Name: Brad Hall
ISDS 7024
Homework 2

1) Given the partial results from a linear regression model below and a sample size of 200, what is the t-statistic for β_1? (Round to 1 decimal place.)

Coefficients				
Source	**Estimate**	**Std Error**	***t*-value**	**Pr(> |*t*|)**
Intercept	−3691.031	1080.676	$\frac{\widehat{\beta_0}}{\text{s.e.}(\widehat{\beta_0})} = -3.4$	
Variable 1	40.121	3.547	$\frac{\widehat{\beta_1}}{\text{s.e.}(\widehat{\beta_1})} = 11.3$	

2) Given the partial results from a linear regression model below and a sample size of 504, what is the F-statistic for the overall model? (Round to 1 decimal place.)

	Degrees of Freedom	**Sum of Squares**	**Mean Square**	***F*-Statistic**
Model	3	SST - SSE = 288	$\frac{\text{SSM}}{\text{df1}} = 96$	$\frac{\text{MSR}}{\text{MSE}} = 4$
Residuals	$n - p - 1 = 500$	12000	$\frac{\text{SSE}}{\text{df2}} = 24$	
Total	$n - 1 = 503$	12288		

(a) What is the p-value for the test? (Round to 3 decimal places.)

Solution: The p-value for the test was found using Python:

$$\texttt{scipy.stats.f.cdf(4, 3, 500)} = 0.008.$$

(b) At 0.05 level of significance, is the model significant? Why or why not?

Solution: Yes, the model is significant because our p-value is smaller than the level of significance; i.e., $0.008 \leq 0.05$. Equivalently, the F-statistic is larger than the critical value of 2.62 found on the F-table, which provides the same conclusion.

3) What is the R^2 for the regression model above? (Round to 4 decimals.)

Solution: The coefficient of determination R^2 for the regression model above is calculated as

$$R^2 = \frac{\text{SSM}}{\text{SST}} = \frac{288}{12288} = 0.0234.$$

4) Use `dataset30.xlsx` to test the regression model $Y = \beta_0 + \beta_1 X + \beta_2 GP$.

(a) What is the value of β_2? (Round to 2 decimal places.)

Solution: The value of β_2 is 818.29.

(b) Is β_2 statistically significant? (Use $\alpha = 0.05$.)

Solution: The value β_2 is statistically significant as indicated by the p-value being less than 0.0001 - way below the maximum threshold of 0.05.

(c) Compare the full model (having predictors X and GP) with the reduced model (having just predictor X).

(1) What is the SSE for the Full Model? (Round to the nearest whole number.)

Solution: The SSE for the Full Model is 7335456.

(2) What is the SSE for the Reduced Model? (Round to the nearest whole number.)

Solution: The SSE for the Reduced Model is 18046973.

(3) Are the two models coincident?

Solution: The two models are not coincident.

(4) Include the F-test statistic. (Round to 3 decimal places.)

Solution: The F-test statistic is 99.995.

(5) What is the p-value? (Round to 3 decimal places.)

Solution: The p-value is 0.

5) Given this regression model: $Y = 100 + 5X_1 + 10X_2 - 15X_1X_2$, what is the predicted value of Y when $X_1 = 2$ and $X_2 = 4$? Assume that all three predictors are statistically significant. (Round to 1 decimal place.)

$$\begin{aligned} Y &= 100 + 5X_1 + 10X_2 - 15X_1X_2 \\ &= 100 + 5(2) + 10(4) - 15(2)(4) \\ &= 30. \end{aligned}$$

6) Assume a regression analysis provides the following model: $Y = \beta_0 + 20X_1 + \epsilon$. What do the residuals represent? Choose the best answer.

Solution: The residuals represent b. The portion of Y that is not associated with X_1.

7) Which of the following statements are true?

Solution: The following statements are true.

- The "variables added last" strategy provides a way to estimate the unique proportion of variance attributable to a single independent variable.

- The full and reduced models approach assumes the reduced model is nested within the full model.
- Use of "dummy" variables is appropriate when the data contains two or more groups that are expected to have different means on the response variable.

Use `M2_HRdat.xlsx` and the variable added last strategy to answer the following questions.

8) Which independent variable explains the most unique variance in Y? Choose the best response.

Solution: The variable that explains the most unique variance in Y is X_1.

- Provide the value of the F-test statistic you obtained for the independent variable selected as explaining the most unique variance in Y. (Round to 2 decimal places.)

  Solution: The F-test statistic found for the independent variable X_1 is 5.22.

- What is the change in R^2 when the variable you selected is deleted? (Round to 3 decimal places.)

  Solution: The change in R^2 is

$$\Delta R^2 = |0.296862 - 0.240373| \approx 0.056.$$

9) Based on the full and reduced model tests, which independent variable(s) do not explain unique variance in Y? Choose all correct responses.

Solution: The independent variables that do not explain unique variance in Y are X_3 and X_4.

10) Which of the following are FALSE with regards to continuous variable interactions? Choose all of the FALSE responses.

Solution: There is one false statement: Dependent variables should be centered.