HW8 - Ming Lin

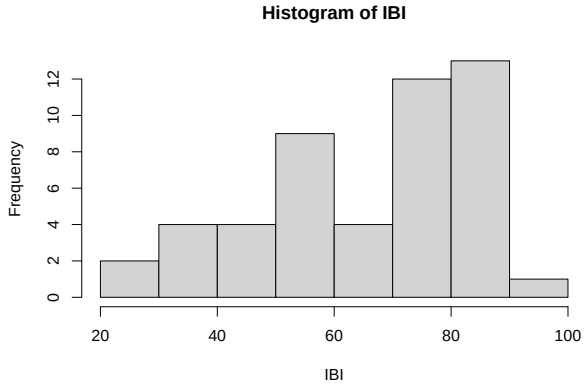
I pledge my honor that I have abided by the Stevens Honor System.

Question 10.32

Part A: Use numerical and graphical methods to describe the variable IBI. Do the same for area. Summarize your results.

Numerical and graphical methods for variable IBI:

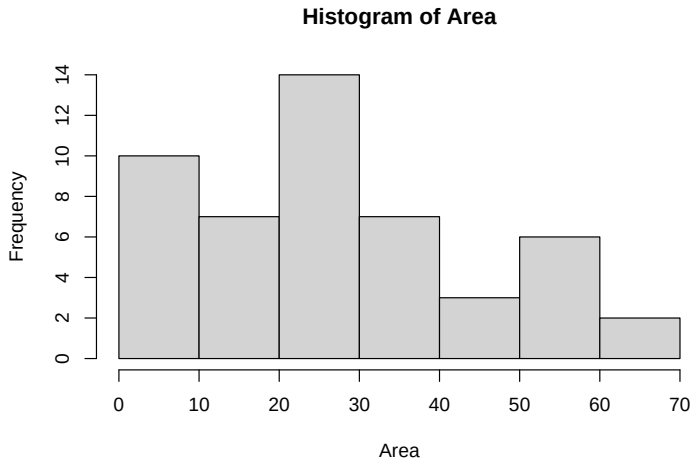
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   29.00   55.00   71.00   65.94   82.00   91.00
```



Summary: Based on the graph above, we can tell that the data for the IBI variable is skewed to the left and has no outliers

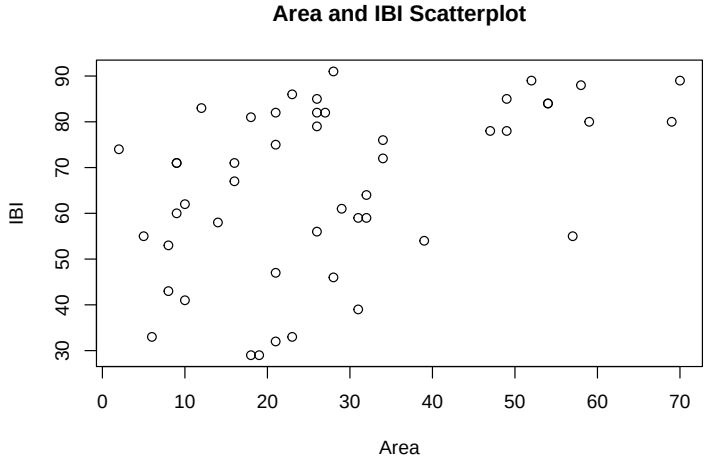
Numerical and graphical methods for area:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.00   16.00   26.00   28.29   34.00   70.00
```



Summary:Based on the graph above, we can tell that the data for the IBI variable is relatively symmetrical.

Part B



Summary: Based on the graph abov, we can tell that the variable IBI has greater variation as the area decreases

Part C

$IBI \ = \ \beta_0 + \beta_1 \left(Area\right) \ + \ \xi_i, i = 1, 2, ..., 49$

$IBI = 52.92 + 0.46 * Area + \epsilon$

Part D

Null Hypothesis: $\beta_1 = 0$

Alternative Hypothesis: $\beta_1 \neq \ 0$

Part E

Significance Level= 0.05

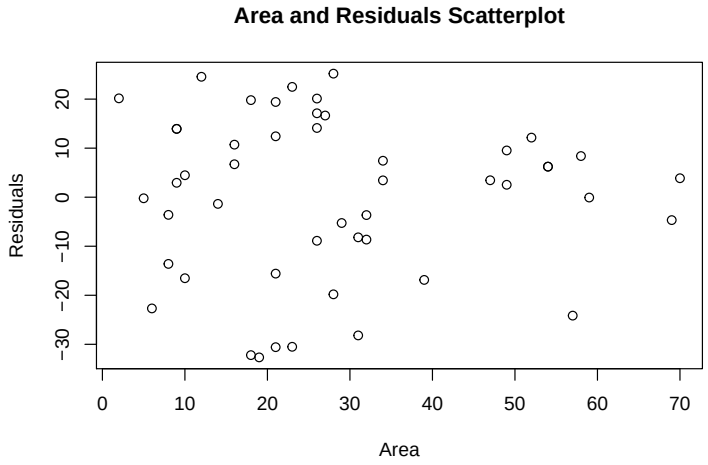
$$T = \frac{\hat{\beta_1}}{SE_{\hat{\beta_1}}} = 3.41$$

$$SE_{\hat{\beta}_1} = \sqrt{\frac{S^2}{\sum_{i=1}^{n} (x_i = \hat{x})^2}} = 0.13$$

$$S^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2} = 273.4$$

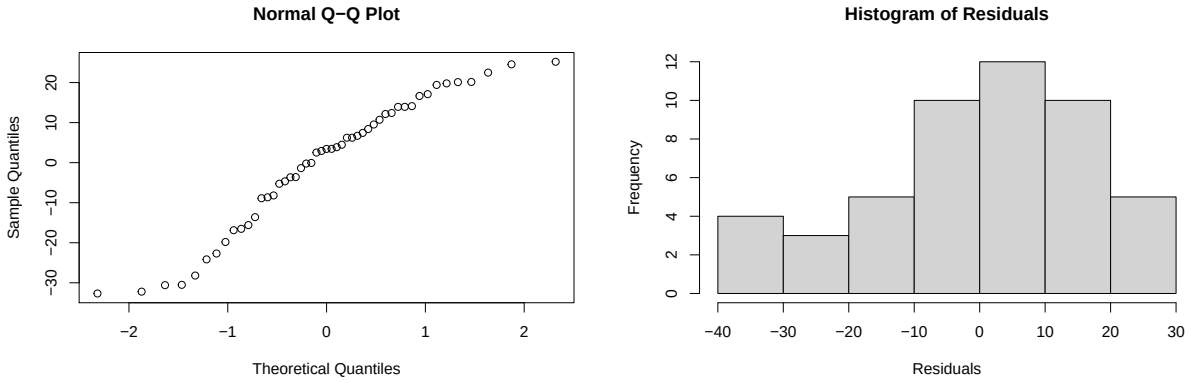
Degrees of Freedom = n-2 = 49 - 2 = 47 P-value = 2*P(T>3.41) = 0.0013 P-value is $0.0013 < \alpha$. We reject H_0.

Part F



Summary: Based on the graph above, we can tell that as the area get smaller the residual starts to approach closer to zero.

Part G



Summary: Based on the graph above, we can tell that the residuals are skewed left and have a normal distribution.

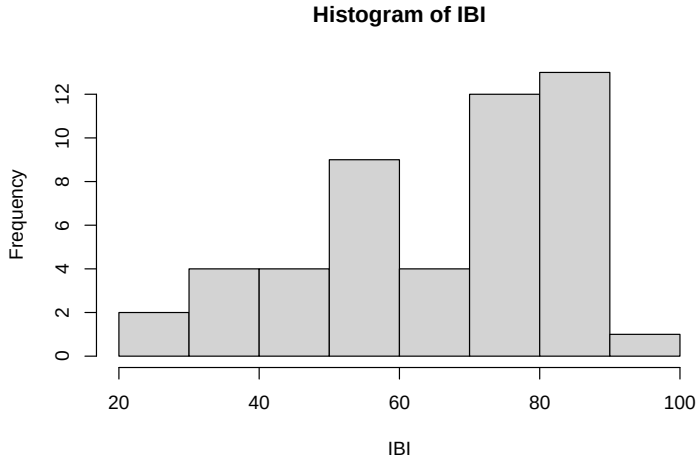
Part H

Based on the graphs in the previous questions and the residual, we can say that the assumptions are reasonable.

Question 10.33

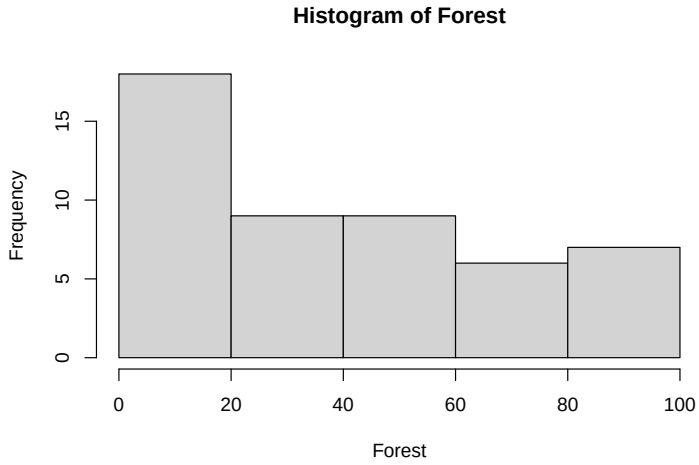
Part A

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   29.00   55.00   71.00   65.94   82.00   91.00
```



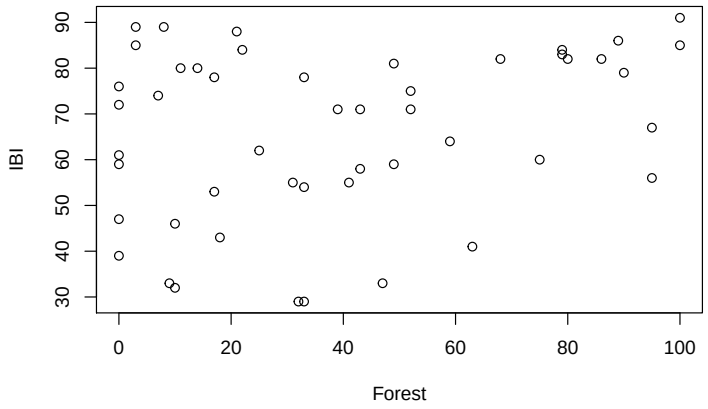
Summary: Based on the graph above, we can tell that the data for the variable IBI is skewed to the left and has no outliers.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   10.00   33.00   39.39   63.00  100.00
```



Summary: Based on the graph above, we can tell the data for Forest is skewed to the right and has no outliers.

Part B



Summary: Based on the graph above, we can tell that as the Forest is decreasing the data for the variable IBI varies more. Also, there is no correlation between Forest and IBI.

Part C

$IBI_i = \beta_0 + \beta_1 \cdot Forest_i + \epsilon_i$ for $i = 1,2,3,\ldots,49

$IBI = 59.91 + 0.153 * Forest + \epsilon$

Part D

Null Hypothesis: $\beta_1 = 0$

Alternative Hypothesis: $\beta \neq 0$

Part E

$\alpha = 0.05$

$$T = \frac{\hat{\beta_1}}{SE_{\hat{\beta_1}}} = \frac{0.153}{0.08} = 1.9$$

$$SE_{\hat{\beta_1}} = \sqrt{\frac{S^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} = \sqrt{\frac{316.4}{49781.6}} = 0.08$$

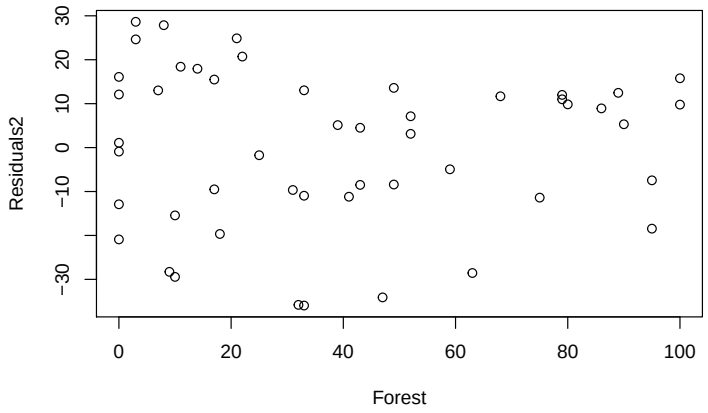
$$S^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y_i})^2}{n-2} = 316.4$$

Degrees of Freedom: n-2 = 47

P-value: P = 2*P(T>1.91) = 0.06

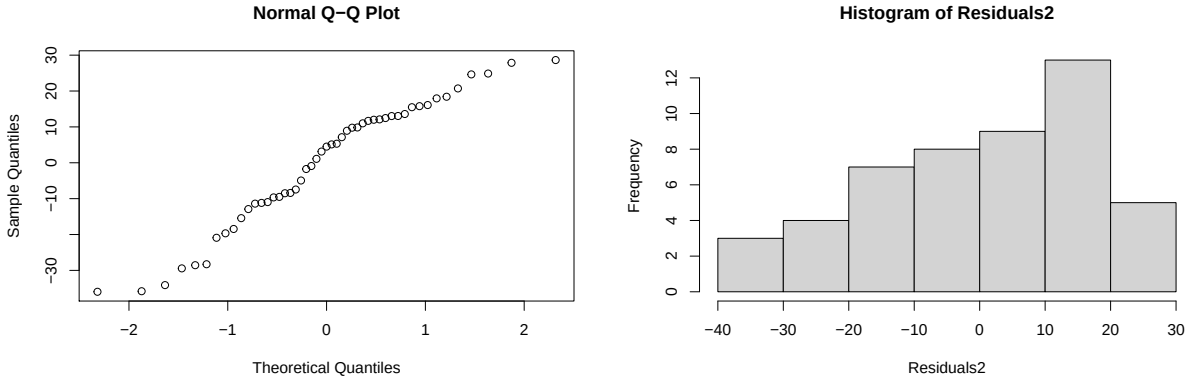
P = 0.06 which is greater than α. There is no significant evidence that there is a linear relationship between IBI and Forest. Hence, we fail to reject the null hypothesis.

Part F



Summary: Based on the graph above, we can tell that the residuals do not have any patterns.

Part G



Summary: Based on the graph above, we can tell that the residuals are skewd to the left and does not have a normal distribution.

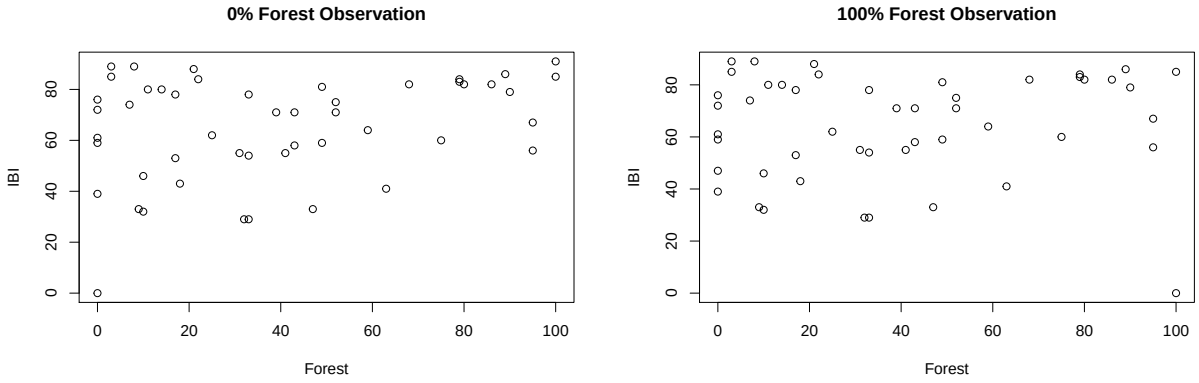
Part H

Based on the graphs in the previous parts, we can say that the assumptions being made are not reasonable and the residuals do not have a normal distribution.

Question 10.34

Based on the first analysis of the area and variable IBI, we can tell that there is significant evidence of a linear relationship. Whereas in the second analysis, we can tell that there is no significant evidence of a linear relation. Also, we can tell that the residuals is skewed to the left and does not have a normal distribution. Between the two anlysises I believe that the first is more appropriate because the regression seems to be more clearly defined and the second does not produce a normal distribution and linear relation.

Question 10.35



At 0% Forest Observation, the Forest and variable IBI relationship is positively associated as the P-value decreases.

At 100% Forest Observation, the Forest and variable IBI relationship is negatively associated as the P-value increases.

Question 10.36

Part A

IBI = 52.923 + 0.46(Area) Standard Error = 16.53.

Based on the IBI formula and standard error above, the 95% confidence interval for the mean response corresponding to an area of 40 km2 is 65.6 and 77.04.

Part B

Based on the IBI formula and standard error, the 95% confidence interval for a future response corresponding to an area of 40 km2 is 37.6 and 105.1.

Part C

From the first part, we can be 95% certain that the area is between 65.61 and 77.04. Using the prediction interval, we can say that we can be 95% certain that the next new observation is between 37.58 and 105.08.

Part D

I think you need to be careful using these results everywhere because different areas may have different results. Similar areas may have similar results.

Problem 10.37

Area: 52.92 + (0.46 * Area) = 57.5 Forest: 59.91 + (0.15 * Forest) = 69.5 The prediction interval is a large range resulting in the forest estimate being greater than the area estimate. This can cause uncertainty.