Applied Statistics for Management (Project Activity)

BPA 2k19

The attached data file includes data on children's test scores and other socio-economic characteristics obtained from the Annual Status of Education Report (ASER)-Pakistan. Data on the following list of variables is available in the dataset attached.

Descriptive Statistics (Rural Sample-2019)

Variable	Obs	Mean	Std. Dev.	Min	Max
age	82816	8.479	3.671	3	16
gender	82685	.408	.491	0	1
child not enrolled	83267	.231	.422	0	1
child dropped out	83267	.036	.186	0	1
child enrolled	83267	.727	.445	0	1
child dropoutclass	2520	3.966	2.293	0	12
current grade	54483	3.73	2.728	0	12
child school govt	83267	.478	.5	0	1
child school pvt	83267	.175	.38	0	1
child school other	83267	0	0	0	0
paid tuition	48949	.108	.31	0	1
house type	83267	1.781	.726	1	3
house ownership	81790	.948	.221	0	1
electricity avail	82401	.653	.476	0	1
mobile avail	80856	.712	.453	0	1
tv avail	76309	.452	.498	0	1
mother age	70539	35.039	6.533	20	83
mother gone toschool	73788	.313	.464	0	1
mother education	22762	8.451	3.349	1	16
father age	70441	39.621	7.877	22	95
father gone toschool	72433	.669	.471	0	1
father education	48122	9.347	3.379	1	20
reading	67150	2.961	1.59	1	5
math	64989	3.554	2.008	1	6
english	65474	3.024	1.622	1	5
hh size1	83267	3.698	1.963	0	20
house kacha	83267	.397	.489	0	1
house semi kacha	83267	.425	.494	0	1
house pakka	83267	.178	.383	0	1
hh size	83213	3.766	2.98	0	99
std reading	67150	0	1	-8.03	3.282
std math	64873	0	.996	-10.33	7.346
std english	64934	0	.996	-11.353	7.814

Important Note: Your answers should be submitted in one ms-word file saved with your name and uploaded to LMS (no email submission). For each question, whatever stata command you used, you need to provide it in the end of each answer. Please follow the this order while answering your questions:

Question#1: *******
Answer:*******(answer may include tables where required)
Stata Command: *****************
Question#2: ********
Answer:*********
Stata command
And so on.

Using the dataset, please attempt the following questions and produce desired results as per instructions.

Question 1 : Conduct a single sample t-test to test the null hypothesis that the <u>population mean of the Reading, Math and English score is equal to 3.0</u>. Use 95% confidence interval to draw your conclusion. Report t-value and two-tailed p-values and draw your conclusion. (Hint! Use Stata command for single sample t-test). (2 points)

Question 2: Now conduct a mean-comparison t-test to test the null hypothesis that std_reading, std_math, and std_English *scores of government school students are not different from private school students*. In order words, test whether government school children perform differently from private school children in terms of standardized reading, math and English? Report t-values and two-tailed p-values and draw your conclusion (Hint! Use stata command for two samples mean comparison test). (5 points)

Question 3: Suppose you are wondering about the impact of *paid-tuition (variable of interest)* on children's standardized test performance (outcome of interest). You have data on a number of variables including, standardized test scores of reading, math, and English (dependent variables) and a number of independent variables (i.e. age, gender, house ownership, mother's education, father education, household size etc).

(i)   Conduct a one-by-one multiple regressions using the following equation and export results onto ms-word using stata command (asdoc). (Your results tables should show, constant, co-efficients of all variables, standard errors, R-squared and number of observations). (5 points)

Equation for Table #1

$$std\_reading = \beta_0 + \beta_1 Tuition + \beta_2 Age + \beta_3 Gender + \beta_4 HHOwnership + \beta_5 HHSize + \beta_6 MotherEducation + \beta_7 FatherEducation + Error$$

Equation for Table#2

$$std\_Math = \beta_0 + \beta_1 Tuition + \beta_2 Age + \beta_3 Gender + \beta_4 HHOwnership + \beta_5 HHSize + \beta_6 MotherEducation + \beta_7 FatherEducation + Error$$

Equation for Table3

$$std\_nglish = \beta_0 + \beta_1 Tuition + \beta_2 Age + \beta_3 Gender + \beta_4 HHOwnership + \beta_5 HHSize + \beta_6 MotherEducation + \beta_7 FatherEducation + Error$$

(ii)   Interpret your results by focusing on each coefficient's statistical and economic significance in table #1. (5)
(iii)   Conduct the following regression diagnostics for robustness of your results (3 points)
    a. Multi-collinearity
    b. Heteroskedasticity
    c. Omitted variables

Question 4: We know that simple OLS is a not a powerful model to draw causal inference of the effect of tuition on standardized test scores after controlling for other variables. We need to adopt a stronger approach to genuinely attribute the learning performance of children to the variable that may cause it. Let's say variable *Tuition (X)* is potentially endogenous and *House_ownership(z) is* exogenous while *Standardize Reading, Math and English* are three outcome variables (Y).

(i) Write 2SLS (instrumental variable approach) equations for evaluating the effect of House Ownership (Z) on standardized test performance(Ys) through tuition (X) while controlling for other variables such as age, gender, HHsize, mothers' education father education etc. (2 points)

(ii) Test the two key assumptions of the 2SLS to decide whether this approach should be adopted? (3 points)

(iii) Produced a Two-stage Least Square (2SLS) results in Table and interpret your result. (5 points).

Good Luck !