## **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 - 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%)**

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: sum pcinc [aw=hhsize]**

**total income**

**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: gen pline=100**

**gen pgap=0**

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

**sum pgap [aw=hhsize]**

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

**A: ifgt pcinc, pline(100) alpha(1) hsize(hhsize)**

**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: gen deflator=1**

**replace deflator=0.9 if region==2**

**replace deflator=0.7 if region==3**

**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 120.

**A: replace pline=120**

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

**sum pgap [aw=hhsize]**

**ifgt rpcinc, pline(120) alpha(1) hsize(hhsize)**

**Exercise 2 (3%)**

* 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 ae\_exp, hsize(hhsize) // ste=1701.506958**

* 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: Only by using the variable strata to initialise the stratification variable of the sampled population \*/**

**svyset, strata(strata)**

**imean ae\_exp, hsize(hhsize) // the ste (1702.926636) is larger than that of the question 1.1, case 2 and case 3. However it is lower than that of case 4.**

**svyset, clear**

**/\* Case 2: Only by using the variable psu to initialise the primary sampling unit variable \*/**

**svyset psu**

**imean ae\_exp, hsize(hhsize) // the ste(1693.014282) is the lowest one.**

**svyset, clear**

**/\* Case 3: Only by using the variable strata and psu \*/**

**svyset psu, strata(strata)**

**imean ae\_exp, hsize(hhsize) // the ste(1699.352783) is larger than that of case 2 and lower than that of all the remaning cases including that of question 1.1.**

**svyset, clear**

**/\* Case 4: Only by using the variable strata, psu and the sampling weight variable \*/**

**svyset psu [pweight=sweight], strata(strata)**

**imean ae\_exp, hsize(hhsize) // the ste(2213.284668) is the largest one.**

* 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: imean ae\_exp, hsize(hhsize) hgroup(region) // mean\_1=59713.667969 and mean\_3=22984.812500, and ste=sqrt(ste\_1^2+4\*ste\_3^2)**

**datest 0, est(13744.04297) ste(6758.71740)**

**// We cannot reject H0: mean\_1-2\*mean\_3 > 0. This is because the statistical error that we make if we reject H0 = 97.90% is greater than the critical level of 5.00%.**

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

**// We cannot reject H0: difference = > 0. This is because the statistical error that we make if we reject H0 = 99.94% is greater than the critical level of 5.00%.**

**Exercise 3 (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 [pw=sweight] // The 2,000 sampled households represent a population size of 5,578,466**

**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: /\* sorting the data by the per capita expenditures \*/**

**sort pcexp**

**/\* generating the variable of popultion share (ps) \*/**

**sum hhsize**

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

**/\* generating the variable percentiles and the quantiles \*/**

**gen p = sum(ps)**

**gen q = pcexp**

**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: line p pcexp, title(The cumulative distribution curve) xtitle(The per capita expenditures (y)) ytitle(F(y)) yscale(range(0 .95))**

**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: c\_quantile pcexp, hsize(hhsize) min(0) max(0.95)**

**// 95% of households have a total expenditures of less than 400,000**

**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, hsize(hhsize) hgroup(zone) min(0) max(0.95)**

**// The average per capita expenditure in urban area is greater than that in rural area**

**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: c\_quantile pcexp, hsize(hhsize) hgroup(sex) yscale(range(0 1000000))**

**// The average per capita expenditure for female headed households is greater than that of male headed households**