## **Assignment weeks 1 and 2**

## *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 1-2-3 - 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 5 11:59 a.m. (*[*Québec time*](https://www.timeanddate.com/worldclock/converter.html?iso=20190205T165900&p1=189)*).*

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

Assume that the population is of composed 10 households that live in regions *A, B and C*.

|  |  |  |  |
| --- | --- | --- | --- |
| *identifier* | *region* | *income* | *hhsize* |
| 1 | A | 310 | 4 |
| 2 | A | 460 | 6 |
| 3 | A | 300 | 5 |
| 4 | A | 220 | 3 |
| 5 | B | 560 | 2 |
| 6 | B | 400 | 4 |
| 7 | C | 140 | 3 |
| 8 | C | 250 | 2 |
| 9 | C | 340 | 2 |
| 10 | C | 220 | 2 |

**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 income of our population.

**A: /\* Estimating the average per capita income \*/**

**sum pcinc [aw=hhsize]**

**/\*Estimating the total incomes \*/**

**total income**

**Q 1.3:** Assume that, the poverty line is equal to 120, 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: /\*generating a new variable per capita poverty gap(pgap)assuming poverty line is equal to 120\*/**

**gen pline = 120**

**gen pgap = 0**

**replace pgap = (pline-pcinc)/pline if (pcinc < pline)**

**/\*estimating the average poverty gap\*/**

**sum pgap [aw=hhsize]**

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

**A: ifgt pcinc, pline(120) alpha(1) hsize(hhsize)**

**Q 1.5:** Assume that the purchasing power in region B is higher than that of region A by 20% and that of region C is higher than that of region A by 40%. 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: /\*generating a deflator\*/**

**gen deflator = 1**

**replace deflator = 0.80 if region == 2**

**replace deflator = 0.60 if region == 3**

**/\*generating a variable real per capita income(rpinc)\*/**

**gen rpcinc = pcinc/deflator**

**Q 1.6:** Redo the question 1.3 and 1.4 using the real per capita income when the poverty line is 110.

**A: sum rpcinc [aw=hhsize]**

**replace pline = 110**

**replace pgap = (pline-rpcinc)/pline if (rpcinc < pline)**

**sum pgap [aw=hhsize]**

**ifgt rpcinc, pline(110) alpha(1) hsize(hhsize)**

**Exercise 2 (3%)**

* 1. Using the file data\_2, 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: sum ae\_exp [aw=hhsize]**

**imean ae\_exp, hsize(hhsize) /\*The statistic is the average household welafe taking into account or adujusting for individual differences such as age and gender and also the economies of scale \*/**

* 1. By using the variables ***strata, psu*** and the sampling weight variable, initialise the sampling design, and then estimate the average per adult equivalent expenditure.

**A: /\*Initializing sampling design\*/**

**svyset psu [pweight=sweight], strata(strata) vce(linearized) singleunit(missing)**

**svydes**

**gen fweight = sweight\*hhsize**

**/\*Estimating average per adult equivalent expenditure\*/**

**sum ae\_exp [aw=fweight]**

* 1. Test whether the average per adult equivalent expenditure in region 1 is higher than the double of that of region 3.

**A: /\*Estimating average expenditure by region\*/**

**imean ae\_exp, hsize(hhsize) hgroup(region)**

**/\*Doubling the estimated expenditure in region3\*/**

**di 2\*21073.082031**

**/\*Testing if 50474.214844 is greater than 42146.164\*/**

**datest 42146.164, est(50474.214844) ste(2973.246582)**

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

**A: /\*Testing whether AE expenditure is higher for male than the female\*/**

**dimean ae\_exp ae\_exp, hsize1(hhsize) test(0) cond1(sex==1 ) hsize2(hhsize) cond2(sex==2 )**

**/\*Brief discussion: This dasp command estimates three test. 1 whether the differences between means of males and females is lower that zero 2. the difference between means of men and females is equal to 0 and lastly the difference betwween means of males and females is greater than zero. Our interest is on the the estimates 3 that is to see if the diffrence between means of men is greater that the means for women. From the results, we see that at P-value of 0.3672 we fail to reject the null hypothesis that the difference between males and females means is greater than 0.\*/**

### Exercise 3 (5.5%)

**Q 3.1** Use the data\_2.dta data file, and then compute the population size of the sampled households.

**A: count**

**sum psu**

**/\*both these show the number of the observations to be 2,000\*/**

**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 population with corresponding per capita expenditures. Based on this, generate the variables percentiles (*p*) and quantiles (*q*).

**A: gen fweight = sweight\*hhsize**

**// Step 1- Ranking percapica expenditure in ascending order**

**sort pcexp**

**list pcexp**

**// step 2- generating populations share (ps)**

**sum pcexp [aw=fweight]**

**gen ps = fweight\*pcexp**

**sum pcexp [aw=fweight]**

**replace ps = ps/r(sum)**

**// Step 3- generating the variable percentile and the quantiles \*/**

**gen p = sum(ps)**

**gen q = pcexp**

**list, sep(0)**

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

**A: line p pcexp, title(The cumulative distribution curve) xtitle(The per capita expenditure (y)) ytitle(F(y))**



**Q 3.4** Plot the quantile curve (X-axis percentiles (0 to 0.95) and Y-axis quantiles), and briefly discuss the results.

**A: line q p , title(The quantile curve) xtitle(the percentile (p)) ytitle(The quantile Q(p))**



**This quantile curve is showing different the level of welfare achieved by people in the economy as such from the curve it shows that 90% of the economy spends less 5,000,000. This is an inverse to the percentile graph**

**Q 3.5** Using DASP, draw the quantile curves by the sex of the household head (percentiles (0 to 0.95)), and briefly discuss the results.

**A: db c\_quantile**

**c\_quantile pcexp, hsize(hhsize) min(0.0) max(0.95)**



**From the graph above, the expenditures of those in urban is greater than those in rural implying that those in urban have higher welfare as expenditure us used as a proxy for welfare**

**Q 3.6** Using DASP, draw the density curves of the per capita expenditures for each of the rural and urban regions (range of per capita expenditures: min=0 and max=1000000), and briefly discuss the results.

**A: db cdensity**

**cdensity pcexp, hsize(hhsize) hgroup(region) popb(1) type(den) min(0) max(1000000)**



First in this economy, both rural and urban consumption is low, but those in urban have better living standards than those in rural.