TP_3

EL_Hadrami_N'DOYE_Ismael_Ramde

29/01/2021

```
source("ScriptsTP3.R")
```

(a) Donnons un intervalle de confiance à 80% pour le nombre total de signatures sur l'ensemble de 676 feuilles.

```
echibarr <- mean(ech)
n <- length(ech)
N <- 676
sigmaybarr <- sqrt(1 - (n / N)) * (sqrt(var(ech)) / sqrt(n))
b.inf <- echibarr - (1.28 * sigmaybarr)
b.sup <- echibarr + (1.28 *sigmaybarr)
N * b.inf
```

```
## [1] 18106
```

```
N * b.sup
```

```
## [1] 21669.84
```

(b) Amelioration de l'estimation du nombre total de signatures sur les 676 feuilles

On calcul l'intervalle de confiance a 80% le nombre total de signature sur l'ensemble des feuilles restantes puis on ajoute sur chaque borne de l'intervalle les 326 feuilles ayant 42 signatures.

```
N42sig <- 326 * 42
Nr <- N - 326
echr <- ech[ech!=42]
echrbar <- mean(ech)
nr <- length(echr)
sybar <- sqrt(1 - (nr / Nr)) * (sqrt(var(echr)) / sqrt(nr))
b.inf <- echrbar - (1.28 * sybar)
b.sup <- echrbar + (1.28 *sybar)
(Nr * b.inf) + N42sig
```

```
## [1] 22901.83
```

```
(Nr * b.sup) + N42sig
```

```
## [1] 25076.17
```

2

(a) Est-ce qu'on gagne beaucoup à connaitre la taille des domaines?

Oui, car cela nous permettra d'avoir une meilleure estimation.

(b) Est-ce que l'un des deux domaines permet une meilleure estimation du total ? Utilisez l'écart type de l'estimateur, ainsi que son coefficient de variation comme critère de qualité.

- Hypothese 1 :les tailles du domaines sont connues

```
# estimation du salaire total pour chaque domaine
estimtot.D1 <- ND.20 * ybar1
estimtot.D1
```

```
## [1] 1071087
```

```
estimtot.D2 <- ND.30 * ybar2
estimtot.D2
```

```
## [1] 5626702
```

```
# estimation de l'ecart type
f.D1 <- nd1 / ND.20
f.D2 <-  nd2 / ND.30
ecartype.estimD1 <- sqrt(1 - f.D1) * (sd1 / sqrt(nd1))
ecartype.estimD2 <- sqrt(1 - f.D2) * (sd2 / sqrt(nd2))
# Coeficient de variation
coefv.D1 <- ecartype.estimD1 / ybar1
coefv.D1
```

```
## [1] 0.06389235
```

```
coefv.D2 <- ecartype.estimD2 / ybar2
coefv.D2
```

```
## [1] 0.03213921
```

Le coefficient de variation du domaine 2 donne la valeur la plus petite.

Donc le domaine 2 donne la meilleure estimation total du salaire.

- Hypothese 2 :les tailles du domaines ne sont pas connues.

```
N <- 200
n <- 50
f <- n / N
estimtot1 <- N * ybar1.prime
estimtot1
```

```
## [1] 1428116
```

```
estimtot2 <- N * ybar2.prime
estimtot2
```

```
## [1] 6317700
```

```
ecartype.estim1 <- sqrt(1 - f) * (sd1.prime / sqrt(n))
ecartype.estim2 <- sqrt(1 - f) * (sd2.prime/ sqrt(n))
# Coeficient de variation
coefv1 <- ecartype.estim1 / ybar1.prime
coefv1
```

```
## [1] 0.2906159
```

```
coefv2 <- ecartype.estim2 / ybar2.prime
coefv2
```

```
## [1] 0.09835826
```

Dans cette hypothese le coefficient de variation du domaine 2 donne la valeur la plus petite.

Donc le domaine 2 donne la meilleure estimation total du salaire que celui du domaine 1.

3. On prélève un échantillon de 25 librairies dans une ville ayant 250 librairies afin d'estimer le nombre total de livres espagnols vendus dans la ville au courant du mois dernier.

(a) Estimons le nombre total de livres espagnols vendus dans la ville au courant du mois dernier, et estimez l'écart type de votre estimateur.

```
N <- 250
n <- 25
estim.totE <- N * mean(y)
# estimation de l'ecart type
f <- (n / N)
estim.ecart <- sqrt(1 - f) * (sqrt(var(y)) / sqrt(n))
estim.ecart
```

```
## [1] 0.4130859
```

Donc on estime que 340 livres espagnols ont éte vendus lors du mois dernier

(b) Supposons maintenant que vous apprenez qu'une étude faite auprès de la population entière a révélé que 175 librairies n'ont pas vendu de livres espagnols. Utilisons cette information pour arriver à une deuxième estimation du nombre total de livres espagnols vendus au courant du mois dernier. Laquelle des deux estimations semble plus précise?

```
y <- 0:8
# effectif
e <- c(175, 3, 2, 4, 0, 0, 0, 1, 1)
y.prime <- rep(y,e)
n.prime <- length(y.prime)
N <- 250
f.prime <- n.prime / N
estim.totE.prime <- N * round(mean(y.prime),2)
estim.totE.prime
```

```
## [1] 45
```

```
estim.ecart.prime <- sqrt(1 - f.prime) * (sqrt(var(y.prime)) / sqrt(n.prime))
estim.ecart.prime
```

```
## [1] 0.03382475
```

Donc on estime que 45 livres espagnols ont éte vendus lors du mois dernier.

Cette estimation semble etre plus precise car elle minimise l'ecart type estimé.

4. Considérons la population de N = 8 unités pour lesquels sont définies deux variables, X et Y.

(a) Montrons que l'estimateur $\hat{R}$ est biaisé, comme on le sait, et exprimons une opinion sur l'importance du biais dans ce cas.

```
# calcul de R
R <- ybar.pop / xbar.pop
R.chapbar <- mean(stats[,3])
R.chapbar - R
```

```
## [1] 0.02453898
```

Donc $\hat{R}$ est un estimateur biaisé, ce biais montre que la moyenne n'est pas toujours la vraie estimateur de l'esperance.

(b) Calul de la variance de $\hat{R}$

```
N <- 8
n <- 3
Rchap <- stats[,3]
varrchap <- var(Rchap)
varrchap
```

```
## [1] 0.04533455
```

```
S.carre <- (1 / (N - 1)) * sum((Rchap - mean(Rchap))^2)
S.carre
```

```
## [1] 0.3562
```

```
sigmarchap <- (1 -(n / N)) * (S.carre / n)
sqrt(sigmarchap)
```

```
## [1] 0.2724121
```

c)La probabilité de se tromper de plus de 20 % dans l'estimation du quotient ?

```
prob <- length(which(Rchap > R + 0.2)) / length(Rchap)
```

Donc la probabilité de se tromper plus de 20% est de 21.42%

(d) Comparons $\sigma_{\hat{R}}$ avec la quantité obtenue par la formule usuelle

```
f <- n / N
expr1 <- sqrt(1 - f) / (xbar.pop * sqrt(n))
expr2 <- sqrt(sum((pop.y - (R*pop.x))^2) / (N - 1))
sigma.estim <- expr1 * expr2
sigma.estim
```

```
## [1] 0.1739178
```

e) Est-ce que s_{xy} est un estimateur sans biais de S_{XY} ?

```
sy <- apply(ech.all[,(n+1):(2*n)],1,sd)
sx <- apply(ech.all[,(2*n+1):(3*n)],1,sd)
sxy <- sy * sx
SXY <- (1 / (N - 1)) * sum((pop.x - xbar.pop) * (pop.y - ybar.pop))
mean(sxy) - SXY
```

```
## [1] 23.58307
```

$\mathrm{E}[s_{xy}] - S_{XY} \neq 0$ donc s_{xy} n'est pas un estimateur sans biais de S_{XY}.