## Chapitre 2 : Plan de sondage aléatoire simple

Un plan de sondage aléatoire simple est caractérisé par un tirage aléatoire avec des probabilités égales, appelé aussi « plan à probabilités égales ». On distingue deux types de tirages : **Avec** et **Sans Remise**.

1. Plan de sondage aléatoire simple sans remise (PESR) :

Un plan de sondage PESR (Probabilités Egales Sans Remises) de taille dans une population de tailleest l’équivalent du tirage sans remise et avec des probabilités égales de boules d’une urne contenant boules. Il y a alors échantillons possibles.

Dans ce cas, un même individu ne peut apparaître qu’une seule fois dans l’échantillon.

On s’intéresse souvent à ce type de plan de sondage (sans remise), car interroger deux fois le même individu n’apporte pas d’informations supplémentaire.

**Remarque :**

* Les données concernant la population toute entière () sont déterministes et inconnues (puisque l’on a pas accès aux informations concernant toute la population.
* En revanche, les valeurs obtenues à partir de l’échantillon sont aléatoires car elles dépendent du hasard puisqu’elles varient d’un échantillon aléatoire à un autre, et elles sont connues puisque l’on dispose des informations nécessaires pour les calculer sur l’échantillon.

Le tableau suivant récapitule les notions relatives à la population et à l’échantillon :

|  |  |  |
| --- | --- | --- |
|  | Population | Echantillon |
| Taille |  |  |
| Moyenne |  |  |
| Total |  |  |
| Variance |  |  |
| Variance corrigée |  |  |

* + 1. Loi de probabilité

On prélève un échantillon de individus suivant un plan de sondage PESR dans une population et soit la variable aléatoire prenant la valeur , où est l’ensemble de tous les échantillons de individus possibles avec un tel plan de sondage.

Dans ce cas, La loi de probabilité de est uniforme (constante) :

**Interprétation :**

Il s’agit de prélever au hasard et simultanément individus de la population pour former un échantillon. Ou encore, prélever au hasard et un à un individus de la population pour former un échantillon, l’ordre n’étant pas pris en compte.

* + 1. Taux de sondage

On appelle le taux de sondage, le réel .

* + 1. Probabilités d’appartenance (ou d’inclusion)
* Pour , la probabilité que l’individu appartienne à est :
* Pour avec ≠ j, la probabilité que les individus et appartiennent à est

Avec

En effet, on utilise le fait que

Où, nombre de combinaisons de individus parmi et nombre de possibilités pour que soit dans l’échantillon (il est égal au nombre de possibilités de prélever individus parmi les autres individus que ).

**Exemple 1:**

Considérons la population dont le revenu de chaque individu est donné par

On veut interroger uniquement deux personnes de la population. On a alors six échantillons de tailles 2 sans remise.

Dans un plan de sondage PESR, les probabilités sont constantes et égales à 1/6.

i.e.

Le taux de sondage est de Où sont les probabilités d’inclusion de premier degré. Les probabilités d’inclusion de second degré sont ,

La moyenne des revenus dans la population est

Pour la moyenne des revenus calculée sur les échantillons est donnée dans le tableau suivant :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Alors, on peut avoir des « bons échantillons » ou des « mauvais ».

* + 1. Estimation des paramètres

Définition

Un estimateur d’un paramètre est une statistique (fonction de ) : , où est la valeur obtenue pour une réalisation de de, appelée estimation de .

Loi d’un estimateur

La loi d’un estimateur est la connaissance des couples pour tout .

En pratique, il est impossible de connaître la vraie loi de sinon on n’aurait pas eu besoin de faire un sondage.

On peut définir,

* L’espérance de est .
* La variance de est .
* La covariance : .

Qualité d’un estimateur

La qualité d’un estimateur est jugé à travers :

* Le biais : .

On préfèrera sans biais ou peu biaisé.

* La variance : , on choisira un estimateur qui a une plus petite variance.
* L’erreur quadratique moyenne :

Qui doit être petite.

* Le coefficient de variation : .

4.1. Estimation aléatoire de la moyenne de la population

Un estimateur aléatoire de la moyenne de la population est donné par

Où , est la fonction indicatrice définie sur .

* 1. 1. Espérance de l’estimateur

L’estimateur est un estimateur sans biais de . D’où,

En effet,

* + 1. Variance de l’estimateur

La variance de l’estimateur est

En effet,

Et en utilisant le fait que :

Et

D’où on obtient :

Le résultat est obtenu en remplaçant

Avec,

* + 1. Erreur Quadratique Moyenne de

L’erreur quadratique moyenne de l’estimateur est le réel

C’est une mesure de l’erreur commise en estimant par

**Remarque :**

* La moyenne empirique est souvent notée , et son estimateur est noté .
* Plus l’échantillon est grand ( est grand), plus la variance de diminue et donc plus l’estimateur est précis. A l’extrême, si alors la variance est nulle. Dans ce cas on a réalisé un recensement et on connaît de façon certaine la vraie moyenne.
* Plus la population est homogène (variance faible), plus est petit et plus l’estimateur est bon et le sondage y est efficace. A l’extrême, si est nul (tous les individus ont le même critère), alors la variance de est nulle et nous aurons besoin d’un seul individu pour connaître µ de manière parfaite. A l’inverse, sonder dans une population très hétérogène nécessite des échantillons de taille importante, ou un découpage au préalable en sous populations homogènes (c’est le principe des sondages stratiﬁés que nous verrons dans le chapitre 3).
  1. Estimation aléatoire du total

Un estimateur aléatoire du total est :

Il est sans biais et de variance

,

où,  **(**la variance corrigée de dans la population), qui peut être estimée par

où, est l’estimateur aléatoire de la variance de la population .

* 1. Estimation aléatoire de la variance de la population

Un estimateur aléatoire de la variance de la population est :

C’est la variance corrigée de dans l’échantillon.

* + 1. L’espérance de **:**

L’estimateur de est un estimateur sans biais.

i.e.

En effet :

Où, et

D’où,

Le résultat est obtenu en remplaçant et par leurs expressions.

**Exemple :**

La moyenne empirique des revenus (voir l’exemple 1) est l’estimateur qui est une variable aléatoire de loi uniforme (constante, car on est dans un plan PERS) :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| Probabilités |  |  |  |  |  |  |

On peut alors calculer l’espérance et la variance de :

et

1. Probabilité d’erreur

La probabilité de se tromper de plus de en estimant par , est le réel

1. Intervalle de confiance pour

* Si la variance corrigée est connue, alors l’intervalle de confiance pour au niveau de confiance , est

Où est le quantile d’ordre de la loi normale centrée réduite et la demi longueur de l’intervalle de confiance , appelée **incertitude absolue**, de au niveau de confiance

Le nombre est appelé **incertitude relative** de l’intervalle de confiance de au niveau de confiance

* Si la variance corrigée est inconnue, on la remplace par son estimation.

Théorème limite de Hajek :

Si et sont suffisamment grands, alors on a

D’où,

L’intervalle de confiance pour au niveau de confiance , est alors

1. Estimation de la proportion (cas particulier d’une moyenne) :

On peut s’intéresser à une proportion d’individus qui vérifie un certain critère Y. Dans ce cas, la variable prendra deux valeurs possibles :

La proportion d’individus appartenant à la catégorie qui nous intéresse sera alors

Une proportion peut être considérée comme un cas particulier de la moyenne.

**Remarque :**

On peut facilement écrire

* (qui est pour grand).
* avec .
* et **.**

**Exemple :**

Considérons le cas sondage électoral. On s’intéresse à la proportion d’individus votant pour un candidat A. On définit alors la variable qui prend deux valeurs :

Le nombre d’individus qui votent pour A est . On en déduit la proportion d’individus qui votent pour A est .

1. Taille de l’échantillon

Le choix de la taille de l’échantillon peut être effectué de deux façons différentes de façon à avoir une incertitude inférieure ou égale à une valeur fixée.

Soit (respectivement ) une incertitude absolue (respectivement relative) de au niveau de confiance

La taille de l’échantillon est alors choisie tel que

* ou

Où et sont des nombres réels fixés.

**Remarque :**

Pour des populations de grande taille, c’est la taille de l’échantillon qui donne la précision et non le taux de sondage.

1. Plan de sondage aléatoire simple avec remise (PEAR) :

Un plan de sondage PEAR (Probabilités Egales Avec Remises) de taille dans une population de tailleest l’équivalent du tirage avec remise et avec des probabilités égales de boules d’une urne contenant boules. Il y a alors échantillons possibles.

Dans ce cas, un même individu peut apparaître plusieurs fois dans l’échantillon.

* + 1. Loi de probabilité

On prélève un échantillon de individus suivant un plan de sondage simple PEAR dans une population et soit la variable aléatoire prenant la valeur , où est l’ensemble de tous les échantillons de individus possibles avec un tel plan de sondage.

Dans ce cas, La loi de probabilité de est uniforme (constante) :

**Interprétation :**

Il s’agit de prélever au hasard et avec remise individus de la population pour former un échantillon. Chaque individu a chances d’être sélectionné (donc la même probabilité qu’un autre d’être sélectionné). Par principe de multiplication on obtient façons différentes de sélectionner individus choisis avec remise parmi les individus de la population.

Par ce principe on calcule les différents paramètres en adoptant les formules obtenues dans le cas PESR à ce cas de plan de sondage.

Cette démarche est intéressante quand est petit ou pour servir d’élément de comparaison avec une situation de type PESR.

* + 1. Probabilités d’appartenance (ou d’inclusion)
* Pour , la probabilité que l’individu appartienne à est :
* Pour avec ≠ j, la probabilité que les individus et appartiennent à est

Avec

En effet, on utilise le fait que

et que ,

où, et nombre de possibilités de choisir, pour chacun des prélèvements, un individu parmi les autres que .

D’un autre côté on a aussi,

Avec

et

D’où,

* + 1. Estimation aléatoire de

Un estimateur aléatoire de la moyenne de la population est donné par

Où

* 1. Espérance de l’estimateur

L’estimateur est un estimateur sans biais de . D’où,

En effet,

* 1. Variance de l’estimateur

La variance de l’estimateur est

En effet,

car les sont des variables indépendantes. Où,

* + 1. Erreur Quadratique Moyenne de

L’erreur quadratique moyenne de l’estimateur est le réel

C’est une mesure de l’erreur commise en estimant par

**Remarque :**

* Plus l’échantillon est grand ( est grand), plus l’estimation est bonne.
* Plus la population est homogène (variance faible), plus l’estimation est bonne.
* L’estimation de par est plus précise pour un plan de sondage PESR que pour un plan de sondage PEAR.

En effet on a :

D’où,

On conclut que l’erreur est plus petite lorsque le plan de sondage est PESR.

* 1. Estimation aléatoire de la variance de la population

Un estimateur aléatoire de la variance de la population est :

C’est la variance corrigée de dans l’échantillon.

* + 1. L’espérance de **:**

L’espérance de est donnée par

En effet on a :

En utilisant le fait que

et , on obtient :

D’où,

Le résultat est obtenu en remplaçant

et

* 1. Estimation aléatoire du total

Un estimateur aléatoire du total est :

Il est sans biais et de variance

,

où,  **(**la variance corrigée de dans la population), qui peut être estimée par

1. Probabilité d’erreur

La probabilité de se tromper de plus de en estimant par , est le réel

1. Intervalle de confiance pour

* Si est une variable de loi normale alors

Où est le degré de liberté de la Student au seuil .

Pour assez grand

L’intervalle de confiance pour au niveau de confiance , est :

Où est le quantile d’ordre de la loi Student à n-1 degré de liberté, et la demi longueur de l’intervalle de confiance , appelée **incertitude absolue**, de au niveau de confiance Le nombre est l’**incertitude relative** de l’intervalle de confiance de au niveau de confiance

* Si est suffisamment grand, on a l’approximation

L’intervalle de confiance pour au niveau de confiance , est alors

Où est le quantile d’ordre de la loi Normale N(0,1).

1. Estimation de la proportion (cas particulier d’une moyenne) :

La proportion d’individus dans la population vérifiant un certain caractère est

C’est un cas particulier de la moyenne.

Dans ce cas on a :

* est sans biais de .
* .
* **.**
* est le quantile d’ordre de la loi N(0,1).

1. Taille de l’échantillon

Soit (respectivement ) une incertitude absolue (respectivement relative) de au niveau de confiance

La taille de l’échantillon est alors choisie tel que

* ou

Où et sont des nombres réels fixés.