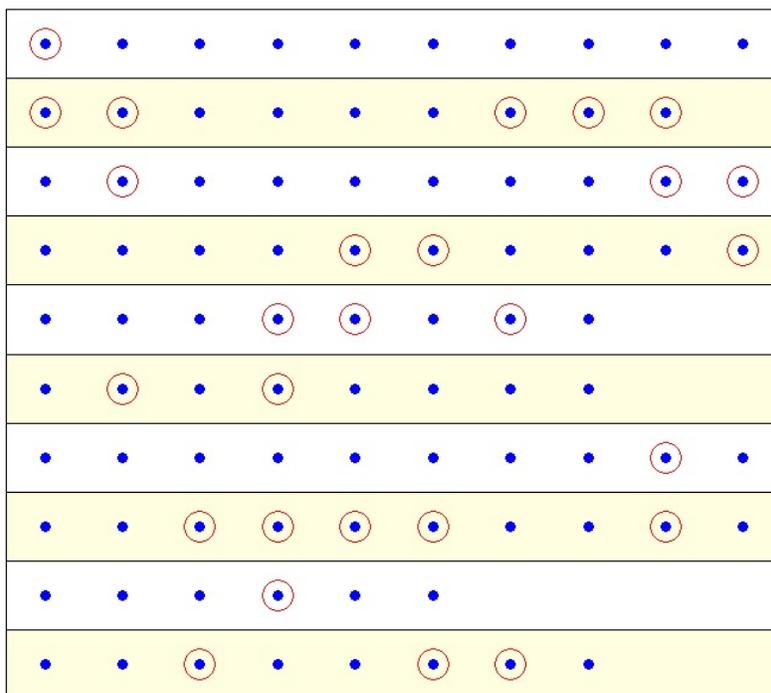


TD n° 5 : Plans de sondage aléatoire ST

Exercice 1. Le graphique ci-dessous illustre un plan de sondage aléatoire de type ST.



Les individus sont représentés par les points bleus, les strates par les bandes jaunes et les individus sélectionnés par les points bleus encadrés de rouge.

1. À partir de ce graphique, préciser les paramètres : $n, H, N_1, \dots, N_H, n_1, \dots, n_H$.
2. Quelle est la probabilité que l'individu représenté par le point bleu dans le coin haut/droit du graphique soit sélectionné ?
3. Quelle est la probabilité que l'individu représenté par le point bleu dans le coin bas/gauche du graphique soit sélectionné ?

Exercice 2. On considère le caractère $Y =$ "poids" en kilogrammes dans la population de 5 individus : $U = \{\text{Rachel, Nargess, Julie, Léa, Lola}\} = \{u_1, \dots, u_5\}$. Pour tout $i \in \{1, \dots, 5\}$, soit y_i la valeur de Y pour l'individu u_i . Les résultats, en kilogrammes, sont :

y_1	y_2	y_3	y_4	y_5
67	58	65	82.5	93

1. Calculer la moyenne-population \bar{y}_U et l'écart-type corrigé-population s_U .
2. Dans un premier temps, on prélève un échantillon de 2 individus suivant 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 .
 - (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, sa variance et son erreur quadratique moyenne : $EQM(\bar{y}_W)$.
3. Dans un deuxième temps, on prélève un échantillon de 2 individus suivant un plan de sondage aléatoire de type ST avec :
- o les 2 strates : $U_1 = \{\text{Rachel, Nargess, Julie}\}$ et $U_2 = \{\text{Léa, Lola}\}$,
 - o un individu par strate.
- (a) Combien d'échantillons peut-on former ? Expliciter les.
 - (b) Pour chaque échantillon ω , calculer la moyenne-échantillon \bar{y}_ω .
 - (c) Soit \bar{y}_W la *var* égale à la moyenne-échantillon stratifié, l'aléatoire étant dans l'échantillon considéré. Déterminer sa loi, puis calculer son espérance, sa variance et son erreur quadratique moyenne : $EQM(\bar{y}_W)$.
4. Quel plan de sondage donne une meilleure précision dans l'estimation de \bar{y}_U ?

Exercice 3. On souhaite étudier la quantité moyenne (en parties par million) d'un polluant particulier dans un vaste terrain U composé de $N = 1000$ parcelles de terre. Ces parcelles sont regroupées en 5 strates U_1, \dots, U_5 . On prélève un échantillon de $n = 50$ parcelles suivant un plan de sondage aléatoire de type STP. On dispose des informations suivantes :

Strate U_h	U_1	U_2	U_3	U_4	U_5
Taille N_h	300	200	100	100	300
Moyenne \bar{y}_{ω_h}	2.1	1.4	2.1	1.9	1.6
Écart-type corrigé s_{ω_h}	0.5	1.1	1.5	1.8	1.7

1. Déterminer le nombre de parcelles dans les échantillons de chaque strate.
2. Donner une estimation ponctuelle de la moyenne-population \bar{y}_U .
3. Donner l'incertitude absolue sur \bar{y}_U au niveau 95%.

Exercice 4. Une population U est partagée en 3 strates U_1, U_2 et U_3 de tailles respectives : $N_1 = 12, N_2 = 28$ et $N_3 = 50$. On prélève un échantillon de $n = 20$ individus suivant un plan de sondage aléatoire de type ST avec :

- o $n_1 = 2$ individus pour U_1 ,
- o $n_2 = 6$ individus pour U_2 ,
- o $n_3 = 12$ individus pour U_3 .

On mesure un caractère quantitatif Y sur chacun d'entre eux. Les résultats obtenus sont :

Pour U_1	1320	1467				
Pour U_2	728	616	942	813	921	833
Pour U_3	211	248	215	271	353	327
	421	358	335	311	273	167

1. Donner une estimation ponctuelle de la moyenne-population \bar{y}_U .
2. Donner une estimation ponctuelle de l'écart-type de l'estimateur de \bar{y}_U .
3. Déterminer un intervalle de confiance pour \bar{y}_U au niveau 99%.

Exercice 5.

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

```
U_1 = c("Bob", "Nico", "Ali", "Fabien", "Malik", "John", "Jean",
"Chris", "Karl")
U_2 = c("Jean", "Bill", "Omar", "Raul", "Mia")
U_3 = c("Paul", "Chael", "Nathan", "Sam", "Tom", "Tim", "Leo", "Kevin")
dat = cbind.data.frame(c(U_1, U_2, U_3), c(rep(1, length(U_1)), rep(2,
length(U_2)), rep(3, length(U_3))))
names(dat) = c("noms", "souspop")
library(sampling)
s = strata(dat, "souspop", size = c(3, 2, 3), method = "srswor")
s
```

Cela renvoie :

	souspop	ID_unit	Prob	Stratum
3	1	3	0.3333333	1
7	1	7	0.3333333	1
9	1	9	0.3333333	1
12	2	12	0.4000000	2
14	2	14	0.4000000	2
15	3	15	0.3750000	3
19	3	19	0.3750000	3
22	3	22	0.3750000	3

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

```
U = c(U_1, U_2, U_3)
U[s[,2]]
```

Cela renvoie :

"Ali" "Jean" "Karl" "Omar" "Mia" "Paul" "Tom" "Kevin"

Exercice 6. On s'intéresse au chiffre d'affaire moyen en euros réalisé par 940 entreprises. Ces entreprises sont regroupées en 5 strates U_1, \dots, U_5 en fonction du nombre de leurs salariés :

Strate U_h	U_1	U_2	U_3	U_4	U_5
Nombre de salariés	0 à 9	10 à 19	20 à 39	40 à 99	plus de 100
Taille N_h	390	250	180	100	20

On prélève un échantillon de $n = 190$ entreprises suivant un plan de sondage aléatoire de type ST.

1. On dispose des informations suivantes :

Strate U_h	U_1	U_2	U_3	U_4	U_5
Taille n_h	90	50	30	15	5
Moyenne \bar{y}_{ω_h}	20000	35000	50000	90000	120000
Écart-type corrigé s_{ω_h}	1800	1200	2000	1800	2700

Déterminer un intervalle de confiance pour le chiffre d'affaire moyen des entreprises au niveau 95%.

2. On considère un plan de sondage aléatoire de type STP. Pour simplifier, on suppose que les écarts-type corrigés associés aux strates sont égaux à ceux présentés à la question 1.
 - (a) Déterminer les tailles des échantillons pour chacune des strates.
 - (b) Donner une estimation ponctuelle de la variance de l'estimateur de la moyenne-population \bar{y}_U .
3. On considère maintenant un plan de sondage aléatoire de type STO. Pour simplifier, on suppose que les écarts-type corrigés associés aux strates sont égaux à ceux présentés à la question 1.
 - (a) Déterminer les tailles des échantillons pour chacune des strates.
 - (b) Donner une estimation ponctuelle de la variance de l'estimateur de la moyenne-population \bar{y}_U .
4. Comparer les résultats 2 (b) et 3 (b) ?

Exercice 7. Une population U est partagée en 4 strates U_1, U_2, U_3 et U_4 . On prélève un échantillon de 77 individus suivant un plan de sondage aléatoire de type ST et on mesure un caractère quantitatif Y sur chacun d'entre eux. On dispose des informations suivantes :

Strate U_h	U_1	U_2	U_3	U_4
Taille N_h	210	120	130	210
Écart-type corrigé s_{U_h}	2.5	3.1	1.5	2.1

1. Quelle est l'effectif total de la population ?

2. On considère un plan de sondage aléatoire de type STP.
 - (a) Déterminer les tailles des échantillons pour chacune des strates.
 - (b) Calculer l'erreur quadratique moyenne de l'estimateur de la moyenne-population.
3. On considère maintenant un plan de sondage aléatoire de type STO.
 - (a) Déterminer les tailles des échantillons pour chacune des strates.
 - (b) Calculer l'erreur quadratique moyenne de l'estimateur de la moyenne-population.
4. Comparer les résultats des 2 plans de sondage considérés.

Exercice 8. Décrire brièvement l'enjeu des commandes R suivantes :

```

U_1 = c("Bob", "Nico", "Ali", "Fabien", "Malik", "John", "Jean",
"Chris", "Karl")
Y_1 = c(167, 171, 170, 169, 168, 171, 176, 161, 163)
U_2 = c("Jean", "Bill", "Omar", "Raul", "Mia")
Y_2 = c(187, 189, 181, 189, 183)
U_3 = c("Paul", "Chael", "Nathan", "Sam", "Tom", "Tim", "Leo", "Kevin")
Y_3 = c(197, 190, 191, 192, 191, 190, 192, 194)
dat = cbind.data.frame(c(U_1, U_2, U_3), c(rep(1, length(U_1)),
rep(2, length(U_2)), rep(3, length(U_3))))
names(dat) = c("noms", "souspop")
U = c(U_1, U_2, U_3)
Y = c(Y_1, Y_2, Y_3)
bar_y_w = NULL
var_bar_y_w = NULL
u = NULL
library(sampling)

for (i in (1:30)) {
s = strata(dat, "souspop", size = c(3, 2, 3), method = "srswor")
s
u = rbind(u,U[s[,2]])
N_h = c(length(U_1), length(U_2), length(U_3))
N = sum(N_h)
n_h = c(3, 2, 3)
y = Y[s[,2]]
bar_y_w_h = unlist(tapply(y, s$souspop, mean))
s_w_h = unlist(tapply(y, s$souspop, sd))
bar_y_w[i] = sum(N_h * bar_y_w_h) / N
var_bar_y_w[i] = (1 / N^2) * sum(N_h^2 * (1-n_h / N_h) *
(s_w_h^2 / n_h))
}

u

```

Cela renvoie :

```
[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] "Fabien" "John" "Jean" "Jean" "Bill" "Sam" "Tom" "Tim"
[2,] "Bob" "Chris" "Karl" "Omar" "Raul" "Paul" "Nathan" "Leo"
[3,] "Malik" "Jean" "Chris" "Raul" "Mia" "Nathan" "Tom" "Tim"
[4,] "Ali" "John" "Jean" "Jean" "Mia" "Chael" "Nathan" "Leo"
[5,] "Fabien" "Jean" "Karl" "Jean" "Mia" "Tom" "Leo" "Kevin"
[6,] "Ali" "Jean" "Karl" "Bill" "Omar" "Chael" "Tim" "Leo"
[7,] "Fabien" "John" "Karl" "Jean" "Raul" "Nathan" "Leo" "Kevin"
[8,] "Bob" "Fabien" "Karl" "Jean" "Mia" "Paul" "Chael" "Kevin"
[9,] "Bob" "Ali" "Malik" "Raul" "Mia" "Chael" "Sam" "Leo"
[10,] "John" "Jean" "Chris" "Jean" "Bill" "Chael" "Nathan" "Tom"
[11,] "Bob" "Ali" "Karl" "Jean" "Bill" "Paul" "Sam" "Leo"
[12,] "John" "Chris" "Karl" "Bill" "Mia" "Nathan" "Tom" "Tim"
[13,] "Ali" "Malik" "Karl" "Bill" "Mia" "Paul" "Sam" "Tim"
[14,] "Bob" "Fabien" "John" "Bill" "Raul" "Nathan" "Sam" "Tom"
[15,] "Bob" "Nico" "Jean" "Jean" "Mia" "Nathan" "Sam" "Tom"
[16,] "Nico" "Malik" "John" "Jean" "Mia" "Paul" "Leo" "Kevin"
[17,] "Nico" "John" "Karl" "Bill" "Mia" "Paul" "Nathan" "Tim"
[18,] "Bob" "Nico" "Malik" "Bill" "Omar" "Paul" "Tim" "Leo"
[19,] "Bob" "Jean" "Karl" "Jean" "Mia" "Paul" "Nathan" "Sam"
[20,] "Bob" "Ali" "Malik" "Jean" "Raul" "Paul" "Sam" "Tim"
[21,] "Fabien" "Malik" "Chris" "Omar" "Raul" "Paul" "Chael" "Tom"
[22,] "Ali" "Malik" "Jean" "Bill" "Raul" "Paul" "Sam" "Tim"
[23,] "Fabien" "Malik" "John" "Jean" "Raul" "Paul" "Nathan" "Kevin"
[24,] "John" "Jean" "Chris" "Jean" "Omar" "Paul" "Nathan" "Tom"
[25,] "Bob" "Ali" "John" "Raul" "Mia" "Nathan" "Tim" "Leo"
[26,] "Fabien" "Jean" "Chris" "Omar" "Raul" "Paul" "Nathan" "Tim"
[27,] "Nico" "Chris" "Karl" "Omar" "Raul" "Paul" "Nathan" "Tom"
[28,] "Fabien" "Jean" "Karl" "Raul" "Mia" "Nathan" "Tom" "Kevin"
[29,] "Bob" "Jean" "Karl" "Bill" "Omar" "Nathan" "Sam" "Leo"
[30,] "Fabien" "John" "Jean" "Jean" "Bill" "Paul" "Nathan" "Tom"
```

```
z = qnorm(1 - (1 - 0.95) / 2)
icmin = bar_y_w - z * sqrt(var_bar_y_w)
icmax = bar_y_w + z * sqrt(var_bar_y_w)
ST = data.frame(bar_y_w, var_bar_y_w, icmin, icmax)

ST
```

Cela renvoie :

```
bar_y_w var_bar_y_w icmin icmax
1 182.5455 0.5420110 181.1025 183.9884
2 179.3030 1.1276400 177.2217 181.3843
3 180.4697 2.3831497 177.4440 183.4954
4 182.0000 0.5358127 180.5653 183.4347
5 181.2576 1.7626263 178.6555 183.8597
6 180.7879 2.1069789 177.9429 183.6329
7 181.2576 0.7398990 179.5717 182.9435
8 180.5152 0.8108356 178.7503 182.2800
9 180.7121 0.4024334 179.4688 181.9555
10 181.3333 2.2095960 178.4199 184.2468
11 181.3333 0.7192378 179.6711 182.9955
12 179.1061 1.3294307 176.8462 181.3659
13 180.7727 1.1205234 178.6980 182.8474
14 181.6667 0.1579431 180.8877 182.4456
15 181.7121 0.8893480 179.8638 183.5605
16 182.2576 0.4100092 181.0026 183.5126
17 181.1970 1.4671717 178.8229 183.5710
18 181.2273 1.0151515 179.2525 183.2020
19 181.3485 2.0573921 178.5372 184.1598
20 181.7727 0.4758953 180.4206 183.1248
21 180.0152 1.5973370 177.5380 182.4923
22 183.2273 1.0027548 181.2646 185.1899
23 182.5455 0.3657025 181.3602 183.7307
24 181.2727 2.7789256 178.0054 184.5400
25 181.0000 0.4676309 179.6597 182.3403
26 181.1061 2.9857668 177.7194 184.4928
27 179.7273 1.8677686 177.0487 182.4059
28 181.3636 1.9359504 178.6366 184.0907
29 180.7424 2.1538108 177.8660 183.6188
30 183.2727 0.8450413 181.4710 185.0744
```

```
bar_y_U = rep(mean(Y), 30)
library(lattice)
x = 1:30
xyplot(ST$icmin + ST$icmax + bar_y_w + bar_y_U ~ x, col = c("red",
"blue", "green", "orange"), type = "o", main = "Sondage aléatoire ST",
xlab = " ", ylab = " ")
```

Cela renvoie :

