

## **Econ 665**

### **Assignment #1:**

**Due Date: October 11, 2017**

#### **General Instructions**

Please use the attached file for this assignment on our mycourse page. It is a data set from the Bangladesh Household Survey that was administered in 1998. The information was collected at the household and community levels. It contains information on 1,129 households. This is the data set that we have already used in class. If you recall, some of these households have members who are participants in a micro-credit project while some households do not have participants. The variable “dfmd” is the indicator variable if the household have a female member who is a micro-credit participant.

Please answer the following sections:. You should submit your answers in a well-written text, using Stata output as well as the do file that you used to get the answers. In your report, you can use another document for your analysis but make sure it answers the questions in each section below. While your answers should be submitted individually, you are allowed to work in cooperation with your classmates. It is expected, however, that you give proper credit to your classmates who facilitated your understanding.

#### **A) Household characteristics (25%)**

Look at how different the household characteristics are between the participants and nonparticipants of microfinance programs. Open hh\_98.dta, which consists of household-level variables.

Please fill in the following table.

	Full sample		Female participants		Households without female participants	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Average household size	_____	_____	_____	_____	_____	_____
Average household assets	_____	_____	_____	_____	_____	_____
Average household landholding	_____	_____	_____	_____	_____	_____
Average age of household head	_____	_____	_____	_____	_____	_____
Average years of education of household head	_____	_____	_____	_____	_____	_____
Percentage of households with male head	_____	_____	_____	_____	_____	_____

Are the sampled households very different among the full sample, participants, and nonparticipants?

---



---



---



---

Is it possible that the Gender of household heads may also affect household characteristics.? Check by filling in the following table

	Male-headed households		Female-headed households	
	Mean	Standard deviation	Mean	Standard deviation
Average household size	_____	_____	_____	_____
Average years of head schooling	_____	_____	_____	_____
Average head age	_____	_____	_____	_____
Average household assets	_____	_____	_____	_____
Average household landholding	_____	_____	_____	_____

Are the sampled households headed by males very different from those headed by females?

---



---



---



---

## **B. Village characteristics (10%)**

Please find information on:

	Mean	Standard deviation
If village is accessible by road	_____	_____
Percentage of village land irrigated	_____	_____

### **C. Prices (15%)**

	Full sample		Participants		Nonparticipants	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Rice	_____	_____	_____	_____	_____	_____
Wheat	_____	_____	_____	_____	_____	_____
Edible oil	_____	_____	_____	_____	_____	_____
Milk	_____	_____	_____	_____	_____	_____
Potato	_____	_____	_____	_____	_____	_____

### **D. Expenditure (20%)**

The data set has household-level consumption expenditure information. Please look at the consumption patterns.

	Per capita expenditure		Per capita food expenditure		Per capita nonfood expenditure	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
<i>By head gender</i>						
Male-headed households						
Female-headed households						
<i>By head education level</i>						
Head has some education						
Head has no education						
<i>By household size</i>						
Large household (> 5)						
Small household (<= 5)						
<i>By land ownership</i>						
Large land ownership (> 50/person)						
Small land ownership or landless						

	Full sample		Female participants		Households without female participants	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Per capita expenditure						
Per capita food expenditure						
Per capita nonfood expenditure						

**Please summarize your findings on per capita expenditure comparison. Any particular insight?**

---



---



---



---

## **E. Statistical Analysis (30%)**

Now that you have taken a general look by providing some general information on the mean and standard deviations of some of the variables in the survey, it is time to assess a little more than just plain descriptive statistics.

For instance, you can describe the distributions by using the graph commands, whether it is one-way histogram, a pie chart or even a two way graphic. (see help “graph”. One can also use a non-parametric density estimation method to graph

density functions. The command is “*kdensity*” and it produces kernel density estimates and graphs the result. The way to think about this command is that it is like the histogram function where the bin number is very large.

In addition to display distribution functions, Stata has some very helpful commands that allow you to test some formal hypotheses. For instance,

- 1) the command for testing the comparison between two means is « *ttest* » .
- 2) the command “*tabulate var1 var2, chi2*” offers a « goodness of fit » test. Indeed the “*chi2*” option calculates and displays the Pearson's chi-squared for the hypothesis that the rows and columns in a two-way table if *var1* and *var2* are independent.
- 3) We can go beyond the means and standards deviation by looking at non-parametric tests that compare distribution functions. One command is « *ksminorv* ». *ksmirnov* performs one- and two-sample Kolmogorov-Smirnov tests of the equality of distributions. A one-sample test compares the distribution of the tested variable with the specified distribution. A two-sample test tests the equality of the distributions of two samples.
- 4) Another similar command is « *ranksum* ». This command *ranksum* tests the hypothesis that two independent samples (that is, unmatched data) are from populations with the same distribution by using the Wilcoxon rank-sum test, which is also known as the Mann-Whitney two-sample statistic (Wilcoxon 1945; Mann and Whitney 1947). The syntax is *ranksum var1, by(var2)*, where *var2* is the category indicator variable.

Use these commands to support the descriptive analysis that you provided in the previous sections. In using these commands, be careful to understand what the testing entails: i.e. what the null hypothesis is, what the p-value reveals and how do the tests and graphs support (or not) the initial analysis you provided previously.