Assignment #2 — Due: Tuesday, October 20, 2020, by 6:00 p.m.

Use R to answer the questions. Your solution should include relevant commands and outputs but not the data set.

1. [20 points] Given are seven observations for two variables x and y:

x	1	2	3	3	4	5	5
y	3	7	5	8	11	14	12

   Suppose that we fitted the linear model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ $(i = 1, ..., 7)$, $\epsilon_i \sim N(0, \sigma^2)$.

   (a) Write the fitted regression equation.

   (b) Find a p-value for testing $H_0 : \beta_1 \leq 2.0$ vs $H_a : \beta_1 > 2.0$.

   (c) Find the 95% confidence interval for β_0, and use the confidence interval to test $H_0 : \beta_0 = 1$.

   (d) Calculate the p-value of F statistics for the hypothesis $H_0 : \beta_1 = 0$. (First, find the value of F statistic and then calculate the p-value by your own coding.)

   (e) Find the 90% confidence interval for $\hat{\beta}_0 + 4\hat{\beta}_1$.

2. [24 points] We are interested in developing a model that describes the gas mileage (in mpg). We will use engine size (in cubic cm), horsepower and weight of the car (in pounds) as explanatory variables. Download `HW2(data).RData` from the *blackboard*.

   (a) Write the estimated regression equation using `mpg` as a response variable and the others as explanatory variables.

   (b) Interpret the estimated coefficient of `weight`.

   (c) Conduct the F-test for the overall fit of the regression. Comment on the results.

   (d) Test each of the regression coefficients. Do the results indicate that any of the explanatory variables should be removed from the model?

   (e) Find the regression model with the explanatory variable(s) identified in part (d) removed.

   (f) Conduct an F test to choose a model between the full model in (a) and reduced model in (e).

3. [26 points, 2 points for each question] The following tables are regression output when Y is regressed on X_1 for 20 observations. Complete the 13 missing numbers. Provide some reasons of your answers. Correct answers without showing the calculation process will not get full credit.

Table 1: (i) ANOVA Table

Source	Sum of Squares	df	Mean Square	F-statistic
Regression	1848.76	(a)	(b)	(c)
Residuals	(d)	(e)	(f)	

Table 2: (ii) Coefficients Table

Variable	Coef	SE	t	p-value
Intercept	–23.4325	12.74	(g)	0.0824
X_1	(h)	0.1528	8.32	< 0.0001
$n =$ (i)	$R^2 =$ (j)	$R_a^2 =$ (k)	$\hat{\sigma} =$ (l)	df = (m)