

AVRIL 2014

CONCOURS INGÉNIEURS DES TRAVAUX STATISTIQUES

ITS Voie B Option Économie

ANALYSE D'UNE DOCUMENTATION STATISTIQUE

(Durée de l'épreuve : 2 heures)

Note : La note finale tiendra compte, de façon non négligeable, des commentaires donnés après chaque résultat.

Un méthodologue du ministère chargé de la statistique doit effectuer un audit du plan de sondage actuel pour une enquête réalisée mensuellement.



1. Le plan de sondage actuel

L'échantillon des entreprises interrogées dans l'enquête mensuelle est tiré dans le champ des entreprises ayant répondu à l'enquête annuelle.

A la date de tirage des échantillons pour l'enquête mensuelle de l'année N (date qui se situe vers la fin du dernier trimestre N-1), les résultats de l'enquête annuelle N-2 sont disponibles. L'échantillon est donc tiré à partir des informations datées de deux ans.

L'unité d'échantillonnage est un produit d'une entreprise.

La méthode de tirage des échantillons de l'enquête mensuelle n'est pas aléatoire. Pour un produit donné, la méthode de sélection consiste à :

- trier les entreprises concernées par le produit par valeurs décroissantes de chiffre d'affaires réalisé pour le produit considéré ;
- retenir les entreprises dans l'échantillon jusqu'à couvrir au moins 75% du chiffre d'affaires total du produit.

Dans cet exercice, l'audit portera sur un seul produit.

2. Description des données

Sur le produit concerné par l'exercice, le tableau joint en annexe vous donne les réponses individuelles des entreprises à l'enquête annuelle pour les années 2011 et 2012.

En 2012, 39 entreprises ont répondu à l'enquête annuelle. L'échantillon pour l'enquête mensuelle 2014 est donc choisi avec ces données.

Question 1 : à la lecture du tableau joint en annexe, indiquer le nombre d'entreprises ayant répondu en 2011 à l'enquête annuelle.

Question 2 : indiquer le nombre d'unités qui n'étaient pas présentes en 2011 dans l'enquête annuelle, le montant total en kE qu'elles représentent en 2012, la part qu'elles représentent en 2012.

Question 3 : compléter le tableau ci-dessous sur la base des entreprises présentes dans la base de sondage des données 2012 et 2011 (c'est-à-dire répondantes sur les deux années) :

	Présents dans l'échantillon 2014 et dans la base de sondage 2012 et 2011	Absents de l'échantillon 2014 et présents dans la base de sondage 2012 et 2011	Total (présents dans la base de sondage 2012 et 2011)
Somme des montants 2012			
Somme des montants 2011			
Taux d'évolution			

Question 4 : indiquer le biais de la méthode actuelle qui est la différence entre l'évolution réelle et celle constatée dans l'échantillon.

Question 5 : commenter les résultats obtenus.



3. Comment assurer une certaine continuité avec la méthode de tirage précédente

L'introduction d'une méthode aléatoire, représentant l'ensemble des unités, dans le tirage de l'échantillon, devrait permettre de produire une estimation du taux d'évolution plus proche de la réalité.

Si l'on souhaite assurer une certaine continuité avec les méthodes de tirage précédentes, il paraît intéressant d'imposer une strate exhaustive couvrant une certaine part du montant total du produit (actuellement la procédure consiste à sélectionner les plus grandes entreprises concernées par un produit jusqu'à couvrir au moins 75% du montant total du produit).

Question 6 : à la lecture du tableau joint en annexe, indiquer la taille actuelle n de l'échantillon 2014.

Question 7 : calculer le seuil de couverture correspondant à un tirage proportionnel au montant avec la taille d'échantillon actuelle selon la formule $n \frac{y_k(2012)}{t_y(2012)} \geq 1$, où $t_y(2012)$ est la somme des $y_k(2012)$ des entreprises présentes. Ici $t_y(2012)$ est égal à 287.994 keuros.

Question 8 : indiquer le nombre d'entreprises sélectionnées par la méthode de la question précédente et le seuil de couverture.

Dans la suite, on a effectué les calculs avec trois taux de couverture différents : 0% (pas de strate exhaustive), 19,1%, 76,2%. Pour notre produit, les nombres d'unités à tirer dans la strate exhaustive sont, en fonction des taux de couverture en montant visés, les suivants :

Taux de couverture pour la strate exhaustive	Seuil de y_k correspondant en kE	Nombre d'unités dans la base de sondage 2012	dont présentes aussi dans la base de sondage 2011
0%	40.000	0	0
19,1%	18.000	2	2
76,2%	7.800	16	15



4. Partie non exhaustive

Dans cette première approche, on sélectionne un échantillon de taille n_{nexh} selon un sondage aléatoire simple dans la partie non exhaustive de façon à ce que l'écart-type dû à l'échantillonnage de l'estimation du taux d'évolution soit de x .

On se limite aux unités de la base de sondage 2012 qui étaient présentes en 2011 et on calcule les u_k :

$$u_k = \frac{1}{t_y(2011)} \left(y_k(2012) - \frac{t_y(2012)}{t_y(2011)} y_k(2011) \right) \text{ avec } t_y(2012) = 275.407 \text{ et } t_y(2011) = 301.708$$

On peut calculer le nombre d'unités à tirer dans la partie non exhaustive n_{nexh} pour obtenir un écart-type dû à l'échantillonnage noté x avec la formule suivante :

$$n_{nexh} = \frac{1}{\frac{x^2}{N_{nexh}^2 S_{unexh}^2} + \frac{1}{N_{nexh}}} \text{ avec } S_{unexh}^2 = \frac{1}{n_{nexh} - 1} \sum_{nexh} (u_k - \bar{u})^2$$

Où \bar{u} désigne la moyenne des u_k et N_{nexh} le nombre d'unités dans la partie non exhaustive.

On obtient les résultats suivants lorsque l'on fait varier le taux de couverture pour la strate exhaustive et l'écart-type dû à l'échantillonnage visé :

Taux de couverture pour la strate exhaustive	Écart-type dû à l'échantillonnage pour l'estimation du taux d'évolution X	Nombre d'unités à tirer parmi les présents-présents dans la strate exhaustive	Nombre d'unités à tirer parmi les présents-présents dans la strate non exhaustive	Nombre d'unités à tirer au total parmi les présents-présents
0%	0,01	0	36	36
19,1%	0,01	2	29	31
75%	0,01	15	16	31
0%	0,02	0	34	34
19,1%	0,02	2	20	22
75%	0,02	15	9	24
0%	0,03	0	32	32
19,1%	0,03	2	13	15
75%	0,03	15	5	20
0%	0,04	0	29	29
19,1%	0,04	2	9	11
75%	0,04	15	3	18
0%	0,05	0	27	27
19,1%	0,05	2	6	8
75%	0,05	15	2	17

Question 9 : avec la taille d'échantillon actuelle que vous avez indiqué à la question 6 et en prenant le taux de couverture que vous avez indiqué à la question 8, donner une estimation de l'écart-type dû à l'échantillonnage du taux d'évolution.

Question 10 : commenter les résultats obtenus.

Liste des entreprises fabricant le produit étudié avec leurs facturations annuelles

NUM K	Enquête mensuelle		Données de l'enquête annuelle			
	SELECTION 2014	PRESENCE ECHANTILLON	MONTANT 2011 (en keuros) y _k (2011)	MONTANT 2012 (en keuros) y _k (2012)	PDSIND 2012	PDSCUM 2012
		2013				
1	1	1	66312	36320	12,6%	12,6%
2	1	1	16345	18782	6,5%	19,1%
3	1	1	15820	16146	5,6%	24,7%
4	1	1	15455	15362	5,3%	30,1%
5	1	1	15228	15228	5,3%	35,4%
6	1	1	12665	14025	4,9%	40,2%
7	1	1	13418	13481	4,7%	44,9%
8	1	1	13357	12350	4,3%	49,2%
9	1	0		11100	3,9%	53,1%
10	1	0	8112	10345	3,6%	56,6%
11	1	0	6798	10260	3,6%	60,2%
12	1	1	11377	10094	3,5%	63,7%
13	1	1	9723	9900	3,4%	67,2%
14	1	1	9336	9736	3,4%	70,5%
15	1	0	8670	8401	2,9%	73,4%
16	1	1	7563	7850	2,7%	76,2%
17	0	0	5654	6382	2,2%	78,4%
18	0	0	5739	5962	2,1%	80,5%
19	0	0	6823	5404	1,9%	82,3%
20	0	1	9951	5403	1,9%	84,2%
21	0	0	5028	5220	1,8%	86,0%
22	0	0	3200	4366	1,5%	87,5%
23	0	0	5507	4336	1,5%	89,0%
24	0	0	3819	3910	1,4%	90,4%
25	0	0	2951	3856	1,3%	91,7%
26	0	0	3320	3678	1,3%	93,0%
27	0	0	3556	3112	1,1%	94,1%
28	0	0	2576	3067	1,1%	95,2%
29	0	0	3301	2984	1,0%	96,2%
30	0	1	2688	2298	0,8%	97,0%
31	0	0	2211	2215	0,8%	97,8%
32	0	0	1998	1881	0,7%	98,4%
33	0	0	1666	1503	0,5%	98,9%
34	0	0		1487	0,5%	99,5%
35	0	0	878	1130	0,4%	99,9%
36	0	0	42	203	0,1%	99,9%
37	0	0	436	102	0,0%	100,0%
38	0	0	145	75	0,0%	100,0%
39	0	0	40	40	0,0%	100,0%

NUM : il s'agit du numéro de l'entreprise permettant de la repérer dans la base de sondage.

SELECTION 2012 : il s'agit d'une indicatrice : 0 signifie que l'entreprise n'est pas sélectionnée dans l'échantillon de l'enquête mensuelle selon la méthode actuelle et 1 qu'elle l'est.

PRESENCE ECHANTILLON 2011 : il s'agit d'une indicatrice : 0 signifie que l'entreprise n'était pas sélectionnée en 2011 dans l'enquête mensuelle et 1 qu'elle l'était.

MONTANT 2011 et MONTANT 2012 : il s'agit des montants de facturations figurant dans l'enquête annuelle.

PDSIND 2012 : il s'agit du poids économique de l'entreprise pour ce produit. Dans ce tableau, les entreprises sont classées selon ce critère.

PDSCUM 2012 : il s'agit du poids économique cumulé.