Pledge: I pledge my honor that I have abided by the Stevens Honor System. - **Eric Altenburg**

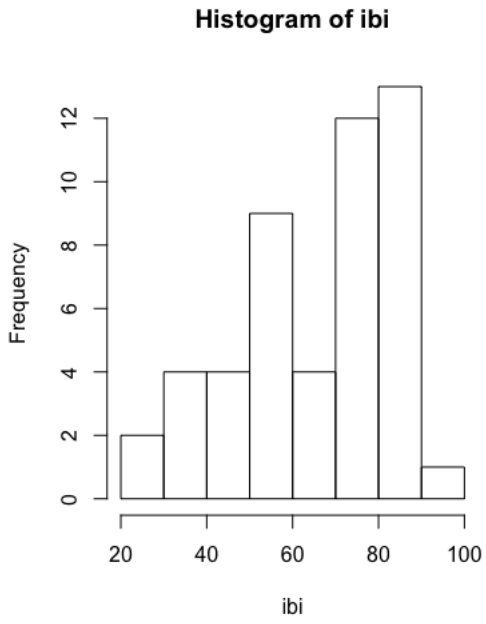
10.32: Page 606

(a)

Numerical data for IBI:

n	Mean	St. Deviation	Minimum	Q1	Median	Q3	Maximum
49	65.94	18.28	29	54.50	71	82	91

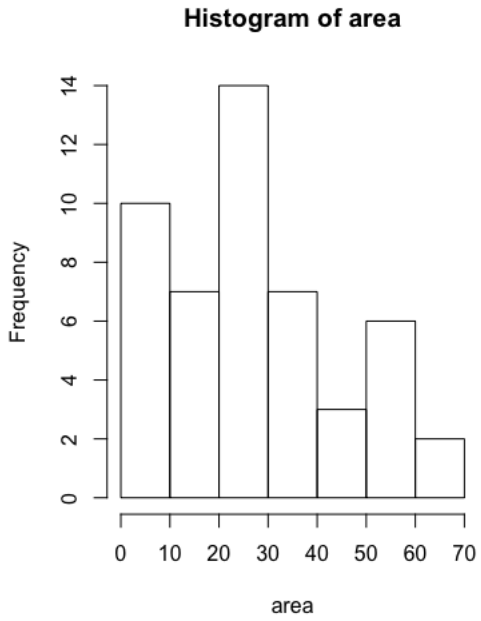
Histogram for IBI:



Based on this, it is left-skewed.

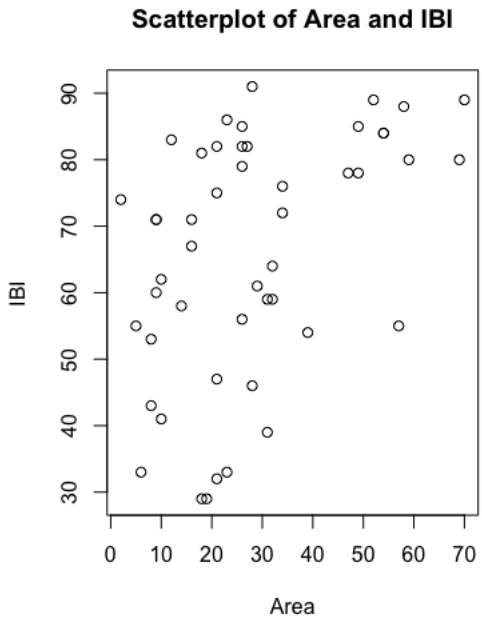
n	Mean	St. Deviation	Minimum	Q1	Median	Q3	Maximum
49	28.29	17.71	2	15	26	36.5	70

Histogram for Area:



From this, we can once again conclude that the data is approximately symmetric.

(b)



Based on the scatter plot, there is little association between the two variables as the points are scattered. No extreme outliers or observations are present in the sample either.

(c)

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

(d)

H_0 : There is no linear association between the area and the IBI

$\beta_1 = 0$

H_1 : There is a linear relationship between the area and the IBI

$\beta_1 \neq 0$

After plugging in the function into R, the corresponding test statistic value is $t = 3.415$ and the $p - value = 0.00132$. Since the p-value is considerably small, we can reject H_0 stating that there is evidence of a linear relationship between the area and the IBI.

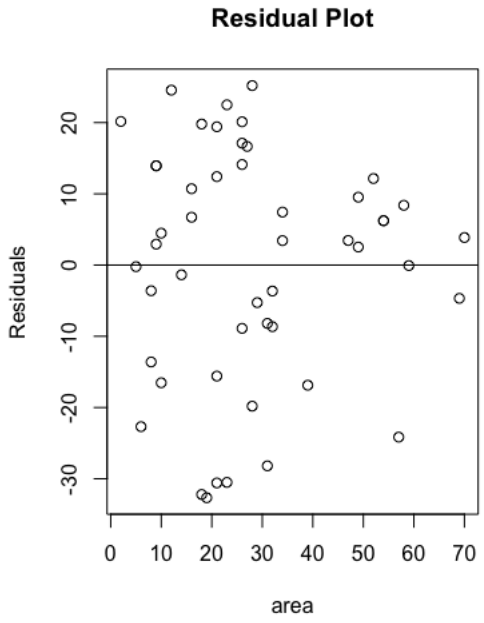
(e)

$IBI = 52.923 + 0.4602(Area)$

$s = 16.5346$

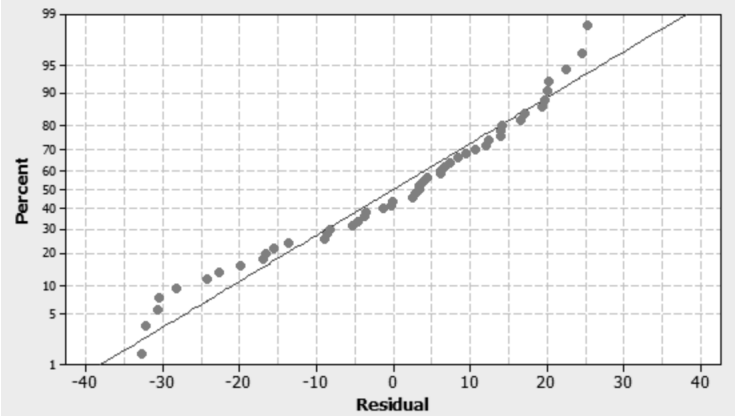
Based on the above line and it's slope, for every one square meter of increment in the area, there will be 0.4602 units of increment in the IBI.

(f)



The observations are scattered over the origin line, and there is no pattern in the above plot. Therefore, the error are independent.

(g)



The residuals are close to the line with no outliers in the sample. Therefore, the residuals are normal.

(h)

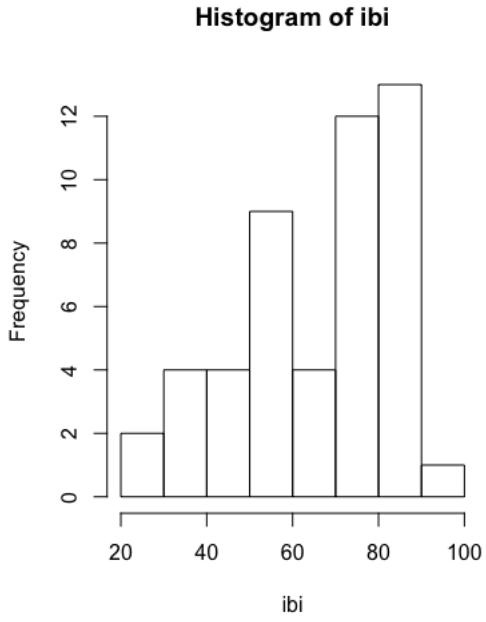
No violations of any assumptions of regression based on the scatterplot.

10.33: Pages 606 - 607

(a)

Numerical data for IBI:

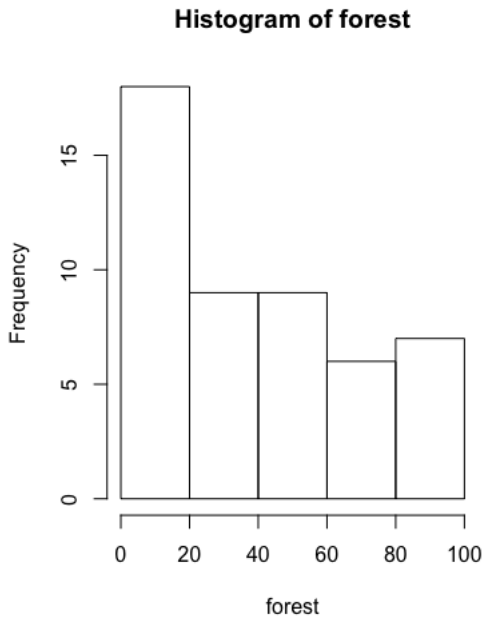
n	Mean	St. Deviation	Minimum	Q1	Median	Q3	Maximum
49	65.94	18.28	29	54.50	71	82	91



Based on this, it is left-skewed.

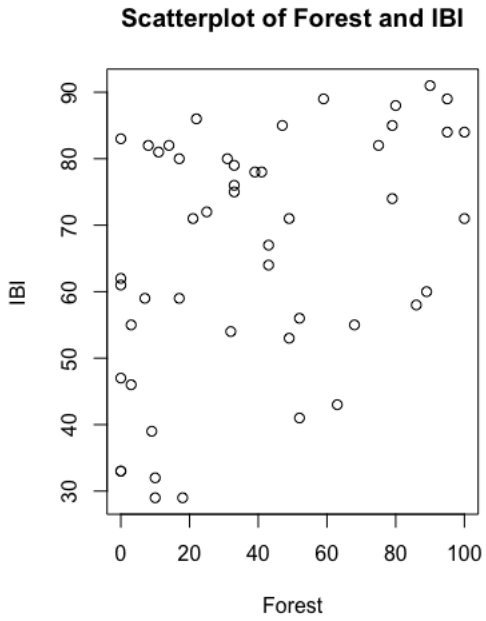
Numerical data for forest percent:

n	Mean	St. Deviation	Minimum	Q1	Median	Q3	Maximum
49	39.39	32.2	0	10	33	63	100



Based on this, it is right-skewed.

(b)



Based on this, the sample does not form any specific pattern, yet the IBI scatters more with smaller percentage values of forest and it tends upwards leading to the conclusion that they do not follow any specific pattern, but the relationship is weak and positive.

(c)

x is forest which is explanatory variable

y is IBI which is response variable

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, ..., 49$ where ε_i are independent $N(0, \sigma)$ variables.

(d)

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

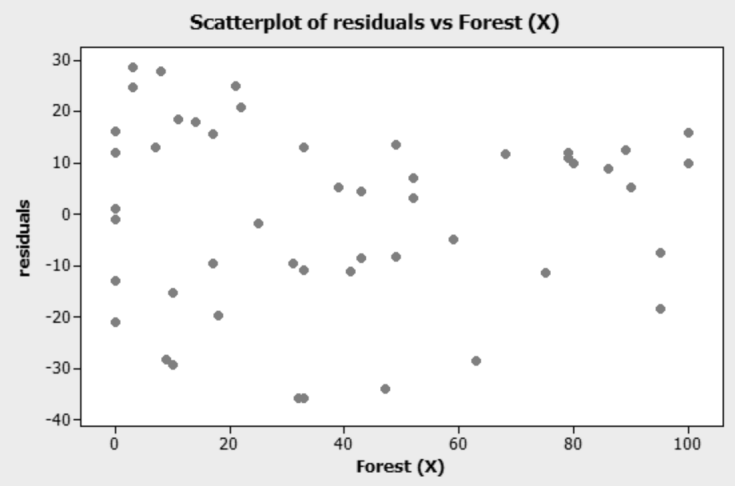
(e)

$\hat{IBI} = 59.90 + 0.153(Forest)$

Standard error is 17.79

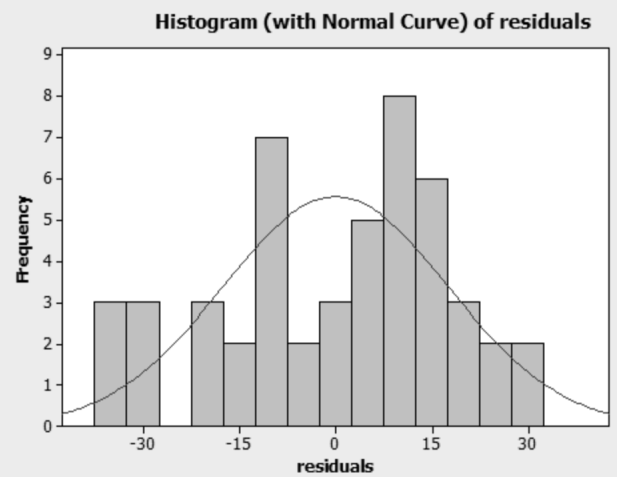
When testing the hypotheses in part (d), the testing statistic $t = 1.92$ and the $p - value = 0.061$ which means the null hypotheses fails to be rejected stating that the slope coefficient β_1 is not significantly different from 0.

(f)



No pattern is observed here, the points are randomly scattered.

(g)



The residuals are left-skewed..

(h)

In part (c) it is assumed that the errors of the residuals are independent and $N(0, \sigma)$, but they are left-skewed as seen in part (g) therefore, the assumption is not reasonable.

10.34: Page 607

The Area as an explanatory variable is more desirable for the analysis because the r^2 value is greater while the $p - value$ is less, which is not the case for when the Forest is the explanatory variable.

10.35: Page 607

The first change ends up decreasing the $p - value$ which means the relationship becomes more significant because it amplifies the positive association between the two.

The second change ends up increasing the $p - value$ which means the relationship is less significant, and it in turn weakens the association.

10.36: Page 607

(a)

Using R, the regression line equation between IBI and the Area is:

$IBI = 52.923 + 0.4602(Area)$

The standard error is 16.5346

The 95% CI for 40 km^2 is (65.61, 77.04)

(b)

The 95% PI for 40 km^2 is (37.58, 105.08)

(c)

Based on the CI in part (a), if one were to sample the number of streams, the expected average IBI would be between 65.61 km^2 and 77.04 km^2. From part (b), the expected IBI value corresponding to 40 km^2 watershed area would be between 37.58 km^2 and 105.08 km^2.

(d)

Yes because all of the assumptions of the regression study are assumed to be satisfiable in the available study. But that does not mean it should be applied to all streams in all states as there might be different factors that can influence the values such as geological location and other factors. The values of the intervals would be different for each state.

10.37: Page 607

$Area = 10, \hat{y} = 57.52$ but for $Forest = 63, \hat{y} = 69.55$.

Both of the predictions though have uncertainty because the prediction intervals area about 70 units wide.