Summary of Simple Linear Regression

Model

$$Y_i = b_0 + b_1 x_i + e_i, \quad e_i \sim \text{iid } N(0, \sigma^2), \quad i = 1, \ldots, n.$$

Model Assumptions

(1) Correct model (2) Independence (3) Homogeneous variance (4) Normal distribution

Statistical Inference

- ANOVA approach: F-test of $H_0 : b_1 = 0$ using the partition of sums of squares $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$
(SSTotal $(n-1)$ = SSRegression (1) + SSError $(n-2)$).

- Standardization approach: $\frac{\odot - \mu_\odot}{\text{s.e.}(\odot)}$ where $\odot$ is a statistic such as $\hat{b}_0, \hat{b}_1, \hat{y}_{\text{est}}, \hat{y}_{\text{pred}}$.

T-test/CI of	Using
b_1	$\hat{b}_1$ and s.e.$(\hat{b}_1)$
b_0	$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}$ and s.e.$(\hat{b}_0)$
$b_0 + b_1 x^*$	$\hat{y}_{\text{est}} = \hat{b}_0 + \hat{b}_1 x^*$ given x^* and s.e.$(\hat{y}_{\text{est}})$
$b_0 + b_1 x^* + e$	$\hat{y}_{\text{pred}} = \hat{b}_0 + \hat{b}_1 x^*$ given x^* and s.e.$(\hat{y}_{\text{pred}})$ [**No test!**]

- Test and CI of σ^2: $s^2_{Y \cdot x}$ = MSError is used to estimate σ^2 and $s_{Y \cdot x} = \sqrt{\text{MSError}}$ is used to estimate σ. e.g., s.e.$(\hat{b}_1) = \frac{s_{Y \cdot x}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$.

Model Fitting

Coefficient of determination $R^2 = \frac{\text{SSRegression}}{\text{SSTotal}}$ where SSTotal is mean-corrected.

Model Diagnostics

How to check model assumptions? Look at residuals $r_i = y_i - \hat{y}_i$ (i.e., obs y – fitted y).

- r_i is a raw residual.
- $\sum_{i=1}^n r_i = 0$.
- $\sum_{i=1}^n r_i^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = \text{SSError}$.
- Plots of r_i versus $\hat{y}_i$ can be used to check model assumptions.
- Normal score plots of r_i can be used to assess the normality assumption.
- If n is small, it might be hard to interpret the residual plots.