

Chapitre 1

Mise en œuvre des méthodes de tirage aléatoire usuelles

1. Introduction

L'objet de ce chapitre 1 est de montrer comment les méthodes de tirage aléatoire usuelles sont mises en œuvre lorsque l'on ne recourt pas à un logiciel de tirage aléatoire, mais plutôt à une calculatrice ou à un tableur comme Excel sans usage de macros ou de programmes. Ce qui est visé est la mise en œuvre manuelle des principales méthodes courantes de tirage aléatoire.

Les méthodes usuelles de tirage aléatoire suivantes sont présentées dans ce chapitre :

- le tirage aléatoire simple avec remise ;
- le tirage aléatoire simple sans remise ;
- le tirage systématique simple ou à probabilités égales ;
- le tirage avec probabilités inégales et avec remise ;
- le tirage systématique à probabilités proportionnelles aux tailles des unités.

En réalité, l'objectif visé est de faire apparaître au grand jour le contenu des boîtes noires que sont les logiciels de tirage aléatoire d'échantillons.

Le tableur Excel sera notre principal outil de calcul dans ce chapitre, sans recours toutefois à des macros ou à des programmes développés sous ce tableur.

La génération de nombres aléatoires issus de la loi uniforme sur l'intervalle $[0,1[$ sera le procédé par lequel nous simulerons le hasard au cours de la mise en œuvre des méthodes de tirage aléatoire d'usage courant présentées dans ce chapitre. Il sera fait appel encore une fois au tableur Excel pour générer de tels nombres aléatoires grâce à la fonction Alea().

Que ce soit au moyen d'un logiciel de tirage aléatoire ou manuellement comme dans ce chapitre, la mise en œuvre des méthodes de tirage aléatoire repose sur l'utilisation de nombres générés de façon aléatoire tout comme c'était le cas avec l'utilisation des tables de nombres au hasard avant l'avènement des calculatrices et de la micro informatique.

2. Principe de la génération des nombres aléatoires

De nombreux logiciels font appel à la fonction Rnd pour générer un nombre aléatoire. Avec le logiciel Excel, un nombre aléatoire se génère au moyen de la fonction Alea(). L'exécution de la fonction Alea() fournit un nombre aléatoire qui obéit la loi uniforme sur l'intervalle $[0,1[$. A partir d'un tel nombre, on peut obtenir tout autre nombre aléatoire répondant à des conditions données.

L'on inscrit le signe = dans une cellule d'une feuille Excel en le faisant suivre de Alea(). En faisant Entrée, l'on voit apparaître un nombre compris entre 0 et 1.

Quand on génère un ensemble de nombres aléatoires d'un même type, c'est-à-dire répondant aux mêmes conditions, on génère un échantillon aléatoire de nombres de ce type.

Tableau 1 : Echantillons de nombres aléatoires entiers de K chiffres associés à un échantillon de nombres aléatoires générés sur [0,1[selon la loi uniforme

| Numéro d'ordre | Nombre aléatoire généré sur [0,1[| Nombre aléatoire entier associé de | | | Nombre aléatoire entier associé compris entre | |
|----------------|-----------------------------------|------------------------------------|------------|------------|---|-----------|
| | | 1 chiffre | 2 chiffres | 3 chiffres | 1 et 63 | 59 et 645 |
| 1 | 0,4826220 | 4 | 48 | 482 | 48 | 482 |
| 2 | 0,7481210 | 7 | 74 | 748 | | 748 |
| 3 | 0,4609711 | 4 | 46 | 460 | 46 | 460 |
| 4 | 0,9188478 | 9 | 91 | 918 | | 918 |
| 5 | 0,6048918 | 6 | 60 | 604 | 60 | 604 |
| 6 | 0,7329880 | 7 | 73 | 732 | | 732 |
| 7 | 0,0810524 | 0 | | | 8 | 81 |
| 8 | 0,2219748 | 2 | 22 | 221 | 22 | 221 |
| 9 | 0,6073915 | 6 | 60 | 607 | 60 | 607 |
| 10 | 0,0558495 | 0 | | | 5 | |
| 11 | 0,6933713 | 6 | 69 | 693 | | 693 |
| 12 | 0,3433434 | 3 | 34 | 343 | 34 | 343 |
| 13 | 0,9388199 | 9 | 93 | 938 | | 938 |
| 14 | 0,0498876 | 0 | | | 4 | |
| 15 | 0,7348286 | 7 | 73 | 734 | | 734 |
| 16 | 0,3412936 | 3 | 34 | 341 | 34 | 341 |
| 17 | 0,5214479 | 5 | 52 | 521 | 52 | 521 |
| 18 | 0,2123699 | 2 | 21 | 212 | 21 | 212 |
| 19 | 0,0865821 | 0 | | | 8 | 86 |
| 20 | 0,5569105 | 5 | 55 | 556 | 55 | 556 |
| 21 | 0,2207015 | 2 | 22 | 220 | 22 | 220 |
| 22 | 0,8723084 | 8 | 87 | 872 | | 872 |
| 23 | 0,3315625 | 3 | 33 | 331 | 33 | 331 |
| 24 | 0,7958014 | 7 | 79 | 795 | | 795 |
| 25 | 0,6602012 | 6 | 66 | 660 | | 660 |
| 26 | 0,1127591 | 1 | 11 | 112 | 11 | 112 |
| 27 | 0,4986080 | 4 | 49 | 498 | 49 | 498 |
| 28 | 0,4159506 | 4 | 41 | 415 | 41 | 415 |
| 29 | 0,2365607 | 2 | 23 | 236 | 23 | 236 |
| 30 | 0,0395867 | 0 | | | 3 | |

La colonne 2 du tableau 1 ci-dessous contient un échantillon de 30 nombres aléatoires générés sur l'intervalle [0,1[selon la loi uniforme.

On déduit de cet échantillon de 30 nombres générés sur l'intervalle [0,1[, les échantillons aléatoires associés de nombres entiers respectivement de 1 chiffre, de 2 chiffres et de 3 chiffres.

L'échantillon aléatoire des nombres de 1 chiffre est obtenu en multipliant les nombres de la colonne 2 du tableau 1 par 10 et en considérant la partie entière des résultats. Les échantillons des nombres de 2 chiffres et de 3 chiffres sont obtenus en considérant la partie entière des nombres de la colonne 2 multipliés respectivement par 10^2 et 10^3 . Les échantillons aléatoires de nombres de 1, 2 et 3 chiffres figurent dans le tableau 1 respectivement dans les colonnes 3, 4 et 5.

Le tableau 1 comporte également deux types de nombres aléatoires entiers associés aux 30 nombres générés sur l'intervalle [0,1[. Ce sont des nombres entiers compris entre 1 et 63 et des nombres entiers compris entre 59 et 645. Ces nombres aléatoires sont obtenus en considérant la

partie entière des 30 nombres de la colonne 2 multipliés au préalable par 10^2 et 10^3 respectivement.

Ces nombres générés entre 1 et 63 ou entre 59 et 645 sont deux exemples de catégories particulières de nombres aléatoires que l'on peut souhaiter obtenir.

Remarque

D'une manière générale, l'on peut générer un nombre aléatoire X d'un intervalle (A, B) suivant la loi uniforme sur cet intervalle, à l'aide de la relation

$$X = A + (B - A) * \text{Alea}()$$

S'il s'agit de la génération d'un nombre entier X compris entre deux nombres entiers A et B , avec A inférieur à B , la formule suivante s'applique

$$X = \text{ENT}(A + (B - A + 1) * \text{Alea}())$$

ENT() étant la fonction donnant la partie entière d'un nombre.

L'ajout du nombre 1 au 2^e membre de la formule assure que le nombre B ne sera pas exclu du processus de la génération.

Le tableau 2 ci-après donne un échantillon de nombres entiers aléatoires compris entre 10 et 25.

Tableau 2 : Echantillon de 120 nombres entiers aléatoires compris entre 10 et 25

| Ligne | Colonne | | | | | | | |
|-------|---------|----|----|----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | 21 | 12 | 23 | 11 | 22 | 18 | 25 | 21 |
| 2 | 10 | 13 | 10 | 12 | 12 | 14 | 25 | 23 |
| 3 | 19 | 15 | 18 | 24 | 14 | 10 | 24 | 15 |
| 4 | 16 | 25 | 22 | 17 | 25 | 18 | 12 | 16 |
| 5 | 17 | 15 | 25 | 25 | 17 | 13 | 19 | 12 |
| 6 | 19 | 24 | 11 | 18 | 11 | 15 | 23 | 24 |
| 7 | 13 | 16 | 17 | 12 | 20 | 18 | 10 | 16 |
| 8 | 19 | 19 | 22 | 12 | 25 | 25 | 10 | 23 |
| 9 | 18 | 16 | 14 | 19 | 19 | 23 | 19 | 16 |
| 10 | 10 | 13 | 10 | 19 | 25 | 16 | 11 | 23 |
| 11 | 23 | 22 | 15 | 12 | 24 | 18 | 21 | 25 |
| 12 | 10 | 13 | 16 | 17 | 14 | 13 | 18 | 25 |
| 13 | 15 | 11 | 11 | 12 | 16 | 16 | 20 | 23 |
| 14 | 22 | 20 | 21 | 10 | 17 | 17 | 20 | 20 |
| 15 | 17 | 16 | 13 | 21 | 16 | 23 | 21 | 16 |

3. Les différents modes de tirage d'échantillon

Deux paramètres sont essentiels dans la définition du mode de tirage d'un échantillon : l'égalité ou non des probabilités de tirage attribuées aux unités de l'univers et la possibilité ou non de remise pour les unités tirées.

Un tirage avec probabilités égales attribue à l'avance une même probabilité à chaque unité statistique de l'univers sondé. Dans le cas où des probabilités inégales sont attribuées aux unités statistiques, on a affaire à un tirage avec probabilités inégales.

Les tirages ont lieu avec remise lorsque, à chaque tirage, l'unité statistique obtenue est remise dans l'univers avant le tirage suivant. C'est le cas des tirages où une même unité statistique peut figurer plus d'une fois dans l'échantillon.

Avec les tirages sans remise, une unité statistique tirée n'est pas remise dans l'univers avant les tirages suivants. Dans ce cas, une unité statistique ne peut figurer plus d'une fois dans l'échantillon.

Les deux modes de tirage d'échantillon les plus simples sont ceux associés au sondage aléatoire simple. Un sondage aléatoire simple (simple random sampling) est une méthode de tirage qui consiste à tirer dans une population finie de M unités, un échantillon de taille fixée m à partir des seuls identifiants des unités de telle façon que chaque unité ait la même probabilité d'appartenir à l'échantillon.

Le tirage aléatoire simple peut être avec remise ou sans remise. Et dans l'un ou l'autre cas, il s'agit de sondage à probabilités égales ou de tirage avec probabilités égales.

Au sujet du tirage aléatoire simple, on fait quelquefois appel aux symboles "PEAR" et "PESR". Le tirage PEAR¹ est le tirage avec probabilités égales et avec remise tandis que le tirage PESR² est le tirage avec probabilités égales et sans remise. PEAR est mis pour "à probabilités égales et avec remise" et PESR pour "à probabilités égales et sans remise"

Dans toute la suite, le tirage PESR se rapporte exclusivement au sondage aléatoire simple sans remise et non pas au tirage systématique avec probabilités égales.

Les autres modes de tirage courants rencontrés sont :

- le tirage avec probabilités inégales et avec remise ou tirage PIAR³ ;
- le tirage systématique à probabilités égales ou tirage SPE⁴ ;
- le tirage systématique avec probabilités inégales, et le cas particulier de ce dernier, qui est le tirage systématique à probabilités proportionnelles aux tailles des unités ou tirage SPPT⁵.

Bien souvent, les probabilités inégales attribuées aux unités par un plan de sondage sont des probabilités proportionnelles à la taille des unités. Si les unités de l'univers sont, par exemple, des villages, le nombre de ménages du village peut servir de mesure de taille pour les villages.

¹ PEAR est mis pour "à Probabilités Egales et Avec Remise" selon Jean-Marie Grosbras

² PESR est mis pour "à Probabilités Egales et Sans Remise" selon Jean-Marie Grosbras

³ PIAR est mis pour "à Probabilités inégales et avec Remise" selon Jean-Marie Grosbras

⁴ SPE est mis pour "Systématique à probabilités égales". Il est suggéré par Julien Amegandjin

⁵ SPPT est mis pour "Systématique à probabilités proportionnelles aux tailles des unités". Il est proposé par Julien Amegandjin

Le tirage systématique à probabilités égales (ou tirage SPE) est une variante du tirage aléatoire simple sans remise pour ce qui est de la constitution de l'échantillon bien que les deux méthodes de tirage ne soient pas totalement équivalentes parce que correspondant à des modèles probabilistes différents même si les probabilités d'inclusion sont identiques.

4. Principe des tirages aléatoires

Soit U un univers composé de M unités et duquel on désire sélectionner au hasard un échantillon de m unités ou un échantillon de taille m . Les M unités statistiques de l'univers sont numérotées de 1 à M et ces numéros constituent les identités ou étiquettes des unités.

Tirer un échantillon aléatoire de taille m de l'univers U consistera simplement à générer à l'aide de la fonction Alea() de Excel, m nombres compris entre 1 et M qui identifieront les unités sélectionnées. La génération des m nombres tiendra compte bien évidemment des spécificités du mode particulier de tirage considéré.

5. Tirage aléatoire simple avec remise de m unités parmi M

Comment mettre en œuvre le tirage d'un échantillon de m unités au sein d'un univers de M unités par tirage aléatoire simple avec remise ou tirage PEAR?

Supposons $m = 10$ et $M = 80$. Nous allons construire un tableau semblable au tableau 1 dans une feuille Excel.

Etape 1

Je génère au hasard un nombre de l'intervalle $[0,1[$. Pour ce faire, j'inscris dans la cellule C4 la formule " $=Alea()$ ". En tapant "Entrée", la valeur du nombre généré apparaît dans cette cellule C4. Mais cette valeur change au moindre déplacement de la souris ou à la moindre action sur la feuille Excel. Mais, il n'est besoin pas se soucier de cette situation. A tout moment, l'on considérera la valeur affichée dans cette cellule.

Etape 2

Je copie la valeur de la cellule C4 sur la plage C5....C33 des cellules de C5 à C33. J'obtiens 29 autres nombres de l'intervalle $[0,1[$. Ces valeurs changent aussi à la suite de toute action de la souris ou de toute autre action sur la feuille Excel.

Etape 3

Je sélectionne la plage C4....C33 des 30 nombres générés. Je fais une copie de cette plage que je colle sur la plage D4....D33 par "Collage spécial" et avec l'option "Valeurs". Cette dernière opération permet de fixer les 30 nombres générés sur l'intervalle $[0,1[$ et éviter qu'ils changent de valeurs à la suite de toute action sur la feuille Excel.

La colonne 2 du tableau 2 ci-après donne les nombres de la plage D4....D33 de la feuille Excel. Les nombres aléatoires entiers de 2 chiffres associés figurent dans la colonne 3 et les nombres aléatoires entiers compris entre 1 et 80 dans la colonne 4.

L'échantillon PEAR de 10 unités tirées d'un univers de 80 unités est constitué par les 10 premiers nombres de la colonne 4. Le rang de tirage de ces unités figure dans la colonne 6. L'on observe que l'unité d'étiquette 30 est tirée deux fois aux rangs de tirages 1 et 10, ce qui est admissible pour ce mode de tirage.

Tableau 2 : Données relatives à un tirage aléatoire simple avec remise ou tirage PEAR

| Numéro d'ordre | Nombre aléatoire généré sur [0,1[| Nombre aléatoire entier associé de 2 chiffres | Nombre aléatoire entier associé compris entre 1 et 80 | Unités tirées | Rang de tirage |
|----------------|-----------------------------------|---|---|---------------|----------------|
| 1 | 0,307897807 | 30 | 30 | X | 1 |
| 2 | 0,882637114 | 88 | | | |
| 3 | 0,163529825 | 16 | 16 | X | 2 |
| 4 | 0,325582103 | 32 | 32 | X | 3 |
| 5 | 0,147365164 | 14 | 14 | X | 4 |
| 6 | 0,396668314 | 39 | 39 | X | 5 |
| 7 | 0,608662078 | 60 | 60 | X | 6 |
| 8 | 0,103592657 | 10 | 10 | X | 7 |
| 9 | 0,207631746 | 20 | 20 | X | 8 |
| 10 | 0,991539945 | 99 | | | |
| 11 | 0,810167633 | 81 | | | |
| 12 | 0,195645243 | 19 | 19 | X | 9 |
| 13 | 0,303625037 | 30 | 30 | X | 10 |
| 14 | 0,700965959 | 70 | 70 | | |
| 15 | 0,981142758 | 98 | | | |
| 16 | 0,120391448 | 12 | 12 | | |
| 17 | 0,192937792 | 19 | 19 | | |
| 18 | 0,310835235 | 31 | 31 | | |
| 19 | 0,618981082 | 61 | 61 | | |
| 20 | 0,288166912 | 28 | 28 | | |
| 21 | 0,002201497 | | | | |
| 22 | 0,251271934 | 25 | 25 | | |
| 23 | 0,988234843 | 98 | | | |
| 24 | 0,956176622 | 95 | | | |
| 25 | 0,460668679 | 46 | 46 | | |
| 26 | 0,13729825 | 13 | 13 | | |
| 27 | 0,347912525 | 34 | 34 | | |
| 28 | 0,895085467 | 89 | | | |
| 29 | 0,352825601 | 35 | 35 | | |
| 30 | 0,445976471 | 44 | 44 | | |

6. Tirage aléatoire simple sans remise de m unités parmi M

Pour mettre en œuvre le tirage d'un échantillon de m unités au sein d'un univers de M unités par tirage aléatoire simple sans remise ou tirage PESR, l'on procèdera comme pour le tirage PEAR.

Supposons $m = 10$ et $M = 375$. Nous allons construire un tableau semblable au tableau 2 dans une feuille Excel, soit le tableau 3 ci-après.

Les nombres aléatoires générés sur l'intervalle $[0,1[$ figurent dans la colonne 2 du tableau 3 et les nombres aléatoires entiers associés compris entre 1 et 375 dans la colonne 4. Les 10 unités tirées sans remise sont celles auxquelles il est associé une croix dans la colonne 5. L'on observe que l'unité d'étiquette 177 apparaît deux fois dans la colonne 4 en positions 1 et 9. Cette unité ne sera considérée qu'une fois parce que nous avons affaire à un tirage sans remise. Le rang du tirage figure dans la colonne 6.

Tableau 3 : Données relatives à un tirage aléatoire simple sans remise ou tirage PESR

| Numéro d'ordre | Nombre aléatoire généré sur [0,1[| Nombre aléatoire entier associé de 3 chiffres | Nombre aléatoire entier associé compris entre 1 et 375 | Unités tirées | Rang de tirage |
|----------------|-----------------------------------|---|--|---------------|----------------|
| 1 | 0,177515712 | 177 | 177 | X | 1 |
| 2 | 0,021340095 | 21 | 21 | X | 2 |
| 3 | 0,548994785 | 548 | | | |
| 4 | 0,910373212 | 910 | | | |
| 5 | 0,808572085 | 808 | | | |
| 6 | 0,603510384 | 603 | | | |
| 7 | 0,325109763 | 325 | 325 | X | 3 |
| 8 | 0,974764376 | 974 | | | |
| 9 | 0,626717369 | 626 | | | |
| 10 | 0,109084937 | 109 | 109 | X | 4 |
| 11 | 0,451126846 | 451 | | | |
| 12 | 0,827157863 | 827 | | | |
| 13 | 0,254146693 | 254 | 254 | X | 5 |
| 14 | 0,938538974 | 938 | | | |
| 15 | 0,037549234 | 37 | 37 | X | 6 |
| 16 | 0,882742515 | 882 | | | |
| 17 | 0,587661062 | 587 | | | |
| 18 | 0,054374701 | 54 | 54 | X | 7 |
| 19 | 0,288302 | 288 | 288 | X | 8 |
| 20 | 0,851023977 | 851 | | | |
| 21 | 0,177760867 | 177 | 177 | | |
| 22 | 0,963343172 | 963 | | | |
| 23 | 0,770359659 | 770 | | | |
| 24 | 0,575858946 | 575 | | | |
| 25 | 0,077539823 | 77 | 77 | X | 9 |
| 26 | 0,800892169 | 800 | | | |
| 27 | 0,746510673 | 746 | | | |
| 28 | 0,951098603 | 951 | | | |
| 29 | 0,788458082 | 788 | | | |
| 30 | 0,061441109 | 61 | 61 | X | 10 |
| 31 | 0,087577996 | 87 | 87 | | |
| 32 | 0,439704783 | 439 | | | |
| 33 | 0,460676317 | 460 | | | |
| 34 | 0,134209019 | 134 | 134 | | |
| 35 | 0,140054433 | 140 | 140 | | |

7. Tirage PIAR d'un échantillon de m unités parmi M

Soit U un univers composé de M unités statistiques d'étiquettes 1, 2, ..., r, ..., M et de tailles respectives $X_1, X_2, \dots, X_r, \dots, X_M$. La taille est souvent l'effectif des exploitations agricoles, l'effectif de la population ou l'effectif des ménages lorsque les unités statistiques sont, par exemple, les villages d'une région.

Un cas classique de tirage PIAR (tirage avec probabilités inégales et avec remise) d'un échantillon d'unités au sein d'un tel univers consiste à prendre les probabilités inégales proportionnelles aux tailles X_r des unités statistiques. C'est ce cas classique de tirage que nous considérerons chaque fois qu'il sera question de tirage PIAR.

Soit C_r la somme des tailles des r premières unités statistiques $1, 2, \dots, r$ de l'univers U . C_r est également appelé le cumul des tailles des r premières unités. Il est défini par la relation

$$C_r = \sum_{j=1}^r X_j$$

Nous désignerons par C le cumul total des tailles ou le cumul des tailles pour l'ensemble des unités de l'univers U , soit

$$C = C_M = \sum_{j=1}^M X_j$$

Pour mettre en œuvre le tirage PIAR, on établit au préalable une partition de l'ensemble des nombres entiers $\{1, 2, \dots, C\}$ en M éléments et on associe l'unité r de l'univers à l'élément r de la partition défini par le sous-ensemble des nombres entiers $\{1+C_{r-1}, \dots, C_r\}$.

Le tableau ci-après présente les caractéristiques de la partition. On note que par construction des grandeurs C_h , l'effectif des nombres entiers de l'élément r de la partition est égal à la taille X_r .

| <u>Unité de l'univers</u> | <u>Élément de la partition</u> | <u>Effectif des nombres de la partition</u> |
|-------------------------------|------------------------------------|---|
| 1 | $\{1, 2, \dots, C_1\}$ | X_1 |
| 2 | $\{1 + C_1, \dots, C_2\}$ | X_2 |
| | | |
| r | $\{1 + C_{r-1}, \dots, C_r\}$ | X_r |
| | | |
| M | $\{1 + C_{M-1}, \dots, C_M\}$ | X_M |

Sélectionner une unité de l'univers U consiste alors à tirer un des M éléments de la partition avec des probabilités proportionnelles aux tailles X_r de ces éléments. Cela revient à tirer au hasard, avec probabilités égales, un nombre compris entre 1 et C , puis à identifier ce nombre avec l'élément de la partition et l'unité de l'univers qui lui correspondent.

Pour obtenir l'échantillon PIAR de taille m , on répète le procédé précédent m fois avec une possibilité de remise pour les tirages.

Exemple d'application

Considérons une région d'un pays dénommée ABCD. Cette région fournit une base de sondage constituée de 20 villages imaginaires dont le nombre de ménages est connu pour chacun. Le nombre de ménages joue le rôle de variable de taille pour les villages. Tirons un échantillon de 4 villages selon le mode PIAR de cette base de sondage. Le tableau 4 ci-dessous contient les données relatives au tirage réalisé.

Le cumul total C prend la valeur 3 640 et les 4 nombres aléatoires tirés avec remise entre 1 et 3 640 inclus sont 515, 1241, 1541 et 3090 comme le montre le tableau 5.

Tableau 4 : Données d'un tirage PIAR de 4 unités de la région ABCD

| Identité du village | Nom du village | Effectif des ménages | Cumul des tailles | Nombre tiré | Unité tirée |
|---------------------|----------------|----------------------|-------------------|-------------|-------------|
| 8121 | Assabe | 136 | 136 | | |
| 8122 | Atal | 240 | 376 | | |
| 8123 | Gorgui | 115 | 491 | | |
| 8124 | Kaka | 127 | 618 | 515 | x |
| 8125 | Kaporo | 181 | 799 | | |
| 8126 | Kipé | 184 | 983 | | |
| 8127 | Lalo | 235 | 1 218 | | |
| 8128 | Loum | 277 | 1 495 | 1 241 | x |
| 8129 | Madina | 211 | 1 706 | 1 541 | x |
| 8130 | Manga | 112 | 1 818 | | |
| 8131 | Mano | 210 | 2 028 | | |
| 8132 | Marsa | 143 | 2 171 | | |
| 8133 | Matoto | 223 | 2 394 | | |
| 8134 | Saga | 252 | 2 646 | | |
| 8135 | Sindaya | 266 | 2 912 | | |
| 8136 | Taouya | 102 | 3 014 | | |
| 8137 | Tilen | 126 | 3 140 | 3 090 | x |
| 8138 | Valo | 203 | 3 343 | | |
| 8139 | Vanta | 148 | 3 491 | | |
| 8140 | Vival | 149 | 3 640 | | |

Tableau 5 : Nombres tirés entre 1 et 3640

| Nombre aléatoire généré sur $[0,1[$ | Nombre entier associé compris entre 1 et 3640 |
|-------------------------------------|---|
| 0,005152965 | 515 |
| 0,012414681 | 1 241 |
| 0,015412267 | 1 541 |
| 0,030900443 | 3 090 |

Avec chacun de ces nombres tirés, comment identifier l'unité correspondante de la région ABCD ? Si Z est l'un des nombres tirés, on calcule successivement la différence $Z - C_r$ en parcourant les valeurs de C_r selon l'ordre croissant jusqu'à ce que la différence devienne négative. Le premier cumul qui rend négative cette différence est le cumul de l'unité correspondant au nombre Z tiré. On observe selon le tableau 4 que les 4 nombres aléatoires générés 515, 1241, 1541 et 3090 correspondent respectivement aux cumuls 618, 1495, 1706 et 3140. Les nombres tirés sont inscrits dans la colonne 5 de ce tableau en face du cumul C_r de l'unité r correspondante. Les unités tirées sont marquées d'une croix dans la colonne 6. Les villages de Kaka, Loum, Madina et Tilen constituent l'échantillon tiré.

8. Tirage systématique à probabilités égales d'un échantillon d'unités

Le tirage SPE ou tirage systématique à probabilités égales est une variante du tirage aléatoire simple sans remise, c'est-à-dire du tirage PESR. Si la constitution des échantillons conduit à des résultats équivalents pour les deux méthodes, ces dernières sont cependant différentes parce qu'elles correspondent à deux modèles probabilistes différents.

Soit U un univers de M unités statistiques d'étiquettes $1, 2, \dots, r, \dots, M$. L'on désire tirer de cet univers, un échantillon de taille m avec probabilités égales et sans remise selon le tirage SPE. Supposons M divisible par m avec

$$M = m k$$

k étant donc un nombre entier. Pour mettre en œuvre le tirage SPE, on répartit au préalable les unités de l'univers U en m groupes de taille k chacun. La méthode consiste alors à tirer avec probabilités égales, une unité de chacun des m groupes d'unités. On commence par tirer une unité du groupe n°1 constituée par l'ensemble $\{1, 2, \dots, k\}$. Si r est l'étiquette de l'unité tirée, avec $r \leq k$, les $m-1$ unités à retenir dans les autres groupes sont ceux dont les étiquettes sont les termes suivants de la progression arithmétique finie de raison k et de premier terme r , soit

$$r, r+k, r+2k, \dots, r+(m-1)k.$$

On parle de tirage systématique de raison k . Dans le cas où M est divisible par m , toutes les unités ont la même probabilité d'être tirée ou d'appartenir à l'échantillon.

Dans le cas où M n'est pas divisible par m et où k n'est donc pas un nombre entier, on ne peut pas constituer des groupes d'unités de même taille comme précédemment. Pour réaliser le tirage systématique avec probabilités égales, on adapte la méthode précédente en générant un nombre u selon la loi uniforme sur $[0, 1[$ et en prenant pour première unité tirée, celle qui a pour étiquette, le nombre $[ku]+1$, où $[ku]$ désigne la partie entière de ku . Et l'échantillon tout entier est donné comme suit :

| <u>Rang du tirage</u> | <u>Nombre généralisé</u> |
|---------------------------|------------------------------|
| 1 | $[ku] + 1$ |
| 2 | $[ku + k] + 1$ |
| 3 | $[ku + 2k] + 1$ |
| | |
| r | $[ku + (r-1)k] + 1$ |
| | |
| m | $[ku + (m-1)k] + 1$ |

Tout revient à considérer une partition de l'ensemble $\{1, 2, 3, \dots, M\}$ en m éléments qui sont les intervalles $]0, k],]k, 2k], \dots,](r-1)k, rk], \dots,](m-1)k, M]$ dont on considère les seuls nombres entiers. L'unité de rang r de l'échantillon, soit le nombre $[ku+(r-1)k] + 1$, est issue de l'élément $] (r-1)k, rk]$ de la partition.

Exemple d'application 1

Considérons la population $U = \{1, 2, \dots, 52\}$ et sélectionnons 6 unités de cette population selon le mode de tirage SPE. Nous avons $m = 6$, $k = 8,66$ et $u = 0,738$.

Les 6 éléments de la partition de l'ensemble des nombres $\{1, 2, \dots, 52\}$ sont les sous-ensembles $\{1, 2, \dots, 8\}$, $\{9, 10, \dots, 17\}$, $\{18, 19, \dots, 26\}$, $\{27, 28, \dots, 34\}$, $\{35, 36, \dots, 43\}$ et $\{44, 45, \dots, 52\}$. Les autres données du tirage sont fournies par le tableau 6 ci-après.

Les 6 éléments de la partition de l'ensemble des nombres $\{1, 2, \dots, 52\}$ sont les sous-ensembles $\{1, 2, \dots, 8\}$, $\{9, 10, \dots, 17\}$, $\{18, 19, \dots, 26\}$, $\{27, 28, \dots, 34\}$, $\{35, 36, \dots, 43\}$ et $\{44, 45, \dots, 52\}$. Les 6 unités tirées sont les nombres 7, 16, 24, 33, 42 et 50.

Tableau 6 : Données d'un tirage SPE de 6 unités de la population {1, 2, ..., 52}

| Rang du tirage r | $ku+(r-1)k$ | $[ku+(r-1)k]+1$ |
|--------------------------|-------------|-----------------|
| 1 | 6,39108 | 7 |
| 2 | 15,05108 | 16 |
| 3 | 23,71108 | 24 |
| 4 | 32,37108 | 33 |
| 5 | 41,03108 | 42 |
| 6 | 49,69108 | 50 |

Exemple d'application 2

Sélectionnons selon le mode de tirage SPE, un échantillon de 3 villages de la région dénommée ABCD du tableau 4. Les données de départ du tirage sont :

$$\begin{aligned}M &= 20 \\m &= 3 \\u &= 0,3914912\end{aligned}$$

Le tableau 7 ci-dessous contient les données du tirage SPE. Les villages de numéros 3, 10 et 16 ont été tirés, c'est-à-dire les villages Gorgui, Manga et Taouya.

Tableau 7 : Données d'un tirage SPE de 3 villages de la région ABCD

| Rang du tirage r | $ku+(r-1)k$ | $[ku+(r-1)k]+1$ |
|--------------------------|-------------|-----------------|
| 1 | 2,60994 | 3 |
| 2 | 9,27661 | 10 |
| 3 | 15,94327 | 16 |

9. Tirage systématique avec probabilités proportionnelles aux tailles des unités (ou tirage SPPT) : une version simplifiée de la procédure de Hartley et Rao

Les sondages avec probabilités inégales et sans remise mis en œuvre sont généralement des sondages utilisant la méthode du tirage systématique. Il existe de nombreuses procédures de tirage systématique avec probabilités inégales.

Soit U une population composée des M unités statistiques d'étiquettes 1, 2, ..., r , ..., M . Généralement, on définit la probabilité inégale P_r de tirage de l'unité d'étiquette r à partir d'une variable auxiliaire X_r connue pour toutes les unités de l'univers U et à peu près proportionnelle à la variable d'étude Y_r du sondage mis en œuvre. Si P_r est donc exactement proportionnel à X_r et que X_r est à peu près proportionnel à Y_r , on aura P_r à peu près proportionnel à Y_r .

Bien souvent, la variable auxiliaire X_r est une variable de taille pour les unités comme pour le tirage PIAR. C'est pourquoi on parle dans ce cas, de tirage systématique avec probabilités proportionnelles aux tailles des unités. Nous présentons ici la méthode de tirage systématique de Hartley et Rao⁶, une procédure de tirage systématique parmi les plus simples.

Soit U une population composée des M unités statistiques d'étiquettes $1, 2, \dots, r, \dots, M$ et de tailles respectives $X_1, X_2, \dots, X_r, \dots, X_M$. La taille des unités est une variable auxiliaire que l'on suppose connue pour toutes les unités de la population. On désire tirer un échantillon de m unités de la population U selon la procédure de tirage de Hartley et Rao.

Désignons par C la somme cumulée totale des tailles X_r , soit

$$C = \sum_{j=1}^M X_j$$

Le cumul total C doit être divisible par m , la taille de l'échantillon. Si tel n'est pas le cas, on multiplie toutes les tailles X_r par une constante de manière à rendre C divisible par m . On prend souvent cette constante égale à m . Cette transformation des tailles X_r et du cumul total C en les multipliant par m ne change rien à la procédure de tirage.

Le cumul total C étant divisible par m , il vient

$$C = m K$$

où K est un nombre entier. L'ensemble des nombres entiers $\{1, 2, \dots, C\}$ est alors subdivisé en m sous-ensembles comportant chacun K nombres.

Tirer m unités de la population U selon le tirage systématique avec probabilités proportionnelles aux tailles X_r , revient à tirer avec probabilités égales et sans remise, un nombre entier dans chacun des m sous-ensembles de K chiffres constitués et à identifier l'unité correspondant à ce nombre. Pour ce faire, on génère au départ un nombre entier compris entre 1 et K inclus. Si r est le nombre généré, les $m-1$ autres nombres tirés des $m-1$ autres sous-ensembles de nombres constitués sont les $m-1$ termes suivants de la suite arithmétique finie de raison K et de premier terme r . L'échantillon des m nombres tirés est alors constitué par l'ensemble $\{r, r+K, \dots, r+(m-1)K\}$. Chacun de ces m nombres tirés peut s'identifier à une unité de la population et à une seule.

En réalité, chacun des m sous-groupes donne lieu au tirage d'une seule unité si $K > X_r$ pour tout $r = 1, 2, \dots, M$. Si pour une unité r , $K < X_r$, alors cette unité risque d'être tirée plus d'une fois. On décide alors, soit de sélectionner d'office cette unité avec la probabilité 1 et de tirer $m-1$ autres unités parmi les autres $M-1$ unités restantes, soit de segmenter cette unité en 2 nouvelles unités ou plus avant les tirages.

Désignons par C_r le cumul des tailles pour les r premières unités de la population, soit

$$C_r = \sum_{j=1}^r X_j \quad \forall r = 1, 2, \dots, M$$

Notons que

$$C = C_M = \sum_{j=1}^M X_j$$

⁶ H. O. Hartley et J. K. N. Rao, Sampling with unequal probabilities without replacement, Annals of Mathematical Statistics, 1962, vol. 33

Comment identifier l'unité de la population qui correspond au nombre r tiré entre 1 et K ? On calcule successivement la différence $r - C_h$ en parcourant selon l'ordre croissant, les cumuls C_h jusqu'à ce que la différence devienne négative. Le premier cumul qui rend négative cette différence est le cumul de l'unité correspondant au nombre r tiré. Le nombre $r + (j-1)K$ tiré du sous-ensemble $\{1 + (j-1)K, \dots, jK\}$ correspond lui à l'unité dont le cumul C_h rend négative le premier, la différence $r + (j-1)K - C_h$ lorsque ces cumuls sont parcourus dans l'ordre croissant.

Exemple d'application

Sélectionnons selon le mode de tirage SPPT, un échantillon de 3 villages de la région des 20 villages dénommée ABCD. La taille des villages est la variable "Nombre_Ménages" du tableau 8. Les tailles X_r sont multipliées par la taille de l'échantillon, soit par 3. Et nous obtenons les données de départ suivantes :

$$\begin{aligned} m &= 3 \\ C &= 10920 \\ K &= 3640 \end{aligned}$$

Tableau 8 : Données d'un tirage SPPT de 3 villages de la région ABCD

| Numéro d'ordre | Identité du village | Nom du village | Nombre de ménages | Cumul des tailles | Cumul modifié des tailles | Nombre tiré | Unité tirée |
|----------------|---------------------|----------------|-------------------|-------------------|---------------------------|-------------|-------------|
| 1 | 8121 | Assabe | 136 | 136 | 408 | | |
| 2 | 8122 | Atal | 240 | 376 | 1 128 | | |
| 3 | 8123 | Gorgui | 115 | 491 | 1 473 | | |
| 4 | 8124 | Kaka | 127 | 618 | 1 854 | | |
| 5 | 8125 | Kaporo | 181 | 799 | 2 397 | | |
| 6 | 8126 | Kipé | 184 | 983 | 2 949 | | |
| 7 | 8127 | Lalo | 235 | 1 218 | 3 654 | 3 293 | x |
| 8 | 8128 | Loum | 277 | 1 495 | 4 485 | | |
| 9 | 8129 | Madina | 211 | 1 706 | 5 118 | | |
| 10 | 8130 | Manga | 112 | 1 818 | 5 454 | | |
| 11 | 8131 | Mano | 210 | 2 028 | 6 084 | | |
| 12 | 8132 | Marsa | 143 | 2 171 | 6 513 | | |
| 13 | 8133 | Matoto | 223 | 2 394 | 7 182 | 6 933 | x |
| 14 | 8134 | Saga | 252 | 2 646 | 7 938 | | |
| 15 | 8135 | Sindaya | 266 | 2 912 | 8 736 | | |
| 16 | 8136 | Taouya | 102 | 3 014 | 9 042 | | |
| 17 | 8137 | Tilen | 126 | 3 140 | 9 420 | | |
| 18 | 8138 | Valo | 203 | 3 343 | 10 029 | | |
| 19 | 8139 | Vanta | 148 | 3 491 | 10 473 | | |
| 20 | 8140 | Vival | 149 | 3 640 | 10 920 | 10 573 | x |

Nous observons que la condition $K > X_r$ est vérifiée pour tout $r = 1, 2, \dots, M$. La procédure de tirage s'applique donc. Le nombre tiré entre 1 et K est le nombre 3293 et les autres nombres tirés sont 6933 et 10573. Les résultats du tirage sont résumés ci-après tandis que le tableau 8 fournit l'ensemble des données du tirage.

| <u>Nombre tiré</u> | <u>Sous-ensemble</u> | <u>Cumul associé</u> | <u>Village tiré</u> |
|--------------------|----------------------|----------------------|---------------------|
| 3293 | {1, 2, ..., 3640} | 3654 | Lalo |
| 6933 | {3641, ..., 7280} | 7182 | Matoto |
| 10573 | {7281, ..., 10920} | 10920 | Vival |

10. Méthode de tirage SPPT basée sur les probabilités d'inclusion d'ordre 1

La procédure du tirage SPPT peut être formulée autrement au moyen des probabilités d'inclusion des unités de la population sondée. Soit U une population composée des M unités statistiques d'étiquettes $1, 2, \dots, r, \dots, M$ et de tailles respectives $X_1, X_2, \dots, X_r, \dots, X_M$. On désire sélectionner m unités de la population U selon le mode de tirage SPPT.

On désigne par π_r la probabilité d'inclusion d'ordre 1 de l'unité r , c'est-à-dire la probabilité pour l'unité r de faire partie de l'échantillon des m unités.

Désignons par X la somme cumulée totale des tailles X_r , soit

$$X = \sum_{j=1}^M X_j$$

La taille relative p_r de l'unité r est définie par la relation

$$p_r = X_r / X \quad \forall r = 1, 2, \dots, M$$

Pour un tirage SPPT, la probabilité d'inclusion π_r est proportionnelle à la taille relative p_r , ce qui s'exprime par la double relation

$$\pi_r = k p_r = k X_r / X \quad \forall r = 1, 2, \dots, M$$

où k est une constante.

On montre que pour tout plan de sondage de taille fixe m (taille de l'échantillon), on a

$$\sum_{r=1}^M \pi_r = m$$

On en déduit

$$m = \sum_{r=1}^M \pi_r = k \sum_{r=1}^M \frac{X_r}{X} = k$$

D'où

$$\pi_r = m p_r = m X_r / X \quad \forall r = 1, 2, \dots, M$$

La procédure de tirage fondée sur les probabilités d'inclusion exige que soit vérifiée la condition :

$$0 < \pi_r < 1 \quad \forall r = 1, 2, \dots, M$$

Si pour une unité r , $\pi_r > 1$, ou bien cette unité est désignée d'office pour faire partie de l'échantillon avec la probabilité 1 et on aura alors à tirer $m-1$ autres unités parmi les $M-1$ unités restantes, ou bien l'unité est segmentée en 2 nouvelles unités ou plus avant les tirages.

La procédure de tirage est fondée sur les cumuls des probabilités d'inclusion à l'instar de la procédure de tirage fondée sur les cumuls des tailles des unités. Désignons par W_r le cumul des probabilités d'inclusion pour les r premières unités de la population U , soit

$$W_r = \sum_{j=1}^r \pi_j \quad \forall r = 1, 2, \dots, M$$

Notons que le cumul total W ou W_M des probabilités d'inclusion vérifie la relation

$$W = W_M = \sum_{j=1}^M \pi_j = m$$

Pour réaliser le tirage, on commence par générer un nombre aléatoire selon la loi uniforme sur l'intervalle $[0,1[$. Si u est le nombre obtenu, les m unités de l'échantillon, soit r_1, r_2, \dots, r_m , sont associés aux m nombres $u, u+1, \dots, u+(m-1)$ selon les conditions ci-après.

La première unité r_1 vérifiera la condition

$$W_{r_1-1} \leq u < W_{r_1}$$

La deuxième unité r_2 , l'unité r_h et l'unité r_m vérifieront respectivement les conditions

$$W_{r_2-1} \leq u+1 < W_{r_2}$$

$$W_{r_h-1} \leq u+(h-1) < W_{r_h}$$

et
$$W_{r_m-1} \leq u+(m-1) < W_{r_m}$$

Tableau 9 : Données d'un tirage SPPT de 3 villages de la région ABCD fondé sur les cumuls des probabilités d'inclusion

| Numéro d'ordre r | Identité du village | Nom du village | Taille de l'unité X_r | Taille relative de l'unité p_r | Probabilité d'inclusion d'ordre 1 π_r | Cumul des probabilités d'inclusion W_r | Nombre généré | Unité tirée |
|-----------------------|---------------------|----------------|----------------------------|-------------------------------------|--|---|---------------|-------------|
| 1 | 8121 | Assabe | 136 | 0,03736 | 0,11209 | 0,11209 | | |
| 2 | 8122 | Atal | 240 | 0,06593 | 0,19780 | 0,30989 | | |
| 3 | 8123 | Gorqui | 115 | 0,03159 | 0,09478 | 0,40467 | | |
| 4 | 8124 | Kaka | 127 | 0,03489 | 0,10467 | 0,50934 | | |
| 5 | 8125 | Kaporo | 181 | 0,04973 | 0,14918 | 0,65852 | 0,61220 | x |
| 6 | 8126 | Kipé | 184 | 0,05055 | 0,15165 | 0,81016 | | |
| 7 | 8127 | Lalo | 235 | 0,06456 | 0,19368 | 1,00385 | | |
| 8 | 8128 | Loum | 277 | 0,07610 | 0,22830 | 1,23214 | | |
| 9 | 8129 | Madina | 211 | 0,05797 | 0,17390 | 1,40604 | | |
| 10 | 8130 | Manqa | 112 | 0,03077 | 0,09231 | 1,49835 | | |
| 11 | 8131 | Mano | 210 | 0,05769 | 0,17308 | 1,67143 | 1,61220 | x |
| 12 | 8132 | Marsa | 143 | 0,03929 | 0,11786 | 1,78929 | | |
| 13 | 8133 | Matoto | 223 | 0,06126 | 0,18379 | 1,97308 | | |
| 14 | 8134 | Saqa | 252 | 0,06923 | 0,20769 | 2,18077 | | |
| 15 | 8135 | Sindaya | 266 | 0,07308 | 0,21923 | 2,40000 | | |
| 16 | 8136 | Taouya | 102 | 0,02802 | 0,08407 | 2,48407 | | |
| 17 | 8137 | Tilen | 126 | 0,03462 | 0,10385 | 2,58791 | | |
| 18 | 8138 | Valo | 203 | 0,05577 | 0,16731 | 2,75522 | 2,61220 | x |
| 19 | 8139 | Vanta | 148 | 0,04066 | 0,12198 | 2,87720 | | |
| 20 | 8140 | Vival | 149 | 0,04093 | 0,12280 | 3,00000 | | |

Exemple d'application

Sélectionnons selon le mode de tirage SPPT fondé sur les cumuls des probabilités d'inclusion d'ordre 1, un échantillon de 3 villages de la région des 20 villages dénommée ABCD.

Nous obtenons les données de départ suivantes :

$$m = 3$$

$$u = 0,61220$$

$$0 < \pi_r < 1 \quad \forall r = 1, 2, \dots, M$$

Les autres données du tirage sont présentées dans le tableau 9 ci-dessus. Nous notons que ce dernier tableau est simplement une adaptation du tableau n° 8 relatif au tirage fondé sur les cumuls des tailles des unités.

Comment identifier l'unité r_1 de la population U qui correspond au nombre u ? On calcule successivement la différence $u - W_h$ en parcourant selon l'ordre croissant, les cumuls W_h jusqu'à ce que la différence devienne négative. Le premier cumul qui rend négative cette différence est le cumul de l'unité correspondant au nombre r_1 tiré, soit l'unité de numéro d'ordre 5, c'est-à-dire le village Kaporo. De même, les nombres r_2 et r_3 tirés correspondent respectivement aux cumuls qui rendent négatives les différences $u+1 - W_h$ et $u+2 - W_h$ les premiers lorsque ces cumuls sont parcourus dans l'ordre croissant. Il leur correspond respectivement les unités de numéros d'ordre 11 et 18, soit les villages Mano et Valo.

Remarque

Les deux formulations reposant sur le cumul des tailles ou le cumul des probabilités d'inclusion d'ordre 1 sont équivalentes. On observe notamment, étant donné que

$$K = X/m$$

que la condition $\pi_r < 1 \quad \forall r = 1, 2, \dots, M$

entraîne la condition $\pi_r = m X_r / X = X_r / K < 1 \quad \forall r = 1, 2, \dots, M$

D'où la condition $X_r < K \quad \forall r = 1, 2, \dots, M$

qui est la condition à laquelle les tailles doivent obéir dans la procédure fondée sur le cumul des tailles des unités.