**Statistical Report Writing Sample No.2.**

**Introduction.** The following data were published in "Repeatability of the petrifilm HEC test and agreement with a hydrophobic grid membrane filtration (HGMF) method for the enumeration of Escherichia coli on beef carcasses," Journal of Food Protection 61: 402-408. Eighteen portions of beef trim were obtained from three Holstein cows, and then contaminated with E. coli. The HEC and HGMF procedures were applied to a portion of each of the 18 samples, and yielded E. coli concentrations in transformed metric (log CFU/ml).

|  |  |  |
| --- | --- | --- |
| RUN | HEC | HGMF |
| 1 | 0.5 | 0.42 |
| 2 | 0.06 | 0.2 |
| 3 | 0.2 | 0.42 |
| 4 | 0.61 | 0.33 |
| 5 | 0.2 | 0.42 |
| 6 | 0.56 | 0.64 |
| 7 | -0.82 | -0.82 |
| 8 | 0.67 | 1.06 |
| 9 | 1.02 | 1.21 |
| 10 | 1.2 | 1.25 |
| 11 | 0.93 | 0.83 |
| 12 | 2.27 | 2.37 |
| 13 | 2.02 | 2.21 |
| 14 | 2.32 | 2.44 |
| 15 | 2.14 | 2.28 |
| 16 | 2.09 | 2.69 |
| 17 | 2.3 | 2.43 |
| 18 | -0.1 | 1.07 |

The columns HEC and HGMF show the transformed metric (log CFU/ml) from the respective methods, which is summarized in the following table. In this study we are interested in determining whether the HEC and HGMF procedures agreed in the level of E. coli concentrations in meat samples.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Mean | S.D | L.Quartile | Median | U.Quartile |
| HEC | 1.009444 | 0.9703757 | 0.275 | 0.8 | 2.0725 |
| HGMF | 1.191667 | 0.9993189 | 0.42 | 1.065 | 2.2625 |

**Data analysis.** We use a linear model with the HEC readings as the dependent variable and HGMF as the explanatory variable because the HGMF procedure is known to be more reliable in determining E. coli concentrations. The result of linear model is summarized in the following table. It shows that the intercept value is not significant (p-value of 0.444), and that the slope value is highly significant (p-value less than 0.001).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter | Estimate | SE | t-statistic | p-value |
| Intercept | -0.09077 | 0.11562 | -0.785 | 0.444 |
| HGMF | 0.92326 | 0.07521 | 12.276 | 1.48E-09 |

In the following figure a fitted regression line is placed in the scatter plot. The first 17 points fell close to the regression, but the 18th reading has a very low value in the HEC reading.

 

An examination of standardized residual plot (see below) indicates that the 18th point falls outside the pattern of the remaining points, and therefore, that it is a potential outlier. Otherwise, the residual plot does not indicate heterogeneity in the variances.

 

**Conclusion.**  The 95% confidence interval for the slope is (0.764, 1.083) indicates that the slope is close to one, and the 95% confidence interval for the intercept is (−0.336, 0.154), which is close to zero. Therefore, data do not support that HEC and HGMF are yielding significantly different. But the estimated slope of 0.923 suggests the possibility of underestimated measurements by HEC compared with HGMF. The proposed linear model would be used to predict HEC readings from HGMF readings. The 95% prediction interval for HEC reading is (0.34, 1.69) when the HGMF reading is 1.2; thus, an HEC reading of 0.6 is not unusual. On the other hand, the 95% prediction interval for HEC reading is (2.46, 4.01) when the HGMF reading is 3.6; thus, an HEC reading of 1.8 is unusual.