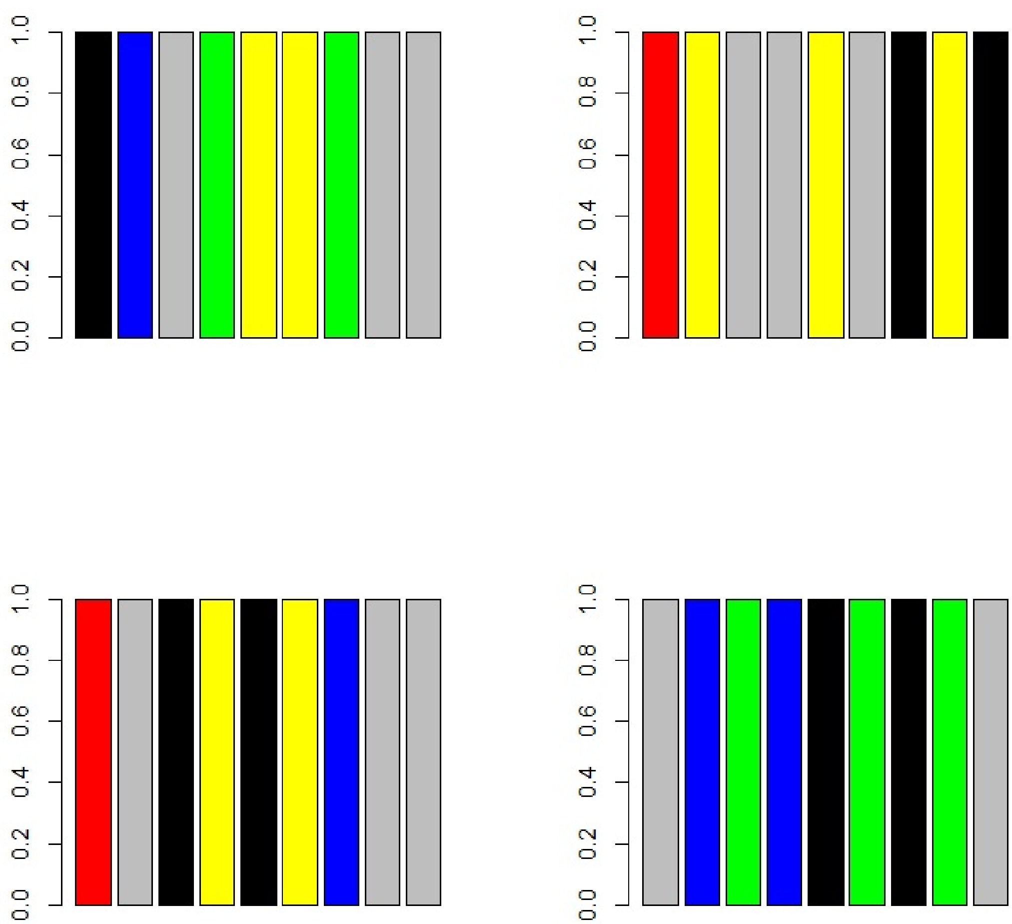
TD n° 3 : Plans de sondage aléatoire PEAR

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 avec remise 2 individus dans U formant ainsi un échantillon (l'ordre est pris en compte).

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

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

```
colors = c("red", "blue", "yellow", "green", "grey", "black")
par(mfrow = c(2, 2))
for (i in 1:4) {
s = sample(colors, size = 9, replace = T)
barplot(rep(1, 9), col = s)
}
```



Exercice 3.

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

```
liste = NULL
for(i in 1:3) {
for (j in 1:3) {
for (k in 1:3) {
liste = rbind(liste, c(i, j, k))
}}}
liste
```

Cela renvoie :

```
      [,1] [,2] [,3]
 [1,]    1    1    1
 [2,]    1    1    2
 [3,]    1    1    3
 [4,]    1    2    1
 [5,]    1    2    2
 [6,]    1    2    3
 [7,]    1    3    1
 [8,]    1    3    2
 [9,]    1    3    3
[10,]    2    1    1
[11,]    2    1    2
[12,]    2    1    3
[13,]    2    2    1
[14,]    2    2    2
[15,]    2    2    3
[16,]    2    3    1
[17,]    2    3    2
[18,]    2    3    3
[19,]    3    1    1
[20,]    3    1    2
[21,]    3    1    3
[22,]    3    2    1
[23,]    3    2    2
[24,]    3    2    3
[25,]    3    3    1
[26,]    3    3    2
[27,]    3    3    3
```

2. On introduit la population $U = \{\text{Bob, Malik, Seb}\}$:

```
U = c("Bob", "Malik", "Seb")
```

(a) À partir de U, modifier simplement les commandes R de la première question afin d'obtenir la liste :

```
      [,1]    [,2]    [,3]
 [1,] "Bob"   "Bob"   "Bob"
 [2,] "Bob"   "Bob"   "Malik"
 [3,] "Bob"   "Bob"   "Seb"
 [4,] "Bob"   "Malik" "Bob"
 [5,] "Bob"   "Malik" "Malik"
 [6,] "Bob"   "Malik" "Seb"
 [7,] "Bob"   "Seb"   "Bob"
 [8,] "Bob"   "Seb"   "Malik"
 [9,] "Bob"   "Seb"   "Seb"
[10,] "Malik" "Bob"   "Bob"
[11,] "Malik" "Bob"   "Malik"
[12,] "Malik" "Bob"   "Seb"
[13,] "Malik" "Malik" "Bob"
[14,] "Malik" "Malik" "Malik"
[15,] "Malik" "Malik" "Seb"
[16,] "Malik" "Seb"   "Bob"
[17,] "Malik" "Seb"   "Malik"
[18,] "Malik" "Seb"   "Seb"
[19,] "Seb"   "Bob"   "Bob"
[20,] "Seb"   "Bob"   "Malik"
[21,] "Seb"   "Bob"   "Seb"
[22,] "Seb"   "Malik" "Bob"
[23,] "Seb"   "Malik" "Malik"
[24,] "Seb"   "Malik" "Seb"
[25,] "Seb"   "Seb"   "Bob"
[26,] "Seb"   "Seb"   "Malik"
[27,] "Seb"   "Seb"   "Seb"
```

(b) Que représente cette liste par rapport à U ?

3. On prélève un échantillon de $n = 3$ individus de U suivant un plan de sondage aléatoire de type PEAR.

   (a) Quelle est la probabilité qu'il contienne uniquement Bob ?

   (b) Quelle est la probabilité qu'il contienne Bob ?

   (c) Quelle est la probabilité qu'il contienne Bob et Malik ?

   (d) Quelle est la probabilité qu'il contienne Bob, Malik et Seb ?

   (e) Quelle est la probabilité qu'il contienne uniquement Bob et Malik ?

   (f) Quelle est la probabilité qu'il contienne uniquement 2 individus différents ?

Exercice 4. *L'objectif de cet exercice est d'illustrer certains résultats théoriques du cours sur les plans de sondage aléatoire de type PEAR avec un exemple.* On étudie un caractère Y dans une population de 3 individus : $U = \{u_1, u_2, u_3\}$. 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
4	7	9

1. Calculer la moyenne-population $\overline{y}_U$ et l'écart-type corrigé-population s_U.

2. On prélève au hasard et avec remise 2 individus de U formant ainsi un échantillon, l'ordre d'apparition étant pris en compte. Un individu peut donc apparaitre plusieurs fois dans 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 PEAR.

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

   (b) Pour chaque échantillon ω, calculer la moyenne-échantillon $\overline{y}_\omega$ et l'écart-type corrigé-échantillon s_ω.

   (c) Soit $\overline{y}_W$ 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.

Exercice 5. Dans une population de N individus $U = \{u_1, \ldots, u_N\}$, on prélève au hasard et avec remise n individus pour former un échantillon. Chaque individu a la même probabilité qu'un autre d'être sélectionné. On est dans le cadre d'un plan de sondage aléatoire de type PEAR (un individu peut apparaître plusieurs fois dans un échantillon). Soit W la *var* égale à l'échantillon obtenu :

$$W = (W_1, \ldots, W_n),$$

où, pour tout $m \in \{1, \ldots, n\}$, W_m est la *var* égale au m-ème individu de l'échantillon.

1. Montrer que, pour tout $m \in \{1, \ldots, n\}$, la loi de W_i est donnée par

$$\mathbb{P}(W_m = u_i) = \frac{1}{N}, \qquad i \in \{1, \ldots, N\},$$

   où $\mathbb{P}$ désigne la probabilité uniforme.

2. Montrer que, pour tout $i \in \{1, \ldots, N\}$,

$$\mathbb{P}(u_i \in W) = 1 - \left(1 - \frac{1}{N}\right)^n.$$

3. Désormais, on étudie un caractère Y dans U. Soient $\overline{y}_U$ la moyenne-population et s_U l'écart-type corrigé-population. Un estimateur aléatoire de $\overline{y}_U$ est
$$\overline{y}_W = \frac{1}{n}\sum_{i=1}^{N} y_i \sum_{m=1}^{n} \mathbb{1}_{\{W_m=u_i\}}.$$
   (a) Montrer que
$$\mathbb{E}(\overline{y}_W) = \overline{y}_U.$$
   (b) Montrer que
$$\mathbb{V}(\overline{y}_W) = \frac{N-1}{N} s_U^2.$$

Exercice 6. Une semaine donnée, le gérant d'un restaurant japonais livrant à domicile souhaite faire une étude sur le temps moyen qui s'écoule entre le moment où un client passe commande par internet et le moment où le client est livré. Sur les 598 commandes internet de la semaine, il prélève un échantillon de 17 commandes suivant un plan de sondage aléatoire de type PEAR. Les temps mesurés donnent une moyenne de 19 minutes et un écart-type corrigé de 1.2. On suppose que le temps considéré en minutes cette semaine est une *var* suivant une loi normale.

Déterminer un intervalle de confiance pour la moyenne des temps des 598 commandes internet de la semaine au niveau 99%.

Exercice 7. On étudie le caractère taille en centimètres sur 8 filles. On suppose que ce caractère se modélise comme une *var* suivant une loi normale.

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 = 40
bar_y_w = NULL
var_bar_y_w = NULL
for (i in 1:m){
t = srswr(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] = s2_w / n
}
q = qt(0.975, n - 1)
icmin = bar_y_w - q * sqrt(var_bar_y_w)
icmax = bar_y_w + q * sqrt(var_bar_y_w)
longueur_ic = icmax - icmin
PEAR = data.frame(bar_y_w, var_bar_y_w, icmin, icmax, longueur_ic)
PEAR
```

Cela renvoie :

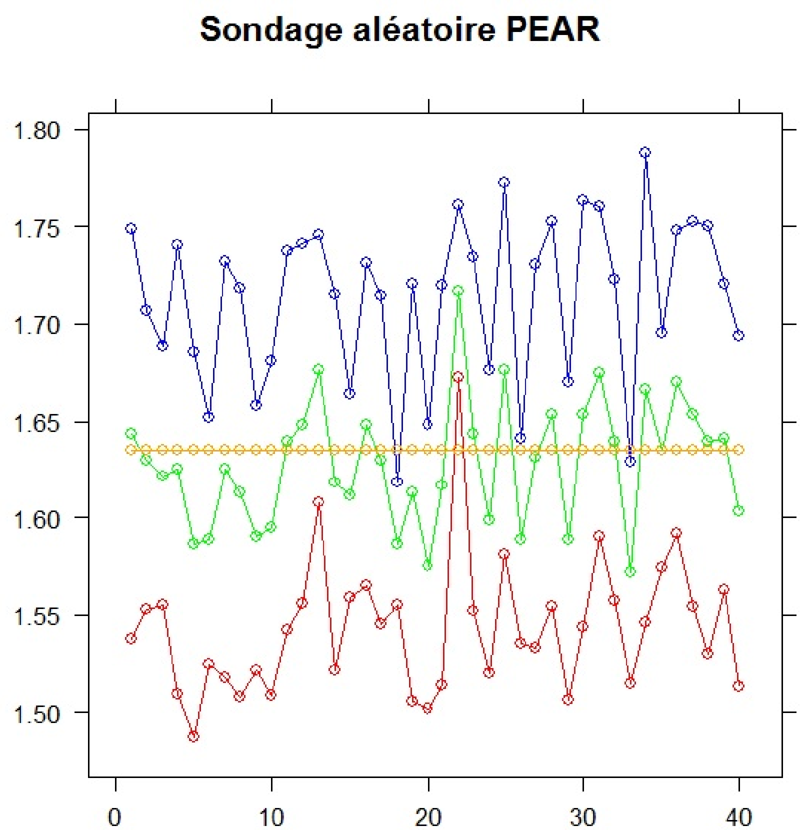
	bar_y_w	var_bar_y_w	icmin	icmax	longueur_ic
1	1.643333	0.0016911111	1.537623	1.749044	0.21142070
2	1.630000	0.0008933333	1.553169	1.706831	0.15366261
3	1.621667	0.0006761111	1.554826	1.688507	0.13368124
4	1.625000	0.0020250000	1.509324	1.740676	0.23135237
5	1.586667	0.0014844444	1.487626	1.685707	0.19808127
6	1.588333	0.0006161111	1.524527	1.652139	0.12761183
7	1.625000	0.0017316667	1.518030	1.731970	0.21394078
8	1.613333	0.0016711111	1.508250	1.718417	0.21016679
9	1.590000	0.0007066667	1.521666	1.658334	0.13666860
10	1.595000	0.0011183333	1.509036	1.680964	0.17192818
11	1.640000	0.0014400000	1.542453	1.737547	0.19509344
12	1.648333	0.0012961111	1.555788	1.740878	0.18508983
13	1.676667	0.0007244444	1.607478	1.745855	0.13837702
14	1.618333	0.0014161111	1.521599	1.715068	0.19346842
15	1.611667	0.0004161111	1.559230	1.664103	0.10487358
16	1.648333	0.0010494444	1.565059	1.731608	0.16654867
17	1.630000	0.0010800000	1.545522	1.714478	0.16895588
18	1.586667	0.0001511111	1.555067	1.618266	0.06319892
19	1.613333	0.0017511111	1.505764	1.720903	0.21513857
20	1.575000	0.0008116667	1.501765	1.648235	0.14647054
21	1.616667	0.0016044444	1.513701	1.719633	0.20593197
22	1.716667	0.0002977778	1.672308	1.761025	0.08871715
23	1.643333	0.0012577778	1.552167	1.734499	0.18233221
24	1.598333	0.0009227778	1.520246	1.676421	0.15617445
25	1.676667	0.0013911111	1.580790	1.772543	0.19175308
26	1.588333	0.0004294444	1.535063	1.641604	0.10654055
27	1.631667	0.0014694444	1.533128	1.730206	0.19707794
28	1.653333	0.0014844444	1.554293	1.752374	0.19808127
29	1.588333	0.0010161111	1.506392	1.670274	0.16388229
30	1.653333	0.0018244444	1.543535	1.763132	0.21959718
31	1.675000	0.0010916667	1.590067	1.759933	0.16986600
32	1.640000	0.0010333333	1.557367	1.722633	0.16526529
33	1.571667	0.0004961111	1.514411	1.628923	0.11451198
34	1.666667	0.0022044444	1.545974	1.787359	0.24138541
35	1.635000	0.0005583333	1.574260	1.695740	0.12148096
36	1.670000	0.0009200000	1.592030	1.747970	0.15593922
37	1.653333	0.0014844444	1.554293	1.752374	0.19808127
38	1.640000	0.0018400000	1.529734	1.750266	0.22053135
39	1.641667	0.0009361111	1.563017	1.720316	0.15729870
40	1.603333	0.0012377778	1.512895	1.693772	0.18087676

2. Quel est le niveau des intervalles de confiance considérés ?

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(PEAR$icmin + PEAR$icmax + bar_y_w + bar_y_U ~ x, col = c("red",
"blue", "green", "orange"), type = "o", main = "Sondage aléatoire PEAR",
xlab = " ", ylab = " ")
```



Retrouver graphiquement le résultat de la question précédente.

Exercice 8. Sur les 75 sacs de farine de maïs d'une petite production, on prélève un échantillon de 25 sacs suivant un plan de sondage aléatoire de type PEAR. On pèse ces 25 sacs. Les valeurs obtenues donnent une moyenne de 13.5 kilogrammes et un écart-type corrigé de 1.3 kilogrammes.

On suppose que le poids en kilogrammes d'un sac de farine de maïs issu de cette production peut être modélisé par une *var* Y suivant une loi normale.

1. Déterminer un intervalle de confiance pour la moyenne des poids des 75 sacs de la production au niveau 95%.

2. Déterminer la taille d'échantillon à choisir pour avoir une incertitude absolue sur la moyenne des poids des 75 sacs inférieure ou égale à 0.5 au niveau 90%.