

TD n° 2 : Plan de sondage aléatoire PESR

Exercice 1. On considère une population U constituée de 4 individus : $U = \{u_1, u_2, u_3, u_4\}$. On prélève au hasard et simultanément 3 individus dans U formant ainsi un échantillon.

Combien d'échantillons peut-on former ? Expliciter les.

Exercice 2. Décrire brièvement l'enjeu des commandes R suivantes :

```
combi = NULL
for(i in 1:4) {
  for (j in (i+1):5) {
    for (k in (j+1):6) {
      for(m in (k+1):7) {
        combi = rbind(combi, c(i, j, k, m))
      }}}
combi
```

Cela renvoie :

	[,1]	[,2]	[,3]	[,4]
[1,]	1	2	3	4
[2,]	1	2	3	5
[3,]	1	2	3	6
[4,]	1	2	3	7
[5,]	1	2	4	5
[6,]	1	2	4	6
[7,]	1	2	4	7
[8,]	1	2	5	6
[9,]	1	2	5	7
[10,]	1	2	6	7
[11,]	1	3	4	5
[12,]	1	3	4	6
[13,]	1	3	4	7
[14,]	1	3	5	6
[15,]	1	3	5	7
[16,]	1	3	6	7
[17,]	1	4	5	6
[18,]	1	4	5	7
[19,]	1	4	6	7
[20,]	1	5	6	7
[21,]	2	3	4	5
[22,]	2	3	4	6
[23,]	2	3	4	7
[24,]	2	3	5	6
[25,]	2	3	5	7
[26,]	2	3	6	7
[27,]	2	4	5	6
[28,]	2	4	5	7
[29,]	2	4	6	7
[30,]	2	5	6	7
[31,]	3	4	5	6
[32,]	3	4	5	7
[33,]	3	4	6	7
[34,]	3	5	6	7
[35,]	4	5	6	7

Exercice 3. L'objectif de cet exercice est d'illustrer certains résultats théoriques du cours sur les plans de sondage aléatoire de type PESR avec un exemple. On étudie un caractère Y dans une population de 5 individus : $U = \{u_1, \dots, u_5\}$. Pour tout $i \in \{1, \dots, 5\}$, soit y_i la valeur de Y pour l'individu u_i . Les résultats sont :

y_1	y_2	y_3	y_4	y_5
10	16	9	26	21

1. Calculer la moyenne-population \bar{y}_U et l'écart-type corrigé-population s_U .
2. On prélève au hasard et simultanément 2 individus dans U formant ainsi un échantillon. Chaque individu a la même probabilité qu'un autre d'être sélectionné. On est donc dans le cadre d'un plan de sondage aléatoire de type PESR.
 - (a) Quel est le taux de sondage ? Combien d'échantillons peut-on former ? Expliciter les.
 - (b) Pour chaque échantillon ω , calculer la moyenne-échantillon \bar{y}_ω et l'écart-type corrigé-échantillon s_ω .
 - (c) Soit \bar{y}_W la *var* égale à la moyenne-échantillon, l'aléatoire étant dans l'échantillon considéré. Déterminer sa loi, puis calculer son espérance et sa variance.
 - (d) Soit s_W la *var* égale à l'écart-type corrigé-échantillon, l'aléatoire étant dans l'échantillon considéré. Calculer l'espérance de s_W^2 .
 - (e) Retrouver les résultats des deux questions précédentes avec les formules du cours.
 - (f) Calculer les erreurs dans l'estimation de \bar{y}_U .
 - (g) Quelle est la probabilité de se tromper de plus de 15% dans l'estimation de \bar{y}_U ?

Exercice 4. Dans une population de 3 individus $U = \{u_1, u_2, u_3\}$, on prélève au hasard 2 individus pour former un échantillon. La *var* W égale à l'échantillon obtenu vérifie :

$$\mathbb{P}(W = \{u_1, u_2\}) = \frac{1}{4}, \quad \mathbb{P}(W = \{u_1, u_3\}) = \frac{1}{4}, \quad \mathbb{P}(W = \{u_2, u_3\}) = \frac{1}{4}.$$

1. Calculer, pour tout $i \in \{1, 2, 3\}$, $\mathbb{P}(u_i \in W)$. Est-ce que l'on a affaire à un plan de sondage aléatoire de type PESR ?
2. On étudie un caractère Y dans U . Pour tout $i \in \{1, 2, 3\}$, soit y_i la valeur de Y pour l'individu u_i . Les résultats sont :

y_1	y_2	y_3
2	5	11

On considère la *var* :

$$\bar{y}_W = \frac{1}{2} \sum_{i=1}^3 y_i \mathbb{1}_{\{u_i \in W\}}.$$

- (a) Déterminer l'ensemble des valeurs possibles de \bar{y}_W .
- (b) Calculer la moyenne-population \bar{y}_U . Est-ce que $\mathbb{E}(\bar{y}_W) = \bar{y}_U$?

Exercice 5. Dans une population de N individus $U = \{u_1, \dots, u_N\}$, on prélève au hasard et simultanément n individus pour former un échantillon. Chaque individu a la même probabilité qu'un autre d'être sélectionné. On est donc dans le cadre d'un plan de sondage aléatoire de type PESR. Soit W la *var* égale à l'échantillon obtenu.

1. Montrer que la loi de W est donnée par

$$\mathbb{P}(W = \omega) = \frac{1}{\binom{N}{n}}, \quad \omega \in W(\Omega),$$

où \mathbb{P} désigne la probabilité uniforme et $W(\Omega)$ désigne l'ensemble de tous les échantillons de n individus possibles.

2. Montrer que

- (a) pour tout $i \in \{1, \dots, N\}$,

$$\mathbb{P}(u_i \in W) = \frac{n}{N}.$$

- (b) pour tout $(i, j) \in \{1, \dots, N\}^2$ avec $i \neq j$,

$$\mathbb{P}((u_i, u_j) \in W) = \frac{n(n-1)}{N(N-1)}.$$

3. Désormais, on étudie un caractère Y dans U . Soient \bar{y}_U la moyenne-population et s_U l'écart-type corrigé-population. Un estimateur aléatoire de \bar{y}_U est

$$\bar{y}_W = \frac{1}{n} \sum_{i=1}^N y_i \mathbb{1}_{\{u_i \in W\}}.$$

- (a) Montrer que

$$\mathbb{E}(\bar{y}_W) = \bar{y}_U.$$

- (b) Montrer que

$$\mathbb{V}(\bar{y}_W) = (1-f) \frac{s_U^2}{n},$$

où f désigne le taux de sondage : $f = n/N$.

4. Un estimateur aléatoire de s_U est

$$s_W = \sqrt{\frac{1}{n-1} \sum_{i=1}^N (y_i - \bar{y}_W)^2 \mathbb{1}_{\{u_i \in W\}}}.$$

Montrer que

$$\mathbb{E}(s_W^2) = s_U^2.$$

Exercice 6. On étudie un caractère Y dans une population de N individus. On prélève au hasard 64 individus pour former un échantillon suivant un plan de sondage aléatoire de type PESR. Les valeurs de Y pour ces individus donnent une moyenne-échantillon de 46 et un écart-type corrigé-échantillon de 4.

Déterminer un intervalle de confiance au niveau 99% pour la moyenne-population avec $N = 1000$, puis avec $N = 100000$.

Exercice 7. On dispose d'une liste de 900 transactions financières en milliers d'euros. Sur un échantillon de 16 transactions prélevé suivant un plan de sondage aléatoire de type PESR, les résultats sont :

50.34	52.62	53.79	54.99	55.82	57.67	51.41	53.13
51.41	53.13	53.89	55.04	55.91	57.99	51.41	53.13

1. Calculer le taux de sondage.
2. Donner une estimation ponctuelle de la moyenne des valeurs des 900 transactions.
3. Donner une estimation ponctuelle de l'écart-type de l'estimateur de la moyenne des valeurs des 900 transactions.
4. Déterminer un intervalle de confiance pour la la moyenne des valeurs des 900 transactions au niveau 99%.
5. Déterminer la taille d'échantillon à choisir pour avoir une incertitude absolue sur la moyenne des valeurs des 900 transactions inférieure ou égale à 1 au niveau 99%.

Exercice 8. Un jour donné, un magasin de vêtements comptent 230 clients. Sur un échantillon de 8 clients prélevé suivant un plan de sondage aléatoire de type PESR, les dépenses, en euros, sont :

20.35	35.60	32.10	55.20	45.80	47.75	51.40	35.15
-------	-------	-------	-------	-------	-------	-------	-------

1. Calculer le taux de sondage.
2. Donner une estimation ponctuelle de la moyenne des dépenses des 230 clients.
3. Donner une estimation ponctuelle de l'écart-type de l'estimateur de la moyenne des dépenses des 230 clients.
4. Déterminer un intervalle de confiance au niveau 95% pour la moyenne des dépenses des 230 clients.
5. Déterminer la taille d'échantillon à choisir pour avoir une incertitude relative sur la moyenne des dépenses des 230 clients inférieure ou égale à 8% au niveau 95%.

Exercice 9. On étudie le caractère $Y = \text{"taille"}$ en centimètres sur 8 filles.

1. Décrire brièvement l'enjeu des commandes R suivantes :

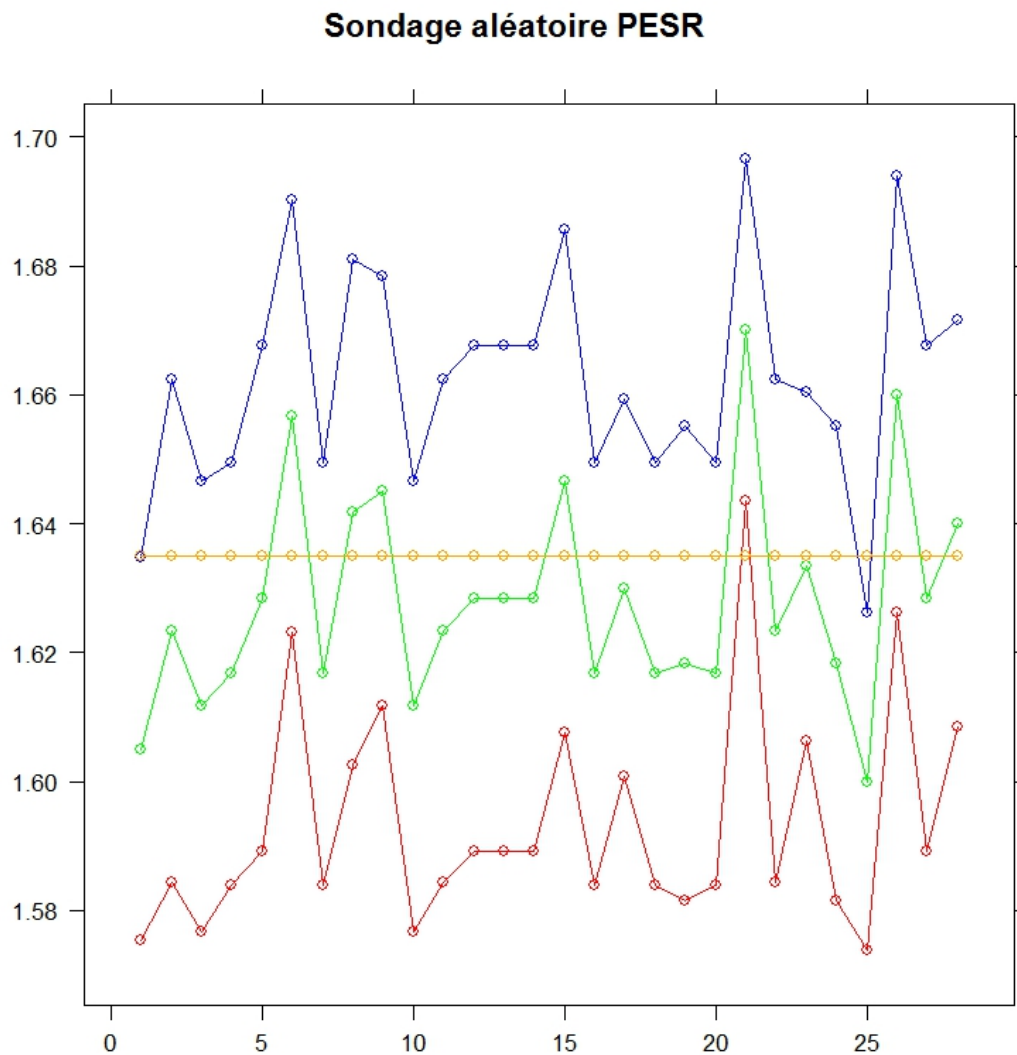
```
U = c("Elea", "Lilly", "Steph", "Rachel", "Fatiha", "Lea", "Anne",
      "Laure")
y = c(1.65, 1.54, 1.76, 1.62, 1.72, 1.52, 1.69, 1.58)
library(sampling)
N = 8
n = 6
M = choose(N, n)
bar_y_w = NULL
var_bar_y_w = NULL
for (i in 1:M){
  t = srswor(n, N)
  bar_y_w[i] = (1 / n) * sum(y * t)
  s2_w = sum((y - bar_y_w[i])^2 * t) / (n - 1)
  var_bar_y_w[i] = (1 - n / N) * (s2_w / n)
}
icmin = bar_y_w - 1.96 * sqrt(var_bar_y_w)
icmax = bar_y_w + 1.96 * sqrt(var_bar_y_w)
longueur_ic = icmax - icmin
PESR = data.frame(bar_y_w, var_bar_y_w, icmin, icmax, longueur_ic)
PESR
```

Cela renvoie :

	bar_y_w	var_bar_y_w	icmin	icmax	longueur_ic
1	1.605000	0.0002295833	1.575302	1.634698	0.05939587
2	1.623333	0.0003961111	1.584324	1.662342	0.07801796
3	1.611667	0.0003173611	1.576750	1.646583	0.06983336
4	1.616667	0.0002811111	1.583805	1.649529	0.06572416
5	1.628333	0.0004006944	1.589099	1.667567	0.07846803
6	1.656667	0.0002944444	1.623034	1.690299	0.06726478
7	1.616667	0.0002811111	1.583805	1.649529	0.06572416
8	1.641667	0.0004006944	1.602433	1.680901	0.07846803
9	1.645000	0.0002895833	1.611646	1.678354	0.06670722
10	1.611667	0.0003173611	1.576750	1.646583	0.06983336
11	1.623333	0.0003961111	1.584324	1.662342	0.07801796
12	1.628333	0.0004006944	1.589099	1.667567	0.07846803
13	1.628333	0.0004006944	1.589099	1.667567	0.07846803
14	1.628333	0.0004006944	1.589099	1.667567	0.07846803
15	1.646667	0.0003961111	1.607658	1.685676	0.07801796
16	1.616667	0.0002811111	1.583805	1.649529	0.06572416
17	1.630000	0.0002233333	1.600709	1.659291	0.05858182
18	1.616667	0.0002811111	1.583805	1.649529	0.06572416
19	1.618333	0.0003540278	1.581455	1.655212	0.07375725
20	1.616667	0.0002811111	1.583805	1.649529	0.06572416
21	1.670000	0.0001833333	1.643461	1.696539	0.05307705
22	1.623333	0.0003961111	1.584324	1.662342	0.07801796
23	1.633333	0.0001894444	1.606356	1.660311	0.05395442
24	1.618333	0.0003540278	1.581455	1.655212	0.07375725
25	1.600000	0.0001783333	1.573826	1.626174	0.05234827
26	1.660000	0.0002983333	1.626146	1.693854	0.06770753
27	1.628333	0.0004006944	1.589099	1.667567	0.07846803
28	1.640000	0.0002583333	1.608497	1.671503	0.06300518

2. Quel est le niveau des intervalles de confiance considérés ? Que faut-il changer dans les commandes pour avoir un niveau de confiance de 99% ?
3. Calculer la moyenne-population. Combien y-a t-il d'intervalles de confiance-échantillon qui recouvrent la moyenne-population ?
4. On illustre graphiquement ces résultats numériques en faisant :

```
bar_y_U = rep(mean(y), M)
library(lattice)
x = 1:M
xyplot(PESR$icmin + PESR$icmax + bar_y_w + bar_y_U ~ x, col = c("red",
"blue", "green", "orange"), type = "o", main = "Sondage aléatoire PESR",
xlab = " ", ylab = " ")
```



Retrouver graphiquement le résultat de la question précédente.