ECON 140A FALL 2020

HW 2

Notes:

1. *Show your work and justify all steps. You will lose points if your reasoning is incomplete.*

Problem 1. OLS in the population.

(a) Consider $f(x_k) = b_0 + b_1 x_k$, the Ordinary Least Squares univariate affine regression function of y_k on a constant and x_k in the population. Explain in words the meaning of "Least Squares". (2 points)

(b) Suppose someone suggests that you should minimize the sum of the prediction error, $\sum_{k=1}^{n}(y_k - b_0 - b_1 x_k)$. Explain why this is a good or bad idea? What are the ramifications for our estimates of the y_k values? (2 points)

(c) Explain in words why it is a good idea to include a constant in a univariate regression. Give a real world example in your explanation. (2 points)

(d) Write down the OLS parameters, b_0 and b_1, for the population regression specification, $y_k = b_0 + b_1 x_k + e_k$. (2 point)

(e) Are the OLS parameters you wrote down above random variables? Explain. (2 points)

Problem 2. Suppose you have a sample of n observations from the population where Y_i represents the dependent variable and X_i represents the independent variable.

(a) Write down expressions for the estimators from a regression of Y_i on a constant and X_i in the sample. That is, the regression specification, $Y_i = b_0 + b_1 X_i + e_i$. (2 points)

(b) Assume that $\lim_{n\to\infty} \hat{\beta}_1 = \beta_1$. Now use the law of large numbers to show that $\lim_{n\to\infty} \hat{\beta}_0 = \beta_0$. *Remember to show each step and provide explanations.* (5 points)

(c) As in the lecture notes, let $\hat{e}_i = (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$, representing the error we make when we use the sample OLS regression to predict Y_i. Show that the sample covariance between X_i and $\hat{e}_i$ is zero. Explain why this would be the case. *Recall that the sample covariance between two variables, W_i and Z_i, is $\frac{1}{n}\sum_{i=1}^{n}(W_i - \bar{W})(Z_i - \bar{Z})$*, where $\bar{W} = \frac{1}{N}\sum_{i=1}^{N} W_i$ and $\bar{Z} = \frac{1}{N}\sum_{i=1}^{N} Z_i$ *Remember to show each step and provide explanations.* (5 points)

Problem 3. Suppose that you are tasked with estimating the primary determinants of annual salaries in the USA. You initially obtain data (a sample) of n observations which contains the following variables for each individual in your sample: $salary_i$, $gender_i$, $education_i$ and $parentSalary_i$. Here $salary_i$ is specified in dollars per year, $gender_i$ is an indicator variable equal to 1 if female and 0 for non-female, age_i is specified in years, $education_i$ is specified in years, and $parentSalary_i$ is the average of the individuals parents salaries specified in dollars per year.

(a) Consider the regression specification $wages_i = \beta_0 + \beta_1 gender_i + e_i$. You use some statistical software and estimate that $\hat{\beta}_0 = 55{,}000$ and $\hat{\beta}_1 = -11,000$. Interpret both estimated coefficients.(2 points)

(b) Now consider instead the regression specification $ln(wages_i) = \beta_0 + \beta_1 gender_i + \beta_2 education_i + e_i$. You use some statistical software and estimate that $\hat{\beta}_2 = 0.053$. Interpret the estimated value of β_2.(2 points)

(c) Now consider instead the regression specification $ln(wages_i) = \beta_0 + \beta_1 gender_i + \beta_2 education_i + \beta_3 ln(parentSalary_i) e_i$. You use some statistical software and estimate that $\hat{\beta}_3 = 0.8$. Interpret the estimated value of β_3.(2 points)

(d) Suppose you obtain more variables. Specifically, you now have n observations and j independent variables, where $n = j - 1$. Suppose you run an OLS regression of Y_i on your $j - 1$ covariates and a constant. What is the R^2? Explain why you get this result and whether you have good explanatory power of the Y_i values in your sample and outside your sample. (2 points)