**Question 1.** Consider the model Yi = α + β ln(Xi) + ui . The interpretation of the slope coefficient β is as follows:

(a) a change in X by 100 × β% is associated with a one unit change in Y.

(b) a change in X by one unit is associated with a 100 × β% change in Y.

(c) a 1% change in X is associated with a change in Y of 0.01×β.

(d) a 0.01×β unit change in X is associated with a change in Y of 1%.

**Question 2** Consider the linear regression model ln Yi = β0 +β1X1i +β2X2i +vi . After running the regression we find V IF1 = V IF2 = 2.25. This implies that the R-squared from a regression of X1 on X2 is approximately,

(a) 0.44

(b) 0.56

(c) 0.67

(d) 0.75

**Question 3.** The Rule of 68-95-99.7,

(a) describes the “fair” distribution of grades following an econometrics exam.

(b) states that exactly 99.7% of the probability mass lies ±3σ of the mean.

(c) suggests that approximately 99.7% of the probability mass lies ±3σ 2 of the mean.

(d) suggests that approximately 99.7% of the probability mass lies ±3σ of the mean.

**For questions 4-5 use the table provided on the next page**.

**Question 4.** The results suggest a diminishing return to years of experience; captured by the negative sign on the squared term. In fact, this diminishing effect is so strong that after approximately years, expected earnings actually fall.

(a) 25

(b) 35

(c) 50

(d) Not enough information.

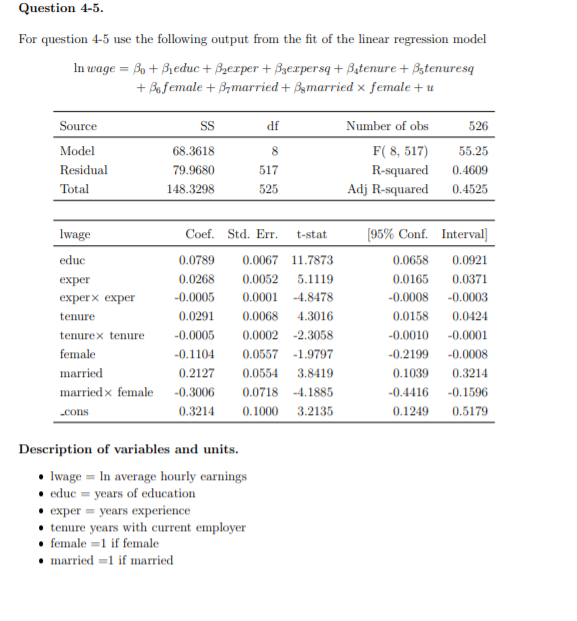
**Question 5.** The results suggest that the expected earnings of married woman:

(a) are approximately 19% less than their married male counterpart, ceteris paribus.

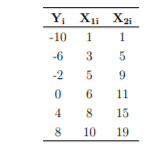
(b) are approximately 30% less than their married male counterpart, ceteris paribus.

(c) are approximately 41% less than their married male counterpart, ceteris paribus.

(d) depend on years of tenure and therefore cannot be determined



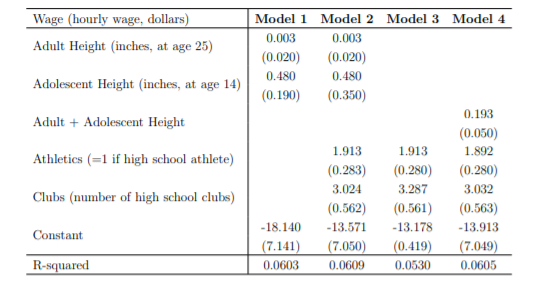
**Question 6** **A.** Answer two out of three of the following questions. Make sure to include the appropriate mathematics, statistics, and diagrams where required. A. Suppose you want to fit the model Yi = β0 + β1X1i + β2X2i + ui to the following data



EXPLAIN why it is not possible to estimate the three unknowns (βb0, βb1, and βb2)

**B**. Suppose your classmate “Bruce” is interested in estimating the relationship between the structural characteristics of a house (sqft, baths and beds) and its sales price (price). To this end, he estimates the linear regression model: ln(pricei) = β0 + β1sqf t gi + β2baths ]i + β3beds gi + ui . For “reasons”, he chose to convert the right-hand side variables into standardized values (similar to what we did when creating t-statistics). Specifically, Xei = (Xi − X)/sx where s 2 x denotes the sample variance on X. Suppose ssqf t = 694, sqf t = 2106 and βb1 = 0.147. Imagine Bruce needs to communicate his results to someone that has never taken econometrics. INTERPRET (in a sentence) the parameter estimate βb1, and CALCULATE and INTERPRET (in a sentence) the marginal effect of sqft on the price of a $200,000 home

**Question 7.** Recently, economists have become interested in how individual physical characteristics affect labor market outcomes. One characteristic that has drawn attention is height. Using a sample of 1, 910 individuals, the following models were estimated (standard errors in parentheses):



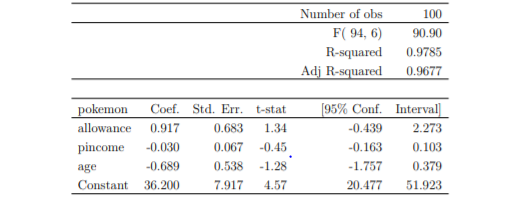
**A.** Consider the following Null hypothesis for Model 2: H0 :βAdult = βAdolescent = 0 HA :βAdult 6= βAdolescent 6= 0 Given the information in the table above, use an F-test to TEST this hypothesis at the 5% significance level

**B.** Consider the following null hypothesis for Model 2: H0 :βAdult = βAdolescent HA :βAdult 6= βAdolescent Given the information in the table above, use an F-test to TEST this hypothesis at the 5% significance level.

**C.** Consider the following null hypothesis for Model 2: H0 :βAtheletics = 0 HA :βAtheletics 6= 0 EXPLAIN whether we have enough information in the table above to construct an F-test to test this hypothesis

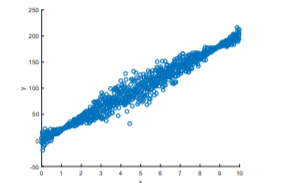
**Question 8.** Answer one of the following questions. Make sure to justify your answer. A. Consider the following regression equation (and Stata output) for spending on Pokemon cards (a function of allowance money, parent’s income, and age),

pokemoni = β0 + β1allowancei + β2pincome + β3age + ui



EXPLAIN whether the above OLS regression results suggest any data problems? If so, what could we do about it?

B. Consider the simple regression equation, Y = α + βX + u. Plotting a scatter plot of X and Y we find



EXPLAIN the likely consequences of assuming that the CLRM assumptions hold and running OLS on the data depicted in the figure. What could we do to address this data issue?

**Question 10**. In this question, you will use the Stata data set PNTSPRD.dta. Along with your written responses to the following prompts, please submit PDF copies of your log and do-files. Let favwin denote a binary variable: equal to 1 when a team favored by the Las Vegas point spread wins a Basketball game. Now, consider a linear probability model to estimate the probability that the favored team wins given the spread: Pr(f avwin = 1|spread) = β0 + β1spread.

A. EXPLAIN why we should expect β0 = 0.5, if spread incorporates all relevant information.

B. Using the PNTSPRD.dta dataset, ESTIMATE the linear probability model described above. Make sure to include a copy of your Stata output in your log file.

C TEST H0 : β0 = 0.5 against a two-sided alternative (5% significance level). Use the appropriate standard errors. Make sure to show your work. 10

D. Is spread statistically significant? CALCULATE the estimated probability that the favored team wins when spread is 10? Show your work.

E. Now, ESTIMATE a probit model for Pr(f avwin = 1|spread). Make sure to include a copy of your Stata output in your log file.

F. EXPLAIN why the null hypothesis H0 : β0 = 0 from the probit model is equivalent to the Null hypothesis H0 : β0 = 0.5 from the linear probability model.

G. INTERPRET βb0 from your probit model and TEST the null hypothesis that the intercept is zero (H0 : β0 = 0). Make sure to show your work. 11

H. Use the estimates from the probit model to CACULATE the estimated probability that the favored team wins when spread=10. Compare this estimate with your LPM estimate from part (D). Make sure to show your work. Hint, the normal command in Stata may be helpful.

I. Add the variables favhome, fav25, and und25 to the linear probability model and TEST the joint significance of these variables. INTERPRET this result, focusing the question of whether spread incorporates all observable information prior to the game

**Question 11.** Let grad denote a dummy variable for whether a student-athlete at a large university graduates in four years (=1 if graduates in four years). Suppose we are interested in modeling this binary outcome as a function of high school GPA (hsGPA), SAT score, and the number of hours spent per week in organized study hall (study). Now, suppose that we estimate the following logit model using 420 student-athletes:

Pr( c grad = 1|hsGP A, SAT, study) = Λ(−1.17 + 0.24hsGP A + 0.00058SAT + 0.73study),

where Λ(Xiβ) = exp(Xiβ)/(1 + exp(Xiβ)) is the logit function.

Holding hsGPA fixed at 3.0 and SAT fixed at 1,200, CALCULATE the estimated difference in graduation probability for someone who spent 10 hours per week in study hall and someone who spent 5 hours per week. *Show your work.*