DATA-231: APPLIED STATISTICAL METHODS
COMPARISON OF SIMPLE AND MULTIPLE LINEAR REGRESSION

	Simple Linear Regression	Multiple Linear Regression
Response Variable type	Categorical Numerical	Categorical Numerical
Explanatory Variable(s) type	Categorical Ordinal Numerical *	Categorical Ordinal Numerical
Model Equation	$Y = \beta_0 + \beta_1 X + \varepsilon$	$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$
Description of model terms	β_0 = intercept β_1 = slope ε = error term	β_0 = intercept $\beta_1, \dots, \beta_k$ = slope coefficients for individual predictors ε = error term
Model Assumptions	· Linearity between X + Y · $\varepsilon \sim N(0, \sigma_\varepsilon)$	· Linearity in the coefficients β_i · $\varepsilon \sim N(0, \sigma_\varepsilon)$
Graphs used to test model assumptions	· X-Y plot · normal prob. plot of resids. · resids vs. fits plot (for constant variance + linearity) · Cook's D for influential observations	· X-Y plot for all X's · normal prob. plot of resids. · resids vs. fits plot · Cook's D for influential obs.
Estimated model equation	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \dots + \hat{\beta}_k X_k$
Equation of residuals	$\hat{\varepsilon}_i = y_i - \hat{y}_i$	$\hat{\varepsilon}_i = y_i - \hat{y}_i$

* So far, we've only seen numerical X in SLR, but in fact we can use any type of variable as X.

	Simple Linear Regression	Multiple Linear Regression
df(Model)	1	k
df(Error)	n - 2	n-k-1
df(Total)	n-1	n-1
SS(Total)	$\sum_{i=1}^{n}(y_i-\bar{y})^2$	$\sum_{i=1}^{n}(y_i-\bar{y})^2$
MS(Model)	$\frac{SSmodel}{df_{model}} = \frac{\sum(\hat{y}-\bar{y})^2}{1}$	$\frac{SSmodel}{df_{model}} = \frac{\sum(\hat{y}-\bar{y})^2}{k}$
MS(Error)	$\frac{SSE}{df_{Error}} = \frac{\sum(y-\hat{y})^2}{n-2}$	$\frac{SSE}{df_{Error}} = \frac{\sum(y-\hat{y})^2}{n-k-1}$
F-statistic	$\frac{msmodel}{mSE} \sim F_{dfmodel,dfError}$	$\frac{msmodel}{mSE} \sim F_{dfmodel,dfError}$
Hypotheses for test of individual coefficient (β_i)	$H_0: \beta_1 = 0$ $H_A: \beta_1 \neq 0$	$H_0: \beta_i = 0$ $H_A: \beta_i \neq 0$
CI for individual coefficient (β_i)	$\hat{\beta}_1 \pm t^*_{n-2} \cdot SE_{\hat{\beta}_1}$	$\hat{\beta}_i \pm t^*_{n-k-1} \cdot SE_{\hat{\beta}_i}$
Hypotheses for ANOVA-based test (in symbols)	$H_0: \beta_1 = 0$ $H_A: \beta_1 \neq 0$	$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ H_A: at least one $\beta_i \neq 0$
Hypotheses for ANOVA-based test (in words)	H_0: model is useless H_A: model is useful	H_0: model is useless H_A: model is useful
R^2 (calculation)	$R^2 = \frac{SSModel}{SSTotal} = 1 - \frac{SSE}{SSTotal}$	same
R^2 (interpretation)	The % of variability in Y that is explained by the linear model. $R^2 = (r)^2$	same $R^2 \neq (r)^2$
R^2_{adj} (calculation)	–	$R^2_{adj} = 1 - \frac{SSE/(n-k-1)}{SSTotal/(n-1)} = 1 - \frac{mSE}{Var(Y)}$
R^2_{adj} (interpretation)	–	R^2 adjusted to account for the # of predictors in the model.