April Ullrich

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

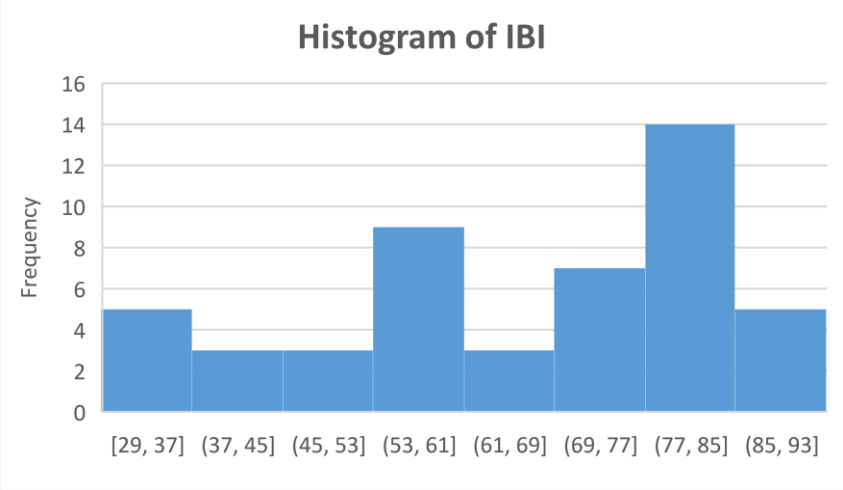
- A. Ullrich

Homework 7

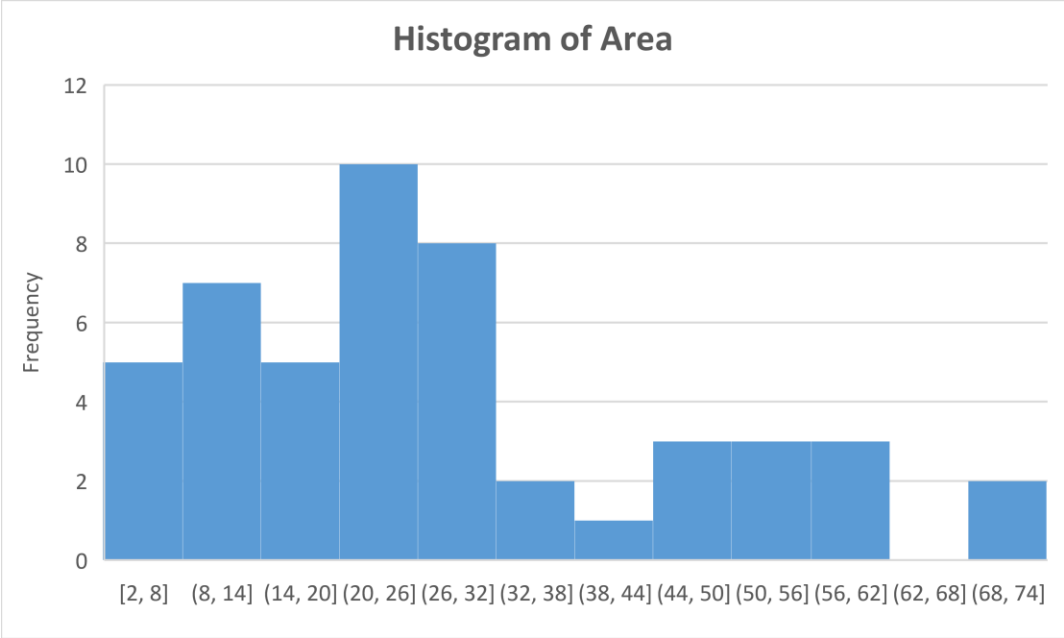
Problem 10.32

a)

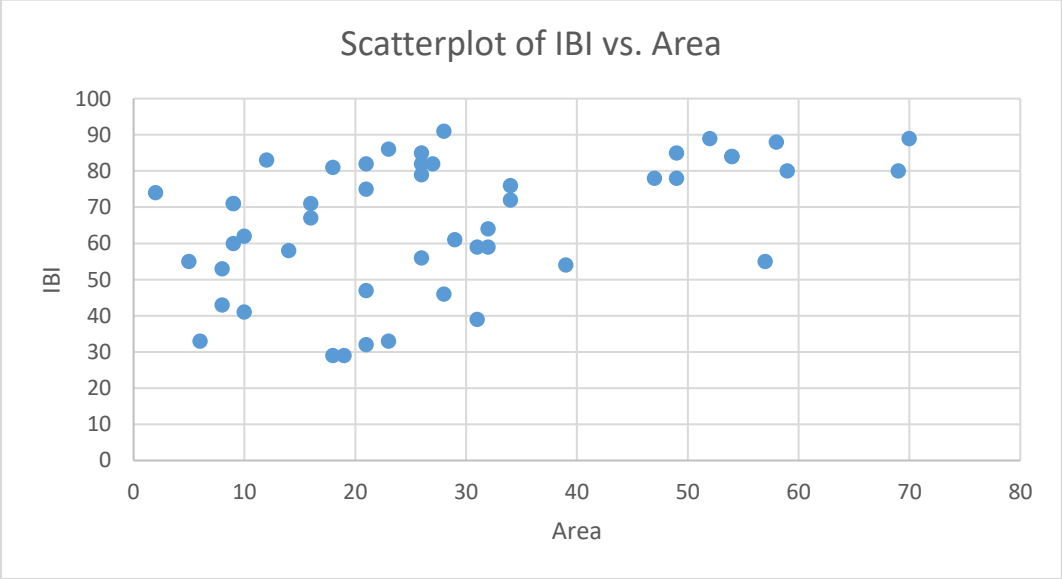
	IBI Statistics
Mean	65.94
Standard Deviation	18.28
Minimum	29
Q1	54.50
Median	71.00
Q3	82.00
Maximum	91.00
Observations	49



	Area Statistics
Mean	28.29
Standard Deviation	17.71
Minimum	2.00
Q1	15.00
Median	26.00
Q3	36.50
Maximum	70.00
Observations	49



b)



- No clear association between the two variables
- No real outliers

c) Model: $IBI = \beta_0 + \beta_1(Watershed\ Area) + \varepsilon_i\ ,\ i = 1,2, \dots 49$

d)

Null Hypothesis (H_0): There is no linear relationship between IBI and Area ($\beta_1 = 0$)

Alternative Hypothesis (H_A): There is a linear relationship between IBI and Area ($\beta_1 \neq 0$)

e)

IBI = 52.9 + 0. 460 (Area)

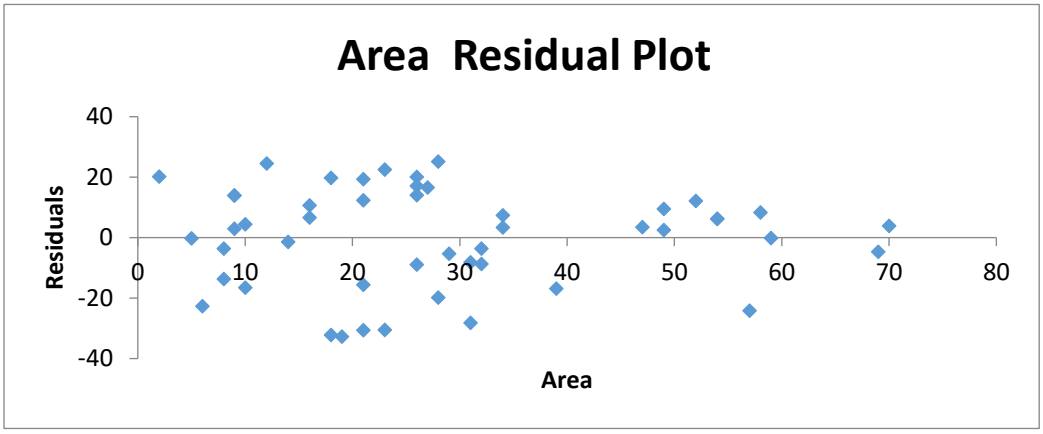
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3189.269734	3189.269734	11.66544488	0.001321791
Residual	47	12849.54659	273.3946084		
Total	48	16038.81633			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	52.9229579	4.483524181	11.80387476	1.16502E-15
Area	0.460155168	0.134726693	3.415471399	0.001321791

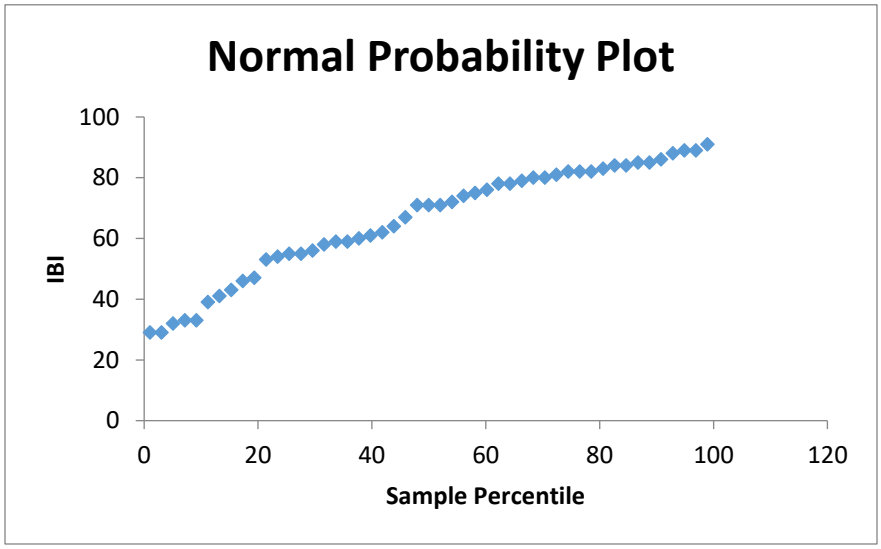
- At a test statistic of 3.42, there is a P-value of 0.001, which is small
- Therefore, we reject the null hypothesis and conclude there is sufficient evidence for the linearity in the regression line.

f)



- There appears to be no pattern in the residuals plot, therefore the errors are independent

g)



- The residuals line is fairly straight thus the residuals are normal

h)

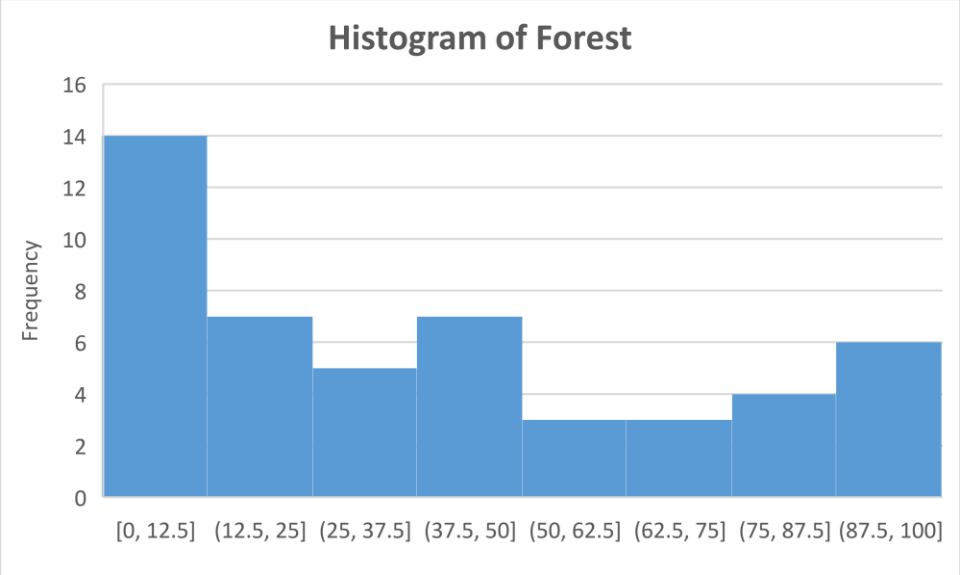
- From the scatterplot, there are no violations of any assumptions of regression

Problem 10.33

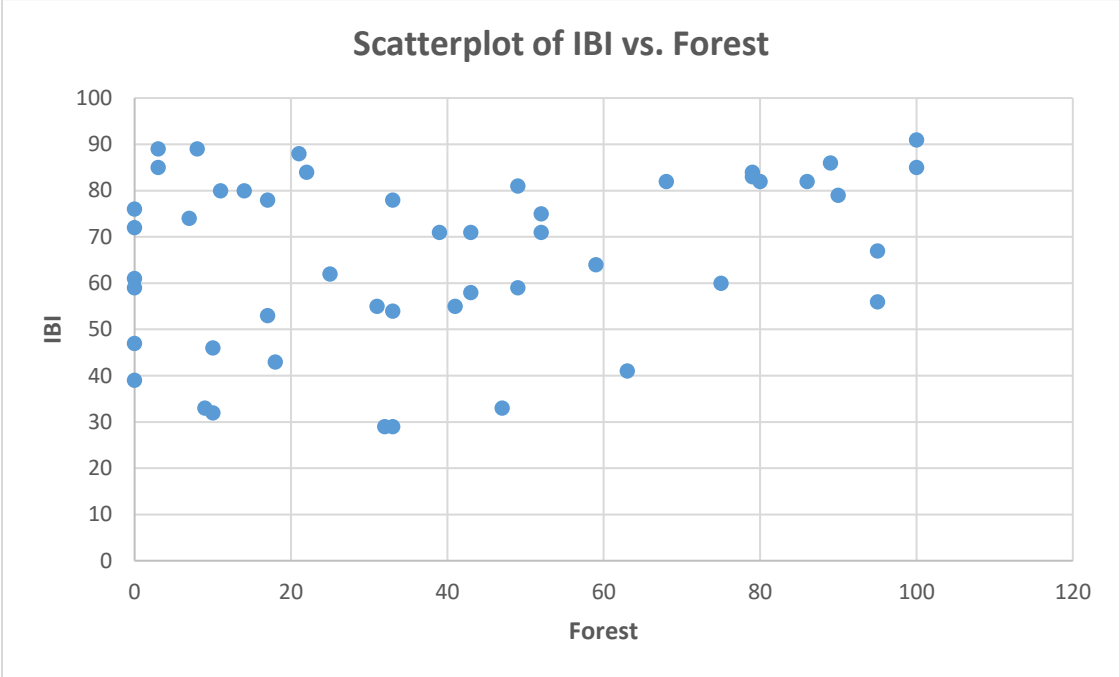
a)

(*IBI already statistics found in 10.32*)

	Forest Statistics
Mean	39.39
Standard Deviation	32.20
Minimum	0
Q1	10
Median	33
Q3	59
Maximum	100
Observations	49



b)



- Based on the scatterplot, the association between the two variables appears to be weak and positive.

c) Model: $IBI = \beta_0 + \beta_1(Forest) + \varepsilon_i\,,\ i = 1,2,\dots 49$

d)

Null Hypothesis (H_0): There is no linear relationship between IBI and Forest ($\beta_1 = 0$)

Alternative Hypothesis (H_A): There is a linear relationship between IBI and Forest ($\beta_1 \neq 0$)

e)

IBI= 59.9 + 0.153 (Forest)

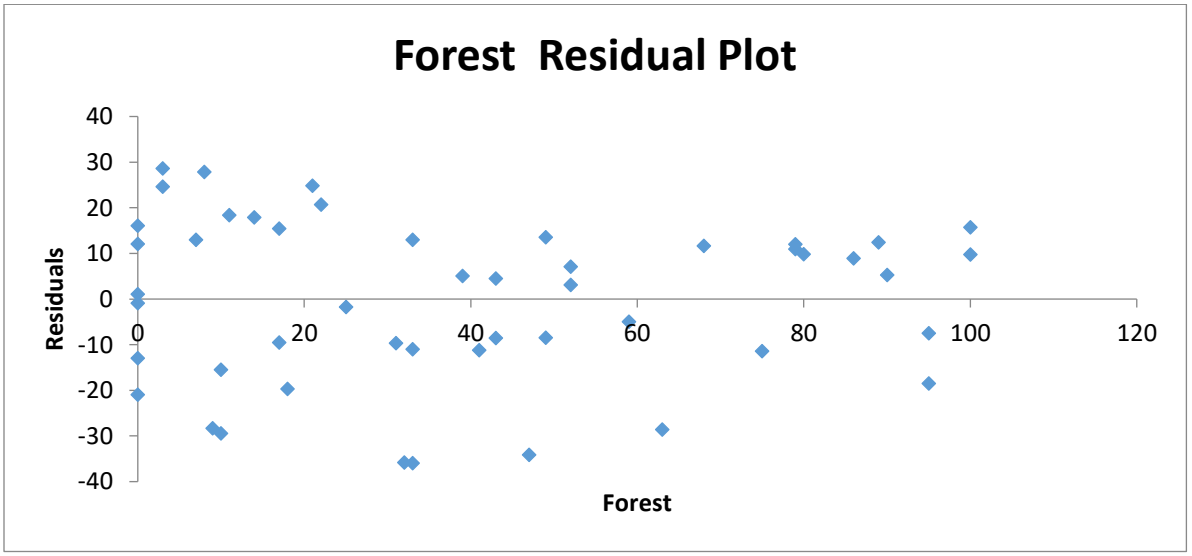
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1167.350588	1167.350588	3.689312044	0.060840299
Residual	47	14871.46574	316.4141646		
Total	48	16038.81633			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	59.90724798	4.039574957	14.83008698	2.38598E-19
Forest	0.153132046	0.079724791	1.920758195	0.060840299

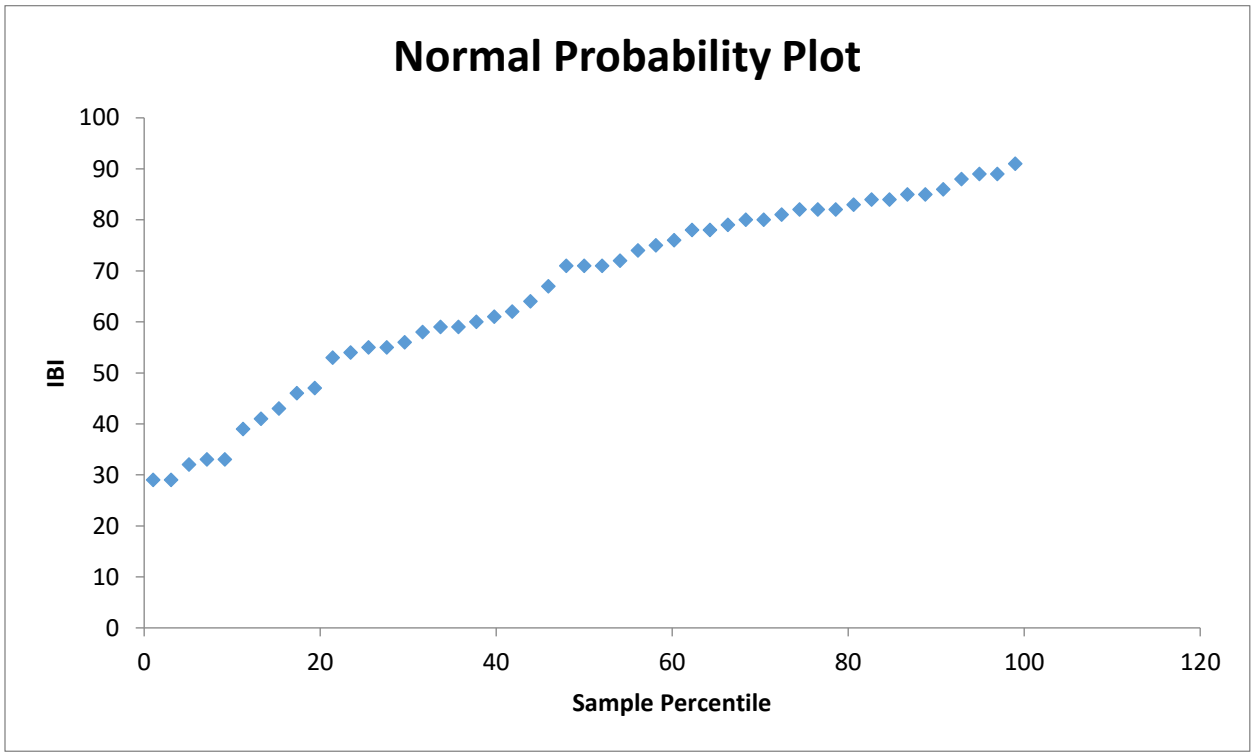
- At a test statistic of 1.92, there is a P-value of 0.06 which compared to a significance level of 0.05 is large.
- Therefore, we fail to reject the null hypothesis and conclude that there is not necessarily a linear relationship between IBI and Forest.

f)



- There appears to be no pattern in the residuals plot, therefore the errors are independent

g)



- The residuals appear to be heavily skewed to the left and are thus not normal

h) Assumptions for regression are violated as they assume residuals to be normally distributed when in fact they are not as described in part g.

Problem 10.34

- After comparing both analyses, Area as an explanatory is more suitable for IBI as it has a higher correlation coefficient (R^2) value as well as produced a statistically significant p-value indicating a linear relationship between Area and IBI. On the other hand, Forest did not produce a statistically significant result as the P-value was too large and thus no linear relationship could be concluded.

Problem 10.35

Case 1: Decrease the IBI to 0.0 for an observation with 0% Forest

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	9343.021041	9343.021041	15.17826236	0.00030861
Residual	47	28930.97896	615.5527438		
Total	48	38274			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	41.65069897	5.634304943	7.392340207	2.0827E-09
Forest	0.433220596	0.11119828	3.895928947	0.00030861

- At a test statistic of 3.896, the p value is low at 0.0003, thus we reject the null hypothesis.
- Therefore, we can conclude that when the lower outliers were removed in this instance, there is a linear relationship between the two variables of Forest and IBI

Case 2: Decrease the IBI to 0.0 for an observation with 100% Forest

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	186.2040065	186.2040065	0.38175037	0.539650418
Residual	47	22924.89803	487.763788		
Total	48	23111.10204			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	64.75585209	5.015479851	12.91119773	4.56027E-17
Forest	-0.061158939	0.098985187	-0.617859507	0.539650418

- At a test statistic of -0.618, the p-value is extremely high at 0.53, thus we fail to reject the null hypothesis.
- Therefore, we conclude that this approach does not improve the regression to demonstrate a linear relationship between the two variables.

Summary: Based on the exercise, I observed that it is possible to eliminate outliers in a data set by setting both their values to zero. However, it is not advantageous to set only one value to 0

when the other is at 100% because this does not support the trend that the two variables would have a linear positive slope as this point would pull the regression equation towards the axis with the data point set at 100.

Problem 10.36

a) 95% Confidence Interval: (65.6, 77.0)

b) 95% Prediction Interval: (37.6, 105.1)

c) If we sample streams from Ozark Highlands, then the expected average IBI for these streams would be between 65.21km^2 to 77.04 km^2
If we sample streams from Ozark Highlands, then the expected IBI for these streams would be between 37.6km^2 to 105.1 km^2

d) Extending the results to other streams in Arkansas may be possible as long as these areas have similar conditions to the Ozark Highlands. I would not extend the results beyond the state as the climate differences and other factors may attribute to very different results.

Problem 10.37

Predicted IBI from Area Model: 57.5

Predicted IBI from Forest Model: 69.5

- Area as the explanatory variable is a much better model to predict IBI than Forest and its model has a statistically significant p-value which means there is in fact a linear relationship between Area and IBI. Also, by comparing the above calculated values by each respective model, we see that the Area Predicted IBI is much closer to the actual IBI value than the Forest Predicted IBI value.