ESTIMATING CONDITIONAL MEANS

Model Assumptions: Linear mean, constant variance, independence, and normality.

Sampling Distribution of Estimate of Conditional Mean:

- \( \hat{E}(Y|x) = \hat{\eta}_0 + \hat{\eta}_1 x \) is our estimate of \( E(Y|x) \). Note that this is a random variable (varying according to our choice of \( y_i \)'s), so has a sampling distribution.

- Since \( \hat{\eta}_0 \) and \( \hat{\eta}_1 \) are linear combinations of the \( y_i \)'s, so is \( \hat{E}(Y|x) \). Hence \( \hat{E}(Y|x) \) has a normal distribution. (Why doesn't this follow just from normality of \( \hat{\eta}_0 \) and \( \hat{\eta}_1 \)?)

- \( E(\hat{E}(Y|x)| x_1, \ldots, x_n) = E(\hat{\eta}_0 + \hat{\eta}_1 x| x_1, \ldots, x_n) \)
  \[ = E(\hat{\eta}_0| x_1, \ldots, x_n) + E(\hat{\eta}_1| x_1, \ldots, x_n)x \]
  \[ = \eta_0 + \eta_1 x = E(Y|x) \]
  So \( \hat{E}(Y|x) \) is an unbiased estimator of \( E(Y|x) \).

- Calculations (left to the interested reader; you need to consider covariances) will show that
  \[ \text{Var}(\hat{E}(Y|x)| x_1, \ldots, x_n) = \sigma^2 \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{SXX} \right) \]

Comments:

1. What does this say when \( x = 0 \)?
2. The further \( x \) is from \( \bar{x} \), the ___________ the variance of the conditional mean estimate.
3. How does \( \text{Var}(\hat{E}(Y|x)) \) depend on \( n \) and the spread of the \( x_i \)'s?

Define the standard error of \( \hat{E}(Y|x) \):

\[ \text{s.e.}(\hat{E}(Y|x)) = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{SXX}} \]

As with \( \hat{\eta}_0 \) and \( \hat{\eta}_1 \), one can show that (under our model assumptions)

\[ \frac{\hat{E}(Y|x) - E(Y|x)}{\text{s.e.}(\hat{E}(Y|x))} \sim t(n-2), \]

so we can use this as a test statistic to do inference on \( E(Y|x) \).