ECON-665  
Homework 1

260535788 Kun Yang

1. **Household characteristics**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Full sample | | Female participants | | Households without female participants | |
|  | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation |
| Average household size | 5.300 | 2.205 | 5.205 | 2.034 | 5.406 | 2.379 |
| Average household assets | 155576.412 | 849719.933 | 75395.110 | 172979.079 | 244917.001 | 1216353.528 |
| Average household landholding | 76.832 | 204.017 | 42.359 | 79.519 | 115.244 | 279.706 |
| Average age of household head | 46.012 | 12.679 | 46.096 | 11.776 | 45.919 | 13.625 |
| Average years of education of household head | 2.317 | 3.476 | 1.753 | 3.015 | 2.946 | 3.833 |
| Percentage of households with male head | 0.908 | 0.289 | 0.901 | 0.299 | 0.916 | 0.278 |

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

The households without female participants have greater average household assets and landholding than the households with female participants.

The households without female participants have similar average household size, household head age and male head proportion with the households with female participants.

The standard deviations are different. The households without female participants have significant greater standard deviation in assets and landholding.

Above all, we can conclude that the sampled households are very different among the full sample, female participants and female nonparticipants.

* STATA code：

tabstat famsize hhasset hhland agehead educhead sexhead, by (dfmfd) stat(mean sd) format(%12.3f) c(s)

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 | 5.414 | 2.147 | 4.183 | 2.460 |
| Average years of head schooling | 2.466 | 3.540 | 0.846 | 2.314 |
| Average head age | 45.630 | 12.713 | 49.779 | 11.745 |
| Average household assets | 159943.056 | 888156.981 | 112539.777 | 250752.131 |
| Average household landholding | 76.695 | 204.830 | 78.185 | 196.773 |

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

The male-headed households have greater average household size, schooling year and household assets than the female-headed households. The female-headed households have greater average head age and household landholding than the male-headed households

The standard deviations are also different. The male-headed households have significant greater standard deviation in household assets and landholding.

Overall, the sampled households headed by males very different from those headed by females.

* STATA code：

tabstat famsize educhead agehead hhasset hhland, by (sexhead) stat(mean sd) format(%12.3f) c(s)

**B. Village characteristics**

|  |  |  |
| --- | --- | --- |
|  | Mean | Standard Deviation |
| If village is accessible by road | 0.835 | 0.371 |
| Percentage of village land irrigated | 0.560 | 0.332 |

It shows that 83.5% of the villages are accessible by road in the fall sample. And 56.0% of the village land is irrigated.

* STATA code：

tabstat vaccess pcirr, stat(mean sd) format(%11.3f) c(s)

**C. Prices**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Full sample | | Participants | | Nonparticipants | |
|  | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation |
| Rice | 10.283 | 1.566 | 10.313 | 1.599 | 10.231 | 1.509 |
| Wheat | 7.467 | 0.847 | 7.456 | 0.809 | 7.485 | 0.910 |
| Edible oil | 39.403 | 4.009 | 39.472 | 4.142 | 39.284 | 3.767 |
| Milk | 10.896 | 3.382 | 10.935 | 3.383 | 10.827 | 3.383 |
| Potato | 6.958 | 1.06 | 6.979 | 1.082 | 6.923 | 1.021 |

Because the homework instruction does not specify the gender of participants in this sector, I included all participants in the participants columns (dmmfd=1 or dfmfd=1) and excluded all participants in the nonparticipants column (dmmfd=0 and dfmfd=0).

As it shown above, the mean and standard deviation of prices are **Not** very different among the full sample, participants, and nonparticipants.

* STATA code：

tabstat rice wheat oil milk potato, stat(mean sd) format(%11.3f) c(s)

tabstat rice wheat oil milk potato if dmmfd==1 | dfmfd==1, stat(mean sd) format(%11.3f) c(s)

tabstat rice wheat oil milk potato if dmmfd==0 & dfmfd==0, stat(mean sd) format(%11.3f) c(s)

**D. Expenditure**

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 |
| By head gender |  | | | | | |
| Male-headed households | 5442.213 | 4041.490 | 3658.636 | 1554.719 | 1783.577 | 3246.287 |
| Female-headed households | 5779.345 | 5023.394 | 3675.514 | 1604.364 | 2103.831 | 3952.394 |
| By head education level |  | | | | | |
| Head has some education | 6603.109 | 4917.749 | 4150.296 | 1736.203 | 2452.813 | 4054.352 |
| Head has no education | 4676.236 | 3265.596 | 3314.451 | 1315.732 | 1361.784 | 2587.073 |
| By household size |  | | | | | |
| Large household (>5) | 5089.209 | 3414.951 | 3424.852 | 1393.446 | 1664.357 | 2579.855 |
| Small household (<=5) | 5738.315 | 4557.831 | 3822.602 | 1644.572 | 1915.713 | 3740.287 |
| By land ownership |  | | | | | |
| Large land ownership (>50/person) | 6370.431 | 4661.499 | 4090.382 | 1783.916 | 2280.050 | 3801.937 |
| Small land ownership or landless | 5076.821 | 3824.351 | 3470.093 | 1407.901 | 1606.728 | 3058.325 |

* STATA code：

tabstat exptot expfd expnfd, by(sexhead) stat(mean sd) format(%11.3f) c(s)

gen edu=0

replace edu=1 if educhead >0

tab educhead edu

tabstat exptot expfd expnfd, by(edu) stat(mean sd) format(%11.3f) c(s)

gen largesize=0

replace largesize=1 if famsize>5

tab famsize largesize

tabstat exptot expfd expnfd, by(largesize) stat(mean sd) format(%11.3f) c(s)

gen largeland=0

replace largeland=1 if hhland>50

tab hhland largeland

tabstat exptot expfd expnfd, by(largeland) stat(mean sd) format(%11.3f) c(s)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Full sample | | Female participants | | Households without female participants | |
|  | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation |
| Per capita expenditure | 5473.268 | 4140.221 | 5439.074 | 4118.625 | 5511.369 | 4167.686 |
| Per capita food expenditure | 3660.191 | 1558.638 | 3627.243 | 1371.990 | 3696.901 | 1743.831 |
| Per capita nonfood expenditure | 1813.078 | 3316.891 | 1811.831 | 3397.786 | 1814.467 | 3227.547 |

* STATA code：

tabstat exptot expfd expnfd, by(dfmfd) stat(mean sd) format(%11.3f) c(s)

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

The expenditure that spend on food are larger than nonfood. The female-headed households have larger per capita expenditure on food and nonfood than the male-headed households. The heads of households with some education have larger per capita expenditure on food and nonfood than the heads with no education. The small size households have larger per capita expenditure on food and nonfood than the large size households. The large landholding households have larger per capita expenditure on food and nonfood than the small landholding households.

Overall, the female, education level, land ownership may positively correlate with the per capita expenditure. And the households size may negatively correlate with the per capita expenditure.

The households without female participants have slightly larger per capita expenditure than the households with female participants. It may because that the households with female participants are usually those poorer ones with less assets and land ownerships.

**E. Statistical Analysis**





From the left graphs, we confirm that most expenditure were spending on food. But we cannot make difference between the female participants and nonparticipants sample. Further tests needed.



* STATA code：

histogram exptot if exptot <15000, kdensity xscale(range(0 15000)) by(, title(Per capita expenditure by female participants)) by(dfmfd, total)

histogram expfd, kdensity xscale(range(0 15000)) by(, title(Per capita food expenditure by female participants)) by(dfmfd, total)

histogram expnfd if expnfd <15000, kdensity xscale(range(0 15000)) by(, title(Per capita nonfood expenditure by female participants)) by(dfmfd, total)

* **comparison between two means**



Null hypothesis: the mean of per capita expenditure with female participants equals the mean of per capita expenditure without female participants

The p-value>0.05, we cannot reject the null.



Null hypothesis: the mean of per capita food expenditure with female participants equals the mean of per capita food expenditure without female participants

The p-value>0.05, we cannot reject the null.



Null hypothesis: the mean of per capita nonfood expenditure with female participants equals the mean of per capita food expenditure without female participants

The p-value>0.05, we cannot reject the null.

Overall, there are **NO** evidence that we can say the mean of per capita total/food/nonfood expenditure with female participants different from the mean without female participants. These results reject my initial analysis in last section D) that the mean is different.

* **goodness of fit test**



The Pearson's chi-squared

Null Hypothesis: female participants and household size are independent.

The p-value>0.05, we cannot reject the null.



The Pearson's chi-squared

Null Hypothesis: female participants and land ownership are independent.

The p-value<0.05, we can reject the null. female participants and land ownership are **NOT** independent.



The Pearson's chi-squared

Null Hypothesis: female participants and education level are independent.

The p-value<0.05, we can reject the null. female participants and education level are **NOT** independent.

Overall, this confirm my initial guess that the female participants are correlated with the land ownership and education level.

* **Distribution test**

**Kolmogorov-Smirnov tests & Mann-Whitney test**



Null hypothesis: the distribution of per expenditure with female participants equals the distribution of per capita expenditure without female participants

The p-value>0.05, we cannot reject the null.



Null hypothesis: the participants and nonparticipants have the same distribution of per expenditure.

The p-value>0.05, we cannot reject the null.

Overall, both Kolmogorov-Smirnov tests and Mann-Whitney test give us the same results. We can say the distribution of per expenditure with female participants equals the distribution of per capita expenditure without female participants.