## **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 12 ménages qui vivent dans les régions *A, B et C*.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *identifier* | *region* | *income* | *hhsize* | *pcinc*  *(Q1.1)* | *Pgap*  *(Q1.3)* | *Deflator*  *(Q1.5)* | *Rpcinc*  *(Q1.5)* | *pgap (Q1.6)* |
| 1 | A | 210 | 4 | 52.5 | .475 | 1 | 52.5 | .5625 |
| 2 | A | 450 | 6 | 75 | .25 | 1 | 75 | .375 |
| 3 | A | 300 | 5 | 60 | .4 | 1 | 60 | .5 |
| 4 | A | 210 | 3 | 70 | .3 | 1 | 70 | .4166667 |
| 5 | B | 560 | 2 | 280 | 0 | .9 | 311.1111 | 0 |
| 6 | B | 400 | 4 | 100 | 0 | .9 | 111.1111 | .074074 |
| 7 | C | 140 | 4 | 35 | .65 | .7 | 50 | .5833333 |
| 8 | C | 250 | 2 | 125 | 0 | .7 | 178.5714 | 0 |
| 9 | C | 340 | 2 | 170 | 0 | .7 | 242.8571 | 0 |
| 10 | C | 220 | 2 | 110 | 0 | .7 | 157.1429 | 0 |
| 11 | C | 360 | 3 | 120 | 0 | .7 | 171.4286 | 0 |
| 12 | C | 338 | 3 | 112.6667 | 0 | .7 | 160.9524 | 0 |

**Q 1.1:** À l’aide de Stata, générez le revenu par habitant (*pcinc*).

**R : Les valeurs de générées (*gen pcinc = income/hhsize* ) sont dans le tableau ci-dessus.**

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

**R : Revenu moyen par habitant = 94.45 ; Revenu total de la population = 3778.**

**Q 1.3:** Supposons que le seuil de pauvreté soit égal à 100. Générez la variable « intensité de la pauvreté par habitant (*pgap*) », puis estimez sa moyenne (l’intensité de la pauvreté par habitant devrait être normalisée par le seuil de pauvreté).

**R :**

* **Les valeurs de *pgap* générées sont dans le tableau ci-dessus.**
* **La moyenne de l'intensité de la pauvreté par habitant est estimée à 0,2225.**

**Q 1.4:** Refaites la question Q 1.3 avec DASP.

**R : Pour ce faire, j’ai utilisé la fonction *ifgt* (voir fichier .do). Le résultat est similaire à la question Q 1.3 : L’estimation de la moyenne de l'intensité de la pauvreté par habitant est 0,2225.**

**Q 1.5:** Supposons que le pouvoir d'achat dans la région B soit supérieur de 10% à celui de la région A et que celui de la région C soit supérieur de 30% 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 de revenu réel par habitant (rpcinc).

**R :**

* **A étant la région de référence, l’indice de déflation des prix pour cette région est 1 ;**
* **Le pouvoir d'achat dans la région B étant supérieur de 10% à celui de la région A alors l’indice de déflation des prix pour la région B est 0,9 (1-0,1) ;**
* **De même, l’indice de déflation des prix pour la région C est 0,7 (1-0,3) ;**
* **Les valeurs de *deflator* et *rpcinc* sont dans le tableau ci-dessus.**

**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 120.

**R : L’estimation de la moyenne de l'intensité de la pauvreté par habitant est 0,271991.**

**Exercice 2 (3%)**

* 1. À l'aide du fichier data\_1, estimez les dépenses moyennes par équivalent adulte sans utiliser le poids de sondage et en utilisant la commande DASP **imean**. À quoi réfère cette statistique?

**R :**

***Remarque : Pour importer le fichier*** *data\_1****, il faudra d’abord indiquer, dans le fichier .do, le répertoire (path) où se trouve le fichier*** *data\_1****.***

**Les dépenses moyennes sont égales à 42964.714844. Cette statistique réfère à l’estimation des dépenses moyennes par équivalent adulte basée sur un échantillonnage aléatoire simple (*simple-random sampling*).**

* 1. Supposez différents cas d'initialisation du plan d'échantillonnage
* CAS1: Seulement en utilisant la variable *strata* pour initialiser la variable de stratification de la population échantillonnée.
* CAS2 : Seulement en utilisant la variable *psu* pour initialiser la variable d'unité primaire d’échantillonnage (primary sampling unit, PSU).
* CAS3: En utilisant la variable *strata* et *psu.*
* CAS4: En utilisant la variable *strata, psu* et la variable de poids de sondage*.*

Pour chacun de ces quatre cas, estimez les dépenses moyennes par équivalent adulte et donnez quelques explications sur le niveau des erreurs-types par rapport à celui de la question 1.1 et à ceux des autres cas.

**R : Pour chacun des cas 1, 2 et 3, les dépenses moyennes ont la même valeur : 42964.714844. Dans le cas 4 les dépenses moyennes sont égales à 41993.101562. La différence est de moyenne est due à l’utilisation des poids de sondage.**

**Les erreurs-types sont différents :**

* **Dans la question 1.1, l’écart type est 1701.506958. Comme le *design* de l’échantillonnage n’a pas été initialisé, un échantillonnage aléatoire simple (*simple-random sampling*) a été appliqué.**
* **Dans le cas 1, l’écart type est 1702.926636. Bien qu’il y ait la stratification, il y a un regroupement dans les strates 8 et 1. En effet, 24.5% des PSUs sont regroupés dans la strate 8 et 16.65% dans la strate 1. Ce qui réduit la précision engendrée par stratification.**
* **Dans le cas 2, l’écart type est 1693.014282. L’erreur type est moins élevé que dans la question 1.1 car le regroupement est moins significatif.**
* **Dans le cas 3, l’écart type est 1699.352783. L’erreur type est légèrement inférieur celui dans la question 1.1 à cause du regroupement déjà énoncé. En effet si la distribution dans les strates étaient moins concentrée, l’on aurait eu plus de précision et donc un écart type moins élevé.**
* **Dans le cas 4, l’écart type est 2213.284668. Cela est dû à l’utilisation des poids de sondages qui accentue l’effet de regroupement et donc diminue la précision.**
  1. Vérifiez si les dépenses moyennes par équivalent adulte dans la région 1 sont supérieures au double de celles de la région 3. Discutez brièvement ce résultat.

**R : L’hypothèse nulle est que la différence entre les dépenses moyennes de la région 1 et le double des dépenses moyennes de la région 3 est supérieure à zéro.**

**L’erreur statistique que nous ferions si l’on rejette l’hypothèse nulle est 97,91% (0.9791). Elle est supérieure à 5%. Ainsi, à un seuil de confiance de 5%, on ne peut pas rejeter l’hypothèse nulle. En d’autres termes, les dépenses moyennes de la région 1 sont supérieures au double des dépenses moyennes de la région 3.**

* 1. À l'aide de la commande DASP ***dimean***, évaluez si les dépenses moyennes par équivalent adulte pour les chefs de famille hommes sont plus élevées que celles des femmes chefs de famille. Discutez brièvement ce résultat.

**R : L’hypothèse nulle est que les dépenses moyennes par équivalent adulte pour les chefs de famille hommes (*sex=1*) sont plus élevées que celles des femmes chefs de famille (*sex=2*).**

**L’erreur statistique que nous ferions si l’on rejette l’hypothèse nulle est 99,94% (0.9994) ; elle est donc supérieure à 5%. Ainsi, à un seuil de confiance de 5%, on ne peut pas rejeter l’hypothèse nulle. En d’autres termes, les dépenses moyennes par équivalent adulte pour les chefs de famille hommes sont plus élevées que celles des femmes chefs de famille.**

### Exercice 3 (5.5%)

**Q 3.1** Utilisez le fichier de données data\_1.dta, puis calculez la taille de la population des ménages échantillonnés.

**R : La taille de la population des ménages échantillonnés est 2000.**

**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 échantillonnée avec les dépenses par habitant correspondantes. Sur cette base, générez les variables centiles (*p*) et quantiles (*q*).

**R :**

* **Ordonner les dépenses par habitant en ordre croissant :** **sort pcexp**
* **Générer la variable ps: sum hhsize**

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

* **Générer la variable p : gen p = sum(ps)**
* **Générer la variable q : gen q = pcexp**

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

**R : Notons qu’en principe, la courbe de distribution cumulative existe sur le domaine des centiles (min=0, max=1). La représentation ci-dessous a été tronquée sur le domaine des centiles (min=0, max=0.95) comme demandé dans la question.**

****

**Q 3.4** Tracez la courbe des quantiles (Axe X: centiles et axe Y: quantiles) (domaine des centiles: min = 0 et max = 0,95), et commentez brièvement les résultats.

**R : Notons qu’en principe, la courbe des quantiles existe sur le domaine des centiles (min=0, max=1). La représentation ci-dessous a été tronquée sur le domaine des centiles (min=0, max=0.95) comme demandé dans la question.**



**Q 3.5** En utilisant DASP, dessinez la courbe des quantiles pour chacune des régions rurales et urbaines (domaine des centiles : min = 0 et max = 0,95), et discutez brièvement des résultats.

**R : Les dépenses par habitant sont plus élevées en zone urbaine qu’en zone rurale. Ce qui pourrait être interpréter comme le fait que le niveau de vie est plus élevé en zone urbaine. Aussi, l’écart entre les dépenses des premiers centiles et des derniers centiles est plus élevé en zone urbaine (pente de la courbe plus élevée); ce qui montre qu’en zone urbaine la disparité entre les dépenses de la population est plus élevées.**

****

**Q 3.6** À l'aide de DASP, dessinez les courbes de densité des dépenses par habitant en fonction du sexe du chef de ménage (domaine des dépenses par habitant: min = 0 et maximum = 1000000) et discuter brièvement des résultats.

**R : Les ménages qui font les petites dépenses (*approx*imativement <150000) sont majoritairement les ménages dont le chef de famille est un homme. Les ménages qui font des dépenses entre (*approximativement)* 150000 et 450000 sont majoritairement les ménages dont le chef de famille est une femme. Les ménages faisant les dépenses les plus élevées sont (très) légèrement dominés par les ménages dont le chef de famille est une femme.**

**En somme, les ménages dont le chef de famille est un homme font plus de petites dépenses et moins de grandes dépenses comparativement aux ménages dont le chef de famille est une femme.**

****