## **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 :** gen pcinc = income/hhsize

pcinc | Freq. Percent Cum.

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

35 | 1 8.33 8.33

52.5 | 1 8.33 16.67

60 | 1 8.33 25.00

70 | 1 8.33 33.33

75 | 1 8.33 41.67

100 | 1 8.33 50.00

110 | 1 8.33 58.33

112.6667 | 1 8.33 66.67

120 | 1 8.33 75.00

125 | 1 8.33 83.33

170 | 1 8.33 91.67

280 | 1 8.33 100.00

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

Total | 12 100.00

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

**R : .**

Le revenu moyen par habitant

sum pcinc [aw=hhsize]

Variable | Obs Weight Mean Std. Dev. Min Max

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

pcinc | 12 40 94.45 56.45776 35 280

ou

. mean pcinc [aw=hhsize]

Mean estimation Number of obs = 12

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

| Mean Std. Err. [95% Conf. Interval]

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

pcinc | 94.45 16.29795 58.57845 130.3216

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

Le revenu total de notre population

total income

Total estimation Number of obs = 12

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

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

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

income | 3778 410.9461 2873.514 4682.486

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

**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 :**

gen pline = 100

gen pgap = 0

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

(5 real changes made)

sum pgap [aw=hhsize]

Variable | Obs Weight Mean Std. Dev. Min Max

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

Pgap | 12 40  **.2225** .2368688 0 .65

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

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

Poverty index : FGT index

Household size : hhsize

Parameter alpha : 1.00

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

Variable | Estimate STE LB UB Pov. line

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

pcinc | **0.222500** 0.070494 0.067344 0.377656 100.00

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

**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 :** gen deflator = 1

replace deflator = 0.9 if region == 2

replace deflator = 0.7 if region == 3

gen rpcinc = pcinc/deflator

rpcinc2 | Freq. Percent Cum.

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

50 | 1 8.33 8.33

52.5 | 1 8.33 16.67

60 | 1 8.33 25.00

70 | 1 8.33 33.33

75 | 1 8.33 41.67

111.1111 | 1 8.33 50.00

157.1429 | 1 8.33 58.33

160.9524 | 1 8.33 66.67

171.4286 | 1 8.33 75.00

178.5714 | 1 8.33 83.33

242.8571 | 1 8.33 91.67

311.1111 | 1 8.33 100.00

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

Total | 12 100.00

**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 :**

sum rpcinc [aw=hhsize]

Variable | Obs Weight Mean Std. Dev. Min Max

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

rpcinc2 | 12 40 114.7738 73.95077 50 311.1111

ge pline2 = 120

. gen pgap2 = 0

. replace pgap2 = (pline2-rpcinc2)/pline2 if (rpcinc2 < pline2)

(6 real changes made)

. sum pgap2 [aw=hhsize]

Variable | Obs Weight Mean Std. Dev. Min Max

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

pgap2 | 12 40 .2719907 .2506777 0 .5833333

. ifgt rpcinc2, pline(120) alpha(1) hsize(hhsize)

Poverty index : FGT index

Household size : hhsize

Parameter alpha : 1.00

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

Variable | Estimate STE LB UB Pov. line

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

rpcinc2 | 0.271991 0.073454 0.110320 0.433662 120.00

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

**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 :**

imean ae\_exp , hsize(hhsize)

Index : Mean index

Household size : hhsize

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

Variable | Estimate STE LB UB

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

1: mean\_ae\_exp | 42964.714844 1701.506958 39627.800781 46301.628906

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

* 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 :**

* 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 :**

dimean ae\_exp ae\_exp , hsize1(hhsize) test(100) cond1(strata==3 ) hsize2(hhsize) cond2(strata==1 )

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

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

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

mean\_D1 | 37562.72 2934.558 12.8001 0.0000 31807.61 43317.83

mean\_D2 | 41398.43 3383.56 12.2352 0.0000 34762.76 48034.1

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

diff.| 3835.716 4478.852 .856406 0.3919 -4947.991 12619.42

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

estimate(diff) = estimate(mean\_D2 - mean\_D1) t = 0.8341

Ho: estimate(diff) = 100 degrees of freedom = 1999

Ha: est.(diff) < 100 Ha: est.(diff) != 100 Ha: est.(diff) > 100

Pr(T < t) = 0.2022 Pr(|T| > |t|) = 0.4043 Pr(T > t) = 0.7978

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

* 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.

1. **R :**

. dimean ae\_exp ae\_exp , hsize1(hhsize) test(100) cond1( sex ==2 ) hsize2(hhsize) cond2( sex ==1 )

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

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

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

mean\_D1 | 37180.35 2601.922 14.2896 0.0000 32077.59 42283.11

mean\_D2 | 43274.53 1787.126 24.2146 0.0000 39769.71 46779.35

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

diff.| 6094.179 3156.551 1.93064 0.0537 -96.29548 12284.65

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

estimate(diff) = estimate(mean\_D2 - mean\_D1) t = 1.8990

Ho: estimate(diff) = 100 degrees of freedom = 1999

Ha: est.(diff) < 100 Ha: est.(diff) != 100 Ha: est.(diff) > 100

Pr(T < t) = 0.0289 Pr(|T| > |t|) = 0.0577 Pr(T > t) = 0.9711

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

### 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 :**

**. sum hhsize**

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

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

**hhsize | 2,000 7.347 5.281623 1 38**

**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 :**

gen ps = hhsize/r(sum)

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 des centiles: min = 0 et max = 0,95).

**R :**

line p pcexp , title(The cumulative distribution curve) xtitle(The per capita expenditure (y)) ytitle(F

> (y))



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

c\_quantile q, hsize(hhsize) hgroup(zone) min(0.0) 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.

**R :**