## **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 1-2-3 - Prénom, Nom" et veuillez les* soumettre *par la boîte de dépôt du portail de cours avant mardi le 5 février 11h59 a.m. (*[*heure du Québec*](https://www.timeanddate.com/worldclock/converter.html?iso=20190205T165900&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*).**

**Réponse :** les résultats sont consignés dans les variables générées dans la base de données actualisées (Data\_0)

**Code stata :** gen pcinc=income/hhsize

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

**Réponse :** le revenu moyen par habitantest de **94.45**

le revenu total de la population est de **3778**

(les détails sur les résultats sont consignés dans le fichier de résultat)

**Code stata :**

egen total\_inc=sum(income)

sum pcinc [aw=hhsize]

**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).**

**Réponse :** les résultats sur la variable « intensité de la pauvreté par habitant » sont consignés dans les variables générées dans la base de données actualisées (Data\_0)

La moyenne de l’intensité de la pauvreté par habitant est de **0,2225** (les détails sur les résultats sont consignés dans le fichier de résultat)

**Code stata :**

gen pline=100

gen pgap=0

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

sum pgap [aw=hhsize]

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

**Réponse :** la moyenne de l’intensité de la pauvreté par habitant est de **0,2225** (les détails sur les résultats sont consignés dans le fichier de résultat)

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

**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).**

**Réponse :** les résultats sur la variable « indice de déflation des prix » (deflator) etla variable de revenu réel par habitant (rpcinc) sont consignés dans les variables générées dans la base de données actualisées (Data\_0)

**Code stata :**

gen deflator=1

replace deflator=0.9 if region==2

replace deflator=0.7 if region==3

gen rpcinc=pcinc/deflator

**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.**

**Réponse :** les résultats sur la variable « intensité de la pauvreté par habitant » sont consignés dans les variables générées dans la base de données actualisées (Data\_0)

La moyenne de l’intensité de la pauvreté par habitant est de **0,2719907** pour la question 1.3 et de **0.271991** pourla question 1.4 (les détails sur les résultats sont consignés dans le fichier de résultat)

**Code stata :**

sum rpcinc [aw=hhsize]

replace pline = 120

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

sum pgap [aw=hhsize]

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

**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 ?**

**Réponse :** les dépenses moyennes par équivalent adulte sont estimées à **42964,714844** (les détails sur les résultats sont consignés dans le fichier de résultat)

Cette statistique a ajusté les dépenses moyennes par habitant par la taille et la composition du ménage pour permettre de comparer les dépenses entre les ménages.

**Code stata :** imean ae\_exp, hsize(hhsize)

* 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.

* **CAS 1 : Seulement en utilisant la variable *strata* pour initialiser la variable de stratification de la population échantillonnée.**

**Réponse :** les résultats de l’initialisation de la variable de stratification de la population échantillonnée sont consignés dans le fichier de résultat.

Les dépenses moyennes par équivalent adulte sont estimées à **42964,714844** (les détails sur les résultats sont consignés dans le fichier de résultat)

Leniveau des erreurs-types des dépenses moyennes par équivalent adulte sont estimées à **1702,926636** (les détails sur les résultats sont consignés dans le fichier de résultat)

**Code stata :**

svyset, strata(strata)

imean ae\_exp, hsize(hhsize)

* **CAS 2 : Seulement en utilisant la variable *psu* pour initialiser la variable d'unité primaire d’échantillonnage (primary sampling unit, PSU).**

**Réponse :** les résultats de l’initialisation de la variable de stratification de la population échantillonnée sont consignés dans le fichier de résultat.

Les dépenses moyennes par équivalent adulte sont estimées à **42964.714844** (les détails sur les résultats sont consignés dans le fichier de résultat)

Leniveau des erreurs-types des dépenses moyennes par équivalent adulte sont estimées à **1693,014282** (les détails sur les résultats sont consignés dans le fichier de résultat)

**Code stata :**

svyset psu

imean ae\_exp, hsize(hhsize)

* **CAS 3 : En utilisant la variable *strata* et *psu.***

**Réponse :** les résultats de l’initialisation de la variable de stratification de la population échantillonnée sont consignés dans le fichier de résultat.

Les dépenses moyennes par équivalent adulte sont estimées à **42964,714844** (les détails sur les résultats sont consignés dans le fichier de résultat)

Leniveau des erreurs-types des dépenses moyennes par équivalent adulte sont estimées à **1699,352783** (les détails sur les résultats sont consignés dans le fichier de résultat)

**Code stata :**

svyset psu, strata(strata)

imean ae\_exp, hsize(hhsize)

* **CAS 4 : En utilisant la variable *strata, psu* et la variable de poids de sondage*.***

**Réponse :** les résultats de l’initialisation de la variable de stratification de la population échantillonnée sont consignés dans le fichier de résultat.

Les dépenses moyennes par équivalent adulte sont estimées à **41993,101563** (les détails sur les résultats sont consignés dans le fichier de résultat)

Leniveau des erreurs-types des dépenses moyennes par équivalent adulte sont estimées à **2213,284668** (les détails sur les résultats sont consignés dans le fichier de résultat)

**Code stata :**

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

imean ae\_exp, hsize(hhsize)

**Explications sur le niveau des erreurs-types par rapport à celui de la question 1.1 et à ceux des autres cas**

La question 1.1, et les cas 1, cas 2 et cas 3 ne prennent pas en compte toute l’information sur l’échantillonnage. Les écarts types sont plus faibles et par conséquent, les statistiques estimées sont moins précises compte tenu du regroupement. Le cas 4 intègre toute l’information sur l’échantillonnage, notamment la stratification. L’écart type est plus élevé et la précision des statistiques plus importantes. En conclusion, la stratification augmente la précision des statistiques estimées et le regroupement diminue la précision des statistiques estimées.

* 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.**

**Réponse :** les résultats sont consignés dans le fichier de résultat. Les résultats indiquent que les dépenses moyennes par équivalent adulte dans la région 1 sont significativement supérieures au double de celles de la région 3 au seuil de 10%.

**Code stata :** dimean ae\_exp ae\_exp, hsize1(hhsize) test(45969.62) cond1(region==3) hsize2(hhsize) cond2(region==1)

* 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.**

**Réponse :** les résultats sont consignés dans le fichier de résultat. Les résultats indiquent que les dépenses moyennes par équivalent adulte pour les chefs de famille hommes sont significativement plus élevées que celles des femmes chefs de famille au seuil de 1%.

**Code stata :** dimean ae\_exp ae\_exp, hsize1(hhsize) cond1(sex==2) hsize2(hhsize) cond2(sex==1)

### 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.**

**Réponse :** la taille de la population des ménages échantillonnés de **14694** (voir fichier de résultats et base de données actualisées Data\_1)

**Code stata :** egen taille\_ech=sum(hhsize)

**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*).**

**Réponse :** les résultats sont consignés dans le fichier de résultat et les variables générées dans la base de données actualisées Data\_1

**Code stata :**

sort pcexp

sum hhsize

gen ps=hhsize/r(sum)

gen p=sum(ps)

gen q=pcexp

list, sep(0)

**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).**

NB Il semble avoir une erreur sur domaine des centiles : min = 0 et max = 0,95. Nous avons utilisé : min = 0 et max = 1000000. Il sembla avoir une inversion antre l’axe X et l’axe Y au niveau de la question.

**Réponse :**



**Code stata :** cdf pcexp, hsize(hhsize) min(0) max(1000000) ytitle(les centiles) xtitle(les dépenses par habitant) title(courbe de distribution cumulative)

**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 :**

****

**Commentaire**

Le graphique met en évidence une forte inégalité dans la distribution des dépenses par habitant. Pour espérer trouver une dépense par habitant d’environ 150000, il faut examiner au moins 80% de la population échantillonnée.

**Code stata :** c\_quantile pcexp, hsize(hhsize) min(0) max(0.95) title(courbe des quantiles) xtitle(les centiles (p)) ytitle(les dépenses par habitant q(p))

**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 :**

****

**Commentaire**

Le graphique met en évidence une forte disparité entre le milieu rural et le milieu urbain en termes de distribution des dépenses par habitant. Pour espérer trouver une dépense par habitant d’un montant donné, il faut examiner une proportion de population échantillonnée plus importante en milieu rural par rapport au milieu urbain. Ce qui implique que les dépenses par habitant sont plus élevées en milieu urbain qu’en milieu rural quel que soit le niveau.

**Code stata :** c\_quantile pcexp, hsize(hhsize) hgroup(zone) min(0) max(0.95)

**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.**

**Réponse :**



**Commentaire**

Le graphique indique une forte concentration de la population échantillonnée entre des revenus compris entre 0 et 200000 F. La proportion d’hommes ayant un revenu entre 0 et 100000 est deux fois plus importante par rapport aux femmes, cependant entre 100000 et 200000 la proportion de est plus élevée que celle des hommes.

**Code stata :** cdensity pcexp, hs(hhsize) hgroup(sex) min(0) max(1000000)