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

**R :**

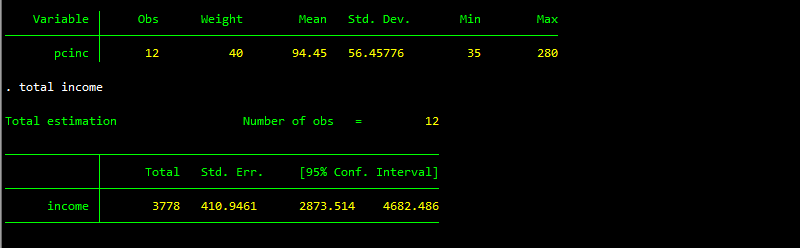
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **identifier** | **region** | **income** | **hhsize** | **pcinc** |
| **1** | **1** | **210** | **4** | **52.5** |
| **2** | **1** | **450** | **6** | **75** |
| **3** | **1** | **300** | **5** | **60** |
| **4** | **1** | **210** | **3** | **70** |
| **5** | **2** | **560** | **2** | **280** |
| **6** | **2** | **400** | **4** | **100** |
| **7** | **3** | **140** | **4** | **35** |
| **8** | **3** | **250** | **2** | **125** |
| **9** | **3** | **340** | **2** | **170** |
| **10** | **3** | **220** | **2** | **110** |
| **11** | **3** | **360** | **3** | **120** |
| **12** | **3** | **338** | **3** | **112.6667** |

**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 : **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 :**

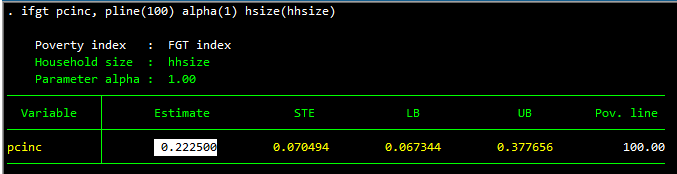
variable « intensité de la pauvreté par habitant (*pgap*) »

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **identifier** | **region** | **income** | **hhsize** | **pcinc** | **pline** | **pgap** |
| **1** | **1** | **210** | **4** | **52.5** | **100** | **.475** |
| **2** | **1** | **450** | **6** | **75** | **100** | **.25** |
| **3** | **1** | **300** | **5** | **60** | **100** | **.4** |
| **4** | **1** | **210** | **3** | **70** | **100** | **.3** |
| **5** | **2** | **560** | **2** | **280** | **100** | **0** |
| **6** | **2** | **400** | **4** | **100** | **100** | **0** |
| **7** | **3** | **140** | **4** | **35** | **100** | **.65** |
| **8** | **3** | **250** | **2** | **125** | **100** | **0** |
| **9** | **3** | **340** | **2** | **170** | **100** | **0** |
| **10** | **3** | **220** | **2** | **110** | **100** | **0** |
| **11** | **3** | **360** | **3** | **120** | **100** | **0** |
| **12** | **3** | **338** | **3** | **112.6667** | **100** | **0** |

Moyenne de pgap **= 0.2225**

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

**R : 0.222500**



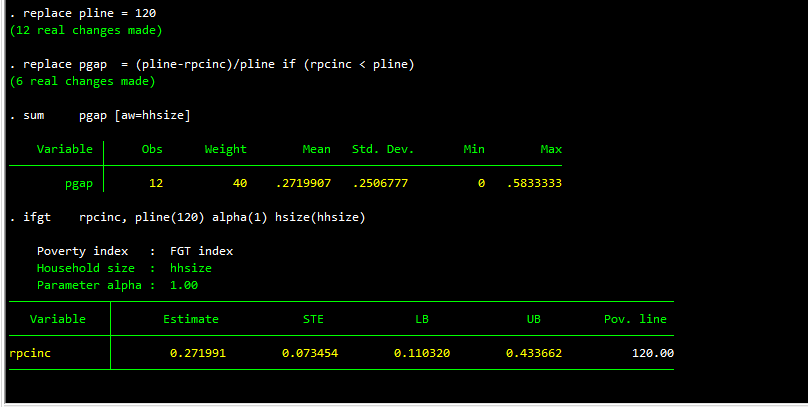
**Q 1.5:** Supposons que le pouvoir d'achat dans la région B soit supérieur de 10% à celui de la région A et que celui de la région C soit supérieur de 30% à celui de la région A. Dans le cas où la région A est la région de référence, générez la variable (deflator) en tant qu'indice de déflation des prix, puis générez la variable de revenu réel par habitant (rpcinc).

**R :**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| identifier | region | income | hhsize | pcinc | pline | pgap | deflator | rpcinc |
| 1 | 1 | 210 | 4 | 52.5 | 100 | .475 | 1 | 52.5 |
| 2 | 1 | 450 | 6 | 75 | 100 | .25 | 1 | 75 |
| 3 | 1 | 300 | 5 | 60 | 100 | .4 | 1 | 60 |
| 4 | 1 | 210 | 3 | 70 | 100 | .3 | 1 | 70 |
| 5 | 2 | 560 | 2 | 280 | 100 | 0 | .9 | 311.1111 |
| 6 | 2 | 400 | 4 | 100 | 100 | 0 | .9 | 111.1111 |
| 7 | 3 | 140 | 4 | 35 | 100 | .65 | .7 | 50 |
| 8 | 3 | 250 | 2 | 125 | 100 | 0 | .7 | 178.5714 |
| 9 | 3 | 340 | 2 | 170 | 100 | 0 | .7 | 242.8571 |
| 10 | 3 | 220 | 2 | 110 | 100 | 0 | .7 | 157.1429 |
| 11 | 3 | 360 | 3 | 120 | 100 | 0 | .7 | 171.4286 |
| 12 | 3 | 338 | 3 | 112.6667 | 100 | 0 | .7 | 160.9524 |

**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.

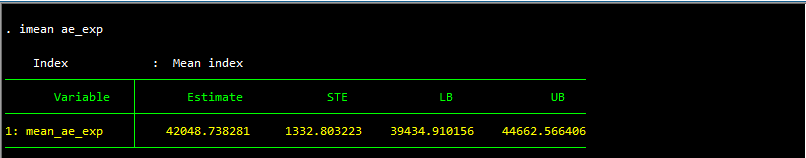
**R :**



**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?

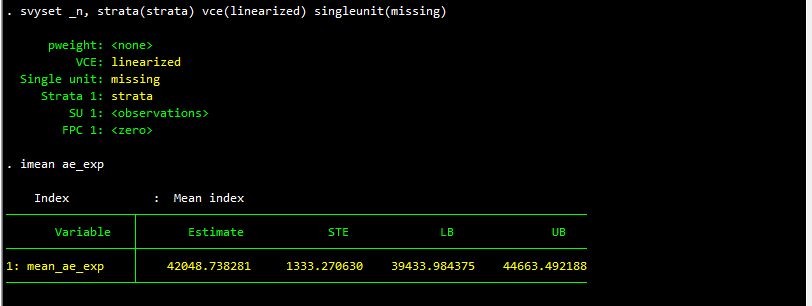
**R : 42048.738281**



* 1. Supposez différents cas d'initialisation du plan d'échantillonnage

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.

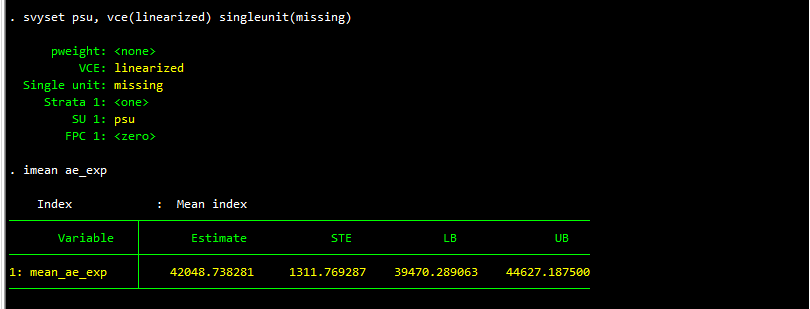
* CAS1: Seulement en utilisant la variable *strata* pour initialiser la variable de stratification de la population échantillonnée.



R : 42048.738281

**Commentaires :** La mesure de l’écart type est plus élevés comparativement au cas où l’on avait pas utilisé le plan de sondage et aux autres cas sauf lorqu’il a eu stratification. Il y a une plus grande variation autour des valeurs centrales. La stratification augmente la précision des statistiques estimées

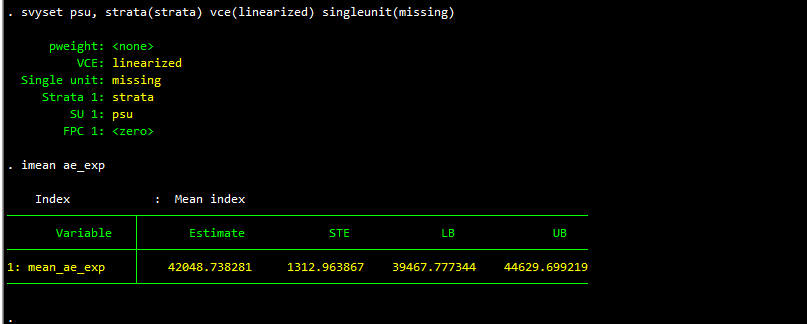
* CAS2 : Seulement en utilisant la variable *psu* pour initialiser la variable d'unité primaire d’échantillonnage (primary sampling unit, PSU).



R : 42048.738281

**Commentaires :** La mesure de l’écart type est plus élevée que pour tous les cas. Il y a une plus grande variation autour des valeurs centrales. Le regroupement diminue la précision des statistiques estimées

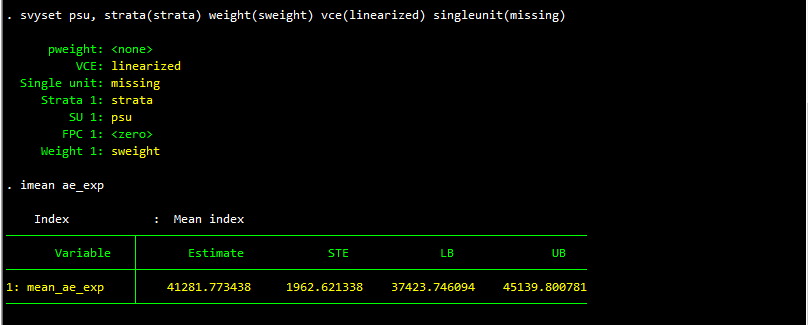
* CAS3: En utilisant la variable *strata* et *psu.*



R : 42048.738281

**Commentaires :** La mesure de l’écart type est la moins élevée que partout ailleurs.. Il y a une moins grande variation autour des valeurs centrales.

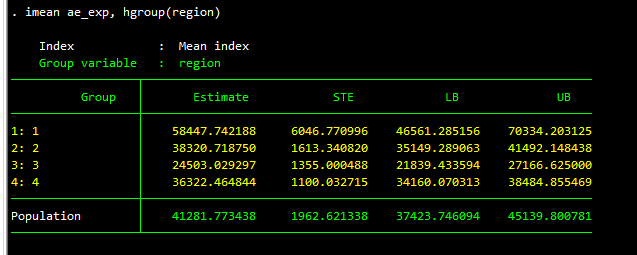
* CAS4: En utilisant la variable *strata, psu* et la variable de poids de sondage*.*



R : 41281.773438

**Commentaires :** L’utilisation du poids affecte sensiblement la moyenne. Elle est plus grande mais l’écart type aussi est élevé. Il y a une plus grande variation autour des valeurs centrales. En raison de l’augmentation de la moyenne, on peut conclure qu’il ya une dispersion de détenteurs de hauts revenus dans toutes les strates ou que les détenteurs de bas revenus sont concentrés dans certaines strates.

* 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.

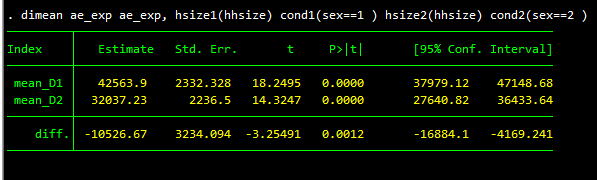


**R :** Les dépenses par équivalent adulte sont deux fois plus élevées dans la région 1 que dans la région 3 (58447.742188 > 2 x 24503.029297).. Ceci confirme notre interprétation précédente. Ceux qui font les plus grosses dépenses sont concentrés dans le premier groupe et les aautres dans les groupes inférieurs.

Les variations de revenus s’observent également dans les classes supérieurs.

* 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.

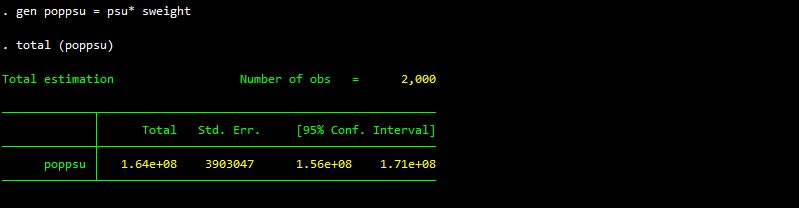
**R :**



**Commentaires :** Au seuil de 95% il existe une différence significative entre les ménages dirigés par les hommes et ceux dirigés par les femmes du point de vue de la moyenne des dépenses par équivalent adulte. Le P value (0.0012) est inférieur à 0,05.

### Exercice 3 (5.5%)

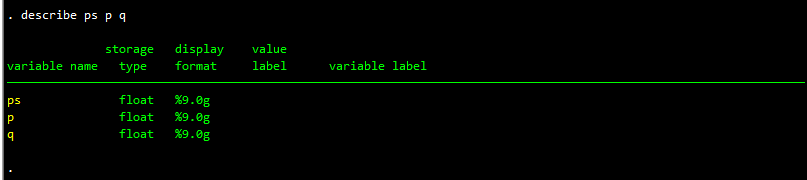
**Q 3.1** Utilisez le fichier de données data\_1.dta, puis calculez la taille de la population des ménages échantillonnés.



**R : 1.64e+08**

**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 :**



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

**R :**

****

**Commentaires : On constate une inégalité dans les dépenses par tête.**

**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 :**



Commentaires : Les dépenses des ménages situés dans les derniers percentiles sont élevées.

**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 :**



**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 :**

### 