Quantitative Reasoning for Management
Individual Problem Set

Instructions: This problem set is worth 15% of your final course grade, and it is **due on November 10. You must submit your solutions to Quercus through the course page no later than 11:59 p.m. (EDT) on the 10th, and please ensure that your solutions include a cover sheet with your name and student number.** The problem set has three questions (each with multiple parts) and there are 100 marks in total for all of the questions.

Question 1 (40 marks)

Your company asks you to analyze the characteristics of people who are good salespeople, so you collect historical data on the weekly sales figures of various people who have been employed by your company. The data contains the following information on workers in the sample: weekly sales (in dollars), education (in years) and tenure at the company (in years). You estimate the following regression:

$$Weekly\ Sales = \beta_0 + \beta_1(Education) + \beta_2(Tenure)$$

The estimation results from this regression are reported below:

	Coefficient	Standard Error of Coefficient
Intercept	3000	150
Education	1000	250
Tenure	800	200

The following questions relate to this regression output.

(a) Bearing in mind that this is a multiple regression, what is the proper interpretation of the coefficient on education? (2 marks)

(b) In this multiple regression setting, how would you test the hypothesis that the effect of education on sales was zero at the 5% level of significance for all workers at this company? (4 marks)

(c) Again, in this multiple regression setting, would we conclude that an additional year of tenure at this company had any effect on sales? How would we formally test this notion at the 5% level of significance? (4 marks)

(Please see the following page for the next part of this question)

(d) Suppose that we now add to this regression a dummy variable entitled "Training Program", which is equal to one if the person participated in a training program offered by the company to teach certain sales techniques, and zero otherwise. The new regression equation is:

$$Weekly\ Sales = \beta_0 + \beta_1(Education) + \beta_2(Tenure) + \beta_3(Training\ Program)$$

And suppose the regression results that included this variable looked like this:

	Coefficient	Standard Error of Coefficient
Intercept	3000	150
Education	100	250
Tenure	800	200
Training Program	500	100

Given these results:

(i) Would we conclude that "Training Program" has an impact on sales at this company? How would we formally test this notion at the 5% level of significance? (5 marks)

(ii) Given the inclusion of the variable representing participation in the training program, would we conclude that education has a meaningful impact on sales at this company? How would we formally test this notion at the 5% level of significance? (5 marks)

(iii) Given what you can see in the results, explain the sign (positive or negative) of the correlation between "Education" and "Training Program". How did this correlation alter the coefficient on "Education"? (10 marks)

(Please see the following page for the next part of this question)

(e) Now suppose that the company provides you with data on the number of times an employee's supervisor has heard about a complaint from a co-worker about the employee. If the variable "Co-worker Complaints" represents the total number of complaints, then suppose that we included this variable in our regression, and obtained the following results:

	Coefficient	Standard Error of Coefficient
Intercept	3000	150
Education	100	250
Tenure	100	200
Training Program	500	100
Co-worker Complaints	-100	10

Intuitively, why did the coefficient on Tenure change after the inclusion of the variable "Co-worker Complaints"? Given this change, should firms value a worker's tenure when assessing the ability to effectively create sales? What does this tell us about the importance of missing variables in regressions? (10 marks)

Question 2 (30 marks)

Suppose that your company manufactures stoves, is attempting to determine the best price to set for a new stove that they are about to release to the public for sale. The company's internal research team runs the following regression using data on some of the current models of stoves that are already on the market:

$$Price = \beta_0 + \beta_1(Number\ of\ Burners) + \beta_2(Number\ of\ Dials) \\ + \beta_3(Maximum\ Temperature)$$

In this case:

(i) "Price" is the price of the stove

(ii) "Number of Burners" is the number of burners on the stove top

(iii) "Number of Dials" is the number of dials on the stove's control panel

(iv) "Maximum Temperature" is the highest possible temperature that can be generated by the stove (in degrees Celsius).

The regression results from the data analyzed by the internal research team are displayed below:

	Coefficient	Standard Error of the Coefficient
Intercept	800	10
Number of Burners	40	16
Number of Dials	8	2
Maximum Temperature	1.5	0.25
Standard Error of the Regression = 100		

(a) Suppose that the new stove that the company will produce will have four burners on the stove top, 4 dials on the control panel, and a maximum temperature of 260 degrees Celsius. If the cost of the stove is \$700 for the company, and the company needs to make a profit of \$500 on each television, then would you advise the company to manufacture this stove? Why or why not? (10 marks)

(Please see the following page for the next part of this question)

(b) Now suppose that you do some work for the company, but you use a slightly different regression than the internal research team. In particular, you estimate:

$$Price = \beta_0 + \beta_1(Number\ of\ Burners) + \beta_2(Number\ of\ Dials) + \beta_3(Maximum\ Temperature) + \beta_4(Stainless\ Steel)$$

In this case, "Stainless Steel" is a dummy variable equal to one if the stove has a stainless steel exterior, and zero if it does not. The results from this regression are displayed below:

	Coefficient	Standard Error of the Coefficient
Intercept	800	10
Number of Burners	40	16
Number of Dials	1	2
Maximum Temperature	1.5	0.25
Stainless Steel	150	10
Standard Error of the Regression = 100		

In this case, the coefficient on "Number of Dials" changed in this regression compared to the regression in part (a). What is the proper explanation (which uses your statistical training) for why this change occurred? (10 marks)

(c) Suppose that the stove that the company intends to manufacture is stainless steel, but has all of the same specifications as were described in part (a). If the cost of the television is still \$700 for the company, and if the company still needs to make a profit of \$500 on it, would you advise the company to do so, given your results in part (b)? How would your advice differ (if at all) from part (a)? (10 marks)

Question 3 (30 marks)

(a) Suppose that you are investigating the difference in productivity between two different (but equally-sized) teams of workers in the same manufacturing plant. Ideally, you would collect data an indefinite number of shifts (of equal duration) for both teams, but because of practical considerations, you collect a sample of data on the output produced by both teams on a set of different shifts. You then use this data to estimate the following regression:

$$Output = \beta_0 + \beta_1(Team\ \#1)$$

In this case, "Output" represents the number of units of output made by a team of workers on a particular shift, and "Team #1" is a dummy variable equal to one if the data was collected from team #1, and zero if the data was collected from team #2. Your regression results are listed below:

	Coefficient	Standard Error of Coefficient
Intercept	400	80
Team #1	-50	5

(i) Interpret the meaning of the intercept in the above regression. (5 marks)
(ii) Interpret the meaning of the coefficient on the variable "Team #1" in the above regression. (5 marks)
(iii) Use these regression results to determine the average output produced by team #1 on a given shift. (5 marks)
(iv) Use your statistical training to rigorously test whether or not the two teams have similar or different levels of output. (5 marks)

(b) Suppose that instead of running the regression listed above, you instead estimate the following regression:

$$Output = \beta_0 + \beta_1(Team\ \#2)$$

In this regression, "Output" is still defined in the same way as before, but "Team #2" is a dummy variable equal to one if the data was collected from team #2, and zero if the data was collected from team #1. In this case:

(i) Use the estimated coefficients from part (a) to determine the value of the intercept term in this regression (here in part (b)). Interpret the meaning of the intercept in this case. (5 marks)
(ii) Use the estimated coefficients from part (a) to determine the value of the coefficient on the dummy variable "Team #2" in this regression (here in part (b)). Interpret the meaning of the coefficient on "Team #2" in this case. (5 marks)