## **Exercices semaines 1 et 2**

## *Pour répondre à toutes les questions ci-dessous, vous devez utiliser Stata (et, spécifiquement, DASP, si demandé). Soyez concis(es) et clair(e)s dans vos réponses.*

## *L’examen est divisé en trois exercices (les points assignés à chaque exercice sont indiqués à côté de chaque exercice). Veuillez répondre (R) directement dans ce fichier après chaque question (Q) et veuillez joindre le fichier \*.do (do-file) que vous avez généré. Renommez ces deux fichiers en : "Exercice semaines 1-2 - Prénom, Nom" et veuillez les soumettre* *par la boîte de dépôt du portail de cours avant mardi le 2 février 23h59 (*[*heure du Québec*](https://www.timeanddate.com/worldclock/converter.html?iso=20210203T045900&p1=189)*).*

## **Exercice 1 (4%)**

Supposons que la population est composée de 10 ménages vivant dans les régions *A, B et 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:** À l'aide de Stata, générez le revenu par habitant (*pcinc*).

**R : On insère d’abord les données du tableau dans stata,**

**ensuite on génère la variable le revenu/tête *pcinc*  (voir dernière colonne du tableau ci-dessous:**

**gen pcinc=income/hhsize**

**hhid region income hhsize pcinc**

**1 1 310 4 77.5**

**2 1 460 6 76.66666**

**3 1 300 5 60**

**4 1 220 3 73.33334**

**5 2 560 2 280**

**6 2 400 4 100**

**7 3 140 3 46.66667**

**8 3 250 2 125**

**9 3 340 2 170**

**10 3 220 2 110**

**Q 1.2:** À l'aide de Stata, estimez le revenu moyen par habitant et le revenu total de notre population.

**sum pcinc [aw=hhsize]**

**R : **

**Le revenu moyen par habitant= 96.9697**

**Le revenu total de notre population=revenu moyen par habitant\* nombre de ménages\* poids=96.9697\*10\*33**

**= 3200**

**Ou dans stata avec la commande *di r(sum)* on trouve 3200**

**ou la commande *gen revtot=sum(income)* et on retient la dernière valeur cumulée de la colonne contenant la vaiable, soit 3200 (voir dernière colonne du tableau ci-dessous)**

**hhid income revtot**

**1 310 310**

**2 460 770**

**3 300 1070**

**4 220 1290**

**5 560 1850**

**6 400 2250**

**7 140 2390**

**8 250 2640**

**9 340 2980**

**10 220 3200**

**Q 1.3:** En supposant que le seuil de pauvreté est égal à 120, générez la variable intensité de la pauvreté par habitant (*pgap*), puis estimez sa moyenne (l'intensité de la pauvreté par habitant doit être normalisée par le seuil de pauvreté).

**R : la variable intensité de la pauvreté par habitant (*pgap*) generée *(dernière colonne du tableau ci dessous)*:**

**gen pline=120**

**gen pgap=0**

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

**hhid pcinc pline pgap**

**1 77.5 120 .3541667**

**2 76.66666 120 .3611111**

**3 60 120 .5**

**4 73.33334 120 .3888889**

**5 280 120 0**

**6 100 120 .1666667**

**7 46.66667 120 .6111111**

**8 125 120 0**

**9 170 120 0**

**10 110 120 .0833333**

**L’intensité de la pauvreté moyenne par habitant**

**sum pgap [aw=hhsize]**

****

**L’intensité de la pauvreté moyenne par habitant=0.3005051**

**Q 1.4:** Refaites la question Q 1.3 en utilisant DASP.

**R :**

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

****

**Avec DASP, on obtient la même moyenne, soit 0.3005051**

**Q 1.5:** Supposons que le pouvoir d'achat dans la région B soit supérieur de 20% à celui de la région A et que celui de la région C soit supérieur de 40% à celui de la région A. Dans le cas où la région A est la région de référence, générez la variable (*deflator*) en tant qu'indice de déflation des prix, puis générez la variable revenu réel par habitant (*rpcinc*).

**R : \*\*generer la variable *deflator***

**gen deflator=1**

**replace deflator=0.8 if region==2**

**replace deflator=0.6 if region==3**

**\*\*generer la variable revenu reel par habitant rpcinc**

**gen rpcinc=pcinc/deflator**

**ce tableau indique la variable *deflator* pour chaque région ainsi que le revenu réel *rpcinc***

**hhid region deflator rpcinc**

**1 1 1 77.5**

**2 1 1 76.66666**

**3 1 1 60**

**4 1 1 73.33334**

**5 2 .8 350**

**6 2 .8 125**

**7 3 .6 77.77778**

**8 3 .6 208.3333**

**9 3 .6 283.3333**

**10 3 .6 183.3333**

**Q 1.6:** Refaites les questions 1.3 et 1.4 en utilisant le revenu réel par habitant lorsque le seuil de pauvreté est de 110.

**R : \* refaire les questions 1.3 et 1.4 en utilisant le revenu réel par habitant lorsque le pauvreté est de 110**

**replace pline=110**

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

**hhid pline pgap rpcinc**

**1 110 .2954545 77.5**

**2 110 .3030303 76.66666**

**3 110 .4545455 60**

**4 110 .3333333 73.33334**

**5 110 0 350**

**6 110 .1666667 125**

**7 110 .2929293 77.77778**

**8 110 0 208.3333**

**9 110 0 283.3333**

**10 110 .0833333 183.3333**

**sum pgap [aw=hhsize]**

****

**L’intensité de la pauvreté moyenne par habitant devient 0.2418651**

**Avec DASP on obtient :**

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

****

**L’intensité de la pauvreté moyenne par habitant est 0.2166713**

**Exercice 2 (3%)**

* 1. À l'aide du fichier data\_2, estimez les dépenses moyennes par équivalent adulte sans utiliser le poids de sondage et en utilisant la commande DASP ***imean***. À quoi cette statistique réfère-t-elle ?

**R : imean ae\_exp, hsize (hhsize)**



**Cette statistique sert à estimer l’écart type des dépenses moyennes.**

* 1. En utilisant les variables ***strata***, ***psu*** et la variable de poids de sondage, initialisez le plan d'échantillonnage, puis estimez la dépense moyenne par équivalent adulte.

**R :**

* 1. Vérifiez si la dépense moyenne par équivalent adulte dans la région 1 est supérieure au double de celle de la région 3.

**R :**

* 1. En utilisant la commande DASP ***dimean,*** vérifiez si la dépense moyenne par équivalent adulte pour les chefs de ménage hommes est plus élevée que celle des ménages dirigés par des femmes. Discutez brièvement vos résultats.

### Exercice 3 (5.5%)

**Q 3.1** Utilisez le fichier de données data\_2.dta, puis calculez la taille de la population des ménages échantillonnés.

**R :**

**Q 3.2** Ordonnez les dépenses par habitant en ordre croissant et générez ensuite la variable part de population (*ps*) qui comprend la proportion de la population avec les dépenses par habitant correspondantes. Sur cette base, générer les variables centiles (*p*) et quantiles (*q*).

**R : \*ordonner les dépenses par habitant en ordre croissant**

**sort pcexp**

**\*** **générez la variable part de population (ps)**

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

**\*** **générer les variables centiles (p) et quantiles (q)**

**gen p=sum( ps)**

**gen q= pcexp**

**Q 3.3** Dessinez la courbe de distribution cumulative (Axe X: les centiles et axe Y: les dépenses par habitant correspondantes) (domaine de centiles: min = 0 et max = 0,95).

**R : \*** **Dessinez la courbe de distribution cumulative**

**line p pcexp, title(la courbe de distribution cumulative) xtitle(dépense par habitant) ytitle(F(Y))**



**Q 3.4** Tracez la courbe quantile (centiles sur l'axe X (0 à 0,95) et quantiles sur l'axe Y), et discutez brièvement les résultats.

**R : Tracez la courbe quantile**

**line q p, title(la courbe quantile) xtitle(centile (p)) ytitle(quantile (Q(p)))**



**Discussion : on voit bien que les 2 courbes sont inverses.**

**Q 3.5** En utilisant DASP, dessinez les courbes quantiles selon le sexe de la tête du ménage (centiles (0 à 0,95)), et discutez brièvement les résultats.

**R : Avec DASP, dessinez les courbes quantiles selon le sexe de la tête du ménage**

**\*Pour sexe masculin**



**Pour sexe Feminin**



DISCUSSION : chez les ménages dont le chef est un homme, les dépenses totales sont plus élevées que les ménages de sexe féminin. Cependant, ces dernières ont une distribution des dépenses plus égalitaire que chez les ménages de sexe masculin.

**Q 3.6** À l'aide du DASP, dessinez les courbes de densité des dépenses par habitant pour chacune des régions rurales et urbaines (domaine des dépenses par habitant : min = 0 et maximum = 1000000), et discutez brièvement des résultats.

**R : ZONE RURALE**



ZONE URBAINE



Discussion : en zone urbaine, la courbe de densité tend vers une loi normale que celle de la zone rurale.