Quiz 1  Solutions

1. Let \((x_1, y_1), \ldots, (x_n, y_n)\) be n points. Please derive the least squares regression line for the usual simple linear regression: \(y = \beta_0 + \beta_1 x + \epsilon\)

To minimize the sum (over all n points) of squared vertical distances from each point to the line:

\[ SS = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2: \]

\[
\begin{align*}
\frac{\partial SS}{\partial \beta_0} &= 0 \\
\frac{\partial SS}{\partial \beta_1} &= 0 \\
\Rightarrow \quad \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\
\Rightarrow \quad \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\
\Rightarrow \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\
\Rightarrow \quad \sum_{i=1}^{n} x_i (y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i) &= 0 \\
\Rightarrow \hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sum_{i=1}^{n} x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}
\]

\(\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x\) is the least squares regression line.

2. Derive the least squares regression line for simple linear regression through the origin:

\(y = \beta_1 x + \epsilon\)

To minimize

\[ SS = \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2: \]

\[
\frac{\partial SS}{\partial \beta_1} = 0 \Rightarrow \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_1 x_i) = 0 \Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}
\]

\(\hat{y} = \hat{\beta}_1 x\) is the least squares regression line through the origin.