Algebra Review

• The equation of a straight line
  • \( y = mx + b \)
    – \( m \) is the slope – the change in \( y \) over the change in \( x \) – or rise over run.
    – \( b \) is the \( y \)-intercept – the value where the line cuts the \( y \) axis.

\[
\begin{align*}
y = 3x + 2 \\
x & = 0 \quad \Rightarrow \quad y = 2 \quad \text{(y-intercept)} \\
x & = 3 \quad \Rightarrow \quad y = 11 \\
\text{Change in } y (+9) \text{ divided by the change in } x (+3) \text{ gives the slope, 3.}
\end{align*}
\]
Linear Regression

- Example: Tar (mg) and nicotine (mg) in cigarettes.
  - \( y \), Response: Nicotine (mg).
  - \( x \), Explanatory: Tar (mg).
  - Cases: 25 brands of cigarettes.

Correlation Coefficient

- Tar and nicotine
  \[
  r = \frac{\sum z_x z_y}{n - 1} = \frac{+22.9437}{24}
  \]
  - \( r = +0.956 \)

Linear Regression

- There is a very strong positive linear association between tar and nicotine.
- What is the equation of the line that models the relationship between tar and nicotine?
Linear Model

- The linear model is the equation of a straight line through the data.
- A point on the straight line through the data gives a predicted value of $y$, denoted $\hat{y}$.

Residual

- The difference between the observed value of $y$ and the predicted value of $y$, $\hat{y}$, is called the residual.
- Residual = $y - \hat{y}$
Line of “Best Fit”

- There are lots of straight lines that go through the data.
- The line of “best fit” is the line for which the sum of squared residuals is the smallest – the least squares line.

\[ \hat{y} = b_0 + b_1x \]

Least squares slope: \[ b_1 = r \frac{s_y}{s_x} \]

intercept: \[ b_0 = \bar{y} - b_1\bar{x} \]

Least Squares Estimates

<table>
<thead>
<tr>
<th>Tar, x</th>
<th>Nicotine, y</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{x} = 11.92 ) mg</td>
<td>( \bar{y} = 0.908 ) mg</td>
</tr>
<tr>
<td>( s_x = 4.636 ) mg</td>
<td>( s_y = 0.2812 ) mg</td>
</tr>
<tr>
<td>( r = 0.956 )</td>
<td></td>
</tr>
</tbody>
</table>