Assignment #3 — Due: Thursday, November 26, 2020, by 10:15 a.m.

Show all of your workings. Turn in your homework in lecture.

1. [36 points] **[Exercise #5.5 in the textbook (5th Edition)]** Table 5.15 shows a regression output obtained from fitting the model $Y = \beta_0 + \beta_1 X + \epsilon$ to a set of data consisting of n workers in a given company, where Y is the weekly wages in \$100 and X is the gender. The Gender variable is coded as 1 for Males and 0 for Females.

Table 5.15 Regression Output from the Regression of the Weekly Wages, Y, on X (Gender: 1 = Male, 0 = Female)

ANOVA Table				
Source	Sum of Squares	df	Mean Square	F-Test
Regression	98.8313	1	98.8313	14
Residual	338.449	48	7.05101	
Coefficients Table				
Variable	Coefficient	s.e.	t-Test	p-value
Constant	15.58	0.54	28.8	< 0.0001
X	–2.81	0.75	–3.74	0.0005

(a) How many workers are there in this data set?

(b) Compute the variance of Y?

(c) Given that $\bar{X} = 0.52$, what is $\bar{Y}$?

(d) Given that $\bar{X} = 0.52$, how many women are there in this data set?

(e) What percentage of the variability in Y can be accounted for by X?

(f) Compute the correlation coefficient between Y and X?

(g) What is your interpretation of the estimated coefficient $\hat{\beta}_1$?

(h) What is the estimated weekly wages of a man chosen at random from the workers in the company?

(i) What is the estimated weekly wages of a woman chosen at random from the workers in the company?

(j) Construct a 95% confidence interval for β_1.

(k) Test the hypothesis that the average weekly wages of men is equal to that of women. [Specify (a) the null and alternative hypotheses, (b) the test statistics, (c) the critical value, and (d) your conclusion.]

2. [24 points] **[Exercise #5.6 in the textbook (5th Edition)]** The price of a car is thought to depend on the horsepower of the engine and the country where the car is made. The variable Country has four categories:USA, Japan, Germany, and Others. To include the variable Country in a regression equation, three indicator variables are created, one for USA, another for Japan, and the third for Germany. In addition, there are three interaction variables between the horsepower and each of the three Country categories (HP*USA, HP*Japan, and HP*Germany). Some regression outputs when fitting three models to the data is shown in Table 5.16. The usual regression assumptions hold.

   (a) Compute the correlation coefficient between the price and the horsepower.

   (b) What is the least squares estimated price of an American car with a 100 horsepower engine?

   (c) Holding the horsepower fixed, which country has the least expensive car? Why?

   (d) Test whether there is an interaction between Country and horsepower. Specify the null and alternative hypotheses, test statistics, and conclusions.

   (e) Given the horsepower of the car, test whether the Country is an important predictor of the price of a car. Specify the null and alternative hypotheses, test statistics, and conclusions.

   (f) Would you recommend that the number of categories of Country be reduced? If so, which categories can be joined together to form one category?

3 points each for 1-(a) through (j). 6 points for 1-(k). 4 points each for 2-(a) through (f).

Table 5.16 Some Regression Outputs When Fitting Three Models to the Car Data

Model 1

Source	Sum of Squares	df	Mean Square	F-Test
Regression	4604.7	1	4604.7	253
Residual	1604.44	88	18.2323	

Variable	Coefficient	s.e.	t-Test	p-value
Constant	–6.107	1.487	–4.11	0.0001
Horsepower	0.169	0.011	15.9	0.0001

Model 2

Source	Sum of Squares	df	Mean Square	F-Test
Regression	4818.84	4	1204.71	73.7
Residual	1390.31	85	16.3566	

Variable	Coefficient	s.e.	t-Test	p-value
Constant	–4.117	1.582	–2.6	0.0109
Horsepower	0.174	0.011	16.6	0.0001
USA	–3.162	1.351	–2.34	0.0216
Japan	–3.818	1.357	–2.81	0.0061
Germany	0.311	1.871	0.166	0.8682

Model 3

Source	Sum of Squares	df	Mean Square	F-Test
Regression	4889.3	7	698.471	43.4
Residual	1319.85	82	16.0957	

Variable	Coefficient	s.e.	t-Test	p-value
Constant	–10.882	4.216	–2.58	0.0116
Horsepower	0.237	0.038	6.21	0.0001
USA	2.076	4.916	0.42	0.6740
Japan	4.755	4.685	1.01	0.3131
Germany	11.774	9.235	1.28	0.2059
HP*USA	–0.052	0.042	–1.23	0.2204
HP*Japan	–0.077	0.041	–1.88	0.0631
HP*Germany	–0.095	0.066	–1.43	0.1560