Simple Linear Regression

\[ \text{cor}(x, y) = 1 \iff y = \beta_0 + \beta_1 x \text{ for some } \beta_0, \beta_1 \]

\[ \Delta y = (\beta_0 + \beta_1 (x+\Delta x)) - (\beta_0 + \beta_1 x) \]

\[ \frac{\Delta y}{\Delta x} = \beta_1 = \text{ "slope"} \]

More typically, \[ |\text{cor}(x, y)| < 1 \]

\[ y \]

"perfect" linear relation

\[ x \]

"noisy" linear relation

\[ y = \beta_0 + \beta_1 x \]

\[ y_i = \beta_0 + \beta_1 x_i \]

\[ y = \beta_0 + \beta_1 x + \epsilon \]

\[ y_i = \beta_0 + \beta_1 x_i + \epsilon_i \]

What values of \( \beta_0, \beta_1 \) provide the "best fit"?

ODES Regression Line

\[ \text{RSS}(\beta_0, \beta_1) = \sum (y_i - (\beta_0 + \beta_1 x_i))^2 = \| y - (\hat\beta_0 + \hat\beta_1 x) \|^2 \]

\[ = \sum \epsilon_i^2 \]

\[ = \| \epsilon \|^2 \]

\( \epsilon_i = y_i - (\beta_0 + \beta_1 x_i) \) is residual for the particular values of \( \beta_0, \beta_1 \)

Defn: The OLS values of \( (\beta_0, \beta_1) \) are the values \( (\hat\beta_0, \hat\beta_1) \) that minimize \( \text{RSS}(\beta_0, \beta_1) \)

\[ (\hat\beta_0, \hat\beta_1) = \arg \min_{(\beta_0, \beta_1)} \text{RSS}(\beta_0, \beta_1) \]
Optimization:

Method 1: geometry (will do later)
Method 2: calculus

\[
\text{RSS}(\beta_0, \beta_1) \text{ is a quadratic (convex) function of } (\beta_0, \beta_1)
\]

\[
\text{global minimizer occurs where derivative (gradient) is zero.}
\]

\[
\frac{d\text{RSS}}{d\beta_0} = -2 \sum (y_i - (\beta_0 + \beta_1 x_i)) = 0 \implies \sum (y_i - (\beta_0 + \beta_1 x_i)) = 0
\]

\[
\frac{d\text{RSS}}{d\beta_1} = -2 \sum x_i (y_i - (\beta_0 + \beta_1 x_i)) = 0 \implies \sum x_i (y_i - (\beta_0 + \beta_1 x_i)) = 0
\]

The OLS values \((\hat{\beta}_0, \hat{\beta}_1)\) therefore satisfy:

\(\sum (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = 0\)

\(\sum x_i (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = 0\)

These are called the Normal equations for simple linear regression.

Why "normal equations"? at \(\hat{e}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = "\text{resid at OLS val."} \text{ or just "residual"}

then
\(\sum \hat{e}_i = 0 \quad \hat{e}_i \cdot 1 = 0\)
\(\sum x_i \hat{e}_i = 0 \quad \hat{e}_i \cdot x = 0\)

\(\hat{e}_i \) is normal (orthogonal) to the vectors \(1, x\).

Also, we have from C that \(\sum y_i = \sum (\hat{\beta}_0 + \hat{\beta}_1 x_i)\)

\(\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}\)

Regression line goes through the point \((\bar{x}, \bar{y})\).
Solving the normal equations

1. \[ \Sigma(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = 0 \]
   \[ \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \]
   \[ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \]

Plug this into (2):

\[ \Sigma x_i (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = 0 \]
\[ \Sigma x_i (y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i) = 0 \]
\[ \Sigma x_i (y_i - \bar{y}) = \hat{\beta}_1, \Sigma x_i (x_i - \bar{x}) = \hat{\beta}_0 \]

Now note the following trick:

\[ \Sigma (x_i - \bar{x})(y_i - \bar{y}) = \Sigma x_i (y_i - \bar{y}) + \Sigma \bar{x}(y_i - \bar{y}) \]

Similarly:

\[ \Sigma x_i (x_i - \bar{x}) = \Sigma x_i (x_i - \bar{x}) \]

\[ \Sigma (x_i - \bar{x})^2 \]

Let

\[ S_{XX} = \Sigma (x_i - \bar{x})^2 \]
\[ S_{XY} = \Sigma (x_i - \bar{x})(y_i - \bar{y}) \]

Then, \[ \hat{\beta}_1 \] says:

\[ S_{XY} = \hat{\beta}_1 S_{XX} \]
\[ \hat{\beta}_0 = \frac{S_{XY}}{S_{XX}} \]

Finally, the OLS values are:

\[ \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 &= \frac{S_{XY}}{S_{XX}} \end{bmatrix} \]

Relation to Correlation

\[ \hat{\beta}_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} \]

- looks like covariance or correlation

\[ \hat{\beta}_1 = \frac{S_{XY}}{S_{XX}^{1/2}} \times \frac{S_{XY}}{S_{XX}^{1/2}} \]

- looks like variance of \( x \)

\[ \hat{\beta}_1 = \left( \frac{S_{XY}}{S_{XX}} \right) \times \cos(\theta) \]

\[ \theta = \frac{1}{2} \log \left( \frac{S_{XY}}{S_{XX}^{1/2}} \right) \]

\[ = \left( \frac{S_Y}{S_X} \right) \times \cos(\theta) \]
What you need to find \( \hat{b}_o, \hat{b}_1 \) is:

\[
\begin{bmatrix}
\bar{X}, \bar{Y} \\
S_X, S_Y \\
\text{Cor}(X,Y)
\end{bmatrix}
\]

**Regression Line**

\[
Y_i = \hat{b}_0 + \hat{b}_1 X_i + \hat{e}_i
\]

\[
\hat{Y} = \hat{b}_0 + \hat{b}_1 X = \text{"predicted" or "fitted" value, at } X_i,
\]

\[
\hat{Y} = \bar{Y} - \hat{b}_1 (X - \bar{X})
\]

\[
\hat{Y} = \bar{Y} + \frac{S_Y}{S_X} \times \text{Cor}(X,Y) \times (X - \bar{X})
\]

**Questions**

1) How does the line change if we change the **location** of the \( x \)'s?

2) How does the line change if we change the **scale** of the \( x \)'s?

3) What aspects of the regression line are **invariant** to:
   a) **location** change?
   b) **scale** changes?
   c) **loc. + scale** changes?