## **Assignment weeks 1 and 2 11.5%**

## *To answer all the questions below, you must use Stata (and, specifically, DASP, if requested). Be concise and clear in your answers.*

## *The assignment is divided into three exercises (the points assigned to each exercise are indicated next to each exercise). Please answer (A) directly in this file after each question (Q) and please attach the \*.do file (do-file) that you generated. Rename both files as: “Assignment weeks 1-2 - Name, Surname”. Please submit this completed file and the \*.do through the virtual drop box (boîte de dépôt) in the course portal, no later than Tuesday, February 2 11:59 p.m. (*[*Québec time*](https://www.timeanddate.com/worldclock/converter.html?iso=20210203T045900&p1=189)*).*

## **Exercise 1 (4%)------4%**

Assume that the population is of composed 12 households that live in regions *A, B and C*.

|  |  |  |  |
| --- | --- | --- | --- |
| *identifier* | *region* | *income* | *hhsize* |
| 1 | A | 210 | 4 |
| 2 | A | 450 | 6 |
| 3 | A | 300 | 5 |
| 4 | A | 210 | 3 |
| 5 | B | 560 | 2 |
| 6 | B | 400 | 4 |
| 7 | C | 140 | 4 |
| 8 | C | 250 | 2 |
| 9 | C | 340 | 2 |
| 10 | C | 220 | 2 |
| 11 | C | 360 | 3 |
| 12 | C | 338 | 3 |

**Q 1.1:** Using Stata, generate per capita income (*pcinc*).

**A: gen pcinc = income/hhsize**

**Q 1.2:** Using Stata, estimate the average per capita income and the total incomes of our population.

**A: Average per capita income = 94.45; total income of the population = 3,778**

**Q 1.3:** Assume that, the poverty line is equal to 100, generate the variable “per capita poverty gap (*pgap*)”, and then estimate its average (the per capita poverty gap should be normalized by the poverty line).

**A: Average *pgap* = 0.2225**

**Q 1.4:** Redo question Q 1.3 using DASP.

**A: 0.2225**

**Q 1.5:** Assume that the purchasing power in region B is higher than that of region A by 10% and that of region C is higher than that of region A by 30%. In the case where the region A is the region of reference, generate the variable (deflator) as a price deflator index, and then generate the variable real per capita income (r*pcinc*).

**A: 82.76049**

**Q 1.6:** Redo the question 1.3 and 1.4 using the real per capita income when the poverty line is 120.

**A: *pgap = 0.370877***

**Exercise 2 (3%) 2.5%**

* 1. Using the file data\_1, estimate the average per adult equivalent expenditures without using the sampling weight and by using the DASP command **imean**. What does this statistic refer to?

**A: imean = 42048.74**

****

**The statistic refers to mean per capita expenditure adjusted for individual differences, for example, in age, sex, geographical zones, household sizes, etc., amounts to 42,048.74 currency unit**

* 1. Assume different cases for initialising the sampling design
* CASE1: Only by using the variable *strata* to initialise the stratification variable of the sampled population.
* CASE2: Only by using the variable *psu* to initialise the primary sampling unit variable.
* CASE3: By using the variable *strata* and *psu.*
* CASE4: By using the variable *strata, psu* and the sampling weight variable*.*

For each of these four cases, estimate the average per adult equivalent expenditures and give some explanation on the level of the standard errors compared to that of the question 1.1 and to those of the other cases.

**A:**

**Case 1: svyset strata**

****

**Case 2: svyset psu**

****

**Case 3:**

****

**Case 4:**

****

Note: It seems the mean is the same in the first 3 cases but the mean changes when sampling weight is taken into considerations. The STE for case for, on the other hand, seems to be different from the first 3 cases

* 1. Test whether the average per adult equivalent expenditures in region 1 is higher than the double of that of region 3. Briefly discuss the result.

A: Done using DASP **dimean** command

**gen region\_1=0**

**replace region\_1= ae\_exp if region==1**

**generate region\_3 = 0**

**replace region\_3 = ae\_exp if region==3**

**dimean region\_1 region\_3**

****

Note: mean\_D1 = region 1; Mean\_D2 = region 2

The average per adult equivalent expenditures in region 1 is higher than average per adult equivalent expenditures in region 1 and the difference is statistically significant

* 1. Using the DASP command ***dimean*** test whether the average per adult equivalent expenditures for male household heads is higher than that of female households headed. Briefly discuss the result.

**A:**

**generate mae\_exp =ae\_exp (*mae\_exp =ae\_exp for male headed households)***

**replace mae\_exp = 0 if sex==2**

**generate fae\_exp = ae\_exp (*fae\_exp =ae\_exp for female headed households)***

**replace fae\_exp = 0 if sex==1**

**dimean mae\_exp fae\_exp**

****

\* mean\_D1 = mean for male-headed households

mean\_D2 = mean for female-headed households

The average per adult equivalent expenditures for male-headed household is higher than the average per adult equivalent expenditures for female headed households and the difference is statistically significant

**Exercise 3 (5.5%)\*\*\*\*\*\*\*\*\*\*\*\*\*5%**

**Q 3.1** Use the data\_1.dta data file, and then compute the population size of the sampled households.

**A: total(hhsize) = 14,694**

**Q 3.2** Rank the per capita expenditures in ascending order and then generate the variable population share (*ps*) that includes the proportion of the sampled population with corresponding per capita expenditures. Based on this, generate the variable percentiles (*p*) and quantiles (*q*).

**A:**

**sort pcexp**

**sum hhsize**

**gen ps = hhsize/ r(sum)**

**gen p = sum(ps)**

**gen q = pcexp**

**list, sep(0)**

**Q 3.3** Draw the cumulative distribution curve (X-Axis: the corresponding per capita expenditures and Y-Axis: the percentiles) (range of percentiles: min=0 and max=0.95).

**A: twoway (line p pcexp if 0<= p <=95), ytitle(The percentiles) xtitle(Per capita expenditures) title(Cumulative distribution curve)**

****

Almost 95% of the sample population have per capita expenditure of less than 2000000. It also suggests that there are some outliers at the top of distribution.

**Q 3.4** Plot the quantile curve (X-axis: percentiles and Y-axis: quantiles) (range of percentiles: min=0 and max=0.95), and briefly discuss the results.

**A: line q p , title(The Quantile Curve) xtitle(the percentiles (p) if 0<= p <=95) ytitle(The Quantiles Q(p))**



The curve shows the level of per capita expenditure that corresponds to the percentile (p). The curve seems to suggest that the maximum expenditure of the bottom 95% of the population is less than 2000000.

It also suggests that there are some outliers at the top of distribution.

**Q 3.5** Using DASP, draw the quantile curve for each of the rural and urban regions (range of percentiles: min=0 and max=0.95), and briefly discuss the results.

**A:**

**c\_quantile pcexp, hgroup(zone) min(0) max(.95)**



Interpretation: The urban per capita expenditure distribution seems to be more equal than the rural per capita expenditure distribution.

**Q 3.6** Using DASP, draw the density curves of the per capita expenditures by the sex of the household head (range of per capita expenditures: min=0 and max=1000000) and briefly discuss the results

**A:**

**cdensity pcexp, hgroup(sex) min(0) max(1000000)**



Both distributions are not symmetric, and they are skewed to the right. The distribution of per capita expenditure by females is less skewed than that of males.