

TIRAGE 3.0

Manuel de l'utilisateur

par
Julien Amegandjin

Sommaire

Avant-propos	2
Présentation du logiciel	3
1. Principe des tirages aléatoires et de la génération des nombres aléatoires	3
2. Les différents modes de tirage d'échantillons	3
3. Les différents groupes de modules du logiciel	4
4. Comment utiliser Tirage 3.0 ?	5
5. Remarques	6
6. Noms des modules dans les noms des fichiers Excel créés	8

Avant-propos

Le logiciel TIRAGE est l'œuvre de Julien Amegandjin, un spécialiste des sondages et un habitué des tirages aléatoires. Sa présente version dénommée Tirage 3.0 succède à deux premières versions Tirage 1.0 et Tirage 2.1 parues en 2005 et 2009 respectivement.

Les deux premières versions de TIRAGE ont été conçues avec les logiciels Visual Basic 6 et Access 2003. La version Tirage 2.1 est toujours utilisée mais par seuls ceux qui possèdent encore Access 2003 et Excel 2003. Ce fait constitue une contrainte aujourd'hui. Pour en avoir perdu le code, je n'ai pas envisagé de faire une mise à jour de Tirage 2.1 avec Visual Basic.Net et une version récente d'Access. J'ai plutôt voulu tirer parti de mes connaissances du langage C pour aller à la découverte du langage C Sharp (C#) dont je me suis servi pour réaliser cette nouvelle version de TIRAGE.

Contrairement aux deux versions précédentes, la création de Tirage 3.0 s'est affranchie de l'utilisation du gestionnaire de bases de données Access pour recevoir les bases de sondage et les échantillons issus des tirages. Dans Tirage 3.0, les bases de sondage et les échantillons aléatoires sont des tableaux d'Excel, ce qui est une grande simplification par rapport aux versions antérieures.

La version Tirage 3.0 du logiciel TIRAGE conserve le même objet des versions antérieures. C'est un logiciel destiné à simuler le hasard. Il permet notamment,

- la génération de nombres aléatoires ;
- le tirage d'échantillons aléatoires au sein d'une population finie ;
- la simulation de processus aléatoires ou phénomènes faisant appel au hasard.

Le logiciel Tirage 3.0 est mis à disposition gratuitement. Il ne peut donc faire l'objet d'aucune transaction commerciale, qu'il s'agisse d'une vente ou de toute autre opération lucrative. Je l'offre aux générations actuelles des statisticiens en activité en reconnaissance de ce que la vie m'a apporté dans le domaine du savoir. C'est le prolongement naturel de mes actions antérieures de partage du savoir accomplies à travers :

- des cours photocopiés produits et offerts à des étudiants entre 1968 et 1985 ;
- des documents techniques produits en faveur du renforcement des capacités dans des projets de développement des statistiques agricoles que j'ai dirigés entre 1986 et 1999 pour le compte de la FAO ou dans le cadre d'ateliers de formation sur les sondages que j'ai animés en rapport avec mon manuel¹ de sondage.

Fait à Lomé, le 21 Novembre 2023

J. Amegandjin

¹ Pratique des sondages, Julien Amegandjin, Editions AFRISTAT, Série Méthodes N°12, Bamako, Décembre 2013

Présentation du logiciel

1. Principe des tirages aléatoires et de la génération des nombres aléatoires

Le langage C# permet de générer des nombres aléatoires au moyen de l'objet **Random** et de la fonction **Next**. Il permet d'obtenir notamment des nombres aléatoires compris entre deux entiers M et Z inclus ou des nombres aléatoires compris entre 0 et 1 (le nombre 1 étant exclu).

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 consiste simplement à générer au moyen de l'objet Random et de la fonction Next, m nombres compris entre 1 et M inclus qui identifieront les unités sélectionnées. La génération des m nombres tiendra compte bien évidemment des spécificités du mode de tirage particulier considéré.

2. Les différents modes de tirage d'échantillon

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

Un sondage aléatoire simple (random simple sampling) est une méthode de tirage aléatoire qui consiste à sélectionner 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 sondage aléatoire simple peut être avec remise ou sans remise. Il donne donc lieu à deux modes de tirage classiques :

- le sondage aléatoire simple avec remise qui utilise des tirages avec probabilités égales et avec remise ;
- le sondage aléatoire simple sans remise qui utilise des tirages avec probabilités égales et sans remise.

Pour des raisons de commodité, le logiciel Tirage 3.0 utilise des abréviations pour désigner les modes de tirage, soit des sigles qui apparaissent comme des qualificatifs du mot "Tirage". Ainsi, "Tirage PEAR" est mis pour "Tirage avec probabilités égales et avec remise" et "Tirage PESR" est mis pour "Tirage avec probabilités égales et sans remise".

En plus du tirage PEAR et du tirage PESR, le logiciel utilise également les autres modes de tirage classiques suivants :

- le tirage systématique avec probabilités égales (SPE) ;
- le tirage avec probabilités proportionnelles aux tailles des unités et avec remise (PPTUAR) ;
- le tirage systématique avec probabilités proportionnelles aux tailles des unités et par cumul des tailles des unités (SPPTU ctu) ;
- le tirage systématique avec probabilités proportionnelles aux tailles des unités et par cumul des probabilités d'inclusion (SPPTU cpi) ;

Le tableau ci-dessous donne la liste des sigles utilisés.

Tableau : Liste des sigles

Tirage PEAR ²	Tirage avec probabilités égales et avec remise
Tirage PESR	Tirage avec probabilités égales et sans remise
Tirage PPTUAR	Tirage avec probabilités proportionnelles aux tailles des unités et avec remise
Tirage SPE	Tirage systématique avec probabilités égales
Tirage SPPTU	Tirage systématique avec probabilités proportionnelles aux tailles des unités
Tirage SPPTU ctu	Tirage systématique avec probabilités proportionnelles aux tailles des unités et par cumul des tailles des unités
Tirage SPPTU cpi	Tirage systématique avec probabilités proportionnelles aux tailles des unités et par cumul des probabilités d'inclusion

Les tirages SPE, PPTUAR, SPPTU cpi et SPPTU ctu ne sont pas présentés dans le présent manuel de l'utilisateur. Toutefois, l'on peut se référer au chapitre 1 du livre "Pratique des Sondages" qui traite des fondements des différentes méthodes de tirage usuelles de la théorie des sondages. Une copie du chapitre 1 dudit manuel est fournie avec le logiciel. L'on notera que les appellations PIAR et SPPT dudit chapitre 1 sont devenues PPTUAR et SPPTU dans le logiciel Tirage 3.0.

3. Les différents groupes de modules du logiciel

Le logiciel Tirage 3.0 comporte 17 modules organisés en trois groupes :

- Tirages avec probabilités égales ;
- Tirages avec probabilités inégales ;
- Sondage à deux degrés avec un tirage PESR au 2^e degré.

Les menus et les sous-menus se présentent comme suit.

² Les sigles PEAR, PESR et PIAR viennent de Jean-Marie Grosbras, Méthodes statistiques des sondages, Dunod, 1982. Les autres abréviations PPTUAR, SPE, SPPTU, SPPTU cpi et SPPTU ctu sont proposées par Julien Amegandjin.

MENUS ET SOUS-MENUS

Tirages avec probabilités égales

Générer un échantillon PEAR de nombres aléatoires compris entre M et Z inclus
Générer un échantillon PESR de nombres aléatoires compris entre M et Z inclus

Tirage avec probabilités égales et avec remise (PEAR)
Tirage avec probabilités égales et sans remise (PESR)
Tirage systématique avec probabilités égales (SPE)

Générer un échantillon PEAR de nombres aléatoires de l'intervalle [0, 1[
Réordonner au hasard les unités d'une base de sondage

Tirages avec probabilités inégales

Tirage avec probabilités proportionnelles aux tailles des unités et avec remise (PPTUAR)
Tirage systématique avec probabilités proportionnelles aux tailles des unités et par cumul des tailles des unités (SPPTU ctu)

Tirage SPPTU par cumul des probabilités d'inclusion (SPPTU cpi)
Contrôle des probabilités d'inclusion dans un tirage SPPTU cpi

Sondage à deux degrés avec un tirage PESR au 2^e degré

L'échantillon du 1^{er} degré est connu
Tirage PESR réalisé au 1^{er} degré
Tirage SPE réalisé au 1^{er} degré

Tirage PPTUAR réalisé au 1^{er} degré
Tirage SPPTU par cumul des tailles des unités réalisé au 1^{er} degré
Tirage SPPTU par cumul des probabilités d'inclusion réalisé au 1^{er} degré

4. Comment utiliser Tirage 3.0 ?

Lorsque le logiciel Tirage 3.0 est lancé, un formulaire d'information s'affiche. C'est la page d'accueil du logiciel. En cliquant sur le bouton de commande "Continuer", l'on accède à la barre des menus qui donne accès aux différents modules. A l'appel du module choisi, un autre formulaire apparaît avec différents contrôles : zones de texte, étiquettes, boutons de commande, boutons de liste déroulante ou tableaux d'affichage. L'utilisateur répond aux demandes du formulaire en parcourant les contrôles de la gauche vers la droite et du haut vers le bas. Les résultats des étapes intermédiaires de l'exécution d'un module tout comme le résultat final s'affichent sur les tableaux du formulaire.

Seul un tableau Excel est candidat pour servir de base de sondage avec le logiciel Tirage 3.0. Les bases de sondage sont logées dans le répertoire BASES du disque C. Le résultat de l'exécution de n'importe lequel des 17 modules est également un tableau Excel qui est enregistré dans un dossier du choix de l'utilisateur. Il porte un nom à trois composantes qui sont dans l'ordre, **le nom du module utilisé, le nom de la base de sondage utilisée et l'identité du tirage retenue**. Par exemple, TirageSPE_ABCD_Golfe est le nom du fichier Excel résultant de l'exécution du module Tirage SPE, ABCD et Golfe étant respectivement la base de sondage utilisée et l'identité du tirage choisie.

Si le module utilisé ne fait pas appel à une base de sondage comme dans un cas de génération de nombres aléatoires, le fichier Excel issu de l'exécution du module porte un nom composé du nom du module et du nom de l'identité du tirage pris dans cet ordre. Par exemple, le nom GénérationPesrMZ_Golfe est celui du fichier fourni par l'exécution du module relatif à la génération PESR de nombres aléatoires compris entre M et Z inclus et à l'identité du tirage nommée "Golfe".

La dernière étape de la mise en œuvre d'un module du logiciel consiste à exporter le résultat final du tirage vers Excel. Lors de cette opération d'exportation, **le formulaire affiche le nom du fichier Excel créé** et demande à l'utilisateur de choisir le dossier où il doit être enregistré.

5. Remarques

Remarque 1

Les sondages à deux degrés traités par Tirage 3.0 font appel presque tous à un tirage PESR au 2^e degré avec **un nombre constant d'unités secondaires** à tirer par unité primaire.

Remarque 2 : comment sélectionner un nombre variable d'US par UP ?

La troisième barre de menu comporte un module dénommé "L'échantillon du 1^{er} degré est connu". Il s'agit du module n°12. C'est un cas de sondage à deux degrés où la source du tirage du 2^e degré (ou la base de sondage des unités secondaires) est connue. Mais ici, **le nombre d'unités secondaires à tirer par unité primaire est variable**. Cette source du tirage du 2^e degré doit comporter impérativement, entre autres variables, les trois variables suivantes :

- la variable désignant l'identité des unités primaires ;
- la variable désignant la taille des unités primaires ;
- la variable donnant le nombre d'unités secondaires à tirer par unité primaire.

Le fichier "Tableau 200" du dossier BASES du disque C est un exemple de fichier qui convient pour servir de source du tirage du 2^e degré. Il comporte les variables suivantes :

- ID_UP (identité de l'unité primaire) ;
- NbUS (nombre d'unités secondaires de l'UP ou taille de l'UP) ;
- NbUS_A_Tirer (nombre d'unités secondaires à tirer par unité primaire) ;
- Nom_UP (nom de l'unité primaire) ;
- Code_UP (code de l'unité primaire).

L'utilisateur nommera à sa guise toutes les variables de la source du tirage du 2^e degré comme dans le fichier "Tableau 300" du répertoire BASES du disque C où Num et NbMen désignent respectivement l'identité et la taille de l'unité primaire.

Ce module de tirage n°12 est particulièrement indiqué lorsque l'on souhaite sélectionner **un nombre variable d'unités secondaires par unité primaire** dans un sondage à deux degrés.

Remarque 3

Parfois un bouton "Parcourir" apparaît parmi les boutons de commande du formulaire. Il sert à sélectionner la base de sondage dans le dossier BASES du disque C.

Remarque 4

Dans les sondages à deux degrés, à l'issue du tirage des unités primaires, le formulaire réclame la variable qui désigne l'identité des unités primaires tirées. C'est la variable "**Unité_tirée**" de l'échantillon des unités primaires qui fournit cette identité.

Remarque 5 : identité du tirage

La variable "**Identité du tirage**" sert à distinguer dans un même dossier, les fichiers qui représentent les résultats de deux ou plusieurs tirages portant sur un même module et une même base de sondage. Ainsi, deux tirages d'échantillon utilisant le module PESR et la base de sondage "Gaya", pourront recevoir les identités de tirage "Golfe" et "Océan". Et les fichiers désignant les échantillons créés prendront alors les noms "TiragePesr_Gaya_Golfe" et "TiragePesr_Gaya_Océan" respectivement.

Remarque 6 : gestion des unités atypiques avec le tirage SPPTU

Il est important de noter que le tirage SPPTU (tirage systématique avec probabilités proportionnelles aux tailles des unités) s'applique à la seule condition que la probabilité d'inclusion p_i de toute unité U_i de la base de sondage **soit strictement inférieure à 1**. Cela indépendamment du fait que le mode de tirage utilise le cumul des tailles des unités (module 16) ou le cumul des probabilités d'inclusion (module 17).

On rappelle qu'avec le tirage SPPTU, la probabilité d'inclusion p_i d'une unité U_i de la base de sondage a pour expression $p_i = m * N_i / N$, m étant la taille de l'échantillon et N_i / N , la taille relative de l'unité U_i dans la base.

Le module 11 dénommé "Contrôle des PI" permet de calculer les probabilités d'inclusion p_i des unités d'une base de sondage donnée et pour une taille d'échantillon donnée, puis affiche la liste des unités dont la probabilité d'inclusion est supérieure ou égale à 1. De telles unités seront appelées ici **des unités atypiques**.

Si toutes les probabilités d'inclusion p_i sont strictement inférieures à 1, alors le tirage SPPTU s'applique indifféremment au moyen du module SPPTUctu ou du module SPPTUcpi. Dans le cas contraire, les unités atypiques sont choisies d'office pour faire partie de l'échantillon à tirer, ce qui revient à dire qu'elles ont été tirées de façon certaine et donc avec une probabilité égale à 1. Le reste de l'échantillon doit être tiré de la base de sondage résiduelle. L'on doit pour cela s'assurer encore, grâce au module "Contrôle des PI", qu'il n'existe pas d'unités atypiques à cette 2^e étape des tirages dans la base résiduelle.

Considérons la base de sondage ABCD2 du dossier BASES du disque C. Elle représente un ensemble de 20 villages d'une région. La variable "NbMen" de cette base désigne la taille du village en termes d'effectif des ménages. L'application du module "Contrôle des PI" à cette base montre que pour des échantillons de taille 1, 2 ou 3 on ne trouve aucune unité atypique. En revanche, pour des échantillons de taille 4 et 5 unités, le nombre d'unités atypiques trouvées est égal à 1 et à 3 respectivement.

6. Noms des modules dans les noms des fichiers Excel créés

Le tableau ci-dessous présente les noms des modules figurant dans les noms des fichiers Excel qui définissent les résultats des tirages.

Tableau : Noms des modules dans les noms des fichiers Excel créés

Numéro du module	Nom du module
1	GénérationPearMZ
2	GénérationPesrMZ
3	TiragePear
4	TiragePesr
5	TirageSpe
6	GénérationNB01
7	Réordonner
8	TiragePptuar
9	TirageSpptuCtu
10	TirageSpptuCpi
11	ContrôleDesPi
12	SDDDegré1Connu
13	SDDPesr
14	SDDSpe
15	SDDPptuar
16	SDDSpptuCtu
17	SDDSpptuCpi