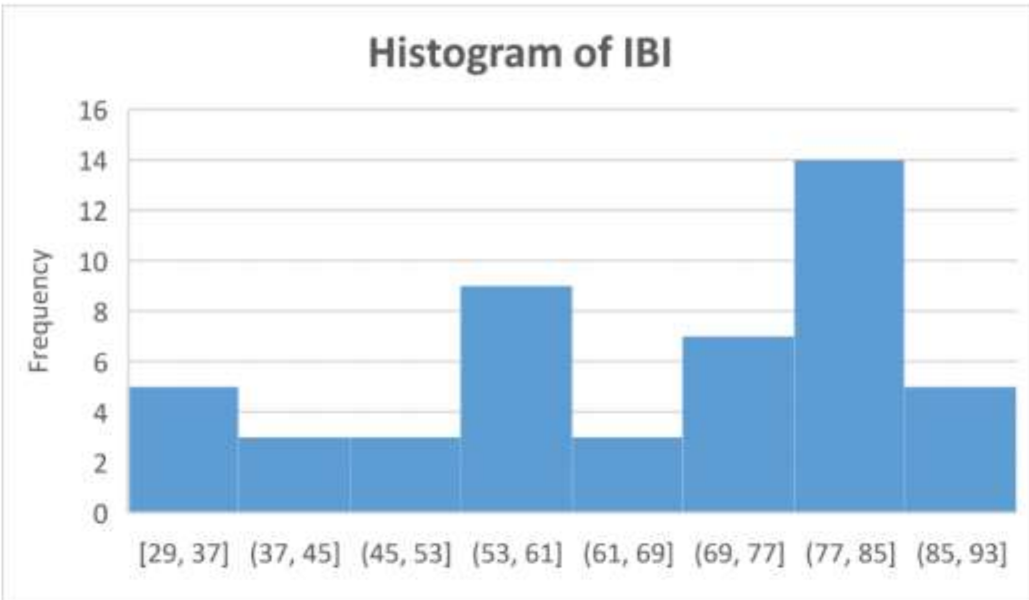
Theodore Jagodits / MA 331 / HW #7

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

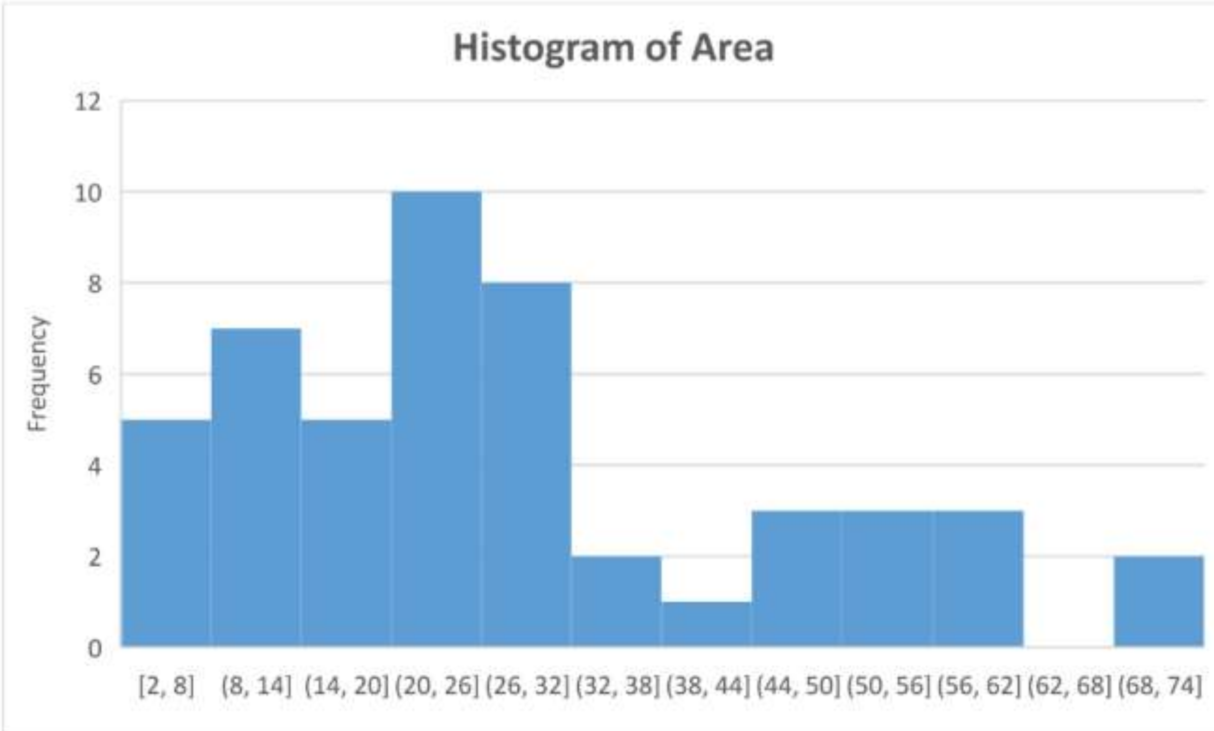
10.32

a)

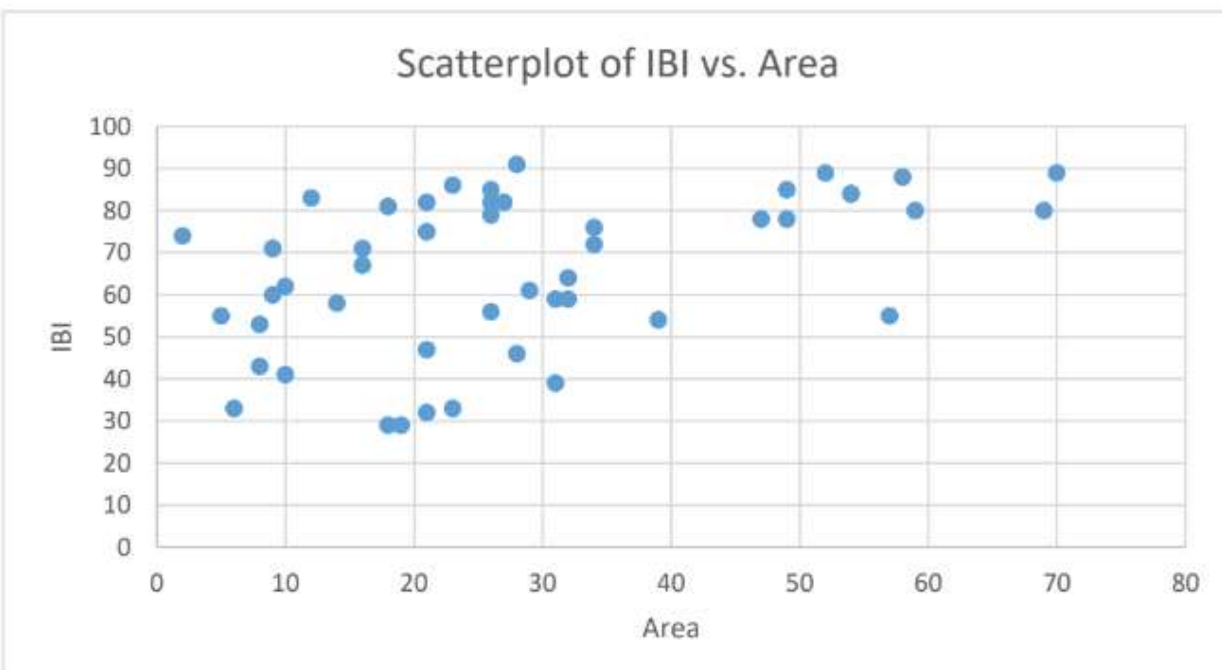
	Mean	Std. Dev.	Min.	Q1	Median	Q3	Max	Obs.
IBI Stat	65.94	18.28	29	54.5	71	82	91	49



	Mean	Std. Dev.	Min.	Q1	Median	Q3	Max	Obs.
Area Stat	28.29	17.71	2	15	26	36.5	70	49



b)



There is no clear association between variables or any noticeable outliers. Does not really show anything.

c)

Model: IBI = $\beta 0 + \beta 1$(Watershed Area) + εi, i = 1,2, ... 49

d)

Null Hypothesis: There is no linear relationship between and IBI and Area $\beta 1 = 0$

Alternative Hypothesis: There is a linear relationship between IBI and Area, $\beta 1$ not equal 0

e)

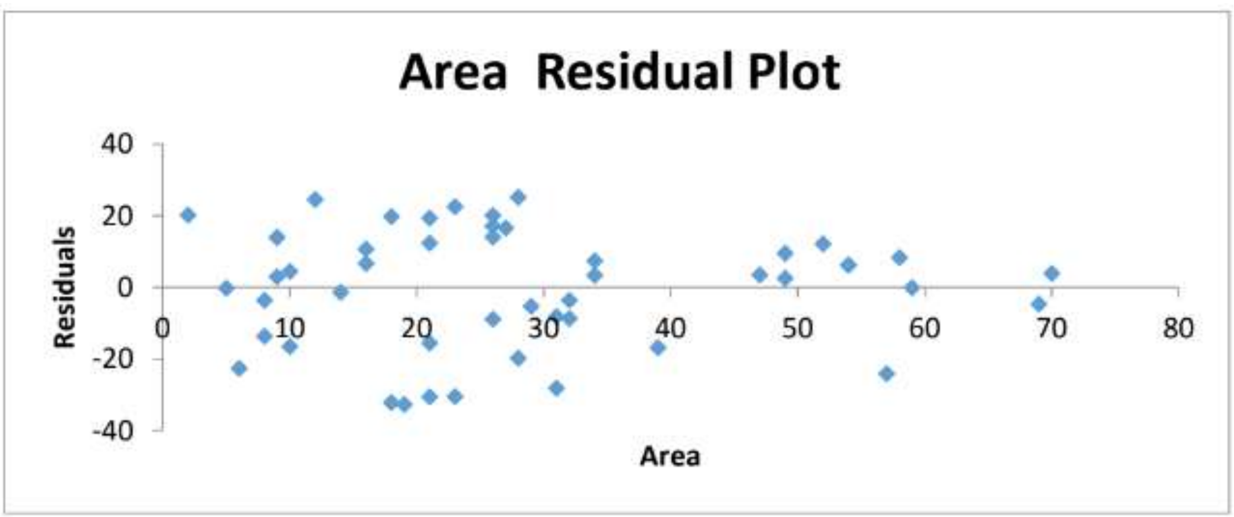
IBI = 52.9 + .46(Area)

ANOVA	df	SS	MS	F	Sign. F
Regression	1	3189.27	3189.27	11.67	.001321
Residual	47	12849.55	273.4		
Total	48	16038.82			

	Coefficients	Standard Error	t-Stat	p-Value
Intercept	52.92	4.48	11.80	1.16E-15
Area	0.46	0.135	3.41	.0013

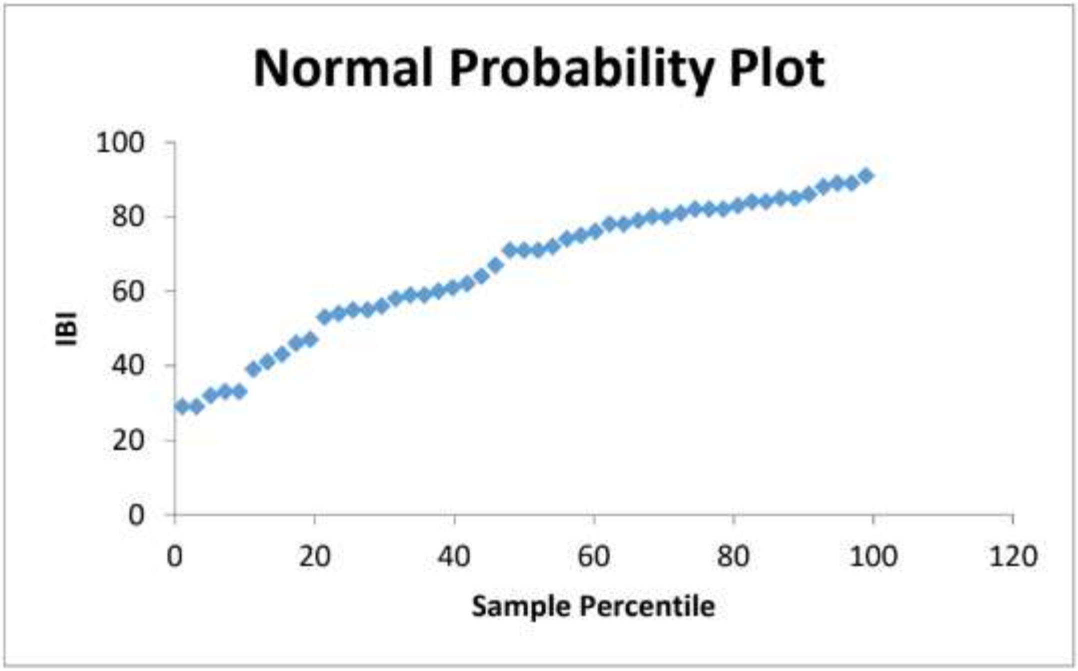
At t-Stat 3.42 we find a P-value of 0.001 which is very small so we can therefore reject the null hypothesis since there is sufficient evidence for the linearity in the regression line.

f)



It looks like there is no pattern in the residuals plot, so no errors appear to be independent.

g)



The residual line follows an almost straight line, so the residuals are normal.

h)

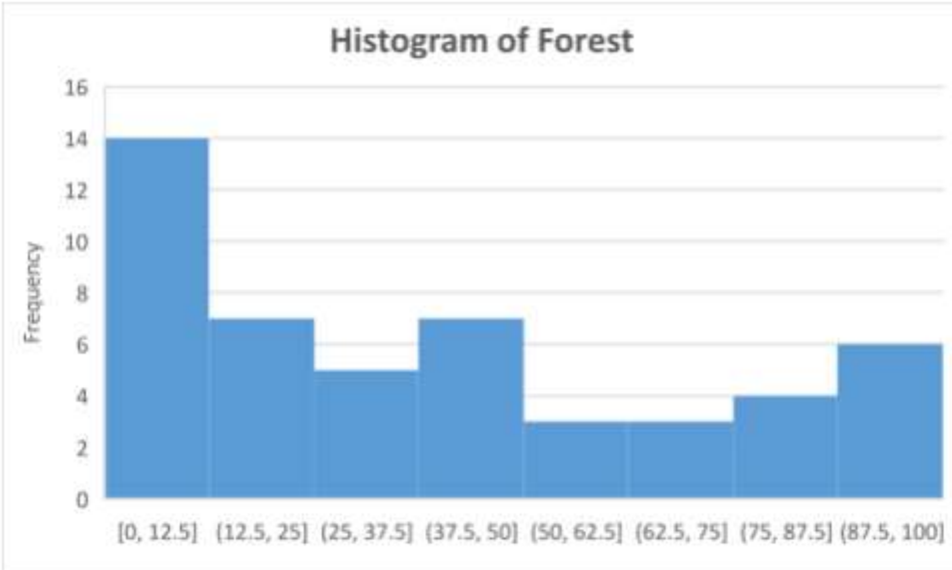
looking at the scatterplot, there are no violations in any of the regression assumptions.

10.33

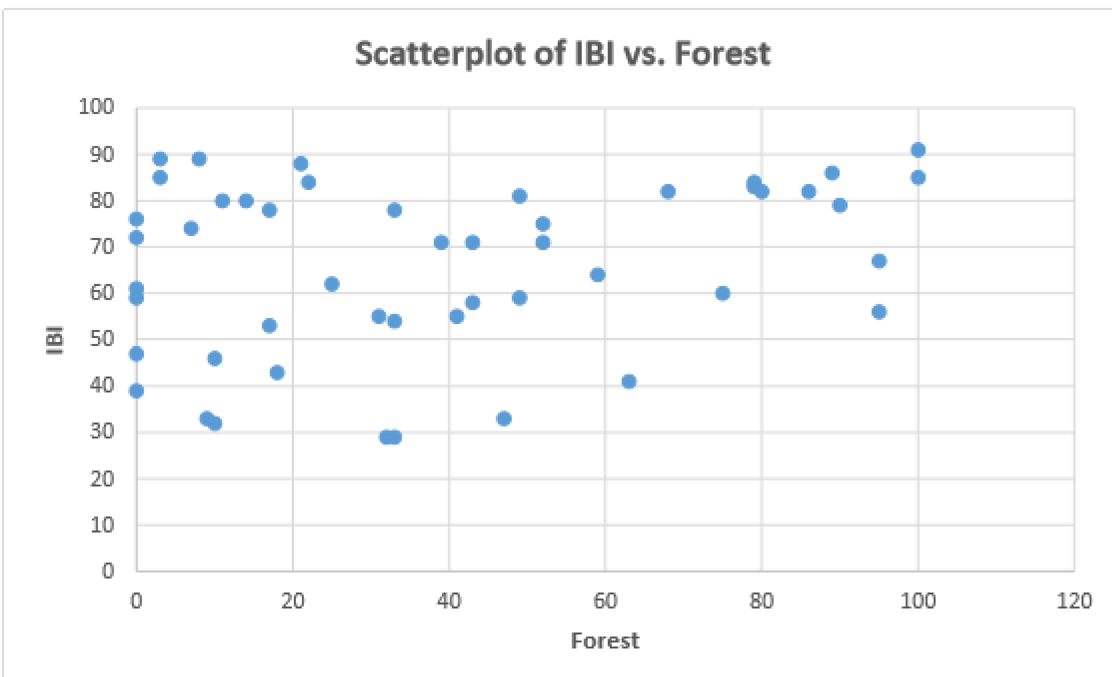
a)

statistics found in 10.32

	Mean	Std. Dev.	Min.	Q1	Median	Q3	Max.	Obs.
Forest Statistics	39.39	32.20	0	10	33	59	100	49



b)



The scatterplot indicates that there is barely any association, so weak and positive.

c)

IBI = $\beta 0 + \beta 1$(Forest) + εi, i = 1,2, ... 49

d)

Null Hypothesis: There is no linear relationship between IBI and Forest.

Alternative Hypothesis: There is a linear relationship between IBI and Forest.

e)

IBI=59.9 + 0.153(Forest)

ANOVA

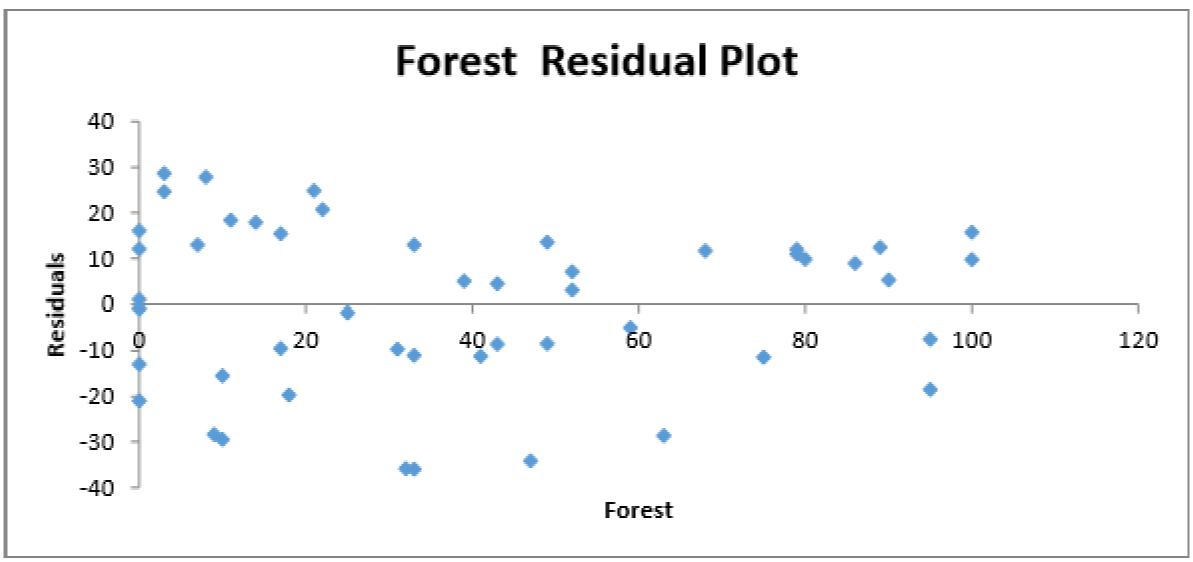
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1167.350588	1167.350588	3.689312044	0.060840299
Residual	47	14871.46574	316.4141646		
Total	48	16038.81633			

	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

The test statistic is at 1.92, and there is a P-value of 0.06 which is pretty large if you compare at the significance level of 0.05.

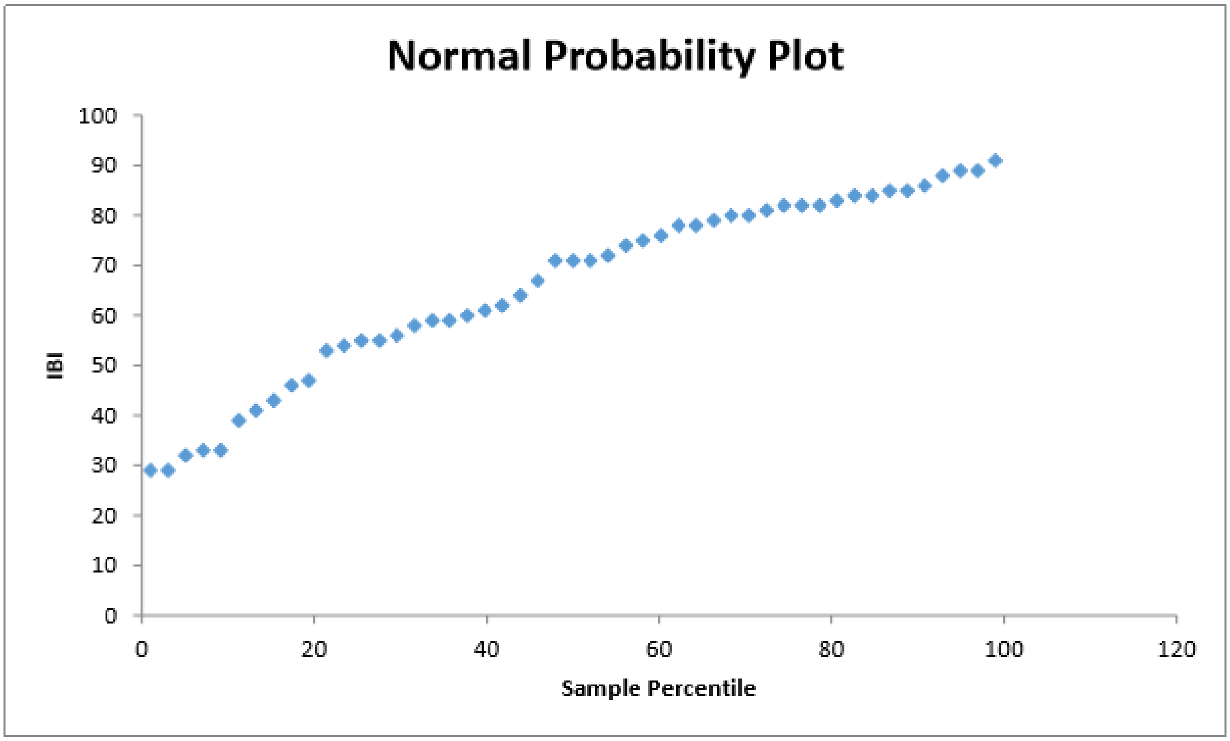
We fail to reject the hypothesis since there is not enough evidence to indicate there is a linear relationship between IBI and Forest.

f)



It looks like there is no pattern to the residual plot, so the errors are assumed to be independent.

g)



The residuals look they are skewed left on the graph so therefore not normal.

h)

One of the assumptions was that residuals were normally distributed, this was violated and we cannot assume this for regression.

10.34

After looking at both analyses, the Area one is better suited for IBI because it has a higher correlation coefficient value as well as produced a statistically significant p-value showing a linear relationship between Area and IBI. The forest experiment did not produce a statistically significant result because the P-Value was too large, and no linear relationship could be concluded.

10.35

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

ANOVA	df	SS	MS	F	Sign. F
Regression	1	9343.021	9343.021	15.17	.0003
Residual	47	28930.97	615.55		
Total	48	38274			

	Coefficients	Standard Error	T - Stat	P – Value
Intercept	41.665	5.63	7.39	~0
Forest	0.433	.1111	3.89	0.0003

At the test statistic 3.89, the P – Value is very low at 0.0003 so we reject the null hypothesis. We can conclude that when lower outliers are removed from this situation, there is a linear relationship between forest and IBI.

Case 2: Decrease IBI to 0 and have 100% Forest observation

ANOVA	df	SS	MS	F	Sign. F
Regression	1	186.2	186.2	.38	.539
Residual	47	22924.89	487.76		
Total	48	23111.1			

	Coefficients	Standard Error	T - Stat	P – Value
Intercept	64.75	5.01	12.91	~0
Forest	-.0611	.09	-.62	0.53

The test statistic is -.618, the p value is pretty high at .53, so we fail to reject the null hypothesis. Therefore, we can conclude that this not help regression at all to show a linear relationship between the two variables.

In summary, I have concluded that elimination outliers in a data set can help improve the model. However, setting one to zero and leaving the other will skew the data towards the outlier and will not give correct results.

<u>10.36</u>

a) 95% CI : [65.6,77.0]

b) 95% PI : [37.6,105.1]

c) If we sample from the Ozarks streams, then the expected average IBI would be between 65.21 km^2 and 77.04 km^2. If we just sample the for the IBI it would be between 37.6 km^2 and 105.1 km^2.

d) These results would only have relevance in areas in Arkansas that have the same geographic conditions, I do not believe that these results can be used outside of the state since conditions vary widely.

<u>10.37</u>

Predicted IBI from Area Model: 57.5

Predicted IBI from Forest Model: 69.5

The area as the explanatory variable is probably a better model to predict the IBI than the Forest variable, because it has a statistically significant p value which means there is a linear relationship between area and IBI. We can also see by computing from the calculated values above that Area predicted IBI is closer to the actual IBI value than the other model.