|  |  |
| --- | --- |
| **Module:** | ST4400 |
| **Name:** | James Henry Hehir |
| **Student Number:** | 120224791 |
| **Chapter:** | 3 |

**Maximum 2 pages! Do not delete the page number in the footer.**

1. A residual by definition is the observed value of y minus the fitted value of y. The model over-estimates by 8899.441.

Text

Description automatically generated



1. An outlier is detected when the corresponding residual to the case, is large or if the modulus of the studentized residual exceeds 2. So for a normal distribution most values 95% are contained within + or – 1.96 standard deviations and rounding that to 2, most values of standard normal distribution would be between + and -2. (2,-2) Any case outside this is considered unusual or an outlier.

5% of 166 cases = 8.3 = 8 expected cases

Text

Description automatically generated

There was 7 cases where an outlier was found. These cases are shown in the screenshot above.

1. A case is said to be of high leverage in multiple regression if the corresponding x vector is far from the mean vector of the x values. This makes the predictor value unusually large or unusually small.

Chart, histogram

Description automatically generated

The graph above shows that case 43 has an unusually large predictor value. This means that this case’s value is far from the mean of the rest of the cases’ values. This case is extreme because it unusually large compared to the other cases.

1. The case has high influence in determining the values of the intercept and slope in a simple regression model. So if we omit the case it would result in large changes in the values of the least squares estimate of Beta0 and Beta1.

Chart, histogram

Description automatically generated

As you can see from the graph above the case with the highest influence is case 43. The value of Cook’s distance for case 43 is 0.0898766715. Case 43 has high influence because it also has high leverage.

1. I would definitely recommend investigating case 43 and try to identify why the case has a high leverage and high influence. Case 43 seems to be an extreme case and needs more investigation. It could be an error in data collection or maybe another issue.