Test Your Understanding 6

Summarize the pattern of variation for each of the six scatterplots shown.

Solution:

*In clockwise order from top left:*

1. Randomly scattered.
2. Positive linear reasonably strong association.
3. Negative linear reasonably strong association.
4. Negative nonlinear strong association Increasing variation.
5. Nonlinear reasonably strong association. Negative for $x < 0$, positive for $x > 0$.
6. Positive linear strong association.
Test Your Understanding 7

The Pearson correlation between $X$ and $Y$ is 0.51. An understanding of Pearson correlation, but no calculation, is needed to answer the following questions.

a. If $W = 6 \times X$, what is the Pearson correlation between $W$ and $Y$?
   \textbf{Solution:} 0.51

b. If $U = X - 8$, what is the Pearson correlation between $U$ and $Y$?
   \textbf{Solution:} 0.51

c. If $V = -6 \times Y$, what is the Pearson correlation between $V$ and $Y$?
   \textbf{Solution:} -0.51
Test Your Understanding 8

Obtain the least squares estimates of the slope and intercept for the data

<table>
<thead>
<tr>
<th>X</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Solution: Since \( \overline{y} = \overline{x} = 0 \), the LSE of slope is

\[
\beta_1 = \frac{\sum_{i=1}^{3} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{3} (x_i - \overline{x})^2} = \frac{\sum_{i=1}^{3} x_i y_i}{\sum_{i=1}^{3} x_i^2} = \frac{(-1)(-2) + (0)(1) + (1)(1)}{(-1)^2 + 0^2 + 1^2} = \frac{3}{2},
\]

and the LSE of intercept is

\[
\beta_0 = \overline{y} - \beta_1 \overline{x} = 0.
\]
Test Your Understanding 9

SAS regression output for the regression of highway miles per gallon on engine displacement for a sample of new cars, is shown in the figure on the reverse.

a. Evaluate the fit of the model in terms of the residuals.
   Solution: The model is a linear function and does not seem adequate to model the nonlinear trend.

b. Evaluate the fit of the model in terms of $r^2$. Interpret $r^2$.
   Solution: $r^2 = 0.3929$, meaning that the model (i.e., the fitted linear function of DISPLACE) accounts for 39.29% of the variation in HIGHMPG. The model does not seem very successful in accounting for the variation in HIGHMPG.

c. What is the Pearson correlation between the response and the predictor?
   Solution: $r = \sqrt{0.3929} = 0.6268$. 
**Model Equation**

\[ \text{HIGHMP} = 37.4802 - 3.2215 \times \text{DISPLACE} \]

**Parametric Regression Fit**

<table>
<thead>
<tr>
<th>Model</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curve</td>
<td>Degree(Polynomial)</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Summary of Fit**

- Mean of Response: 29.0886
- R-Square: 0.3929
- Root MSE: 4.1772
- Adj R-Sq: 0.3862

**Parameter Estimates**

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Estimate</th>
<th>Std Error</th>
<th>T Stat</th>
<th>Prob &gt;</th>
<th>Tolerance</th>
<th>Var Inflation</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>1</td>
<td>37.4802</td>
<td>1.2008</td>
<td>31.177</td>
<td>0.0001</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>DISPLACE</td>
<td>1</td>
<td>-3.2215</td>
<td>0.4198</td>
<td>-7.673</td>
<td>0.0001</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

**DISPLACE**

\[ 3.2215 - 37.4802 = \text{HIGHMP} \]

**Model Equation**

\[ \text{HIGHMP} = 37.4802 - 3.2215 \times \text{DISPLACE} \]
Test Your Understanding 10

Refer again to the regression output from TYU 9 (shown on the reverse).

a. What is the fitted model?
   \[ \text{Solution: } \text{HIGHMPG} = 37.6802 - 3.2215 \text{ DISPLACE} \]

b. Does the fitted intercept have a meaning of its own? Why or why not?
   \[ \text{Solution: No, since } \text{DISPLACE} = 0 \text{ is outside the range of the data, and indeed, makes no sense.} \]

c. Interpret the slope of this line in terms that have meaning for car shoppers.
   \[ \text{Solution: The model estimates that average highway MPG is lower for cars with higher engine displacements, and that the rate of decrease is 3.2215 per 1 cubic inch increase in displacement.} \]

d. Estimate the variance of the random errors.
   \[ \text{Solution: MSE=17.4487} \]
Model Equation
HIGHMP = 37.6802 - 3.2115 DISPLACE

<table>
<thead>
<tr>
<th>Model</th>
<th>Error</th>
<th>Curve</th>
<th>Degree</th>
<th>Polynomial</th>
<th>DF</th>
<th>Mean Square</th>
<th>DF</th>
<th>Mean Square</th>
<th>R-Square</th>
<th>F Stat</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1027.8817</td>
<td>1</td>
<td>17.4687</td>
<td>0.3929</td>
<td>58.8859</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Summary of Fit
Mean of Response 29.0468  R-Square 0.3929
Root MSE 4.1772  Adj R-Sq 0.3862

Parameter Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Estimate</th>
<th>Std Error</th>
<th>T Stat</th>
<th>Prob &gt;</th>
<th>Tolerance</th>
<th>Var Inflation</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>1</td>
<td>37.6802</td>
<td>1.2006</td>
<td>31.3172</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>DISPLACE</td>
<td>1</td>
<td>-3.2115</td>
<td>0.4198</td>
<td>-7.6737</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Test Your Understanding 11

Using the regression output from TYU 9 (shown on the reverse), obtain 95% confidence intervals for $\beta_0$ and $\beta_1$. Interpret these intervals in terms of the model.

**Solution:** Since $t_{0.025} = 1.9867$ is not tabulated, we will use $t_{0.0.975} = 1.9867$ as an approximation (Note that we could also have used $z_{0.975} = 1.96$). Then the confidence intervals are:

$$
\beta_0 : \hat{\beta}_0 \pm \hat{\sigma}(\hat{\beta}_0)t_{0.975} \approx 37.6802 \pm (1.2008)(1.9867) = (35.2946, 40.0658)
$$

$$
\beta_1 : \hat{\beta}_1 \pm \hat{\sigma}(\hat{\beta}_1)t_{0.975} \approx -3.2215 \pm (0.4198)(1.9867) = (-4.0555, -2.3875)
$$

With 95% confidence, we estimate that the true intercept lies between 35.2946 and 40.0658. With 95% confidence, we estimate that the true slope lies between $-4.0555$ and $-2.3875$. Of particular note, is that we are 95% confidence there is a negative relationship between displacement and mean highway mpg, and we estimate the decline in mean highway mpg to be between 2.3875 and 4.0555 per cubic inch increase in displacement.
**Model Equation**

\[ \text{HIGHMP} = 37.6401 - 3.2113 \text{ DISPLACE} \]

**Parameter Estimates**

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Estimate</th>
<th>Std Error</th>
<th>T Stat</th>
<th>Prob &gt;</th>
<th>Tolerance</th>
<th>Var Inflation</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>1</td>
<td>37.6401</td>
<td>1.208</td>
<td>31.317</td>
<td>0.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>DISPLACE</td>
<td>1</td>
<td>-3.2113</td>
<td>0.1198</td>
<td>-27.376</td>
<td>0.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

**Summary of Fit**

- **Mean of Response**: 23.0468
- **R-Square**: 0.3929
- **Root MSE**: 4.1772
- **Adjusted R-Sq**: 0.3862

**Model Regression Fit**

<table>
<thead>
<tr>
<th>Model</th>
<th>Error</th>
<th>R-Square</th>
<th>F Stat</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>17.4487</td>
<td>4.3963</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

**Table: Model Fit**

<table>
<thead>
<tr>
<th>Curve</th>
<th>Degree of Polynomial</th>
<th>DF</th>
<th>Mean Square</th>
<th>DF</th>
<th>Mean Square</th>
<th>R-Square</th>
<th>F Stat</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>1227.2817</td>
<td>81</td>
<td>17.4487</td>
<td>0.3963</td>
<td>0.0001</td>
<td></td>
</tr>
</tbody>
</table>