

RAPPORT

CETE
Normandie Centre

DADT
Département
Aménagement Durable
des Territoires

Intervalles de confiance pour Les estimations des indicateurs donnés par Les EMD

Ressources, territoires, habitats et logement
Énergie et climat
Prévention des risques
Développement durable
Infrastructures, transports et mer

**Présent
pour
l'avenir**



Ministère de l'Ecologie, du Développement durable,
des Transports et du Logement

www.cete-normandie-centre.developpement-durable.gouv.fr

Historique des versions du document

Version	Auteur	Commentaires
1	M. BEN AÏSSA Sami	INSA de Rouen – Département Génie Mathématique

Affaire suivie par

Mathieu RABAUD – Département Aménagement Durable des Territoires
Tél. 0235688117 / fax 0235688252
Mél. mathieu.rabaud@developpement-durable.gouv.fr

Référence Internet

http://www.cete-normandie-centre.developpement-durable.gouv.fr

Sommaire

INTRODUCTION.....	4
INTERVALLES DE CONFIANCE AVEC L'HYPOTHÈSE D'UN TIRAGE ALÉATOIRE.....	5
Intervalle de confiance approximatif de niveau $1-\alpha$ pour l'estimation d'une proportion p	5
Intervalle de confiance approximatif de niveau $1-\alpha$ pour l'estimation d'une moyenne	6
INFLUENCE DE LA STRATIFICATION ET DU TIRAGE EN GRAPPES.....	8
Influence de la stratification	8
Influence du tirage en grappes	8

Introduction

La méthode d'échantillonnage pour les Enquêtes Ménages Déplacements est la suivante :
L'agglomération est divisée en plusieurs secteurs (on parle de stratification). Dans chaque secteur, il y a un tirage aléatoire sans remise des ménages. Toutes les personnes de chaque ménage de plus de 5 ans et tous leurs déplacements sont recensés (il y a donc un effet de grappe).

D'abord, nous allons donner les intervalles de confiance en considérant un échantillonnage aléatoire. Ensuite, nous allons discuter l'influence de la stratification et du tirage en grappes.

Remarque : les calculs se font sur les données non redressées.

Intervalles de confiance avec l'hypothèse d'un tirage aléatoire

Il y a deux types de tirage aléatoire : tirage avec remise et tirage sans remise. Dans les EMD, le tirage se fait sans remise. Dans le calcul des intervalles des confiances, la différence entre les deux méthodes est négligeable si le taux de sondage (n/N) est faible, ce qui est le cas ici (n : taille de l'échantillon ; N : taille de la population totale). En effet, dans le tirage sans remise, les bornes sont multipliées par $\frac{N-n}{N} \sim 1$ ce qui réduit très légèrement la taille de l'intervalle. Par conséquent, nous allons donner les intervalles en considérant un tirage avec remise.

Intervalle de confiance approximatif de niveau $1-\alpha$ pour l'estimation d'une proportion p

Il faut avoir $n > 100$ et $np(1-p) > 5$.

$$\left[f - u_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n-1}}, f + u_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n-1}} \right]$$

f est la proportion dans l'échantillon de l'indicateur (estimation ponctuelle) et $u_{1-\frac{\alpha}{2}}$ est le fractile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite.

Voici un tableau qui donne les fractiles les plus utilisés :

α	$u_{1-\frac{\alpha}{2}}$
1,00%	2,576
5,00%	1,960
10,00%	1,645
20,00%	1,282

Fonction Excel : LOI.NORMALE.STANDARD.INVERSE($1 - \alpha/2$)

Exemples :

- La proportion des scolaires dans l'agglomération de Rouen en 2007 :

Niveau de confiance	n	Nombre de scolaires dans l'échantillon	f
95,00%	110452	2010	18,20%

Borne inférieure	Borne supérieure	Rayon
17,48%	18,92%	0,72%

- Proportion du motif « Études » dans les déplacements des scolaires à Rouen 2007 :

Niveau de confiance	n	Nombre de motifs « Etudes » dans l'échantillon	f
95,00%	67682	2234	33,01%

Borne inférieure	Borne supérieure	Rayon
31,89%	34,13%	1,12%

- Proportion du mode « Marche » dans les déplacements des scolaires à Rennes 2007 :

Niveau de confiance	n	Nombre de mode « Marche » dans l'échantillon	f
95,00%	36571	1204	32,92%

Borne inférieure	Borne supérieure	Rayon
31,40%	34,45%	1,52%

Intervalle de confiance approximatif de niveau $1-\alpha$ pour l'estimation d'une moyenne

On doit avoir $n > 30$.

$$\left[\bar{X} - t_{1-\frac{\alpha}{2}}(n-1) \frac{s}{\sqrt{n}}, \bar{X} + t_{1-\frac{\alpha}{2}}(n-1) \frac{s}{\sqrt{n}} \right]$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad : \text{estimation ponctuelle de la moyenne}$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad : \text{estimation de l'écart-type}$$

$$t_{1-\frac{\alpha}{2}}(n-1) \quad : \text{fractile d'ordre } 1 - \alpha/2 \text{ de la loi de Student à } n-1 \text{ degrés de liberté.}$$

Fonction Excel : LOI.STUDENT.INVERSE(α ; n-1)

Exemples :

- Mobilité des scolaires à Tours 2007 :

Niveau de confiance	n	Estimation de la moyenne	s
95,00%	9583	3,42	1,76

Borne inférieure	Borne supérieure	Rayon
3,31	3,53	0,11

- Durée moyenne d'un déplacement pour les scolaires à Rouen 2007 :

Niveau de confiance	n	Estimation de la moyenne	s
95,00%	67691	16,25	19,44

Borne inférieure	Borne supérieure	Rayon
15,79	16,71	0,46

- Taux de motorisation des ménages dans la première couronne à Rouen 2007 :

Niveau de confiance	n	Estimation de la moyenne	s
95,00%	18301	1,23	0,8

Borne inférieure	Borne supérieure	Rayon
1,19	1,27	0,04

Influence de la stratification et du tirage en grappes

Influence de la stratification

On suppose que le nombre de ménages enquêtés est le même dans chaque secteur.

La perte de précision qui résulte de cette méthode de tirage par rapport à un tirage proportionnel sur un indicateur au niveau de l'agglomération est donnée par :

$$\frac{\sigma^2}{\sigma_{\text{prop}}^2} = n_s \sum_i p_i^2 \quad \text{où } n_s \text{ est le nombre de secteurs et } p_i = \frac{N_i}{N} \text{ et en supposant que } \sigma_i = \text{cste}$$

Exemple issue du document du Certu sur la précision des EMD :

Pour le fichier de l'enquête ménages déplacements de Marseille 1976, on a $n_s = 41$ zones,

$$\sum_i p_i^2 = 0,0290$$

$$\frac{\sigma^2}{\sigma_{\text{prop}}^2} = 1,19$$

$$\frac{\sigma}{\sigma_{\text{prop}}} = 1,09$$

La perte de précision est donc dans ce cas inférieure à 10% en valeur relative. Ce résultat montre que la méthode de tirage adoptée (effectifs constants dans chaque strate) n'influe pas de manière sensible sur la précision des indicateurs.

Influence du tirage en grappes

On va donner quelques résultats issus du document du Certu qui vont permettre d'apprécier l'ordre de grandeur de cet effet.

On donne le rapport entre l'écart-type de l'estimation compte tenu de la méthode en grappes, et l'écart-type de l'estimation faite sur un échantillon aléatoire de même taille. Ce rapport est appelé coefficient de grappe.

Base statistique	Indicateur	Coefficient de grappe
Échantillon de personnes	<ul style="list-style-type: none"> • taux d'activité • % de femmes • % de personnes entre 20 et 30 ans • mobilité (nombre de moyen de déplacements par personne et par jour) • tous motifs - tous modes • tous motifs - modes motorisés • tous motifs - VP • tous motifs - TC • tous motifs - 2R • tous motifs - MAP • travail - tous modes • autres motifs - tous modes • secondaires - tous modes 	1 0,7 1,2 à 1,3 1,05 à 1,15 1 à 1,2 1,1 à 1,2 1 à 1,15 1 à 1,15 1 à 1,2 1 à 1,1 1 à 1,2 1 à 1,2
Échantillon de déplacements	% de déplacements effectués en VP (tous motifs) % de déplacements effectués en TC (tous motifs) % de déplacements effectués en 2R (tous motifs) % de déplacements effectués en MAP (tous motifs) Coefficient de pointe (tous modes, tous motifs)	1,8 à 2 1,6 à 1,8 1,8 à 2,2 1,8 à 2,2 1

Pour toutes les variables évaluées à partir d'un échantillon de grappes de personnes, l'effet du triage en grappes est peu sensible, puisqu'il conduit à multiplier l'erreur sur les estimations par un facteur de l'ordre de 1,1 par rapport à un tirage aléatoire.

Pour les variables évaluées à partir d'un échantillon de grappes de déplacements, l'effet est par contre important pour les pourcentages de répartition modale, le coefficient étant de l'ordre de 2. Il faut donc corriger les résultats donnés en première partie en multipliant la précision par un facteur de l'ordre de 2.

Ces résultats ont été établis à partir de tests sur le fichier de l'enquête de Grenoble 1973. La liste des indicateurs rassemblés dans le tableau ci-dessus n'est absolument pas exhaustive et mériterait d'être complétée par d'autres variables. Les mêmes calculs devraient également être effectués sur d'autres enquêtes, afin de vérifier que l'on retrouve bien les mêmes valeurs ou du moins des valeurs très proches des différents coefficients de grappe.

CETE Normandie Centre
10 Chemin de la Poudrière
76121 Le Grand-Quevilly cédex
téléphone : 02 35 68 81 00
courriel : DADT.CETE-NC@developpement-durable.gouv.fr
<http://www.cete-normandie-centre.developpement-durable.gouv.fr>

Ressources, territoires, habitats et logement
Énergie et climat Développement durable
Prévention des risques Infrastructures, transports et mer

**Présent
pour
l'avenir**