WHAT YOU WILL LEARN

• Correlation and regression
Section 8

Linear Correlation and Regression
Correlation

(a) \( r = 1 \); strongest positive correlation

(b) \( r > 0 \); strong positive correlation

(c) \( r > 0 \); moderate to weak positive correlation

(d) No correlation
Correlation

(e) $r = -1$; strongest negative correlation
(f) $r < 0$; strong negative correlation
(g) $r < 0$; moderate to weak negative correlation
(h) No correlation
Linear Regression

- *Linear regression* is the process of determining the linear relationship between two variables.
- The *line of best fit* (*regression line* or the *least squares line*) is the line such that the sum of the squares of the vertical distances from the line to the data points (on a scatter diagram) is a minimum.
The Line of Best Fit

- **Equation:**

\[ y = mx + b, \quad \text{where} \]

\[
m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}, \quad \text{and} \quad b = \frac{\sum y - m(\sum x)}{n}
\]
Example

- Use the data in the previous example to find the equation of the line that relates the number of words per minute and the number of mistakes made while typing a chart.

- Graph the equation of the line of best fit on a scatter diagram that illustrates the set of bivariate points.
Solution

From the previous results, we know that

\[ m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \]

\[ m = \frac{5(2,281) - (219)(50)}{5(10,711) - 219^2} \]

\[ m = \frac{455}{5594} \]

\[ m \approx 0.081 \]
Solution

- Now we find the $y$-intercept, $b$.

\[
b = \frac{\sum y - m(\sum x)}{n}
\]

\[
b = \frac{50 - 0.081(219)}{5}
\]

\[
b = \frac{32.261}{5} \approx 6.452
\]

Therefore the line of best fit is $y = 0.081x + 6.452$
Solution (continued)

- To graph \( y = 0.081x + 6.452 \), plot at least two points and draw the graph.

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>7.262</td>
</tr>
<tr>
<td>20</td>
<td>8.072</td>
</tr>
<tr>
<td>30</td>
<td>8.882</td>
</tr>
</tbody>
</table>
Solution (continued)

![Graph: Words Per Minute Versus Mistakes](image)

- The graph shows the relationship between words per minute and mistakes.
- As words per minute increase, the number of mistakes also increases.
- The trend line indicates a positive correlation between the two variables.