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.

\[ \hat{y} = b_0 + b_1 x \]

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

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

Least Squares Estimates

\[
\begin{array}{cc}
\text{Tar, } x & \text{Nicotine, } y \\
\bar{x} = 11.92 \text{ mg} & \bar{y} = 0.908 \text{ mg} \\
s_x = 4.636 \text{ mg} & s_y = 0.2812 \text{ mg} \\
r = 0.956 & \\
\end{array}
\]
Least Squares Estimates

\[ b_1 = 0.956 \frac{0.2812}{4.636} = 0.058 \]
\[ b_0 = 0.908 - 0.058(11.92) = 0.217 \]
\[ \hat{y} = 0.217 + 0.058x \]

Interpretation

• Estimated slope – for every 1 mg increase in tar, the nicotine content increases, on average, 0.058 mg.
• The average change in nicotine for a 1 mg change in tar.

Interpretation

• Estimated intercept – there is not a reasonable interpretation of the intercept in this context because one wouldn’t see a cigarette with 0 mg of tar.
### Prediction

- Least squares line
  
  \[
  \hat{y} = 0.217 + 0.058x
  \]
  
  \(x = 13\)
  
  \[
  \hat{y} = 0.217 + 0.058(13) = 0.97
  \]

### Residual

- Brand: Multi-Filter
- Tar, \(x = 13\) mg
- Nicotine, \(y = 0.8\) mg
- Predicted, \(\hat{y} = 0.97\) mg
- Residual, \(y - \hat{y} = 0.8 - 0.97 = -0.17\) mg
Residuals

- Residuals help us see if the linear model makes sense.
- Plot residuals versus the explanatory variable.
  - If the plot is a random scatter of points, then the linear model is the best we can do.

![Plot of Residuals vs. Tar Content](image)

Interpretation of the Plot

- The residuals are scattered randomly. This indicates that the linear model is an appropriate model for the relationship between tar and nicotine content of cigarettes.