## **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 : Le revenu par habitant est généré par la commande : *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.

**R : Estimation du revenu moyen à l’aide de stata : *sum pcinc [aw=hhsize]*. Il est de 96,96.**

**Variable | Obs Weight Mean Std. Dev. Min Max**

**-------------+-----------------------------------------------------------------**

**pcinc | 10 33 96.9697 57.94032 46.66667 280**

**Estimation du revenu total de la population : *total income*. Il est de 3200.**

**Total estimation Number of obs = 10**

**--------------------------------------------------------------**

**| Total Std. Err. [95% Conf. Interval]**

**-------------+------------------------------------------------**

**income | 3200 396.9327 2302.076 4097.924**

**--------------------------------------------------------------**

**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 : Générer intensite de pauvreteé :**

***gen pline = 120***

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

***replace pgap = 0 if pgap==.***

**Estimtion moyenne de l’intensite de pauvrete : *sum pgap [aw=hhsize]*. pgap =0,30**

**Variable | Obs Weight Mean Std. Dev. Min Max**

**-------------+-----------------------------------------------------------------**

**pgap | 10 33 .3005051 .2054157 0 .6111111**

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

**R : intensite de pauvrete avec DASP : *ifgt pcinc, pline(120) alpha(1) hsize(hhsize).* Il est de 0.30**

**Poverty index : FGT index**

**Household size : hhsize**

**Parameter alpha : 1.00**

**----------------------------------------------------------------------------------------------**

**Variable | Estimate STE LB UB Pov. line**

**-------------+--------------------------------------------------------------------------------**

**pcinc | 0.300505 0.061199 0.162064 0.438946 120.00**

**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 : Générer la variable *deflator :***

***gen deflator = 1***

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

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

**Generer le revenu reel :**

***gen rpcinc = pcinc/deflator***

***sum rpcinc [aw=hhsize].* Le revenu réel par habitant est de 123,43.**

**Variable | Obs Weight Mean Std. Dev. Min Max**

**-------------+-----------------------------------------------------------------**

**rpcinc | 10 33 123.4343 88.16636 60 350**

**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 : Estimation intensité de pauvreté avec stata :**

***replace pline = 110***

***gen pgap1 = (pline-rpcinc)/pline if (rpcinc<pline)***

***replace pgap1 = 0 if pgap1==.***

***sum pgap1 [aw=hhsize].* L’intensité de pauvreté est 21,67.**

**Variable | Obs Weight Mean Std. Dev. Min Max**

**-------------+-----------------------------------------------------------------**

**pgap1 | 10 33 .2167126 .1811048 0 .4545455**

**Estimation de l’intensité de pauvreté avec DASP**

***ifgt rpcinc, pline(110) alpha(1) hsize(hhsize).* L’intensité de pauvreté est 21,67.**

***Poverty index : FGT index***

***Household size : hhsize***

***Parameter alpha : 1.00***

***-----------------------------------------------------------------------------------------------***

***Variable | Estimate STE LB UB Pov. line***

***--------------+--------------------------------------------------------------------------------***

***rpcinc | 0.216713 0.059408 0.082322 0.351103 110.00***

**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 : La dépense moyenne par équivalent adulte : *imean pcexp, hsize(hhsize).* Cette statistique se réfère aux informations sur l’échantillonnage et permet d’estimer les erreur standart.**

Index : Mean index

Household size : hhsize

-----------------------------------------------------------------------------------

Variable | Estimate STE LB UB

------------------+----------------------------------------------------------------

1: mean\_ae\_exp | 40428.652344 974.026428 38518.437500 42338.867188

----

* 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 : Initialiser le plan d’échantillonnage**

***svyset \_n [pweight=sweight], strata(strata)***

***gen numerateur = hhsize\*eaexp***

***gen denominateur = hhsize***

***svy: ratio numerateur/denominateur***

***Number of strata = 10 Number of obs = 2,000***

***Number of PSUs = 2,000 Population size = 787,499.95***

***Design df = 1,990***

***\_ratio\_1: numerateur/denominateur***

***--------------------------------------------------------------***

***| Linearized***

***| Ratio Std. Err. [95% Conf. Interval]***

***-------------+------------------------------------------------***

***\_ratio\_1 | 38796.24 1256.984 36331.1 41261.38***

***--------------------------------------------------------------***

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

**R : Estimation de la dépense moyenne par région**

**. imean ae\_exp, hgroup(region)**

**Index : Mean index**

**Sampling weight : sweight**

**Group variable : region**

**-----------------------------------------------------------------------------------**

**Group | Estimate STE LB UB**

**------------------+----------------------------------------------------------------**

**1: 1 | 56417.695313 4115.376465 48346.796875 64488.593750**

**2: 2 | 35296.605469 1138.113159 33064.585938 37528.621094**

**3: 3 | 22720.632813 1291.645020 20187.515625 25253.751953**

**4: 4 | 40580.050781 1432.108276 37771.460938 43388.636719**

**------------------+----------------------------------------------------------------**

**Population | 41087.519531 1577.828857 37993.148438 44181.886719**

**La moyenne de la region 1 étant 56417,69 et celle de la region 3 est 22720,63, l'hypothèse revient à verifier si la différence de dépense moyenne entre la région 1 et la région 3 est supérieure à 22720,63.**

**dimean ae\_exp ae\_exp, hsize1(hhsize) test(22720.63) cond1(region==3) hsize2(hhsize) cond2(region==1) conf(ub)**

**------------------------------------------------------------------------------**

**Index | Estimate Std. Err. t P>|t| [95% Conf. Interval]**

**---------+--------------------------------------------------------------------**

**mean\_D1 | 21073.08 1445.779 14.5756 0.0000 -infinity 23452.28**

**mean\_D2 | 50474.22 2814.163 17.9358 0.0000 -infinity 55105.26**

**---------+--------------------------------------------------------------------**

**diff.| 29401.13 3163.809 9.29295 0.0000 -infinity 34607.56**

**------------------------------------------------------------------------------**

**estimate(diff) = estimate(mean\_D2 - mean\_D1) t = 2.1115**

**Ho: estimate(diff) = 22720.63 degrees of freedom = 1990**

**Ha: est.(diff) < 22720.63 Ha: est.(diff) != 22720.63 Ha: est.(diff) > 22720.63**

**Pr(T < t) = 0.0174 Pr(|T| > |t|) = 0.0349 Pr(T > t) = 0.9826**

**Le test statistique est 98,26 supérieur à 5% donc nous acceptons l'hypothèse que la différence de dépense moyenne est > 22720,63. Donc la dépense moyenne par équivalent adulte dans la région 1 est supérieure au double de la région 3**

* 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.

**R : Vérification si la dépense des chefs de ménage homme est plus éléve que celle des chefs ménage femme. Cela revient à tester si la différence de revenu mean\_homme - mean\_femme est > 0**

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

------------------------------------------------------------------------------

Index | Estimate Std. Err. t P>|t| [95% Conf. Interval]

---------+--------------------------------------------------------------------

mean\_D1 | 37429.57 4050.782 9.24009 0.0000 -infinity 44095.62

mean\_D2 | 38866.03 1305.801 29.7641 0.0000 -infinity 41014.88

---------+--------------------------------------------------------------------

diff.| 1436.469 4257.983 .337359 0.7359 -infinity 8443.49

------------------------------------------------------------------------------

estimate(diff) = estimate(mean\_D2 - mean\_D1) t = 0.3374

Ho: estimate(diff) = 0 degrees of freedom = 1990

Ha: est.(diff) < 0 Ha: est.(diff) != 0 Ha: est.(diff) > 0

Pr(T < t) = 0.3679 Pr(|T| > |t|) = 0.7359 Pr(T > t) = 0.6321

**Le test statistique est 63,21 supérieur à 5% donc lhypothese selon laquelle la difference est >0 est acceptée. Donc la depense moyenne par equivalent adulte des hommes est plus élevé que celle des femmes\*/**

### 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 : Calculer la taille de la population : *total hhsize*.**

**total hhsize**

**Total estimation Number of obs = 2,000**

**--------------------------------------------------------------**

**| Total Std. Err. [95% Conf. Interval]**

**-------------+------------------------------------------------**

**hhsize | 14609 224.0282 14169.65 15048.35**

**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 depenses par habitant : *sort pcexp***

**Générer la variable ps :**

***sum hhsize***

***Variable | Obs Mean Std. Dev. Min Max***

***-------------+---------------------------------------------------------***

***hhsize | 2,000 7.3045 5.009424 1 38***

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

**Générer la variable des centiles : *gen p = sum(ps)***

**Générer la variable des quantiles :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 : La courbe de distribution cumulative : *cnpe pcexp, xvar(p) type(npr) min(0) max(0.95) ytitle(Dépenses par habitant (F(p))) xtitle(Les centiles (p)) xscale(range(0 0.95)) title(Courbe de distribution cumulative)***

******

**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 : La courbe des quantiles : *c\_quantile p q, min(0.0) max(0.95) ytitle(Quantiles (Q)) xtitle(Centiles (p)) title(Courbe des quantiles)***

****

**Discussion : la courbe des quantiles ainsi obtenue est identique à celle de distribution cumulative**

**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 : Courbe des quantiles selon le sexe de la tête du ménage : *c\_quantile q, hgroup(sex) min(0.0) max(0.95) ytitle(Quantiles (Q)) xtitle(Centiles (p)) xscale(range(0 0.95)) title(Courbe des quantiles)***

****

**Discussion : On note que la courbe des quantiles des hommes est inférieur à celle des femmes. On pourrait supposer que les dépenses par habitants des femmes seraient alors supérieures à celles des hommes**

**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 : Courbe de densité : *cdensity pcexp, hgroup(zone) min(0) max(1000000) title(Courbe de densité des dépenses par habitant***



**Discussion : Les dépenses dans les zones rurales sont plus concentrées entre 0 et 200000. A partir d’un seuil, les dépenses en milieu urbain sont supérieures à celle des milieux ruraux.**