Consider two variables: \( x \), the temperature 10's of degrees Fahrenheit for a manufacturing process, and \( y \), a measure of the yield of the process.

Suppose that the following sample of data for \( x \) and \( y \) is provided:

<table>
<thead>
<tr>
<th>( x )</th>
<th>( y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5</td>
<td>1</td>
</tr>
<tr>
<td>-4</td>
<td>5</td>
</tr>
<tr>
<td>-3</td>
<td>4</td>
</tr>
<tr>
<td>-2</td>
<td>7</td>
</tr>
<tr>
<td>-1</td>
<td>10</td>
</tr>
<tr>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
</tr>
</tbody>
</table>

A scatter plot of these data with the SLR regression line looks like the following:

Confidence Interval for \( \beta_1 \)

A 100(1 - \( \alpha \))% confidence interval for \( \beta_1 \) is of the form

\[
b_1 \pm t_{\alpha/2, \nu} \frac{s}{\sqrt{SS_{xx}}} \]

where \( \nu = n - 2 \), \( s^2 \) is the residual mean square error from the ANOVA table, and \( SS_{xx} \) is what has already been computed in finding \( b_1 \).

1. Find a 95 percent confidence interval for the \( \beta_1 \) of the data with which we have been working.
The F-distribution Hypothesis Test (F-Test) for the Significance of an SLR Model

The following is a test at the $100(1 - \alpha)\%$ significance level for whether or not an SLR model is statistically significant, that is, that does or does not happen by chance because of the particular sample that is chosen:

\[ H_0 : \text{The model is not significant.} \]
\[ H_1 : \text{The model is significant.} \]

Test statistic: \[ F = \frac{MS_{reg}}{s^2} \]

Rejection region: \[ F > F_{\nu_n, \nu_d, \alpha} \]

where $\nu_n = df_{reg} =$ number of predictor variables = 1 and $\nu_d = df_{residual} = n - 2$.

2. Perform the F-test at the 95% significance level to see if the SLR model for the data that we have been using is statistically significant or not.

3. What does this hypothesis test tell us?

Two-sided Hypothesis Test for $\beta_1$

A $100(1 - \alpha)\%$ hypothesis test of whether a parameter $\beta_1$ equals a certain numerical value, call it $\beta_{10}$, has the form:

\[ H_0 : \beta_1 = \beta_{10} \]
\[ H_1 : \beta_1 \neq \beta_{10} \]

Test statistic: \[ t = \frac{b_1 - \beta_{10}}{s} \frac{1}{\sqrt{SS_{xx}}} \]

Rejection Region: \[ |t| > t_{\nu, \alpha/2} \] (where $\nu = n - 2$)

Conclusion: Not reject $H_0$ or reject $H_0$ (accept $H_1$ as new working hyp.)

Probability that the conclusion is correct: $1 - \alpha$

4. Perform a hypothesis test at the 95% confidence level to see whether or not $\beta_1$ equals 0.
5. What would it mean in terms of the SLR model if $\beta_1$ equals 0? What would this tell you about whether or not $x$ helps to reduce the uncertainty in predicting $y$?

6. How can this be used as a test of whether the SLR model is statistically significant or not?

<table>
<thead>
<tr>
<th>Relation of F-test for the SLR Model and the t-test for $\beta_1 = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>The hypotheses for the two tests for the SLR model are equivalent. Further $F = t^2$ and the $F$ or $t$ values for the rejection regions are similarly related.</td>
</tr>
</tbody>
</table>

7. Verify that the relation between $F$ and $t$ given in the box above is true using the $F$ from 7 and the $t$ from 9.
Correlation of Two Variables

Recall that the definition of correlation for a population is:

\[ \rho_{XY} = \frac{\text{Cov}(x, y)}{\sqrt{\sigma_x^2 \sigma_y^2}} \]

For a sample the two variables would have a correlation defined as:

\[ r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \]

\[ = \frac{n \sum_{i=1}^{n} x_i y_i - \left( \sum_{i=1}^{n} x_i \right) \left( \sum_{i=1}^{n} y_i \right)}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \]

\[ = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} \]

8. Calculate the sample correlation, \( r_{xy} \), for the data we have been using.

9. What does this correlation coefficient, \( r_{XY} \) signify?
Relation of the Correlation Coefficient and the Coefficient of Determination for SLR Models

For SLR models in which \( r_{xy} \) is the sample correlation coefficient of the two variables involved and \( R^2 \) is the coefficient of determination for the SLR model then we have

\[
\left( r_{xy} \right)^2 = R^2
\]

Note: This is true only for SLR models. With more than one predictor variable (more than one X) as in multiple regression, the coefficient of determination, \( R^2 \), is calculated as before but there is no single \( r_{XY} \) to be related to it.

10. You have calculated \( r_{xy} \) in #9 above and you have calculated \( R^2 \) in a previous worksheet for the data that we have been using. Verify that \( \left( r_{xy} \right)^2 = R^2 \) within round-off error for these data.

Relation of Slope Coefficient \( b_1 \) and Correlation Coefficient \( r_{XY} \)

There is a systematic relationship between the slope coefficient \( b_1 \) for an SLR model and for the sample correlation coefficient \( r_{xy} \) for the two variables involved. It can be shown that:

\[
b_1 = \frac{s_y}{s_x} r_{xy}
\]

where \( s_y \) is the sample standard deviation of \( y \) and \( s_x \) is the sample standard deviation of \( x \).

11. You have found \( b_1 \) for the data with which we have been working and you found \( r_{xy} \). Verify that the relation in the box above holds for these data.

Suggested Homework: 11.18abcd, 11.19, 11.25, 11.26, 11.28, 11.35abcd, 11.37, 11.38

Solutions to be Posted: 11.18ac, 11.25, 11.28, 11.35ac, 11.37