## **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 : 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 : /\* Estimating the average per capita income \*/**

**. sum pcinc [aw=hhsize]**

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

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

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

**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 : . gen pline = 120**

**. gen pgap = 0**

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

**(7 real changes made)**

**. sum pgap [aw=hhsize]**

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

**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 : . gen deflator = 1**

**. replace deflator = 1.2 if region == 2**

**(2 real changes made)**

**. replace deflator = 1.4 if region == 3**

**(4 real changes made)**

**. 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 110.

**R :**

**sum rpcinc [aw=hhsize]**

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

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

**rpcinc | 10 33 83.8961 44.54531 33.33334 233.3333**

**. replace pline = 110**

**(10 real changes made)**

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

**(8 real changes made)**

**. sum pgap [aw=hhsize]**

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

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

**pgap | 10 33 .3115571 .1825609 0 .6969697**

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

**Poverty index : FGT index**

**Household size : hhsize**

**Parameter alpha : 1.00**

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

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

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

**rpcinc | 0.311557 0.053215 0.191176 0.431938 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 : . imean ae\_exp, hsize(hhsize)**

**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 : . svyset psu [pweight=sweight], strata(strata)**

**pweight: sweight**

**VCE: linearized**

**Single unit: missing**

**Strata 1: strata**

**SU 1: psu**

**FPC 1: <zero>**

**. gen nominator = hhsize\*ae\_exp**

**. gen denominator = hhsize**

**. svy: ratio nominator/denominator**

**(running ratio on estimation sample)**

**Survey: Ratio estimation**

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

**Number of PSUs = 421 Population size = 787,499.95**

**Design df = 411**

**\_ratio\_1: nominator/denominator**

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

**| Linearized**

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

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

**\_ratio\_1 | 38796.24 1197.018 36443.2 41149.28**

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

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

**R : #delimit;**

**delimiter now ;**

**. dimean ae\_exp ae\_exp, hsize1(hhsize) test(86480) cond1(strata==3 ) hsize2(hhsize) cond2(strata==1 ) conf(ub);**

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

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

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

**mean\_D1 | 43188.58 3370.073 12.8153 0.0000 -infinity 48744.38**

**mean\_D2 | 37507.31 2353.627 15.936 0.0000 -infinity 41387.43**

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

**diff.| -5681.271 4110.59 -1.38211 0.1677 -infinity 1095.322**

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

**estimate(diff) = estimate(mean\_D2 - mean\_D1) t = -22.4204**

**Ho: estimate(diff) = 86480 degrees of freedom = 411**

**Ha: est.(diff) < 86480 Ha: est.(diff) != 86480 Ha: est.(diff) > 86480**

**Pr(T < t) = 1.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 0.0000**

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

Nous ne pouvons pas rejeter Ha: différence => 86480, parce que l'erreur statistique que nous faisons si nous rejetons Ha = 0,00 est inférieure au niveau critique de 5.00%.

* 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 : . #delimit;**

**delimiter now ;**

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

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

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

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

**mean\_D1 | 38866.03 1245.495 31.2053 0.0000 -infinity 40919.32**

**mean\_D2 | 37429.57 4022.335 9.30543 0.0000 -infinity 44060.67**

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

**diff.| -1436.469 4232.079 -.339424 0.7345 -infinity 5540.407**

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

**estimate(diff) = estimate(mean\_D2 - mean\_D1) t = -0.3394**

**Ho: estimate(diff) = 0 degrees of freedom = 411**

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

**Pr(T < t) = 0.6328 Pr(|T| > |t|) = 0.7345 Pr(T > t) = 0.3672**

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

Nous ne pouvons pas rejeter Ha: différence => 0, parce que l'erreur statistique que nous faisons si nous ne rejetons pas Ha est supérieure au niveau critique de 5.00%.

### 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 : gen thhsize=psu\*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 avec les dépenses par habitant correspondantes. Sur cette base, générer les variables centiles (*p*) et quantiles (*q*).

**R :**

**sort pcexp**

**sum hhsize**

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

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

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

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

**. /\* generating the variable percentile and the quantiles \*/**

**. 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 : #delimit ;**

**line p pcexp, title(The cumulative distribution curve) xtitle(the percentile (p)) ytitle(les dépenses par habitant(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 :**

**#delimit ;**

**line q p , title(The quantile curve) xtitle(the percentile (p)) ytitle(The quantile Q(p));**

****

**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 : c\_quantile pcexp, hsize(sex)**

****

**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 : cdensity pcexp, hs(hhsize) band(3) min(0) max(1000000)**

