one slope $\neq 0$

$\alpha = 0.05$

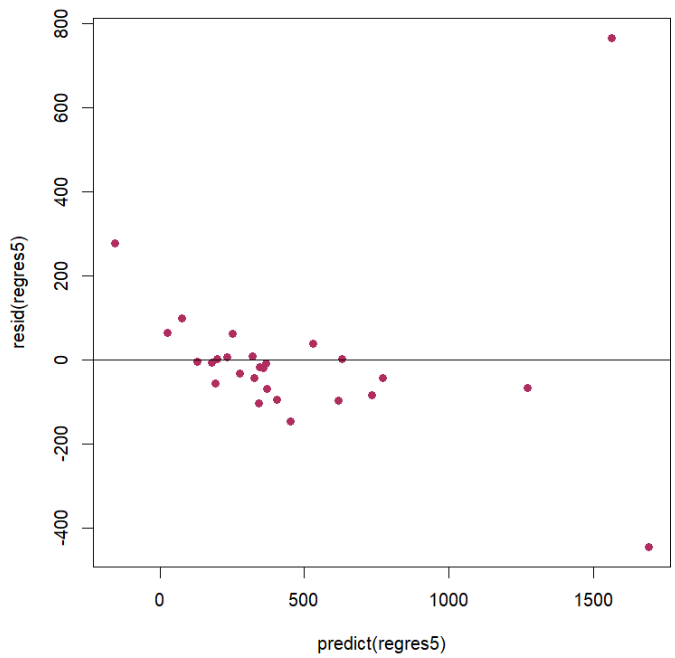
Checks:

Normality:



Probability plot is not linear, especially on the ends where there lies extreme outliers. Proceed with caution.

Equal Variance:



Residual plot does not appear normally distributed. Proceed with caution.

Calculations:

$$F = \frac{R^2/k}{(1-R^2)/(n-(k+1))} \quad \text{w/df} = \frac{k}{n-(k+1)}$$

$$F = \frac{0.8371/3}{(1-0.8371)/(23)} \quad \text{w/df} = \frac{3}{23} = 0.130435$$

F = 39.39697 w/df = 0.130435

p-value = p(F > 39.39697) = 0.000000003155 ≈ 0

Since the p-value (0) < α (0.05), we reject the null hypothesis. Therefore, there is convincing evidence to suggest a useful linear relationship that uses Min Width, Height, and Elongation to predict volume.