

TD n° 4 : Estimation d'autres paramètres ; PEAR

Exercice 1. Une firme immobilière s'intéresse à la valeur totale des coûts pour la construction de 126 maisons. Pour cela, un échantillon de 15 maisons est sélectionné suivant un plan de sondage aléatoire de type PEAR. Les coûts en euros pour chacune de ces 15 maisons sont :

35 505	30 200	28 700
36 450	29 800	35 300
30 200	26 400	38 000
38 100	33 200	30 200
31 600	35 300	38 200

1. Donner une estimation ponctuelle du coût total des 126 maisons.
2. Donner une estimation ponctuelle de l'écart-type de l'estimateur du coût total des 126 maisons.
3. Déterminer un intervalle de confiance au niveau 95% pour le coût total des 126 maisons.

Exercice 2. Dans un premier temps, on considère le caractère $X = \text{"âge"}$ en années dans une population de 5 garçons : $U = \{\text{André, Jean, Baptiste, Matt, Gabin}\} = \{u_1, \dots, u_5\}$. Pour tout $i \in \{1, \dots, 5\}$, soit x_i la valeur de X pour l'individu u_i . Les résultats, en années, sont :

x_1	x_2	x_3	x_4	x_5
18.5	12	17	15	19

On étudie le caractère $Y = \mathbb{1}_{\{X \geq 16\}}$ qui vaut 1 si l'individu considéré a un age supérieur ou égal à 16 ans et 0 sinon. 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
1	0	1	0	1

On s'intéresse à la proportion des individus dont l'âge est supérieur ou égal à 16 ans.

1. Calculer la proportion-population p_U .
2. On prélève au hasard et avec remise 3 individus dans U formant ainsi un échantillon. Chaque individu a la même probabilité qu'un autre d'être sélectionné.
 - (a) De quel type est ce plan de sondage ? Combien d'échantillons peut-on former ?
 - (b) Soit p_W la *var* égale à la proportion-échantillon, l'aléatoire étant dans l'échantillon considéré. Donner, sans calcul, son espérance et sa variance.

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

```

U = c("André", "Jean", "Baptiste", "Matt", "Gabin")
x = c(18.5, 12, 17, 15, 19)
y = as.numeric(x >= 16)
library(sampling)
N = 5
n = 3
M = 40
p_w = NULL
ech = NULL
for (i in 1:M){
t = srswr(n, N)
ech = rbind(ech, c(rep(U[t == 1], 1), rep(U[t == 2], 2),
rep(U[t == 3], 3)))
p_w[i] = (1/n) * sum(y * t)
}
PEAR = data.frame(ech, p_w)
PEAR

```

Cela renvoie :

	X1	X2	X3	p_w
1	Baptiste	André	André	1.0000000
2	Jean	Baptiste	Matt	0.3333333
3	André	Baptiste	Gabin	1.0000000
4	André	Jean	Jean	0.3333333
5	André	Baptiste	Matt	0.6666667
6	Jean	Baptiste	Matt	0.3333333
7	André	Baptiste	Matt	0.6666667
8	Gabin	Baptiste	Baptiste	1.0000000
9	André	Jean	Baptiste	0.6666667
10	André	Baptiste	Gabin	1.0000000
11	André	Baptiste	Baptiste	1.0000000
12	Baptiste	Jean	Jean	0.3333333
13	André	Jean	Jean	0.3333333
14	André	Matt	Gabin	0.6666667
15	Baptiste	Gabin	Gabin	1.0000000
16	André	Jean	Gabin	0.6666667
17	Baptiste	Matt	Gabin	0.6666667
18	Jean	Baptiste	Gabin	0.6666667
19	Baptiste	Matt	Gabin	0.6666667
20	Baptiste	Matt	Gabin	0.6666667
21	Baptiste	Matt	Gabin	0.6666667
22	André	Jean	Matt	0.3333333
23	André	Jean	Gabin	0.6666667
24	Gabin	Matt	Matt	0.3333333
25	Jean	Baptiste	Gabin	0.6666667
26	André	Matt	Gabin	0.6666667
27	Gabin	Matt	Matt	0.3333333
28	Matt	Matt	Matt	0.0000000
29	André	Jean	Gabin	0.6666667
30	André	Jean	Matt	0.3333333
31	Jean	Gabin	Gabin	0.6666667
32	Matt	André	André	0.6666667
33	Jean	Matt	Gabin	0.3333333
34	André	Matt	Gabin	0.6666667
35	Baptiste	Matt	Matt	0.3333333
36	André	Matt	Gabin	0.6666667
37	Baptiste	Matt	Gabin	0.6666667
38	André	Jean	Jean	0.3333333
39	André	Matt	Gabin	0.6666667
40	Jean	Baptiste	Baptiste	0.6666667

Exercice 3. Un échantillon de 1550 étudiants d'une université est sélectionné suivant un plan de sondage aléatoire de type PEAR. Dans cet échantillon, 850 étudiants boivent régulièrement du café. On suppose que cette université compte 10000 étudiants et on note p_U la proportion de buveurs de café au sein des étudiants de cette université.

1. Donner une estimation ponctuelle de p_U .
2. Donner une estimation ponctuelle de la variance de l'estimateur de p_U .
3. Déterminer un intervalle de confiance au niveau 95% pour p_U .
4. Déterminer la taille d'échantillon à choisir pour avoir une incertitude absolue sur p_U inférieure ou égale à 0.12 au niveau 95%.

Exercice 4. Dans un de ses circuits, une agence de voyages propose la visite d'un musée sous la forme d'une option supplémentaire. En prélevant au hasard 60 fiches de clients parmi les 2379 de la période considérée suivant un plan de sondage aléatoire de type PEAR, on observe que seulement 14 fiches comportent cette option. Soit p_U la proportion des 2379 clients ayant pris l'option.

Déterminer un intervalle de confiance pour p_U au niveau 95%.

Exercice 5.

1. Que détermine la fonction R ci-dessous ?

```
icPEAR = fonction(y) {  
  n = length(y)  
  p_w = mean(y)  
  var_p_w = p_w * (1 - p_w) / (n - 1)  
  a = p_w - 1.96 * sqrt(var_p_w)  
  b = p_w + 1.96 * sqrt(var_p_w)  
  print(c(a, b))  
}
```

2. Quelles sont les valeurs renvoyées par les commandes ci-dessous ?

```
icPEAR(y = c(1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0))
```

3. Que détermine la fonction R ci-dessous ?

```
n_ech = fonction(p_w, d0) {  
  n = p_w * (1 - p_w) * 1.96^2 / d0^2  
  print(ceiling(n))  
}
```

4. Quelles sont les valeurs renvoyées par les commandes ci-dessous ?

```
n_ech(p_w = 0.45, d0 = 0.2)
```