Multiple Linear Regression

- Multiple Linear Regression (MLR) is a linear regression model with two or more predictors
- The regression line is represented by the equation

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \cdots + b_kX_k$$

- where
  - a is the intercept
  - $b_1, b_2, \ldots, b_k$ are the slopes
  - k is the number of predictors

The Standardized MLR Model

- The equation for the multiple linear regression line with standardized variables is

$$\hat{z}_Y = \beta_1 z_{X_1} + \beta_2 z_{X_2} + \cdots + \beta_k z_{X_k}$$

- where $Y, X_1, X_2, \ldots, X_k$ are standardized using z-scores

Effect Size in MLR

- The coefficient of determination in MLR is called the coefficient of multiple determination
  - It is the proportion of variance in Y that is accounted for by $X_1, X_2, \ldots, X_k$
  - It is still denoted as R^2
    - To make it clear what predictors are in the model, we will sometimes denote it by $R^2_{Y.X_1 X_2 \ldots X_k}$
    - E.g., for a MLR with three independent variables (X_1, X_2, X_3)
      - $R^2_{Y.X_1 X_2 X_3}$

Squared Semipartial Correlation

- $sr_{X_1}^2$ is the variance accounted for by X_1 beyond that accounted for by X_2 and X_3
- $sr_{X_2}^2$ is the variance accounted for by X_2 beyond that accounted for by X_1 and X_3
- $sr_{X_3}^2$ is the variance accounted for by X_3 beyond that accounted for by X_1 and X_2

Squared Partial Correlation

- $pr_{X_1}^2$: Of the variance in Y that is not explained by X_2 and X_3, what proportion is explained by X_1
- $pr_{X_2}^2$: Of the variance in Y that is not explained by X_1 and X_3, what proportion is explained by X_2
- $pr_{X_3}^2$: Of the variance in Y that is not explained by X_1 and X_2, what proportion is explained by X_3

Hypothesis Testing

- As with SLR, sample data are used to determine the parameter estimates for the OLS regression line

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \cdots + b_kX_k$$

- But we are truly interested in the population model

$$\hat{Y} = \alpha^* + \beta_1^*X_1 + \beta_2^*X_2 + \cdots + \beta_k^*X_k$$

Hypothesis Test for ρ^2

- Step 3: Calculate the Test Statistic
  - This is the F-statistic

Source	SS	df	MS	F
Regression	$SS_{Regression}$	k	MSR	MSR/MSE
Error	SS_{Error}	$n-k-1$	MSE	
Total	SS_{Total}	$n-1$		

Hypothesis Test for $\alpha^*, \beta_1^*, \beta_2^*, \ldots, \beta_k^*$

- Step 3: Calculate the Test Statistic
  - This is the t-statistic

$$t = \frac{a}{s_a} \qquad t = \frac{b_1}{s_{b_1}} \qquad t = \frac{b_2}{s_{b_2}} \quad \ldots \quad t = \frac{b_k}{s_{b_k}}$$

  - where
    - a is the parameter estimate for the intercept and s_a is the estimated standard error for the intercept
    - $b_1, b_2, \ldots, b_k$ are the parameter estimates for the slopes and $s_{b_1}, s_{b_2}, \ldots, s_{b_k}$ are the estimated standard errors for the slopes