

Tirage de Bernoulli (plan BE)

Pierre Duchesne

August 1, 2017

Définition du plan de Bernoulli (plan BE)

- ▶ Pour le mettre en oeuvre, on exécute une expérience de Bernoulli pour chaque unité dans la base de sondage, avec la probabilité $\pi \in (0, 1)$ pour inclusion, et $1 - \pi$ pour la non-inclusion.
- ▶ Les expériences, au total de N , sont effectuées de manière indépendante.
- ▶ L'échantillon est composé des unités k pour lesquelles on a obtenu inclusion.

Algorithme pour exécuter un plan BE

- ▶ On simule N variables aléatoires $U(0, 1)$ indépendamment.
- ▶ Avec un logiciel tel R, on utilise une fonction qui génère des variables aléatoires uniformes. En R, c'est `runif()`.
- ▶ On obtient alors u_1, u_2, \dots, u_N .
- ▶ Si $u_i < \pi$, alors l'unité est choisie.
- ▶ Sinon l'unité est rejetée.
- ▶ L'algorithme respecte la définition car $P(u_i < \pi) = \pi$, puisque $u_i \sim U(0, 1)$.

Illustration du plan BE

k	Ville	Prix
1	Ahuntsic	229 000.00 \$
2	Beaconsfield	275 000.00 \$
3	Beloeil	152 000.00 \$
4	Blainville	314 000.00 \$
5	Boucherville	205 000.00 \$
6	Chomedey	212 000.00 \$
7	Cote-St-Luc	475 000.00 \$
8	Dorval	157 000.00 \$
9	Duvernay	243 000.00 \$
10	Fabreville	169 800.00 \$
11	Hudson	245 000.00 \$
12	Kirkland	198 000.00 \$
13	Lachine	189 000.00 \$
14	Lasalle	175 000.00 \$
15	Lorraine	309 000.00 \$
16	Montreal West	360 000.00 \$
17	Mount Royal	370 000.00 \$
18	Notre-Dame-De-Grace	375 000.00 \$
19	Outremont	600 000.00 \$
20	Pierrefonds	138 000.00 \$
21	Pointe Claire	290 000.00 \$
22	Rosemere	338 000.00 \$
23	St-Bruno-De-Montarville	235 000.00 \$
24	St-Eustache	250 000.00 \$
25	St-Lambert	300 000.00 \$
26	St-Laurent	250 000.00 \$
27	Ste-Therese	255 000.00 \$
28	Terrebonne	165 000.00 \$
29	Vimont	259 000.00 \$
30	Westmount	758 000.00 \$

Utilisation de la fonction R runif()

- ▶ `> set.seed(1) # Fix the seed`
- ▶ `> round(runif(30), digits=3) # Commandes pour le premier échantillon`
[1] 0.163 0.425 0.317 0.646 0.084 0.083 0.203 0.978
0.439 0.272 0.968 0.788 [13] 0.021 0.908 0.904 0.559 0.373
0.798 0.385 0.818 0.525 0.857 0.492 0.348 [25] 0.117 0.216
0.572 0.807 0.859 0.955
- ▶ `> round(runif(30), digits=3) # Commandes pour le second échantillon`
[1] 0.913 0.922 0.863 0.210 0.548 0.472 0.772 0.068
0.052 0.384 0.613 0.404 [13] 0.224 0.151 0.560 0.061 0.099
0.937 0.270 0.620 0.275 0.411 0.617 0.570 [25] 0.001 0.586
0.323 0.326 0.335 0.465
- ▶ `> round(runif(30), digits=3) # Commandes pour le troisième échantillon`
[1] 0.273 0.998 0.056 0.037 0.127 0.032 0.287 0.968
0.003 0.866 0.160 0.353 [13] 0.398 0.703 0.951 0.375 0.220
0.090 0.328 0.512 0.710 0.170 0.437 0.376 [25] 0.984 0.676
0.660 0.355 0.127 0.339

Trois échantillons, $\pi = 14/30 = 0.46$

k	Ville	Prix	u_k	Inclus?	u_k	Inclus?	u_k	Inclus?
1	Ahuntsic	229000	0.163	Oui	0.913	Non	0.273	Oui
2	Beaconsfield	275000	0.425	Oui	0.922	Non	0.998	Non
3	Beloeil	152000	0.317	Oui	0.863	Non	0.056	Oui
4	Blainville	314000	0.646	Non	0.21	Oui	0.037	Oui
5	Boucherville	205000	0.084	Oui	0.548	Non	0.127	Oui
6	Chomedey	212000	0.083	Oui	0.472	Non	0.032	Oui
7	Cote-St-Luc	475000	0.203	Oui	0.772	Non	0.287	Oui
8	Dorval	157000	0.978	Non	0.068	Oui	0.968	Non
9	Duvernay	243000	0.439	Oui	0.052	Oui	0.003	Oui
10	Fabreville	169800	0.272	Oui	0.384	Oui	0.866	Non
11	Hudson	245000	0.968	Non	0.613	Non	0.16	Oui
12	Kirkland	198000	0.788	Non	0.404	Oui	0.353	Oui
13	Lachine	189000	0.021	Oui	0.224	Oui	0.398	Oui
14	Lasalle	175000	0.908	Non	0.151	Oui	0.703	Non
15	Lorraine	309000	0.904	Non	0.56	Non	0.951	Non
16	Montreal West	360000	0.559	Non	0.061	Oui	0.375	Oui
17	Mount Royal	370000	0.373	Oui	0.099	Oui	0.22	Oui
18	Nontre-Dame-De-Grace	375000	0.798	Non	0.937	Non	0.09	Oui
19	Outremont	600000	0.385	Oui	0.27	Oui	0.328	Oui
20	Pierrefonds	138000	0.818	Non	0.62	Non	0.512	Non
21	Pointe Claire	290000	0.525	Non	0.275	Oui	0.71	Non
22	Rosemere	338000	0.857	Non	0.411	Oui	0.17	Oui
23	St-Bruno-De-Montarville	235000	0.492	Non	0.617	Non	0.437	Oui
24	St-Eustache	250000	0.348	Oui	0.57	Non	0.376	Oui
25	St-Lambert	300000	0.117	Oui	0.001	Oui	0.984	Non
26	St-Laurent	250000	0.216	Oui	0.586	Non	0.676	Non
27	Ste-Therese	255000	0.572	Non	0.323	Oui	0.66	Non
28	Terrebonne	165000	0.807	Non	0.326	Oui	0.355	Oui
29	Vimont	259000	0.859	Non	0.335	Oui	0.127	Oui
30	Westmount	758000	0.955	Non	0.465	Oui	0.339	Oui

Les trois échantillons sont:

- ▶ $s_1 = \{1, 2, 3, 5, 6, 7, 9, 10, 13, 17, 19, 24, 25, 26\}$;
- ▶ $s_2 = \{4, 8, 9, 10, 12, 13, 14, 16, 17, 19, 21, 22, 25, 27, 28, 29, 30\}$;
- ▶ $s_3 = \{1, 3, 4, 5, 6, 7, 9, 11, 12, 13, 16, 17, 18, 19, 22, 23, 24, 28, 29, 30\}$.

Exemple d'utilisation du plan BE

Lors d'une édition de l'Enquête sur l'emploi, la rénumération et les heures de Statistique Canada.

Le plan d'échantillonnage est un plan stratifié à deux phases:

- ▶ Phase I: Deux strates par région sont considérées. On fait un plan BE dans chaque strate. On ramasse beaucoup d'information sur les unités choisies (informations pas trop cher à obtenir).
- ▶ Phase II: On fait un tirage aléatoire simple dans les unités choisies. On recueille l'information qui nous intéresse vraiment.

- ▶ La taille de l'échantillon s , que l'on note n_s , est aléatoire. En fait:

$$n_s \sim \text{Bin}(N, \pi)$$

- ▶ Pour montrer cela, on note que l'on peut écrire $n_s = \sum_U I_k$ et les variables I_k sont indépendantes. On a

$$I_k \sim \text{Bernoulli}(\pi)$$

- ▶ On a

$$\pi_k = P(s \ni k) = P(I_k = 1) = \pi,$$

$$\pi_{kl} = P(s \ni k, l) = P(s \ni k)P(s \ni l) = \pi^2.$$