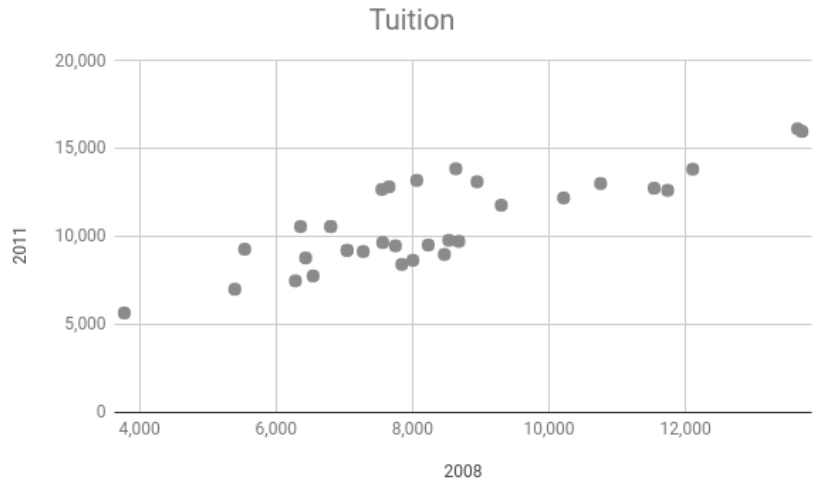
Christina Li
November 30, 2018

"I pledge my honor that I have abided by the Stevens Honor System." –cli50

Homework 7

10.16

a.

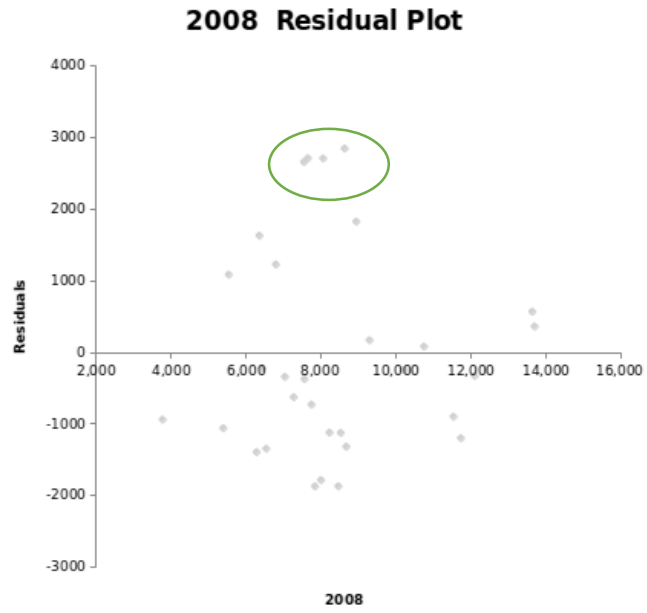


There is a strong, positive and linear correlation between the in-state tuition in 2008 to 2011. There are no outliers or unusual points.

b.

$\widehat{IN11} = 2769.18 + 0.94\text{IN08}$

c.



The 5 points in the green circle are unusually high compared to the rest of the values.

d.
The residuals don't appear to be approximately normal with the constant variance between of the 5 outliers as seen in the graph on part c.

e.
$\widehat{IN11} = 2058.38 + 0.97IN08$
Without the 5 California schools, it can be observed from the equation above that the y-intercept decreased while the slope is increased slightly.

f.
Between the two models, I would use the equation of the least squares regression line without the 5 California schools because with all schools, the residual plot was not approximately normal.

10.17
a.
H_0: $\beta_1 = 0$
H_1: $\beta_1 > 0$

b.
t = 13.94
df = 26
P-value = 0.000 => P < 0.0005
Because the P value is very small, we reject the null hypothesis. Therefore, there is sufficient evidence to conclude that the tuition amount will increase.

c.
$0.9675 \pm 2.056(0.06939) = (0.825, 1.11)$
The 95% confidence interval tells us that the tuition fee will increase between 0.825 and 1.11 percent.

d.
$r^2 = 88.2\%$

e.
β_0 is not interested in this problem because there were no colleges that had $0 tuition fee.

10.18
a.
$\widehat{IN11} = 2058.38 + 0.97IN08$
In 2008, the in-state tuition of State U was $5100.

$\widehat{IN11} = 2058.38 + 0.97IN08$
$= 2058.38 + 0.97(5100)$
$= 2058.48 + 4947$
$= 7005.38$
In 2011, the predicted in-state tuition of State U can increase up to $7,005.38.

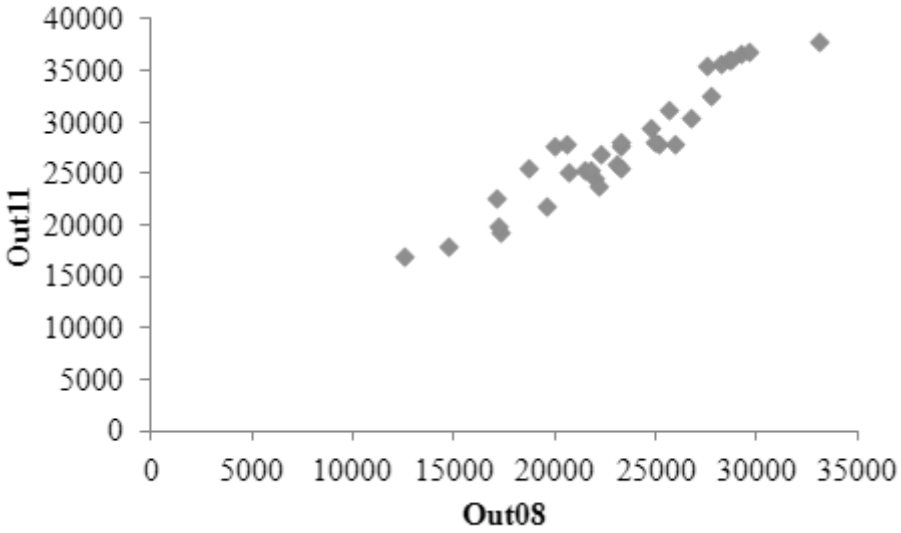
b.
$\widehat{IN11}$ = 2058.38 + 0.97IN08
In 2008, the in-state tuition of Moneypit U was $15,700.

$\widehat{IN11}$ = 2058.38 + 0.97IN08
= 2058.38 + 0.97(15700)
= 2058.38 + 15229
= 17287.38

In 2001, the predicted in-state tuition of Moneypit U can increase up to $17,287.38.

c.
The equation of regression line used the minimum in-state tuition $3,778 and the maximum $13,706. For State U, the in-state tuition was $5,100 in 2008, which is within range. The equation of regression line is then used to the predict the in-state tuition in 2011 for State U. For Moneypit U, the in-state tuition was $15,000 in 2008, which is outside the range. However, if the regression line is used to predict the in-state tuition in 2011 for Moneypit, it can't say that the university with the in-state tuition in 2008 will follow the same pattern. Then, the prediction may not appropriate and will be extrapolation.

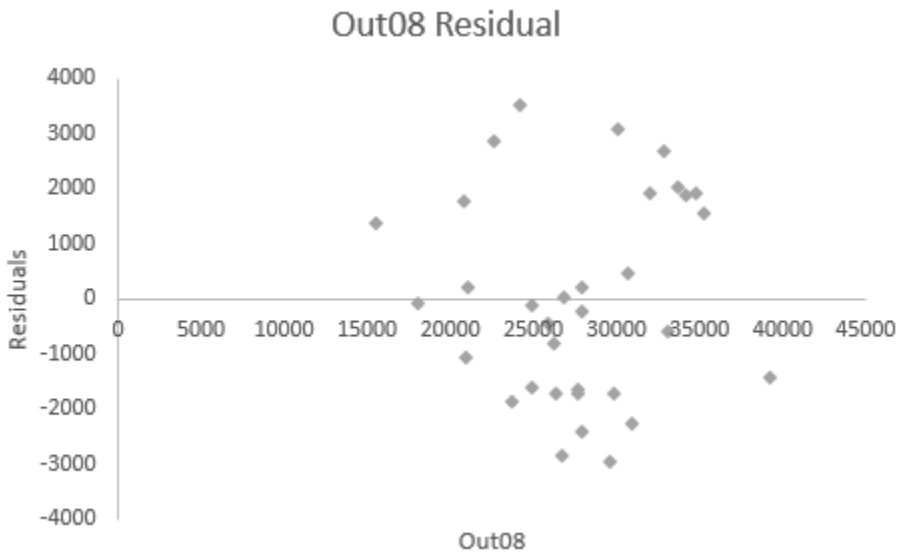
10.19
a.



There is a strong, positive and linear correlation between the two variables. There are no outliers or unusual points.

b.
$\widehat{Out11}$ = 1075.07 + 1.15Out08

c.



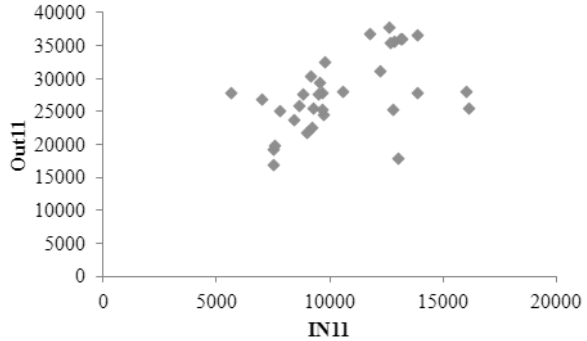
There are no unusual or outlier points observed in the plot.

d.
The residuals appear to be approximately normal with constant variance since all the points seem to be randomly scattered as there is no obvious pattern in the residual plot.

10.20
a.
The 95% confidence interval for the slope is (1.00 and 1.30). Thus, there is a 95% confidence that the tuition fee will increase between 1.00 and 1.30 percent.

b.
Because the samples are not independent, we cannot use the two sample t test. To use the two sample t test, the samples must be independent.

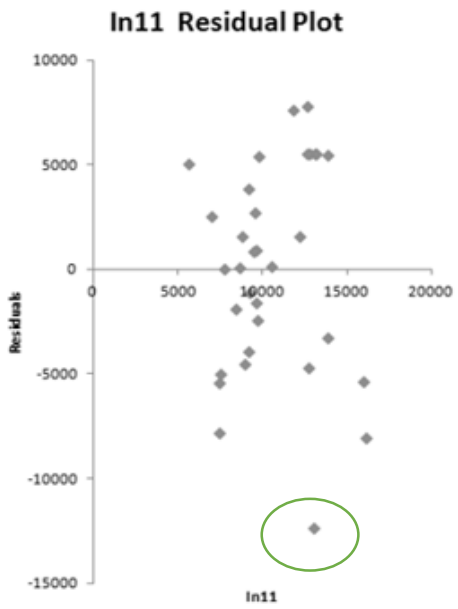
10.21



The points do not follow a straight-line trend. Thus, there is a weak, positive and linear correlation between the two variables.

$\widehat{Out11} = 17159.72 + 1.017\text{IN11}$

Christina Li
November 30, 2018



When comparing in-state and out-of-state tuition, Minnesota appears to be an outlier since the in-state tuition is $13,022 and out-of-state tuition is $18,022. The residual plot against x does not show any obvious problems except for the low outlier(Minnesota) that is shown in the green circle. The residuals appear to be approximately normal with constant variance since the points seem to be randomly scatter as there are no obvious pattern in the residual plot.