

**Ecole Supérieure de la Statistique et de l'analyse de  
l'Information**



# **Enquête sur les jeunes et le téléphone portable**

**Rapport de projet Sondages**

**Bahria Hala**

**Zaouali Omar**

**Hamza Wided**



## I. Table des matières

II.	Introduction .....	2
III.	Mise en œuvre du sondage .....	3
1.	Définition de la population et de la variable d'intérêt.....	3
2.	Introduction de variables auxiliaires .....	3
3.	Mode de collecte .....	3
4.	Le questionnaire.....	3
5.	Exploration des données.....	5
IV.	Réalisation des sondages.....	9
1.	Sondage aléatoire simple .....	9
i.	Plan simple sans remise .....	10
ii.	Plan simple avec remise .....	10
2.	Sondage à probabilité inégales .....	12
i.	Plan de sondage avec remise .....	12
ii.	Plan simple sans remise .....	13
iii.	Synthèse, PISR, PIAR, PEAR, PESR .....	15
3.	Sondage par stratification .....	15
i.	Strates de même taille.....	17
ii.	Strates à allocation proportionnelles.....	18
iii.	Strates à allocation optimale .....	20
	La moyenne des $\bar{y}_{\text{chapeau}} = 2.48203$ .....	20
iv.	Synthèse .....	21
4.	Sondage par grappes et à deux degrés.....	22
i.	Plan de sondage par grappes.....	22
ii.	Plan de sondage par grappe e à deux degrés.....	23
iii.	Synthèse .....	24
V.	Redressement .....	25
VI.	Synthèse .....	26
VII.	Annexe .....	27

## II. Introduction

Dans le cadre du projet de sondage effectué en deuxième année, nous avons mené une recherche sur les jeunes et les téléphones portables. En effet, l'attachement des jeunes à leurs téléphones portables est remarquable, ils ont tendance souvent à les changer à la dernière gamme mise sur le marché. Nous avons choisi de mettre l'accent sur ce comportement en se posant la question combien de téléphone portable a possédé un jeune pendant les dernières trois années ?

Pour mettre en relief ce comportement, nous avons cherché les différents facteurs qui l'expliquent. En outre, les dépenses téléphoniques qu'un jeune est apte à consacrer de son budget mensuel ainsi de suite.

L'objectif est de manipuler les différentes méthodes de sondage vues en cours et comparer leurs précisions. Nous avons alors questionné cent étudiants, comme les étudiants est notre population cible, et obtenir une base sur laquelle nous avons appliqué nos différents méthodes.

### III. Mise en œuvre du sondage

#### 1. Définition de la population et de la variable d'intérêt

Comme notre enquête s'intéresse aux jeunes étudiants, nous avons choisie comme population source les étudiants de l'ESSAIT.

- La population cible : les étudiants.
- La population source : les étudiants de l'ESSAIT.
- La variable d'intérêt : le nombre de téléphones possédé pendant les trois dernières années.

#### 2. Introduction de variables auxiliaires

Afin d'expliquer le comportement des jeunes vis-à-vis la possession d'un téléphone portable, nous nous sommes demandé combien de téléphone portable obtient-il un jeune comme cadeau parmi les téléphones qu'il a possédé ? Ceci explique-t-il notre variable d'intérêt ? Nous avons ajouté aussi les dépenses en téléphonies ainsi que le budget mensuel qui nous a paru liés à ce comportement.

#### 3. Mode de collecte

La collecte de nos informations était effectuée sur terrain, en utilisant un formulaire à remplir

#### 4. Le questionnaire

# Enquête sur les jeunes et le téléphone portable

1) Vous êtes

Homme ☐

Femme ☐

---

2) Combien de téléphones portables avez-vous possédé pendant les 3 dernières années ?

1 ☐

2 ☐

3 ☐

4 ou plus ☐

---

3) Parmi ces téléphones, combien en avez-vous reçu comme cadeau ?

.....

---

4) Quelles ont été vos dépenses en téléphonie le mois dernier ? (téléphone, internet et accessoires)

Moins de 10 dinars ☐

[10,20[ ☐

[20,30[ ☐

plus de 30

dinars ☐

---

5) Quel est votre budget mensuel ? (hormis le loyer)

Moins de 200 dinars ☐

[200,300[ ☐

300 dinars ou plus ☐

---

6) Combien vous a coûté votre téléphone portable actuel ?

Moins de 200 dinars ☐

[200,500[ ☐

[500,800[ ☐

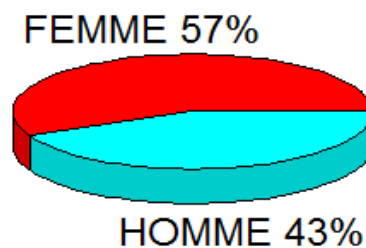
800 dinars ou plus ☐

## 5. Exploration des données

Notre base de données obtenue est constituée de 6 variables et 100 observations :

- Question 1 : Sexe

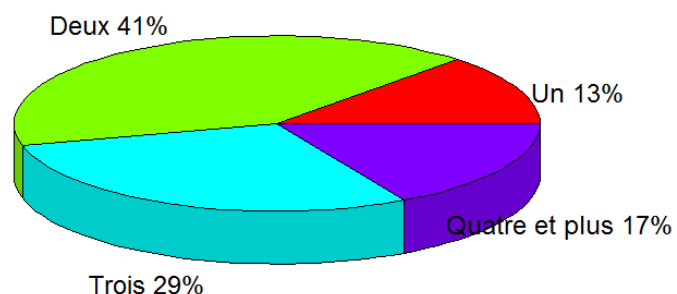
### la répartition des données selon le genre



Notre population est constituée de 57% femme et 43% d'homme, cette répartition est plus ou moins attendu vue que notre population source est constituée de femmes plus que d'hommes.

- Question 2 : Combien de téléphones portables avez-vous possédé pendant les 3 dernières années

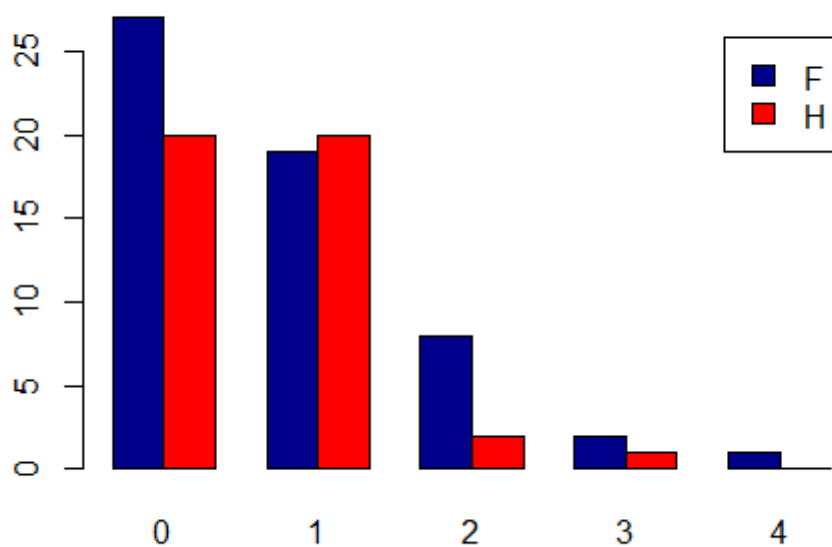
Combien de téléphones portables avez-vous possédez durant les 3 dernières années?



Parmi les 100 étudiants qu'on a interrogés, on compte 13 étudiants qui ont possédé un seul téléphone durant les 3 dernières années, 41 qui ont possédé 2, 29 qui ont possédé 3 et 17 qui ont possédé plus que quatre téléphones.

- Question 3 : Parmi ces téléphones, combien en avez-vous reçu comme cadeau ?

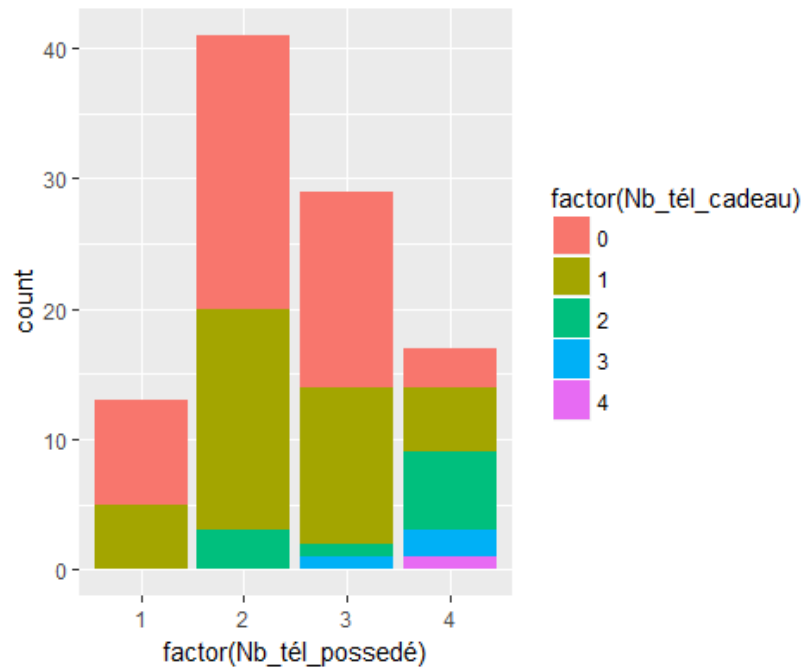
### Nombre de téléphones reçu comme cadeaux pour H/F



Nombre de téléphones cadeaux

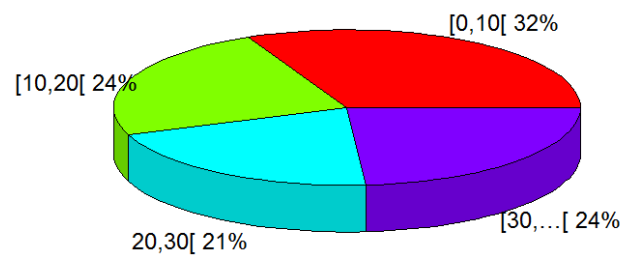
D'après le graphe précédent, le nombre de téléphones reçu est généralement plus important pour les femmes que pour les hommes, ce qui est visible dans notre enquête.

On remarque ci-dessous, que les nombre de téléphones possédés chez par jeunes dépend d'une façon directe du nombre de téléphones qu'il a reçu comme cadeau.



- Question 4 : Quelles ont été vos dépenses en téléphonie le mois dernier ? (téléphone, internet et accessoires) ?

Combien avez-vous dépensé en téléphonie le mois dernier ?

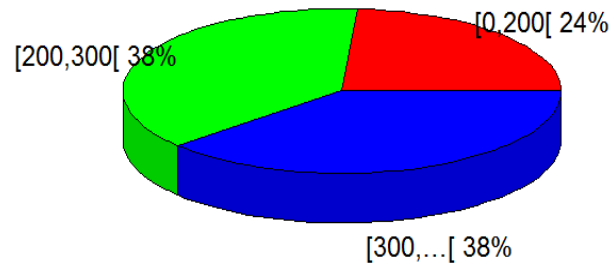


Parmi les 100 étudiants interrogés, on compte 32 étudiants qui ont dépensé moins que 10 DT en téléphonie le mois précédent, 24 qui ont dépensé entre 10 DT et 20 DT, 21 qui ont dépensé entre 20 DT et 30 DT et 24 qui ont dépensé plus que 30 DT.



- Question 5 : Quel est votre budget mensuel ? (hormis le loyer)

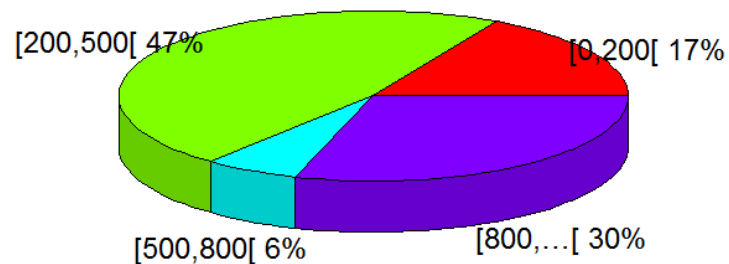
### Répartition de la population cible selon le budget mensuel ?



Parmi les 100 étudiants interrogés, on compte 24 étudiants qui ont un budget mensuel moins que 200 DT, 38 étudiants entre 200 DT et 300 DT et 38 étudiants dépensent mensuellement plus que 300 DT.

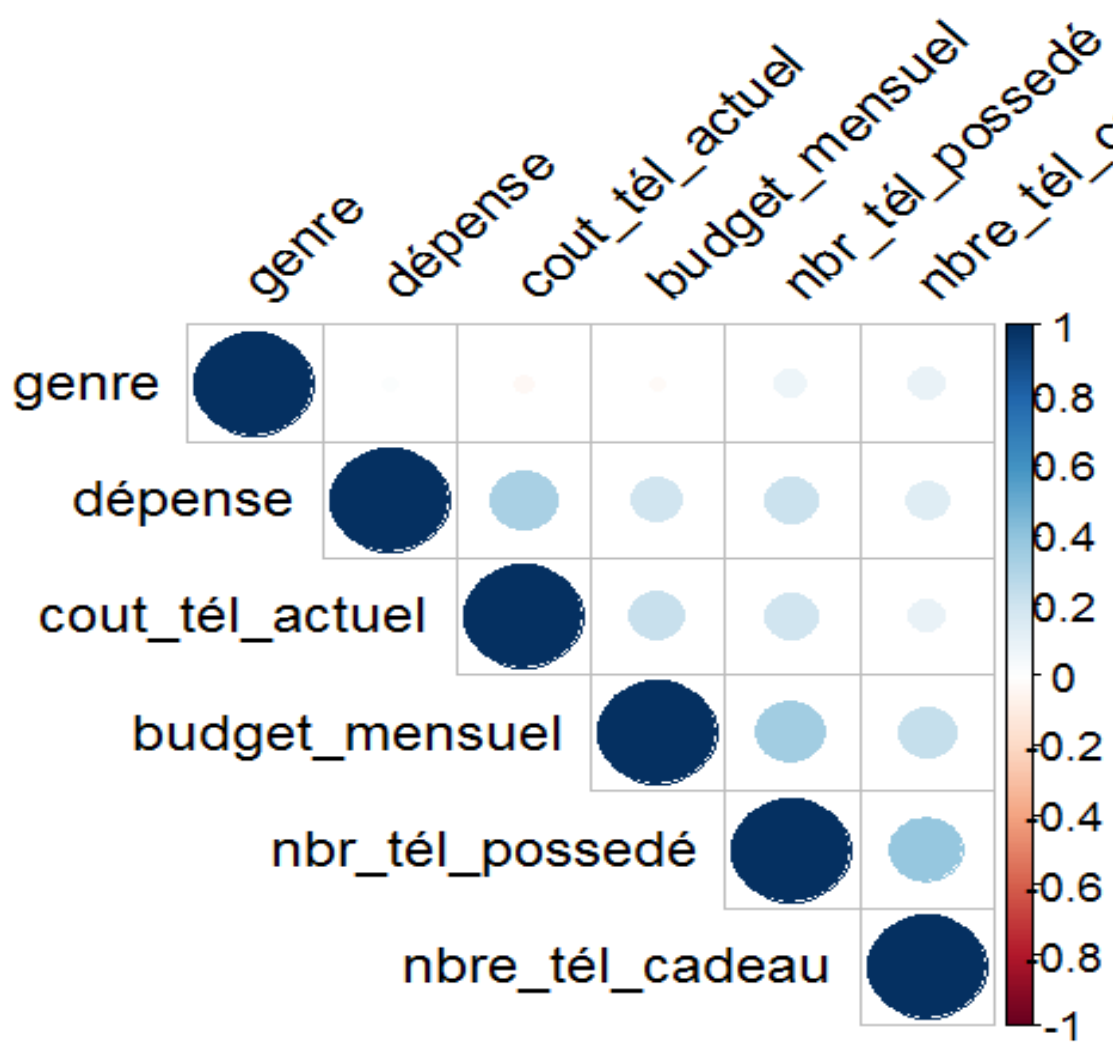
- Question 6 : Combien vous a coûté votre téléphone portable actuel ?

### Combien vous a coûté votre téléphone portable actuel ?



Parmi les 100 étudiants interrogés, on compte 17 étudiants dont le cout de téléphone actuel est moins que 200 DT, 47 entre 200 et 500 DT ,6 entre 500 et 800 DT et 30 étudiants plus que 800 DT.

- Corrélation entre les variables



## IV. Réalisation des sondages

### 1. Sondage aléatoire simple

Un sondage est dit aléatoire simple si tous les échantillons de même taille ont la même probabilité d'être sélectionné autrement dit tous les individus qui constituent l'échantillon ont la même probabilité d'inclusion. Ce type de sondage ne repose pas sur des informations auxiliaires.

Tous les sondages sont basés sur un sondage aléatoire simple.

On a  $N=100$ , on choisira  $n=20$ . Pour chaque cas on a essayé de produire un algorithme qui calcule l'estimateur  $\bar{y}_{\text{hat\_pesr}}$ , sa variance, sa variance estimée ainsi que les bornes des intervalles de confiance et leurs longueurs. Puis on a représenté la longueur de l'intervalle de confiance pour chaque tirage dans les différents cas.

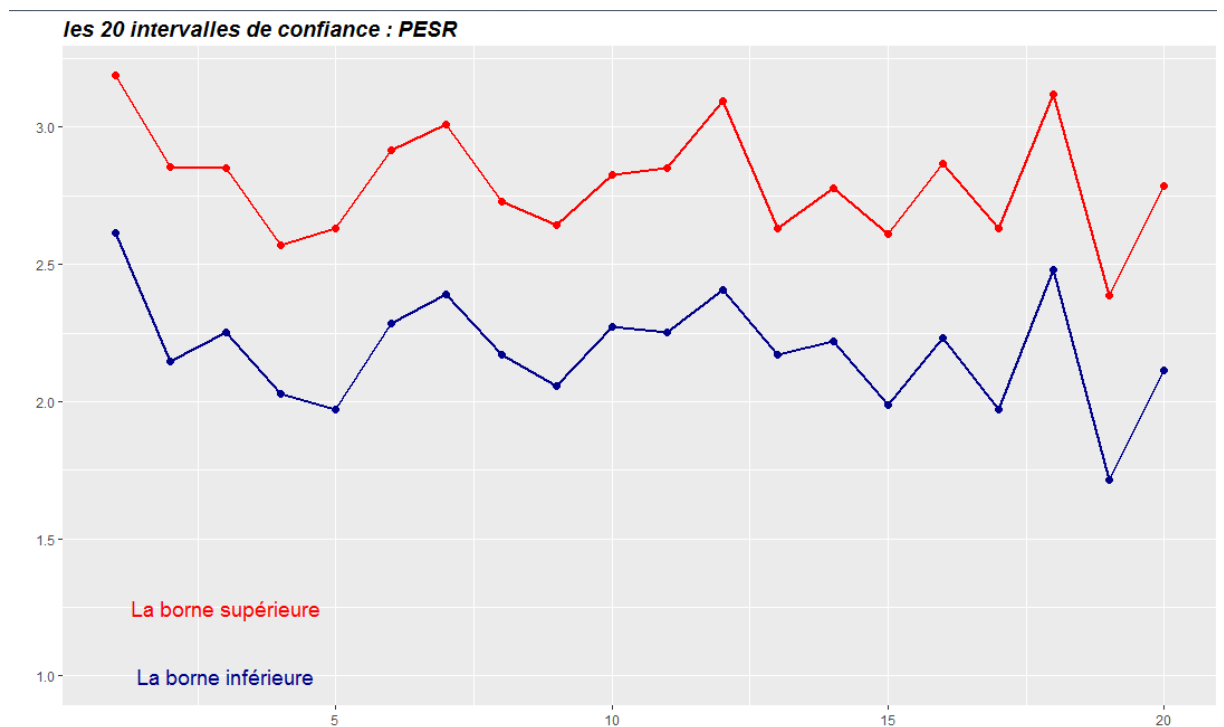
## i. Plan simple sans remise

- Résultats obtenus

	ybar_chapeau_pesr	var_ybar_chapeau_pesr	var_estim_ybar_chapeau_pesr	ICbinf_pesr	ICsup_pesr	long_IC_pesr
8	2.05	0.034	0.0371	2.310971	2.983029	0.6660578
9	2.30	0.034	0.0364	1.970128	2.629872	0.6597444
10	2.70	0.034	0.0604	2.275074	3.124926	0.8498523
11	2.60	0.034	0.0416	2.247352	2.952648	0.7052964
12	2.60	0.034	0.0336	2.283069	2.916931	0.6338619
13	2.45	0.034	0.0379	2.113400	2.786600	0.6732008
14	2.20	0.034	0.0224	1.941227	2.458773	0.5175460
15	2.75	0.034	0.0275	2.463278	3.036722	0.5734444
16	2.30	0.034	0.0364	1.970128	2.629872	0.6597444
17	2.55	0.034	0.0379	2.213400	2.886600	0.6732008
18	2.70	0.034	0.0444	2.335677	3.064323	0.7286458
19	2.90	0.034	0.0236	2.634386	3.165614	0.5312280
20	2.45	0.034	0.0419	2.096083	2.803917	0.7078349

La moyenne de  $\bar{Y}$  chapeau pistr 2.53

- Représentation des intervalles de confiances



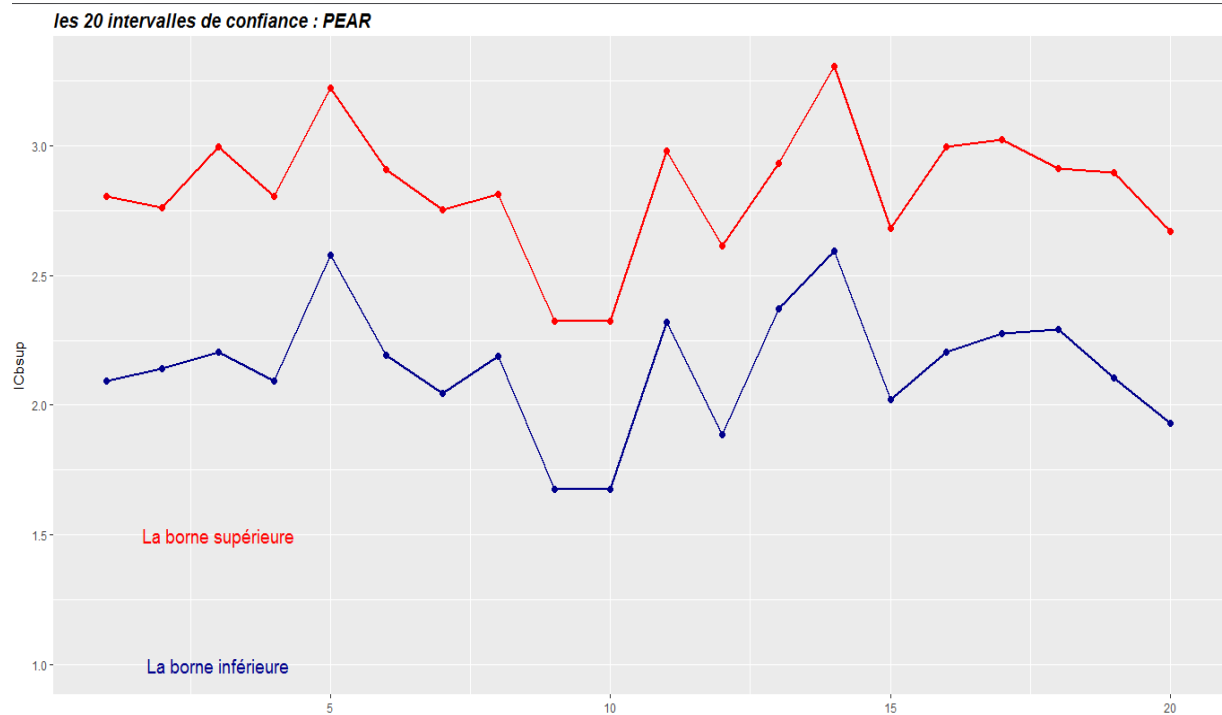
## ii. Plan simple avec remise

- Résultats obtenus

	ybar_chapeau_pear	var_ybar_chapeau_pear	var_estim_ybar_chapeau_pear	ICbinf_PEAR	ICbsup_PEAR	long_IC_PEAR
1	2.55	0.04292929	0.042375	2.194082	2.905918	0.7118358
2	2.75	0.04292929	0.039375	2.406912	3.093088	0.6861756
3	2.30	0.04292929	0.020500	2.052445	2.547555	0.4951103
4	2.40	0.04292929	0.047000	2.025162	2.774838	0.7496765
5	2.30	0.04292929	0.045500	1.931192	2.668808	0.7376166
6	2.50	0.04292929	0.047500	2.123173	2.876827	0.7536536
7	2.35	0.04292929	0.026375	2.069204	2.630796	0.5615924
8	2.10	0.04292929	0.034500	1.778853	2.421147	0.6422950
9	2.40	0.04292929	0.047000	2.025162	2.774838	0.7496765
10	2.50	0.04292929	0.042500	2.143558	2.856442	0.7128850
11	2.35	0.04292929	0.046375	1.977662	2.722338	0.7446753
12	2.35	0.04292929	0.036375	2.020241	2.679759	0.6595178
13	2.65	0.04292929	0.031375	2.343742	2.956258	0.6125152
14	2.80	0.04292929	0.038000	2.462956	3.137044	0.6740883
15	2.55	0.04292929	0.042375	2.194082	2.905918	0.7118358
16	2.45	0.04292929	0.067375	2.001209	2.898791	0.8975825
17	1.95	0.04292929	0.037375	1.615739	2.284261	0.6685218
18	2.10	0.04292929	0.029500	1.803034	2.396966	0.5939310
19	2.55	0.04292929	0.057375	2.135851	2.964149	0.8282975
20	2.45	0.04292929	0.032375	2.138900	2.761100	0.6221998

La moyenne des  $\bar{Y}_{\text{chapeau pear}} = 2.4175$

- Représentation des intervalles de confiances



## 2. Sondage à probabilité inégales

Dans le sondage à probabilités inégales, la probabilité d'inclusion  $\pi_i$  est proportionnelle à une variable auxiliaire  $X$ . Dans notre travail, nous avons choisi le budget mensuel de l'étudiant comme information auxiliaire qui va intervenir au cours du tirage des individus.

### i. Plan de sondage avec remise

Dans ce type de sondage, nous avons utilisé l'estimateur de Hansen & Hurwitz, le tirage s'est effectué comme suit :

- ✓ On génère une variable aléatoire  $u$  suivant la loi uniforme sur  $[0,1]$
- ✓ On sélectionne l'individu  $i$  vérifiant  $v_{ki}-1 \leq u < v_{ki}$

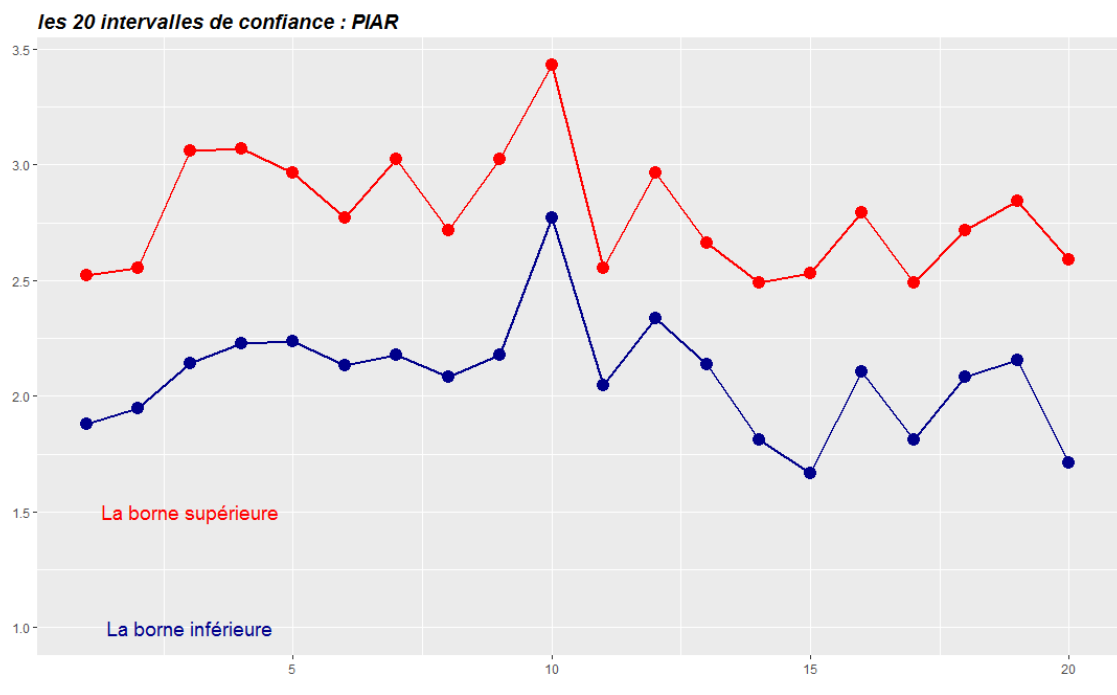
L'opération répétée 20 fois de manière indépendante fournit l'échantillon de taille 20.

#### • Représentation des échantillons obtenus

	ybar_chapeau_piar	var_ybar_chapeau_piar	var_estim_ybar_chapeau_piar	ICbinf_Piar	ICsup_Piar	long_IC_Piar
1	2.078150	0.1061071	0.03074768	1.774970	2.381330	0.6063609
2	2.778650	0.1061071	0.12472682	2.168025	3.389275	1.2212510
3	2.352513	0.1061071	0.06078334	1.926240	2.778785	0.8525449
4	2.866213	0.1061071	0.10563507	2.304261	3.428164	1.1239035
5	2.124850	0.1061071	0.01718886	1.898167	2.351533	0.4533655
6	2.761138	0.1061071	0.22011744	1.949948	3.572327	1.6223786
7	2.568500	0.1061071	0.21234982	1.771752	3.365248	1.5934958
8	2.661900	0.1061071	0.12644858	2.047074	3.276726	1.2296513
9	2.475100	0.1061071	0.04373258	2.113526	2.836674	0.7231486
10	2.819513	0.1061071	0.15074505	2.148213	3.490812	1.3425996
11	2.078150	0.1061071	0.03655860	1.747560	2.408740	0.6611801
12	2.358350	0.1061071	0.09373523	1.828996	2.887704	1.0587085
13	2.054800	0.1061071	0.03353118	1.738194	2.371406	0.6332124
14	2.264950	0.1061071	0.02972897	1.966834	2.563066	0.5962315
15	2.410888	0.1061071	0.08739881	1.899738	2.922037	1.0222985
16	2.434238	0.1061071	0.09815978	1.892534	2.975941	1.0834073
17	2.124850	0.1061071	0.03605642	1.796538	2.453162	0.6566233
18	2.550988	0.1061071	0.08890534	2.035452	3.066523	1.0310718
19	2.556825	0.1061071	0.08249360	2.060227	3.053423	0.9931963
20	1.920538	0.1061071	0.05254761	1.524194	2.316881	0.7926865

*Moyenne des Y\_bar chapeau Piar = 2.412055*

- Représentation des intervalles de confiance



## ii. Plan simple sans remise

Pour le plan de sondage à probabilités inégales sans remise, nous avons utilisé la méthode de Rao Hartley Cochran. Pour cela, nous avons divisé notre population en 20 groupes disjoints de taille 5.

Dans chaque groupe  $g$ , on tire un seul individu proportionnellement à la variable  $X$  (budget mensuel de l'étudiant) de la manière suivante

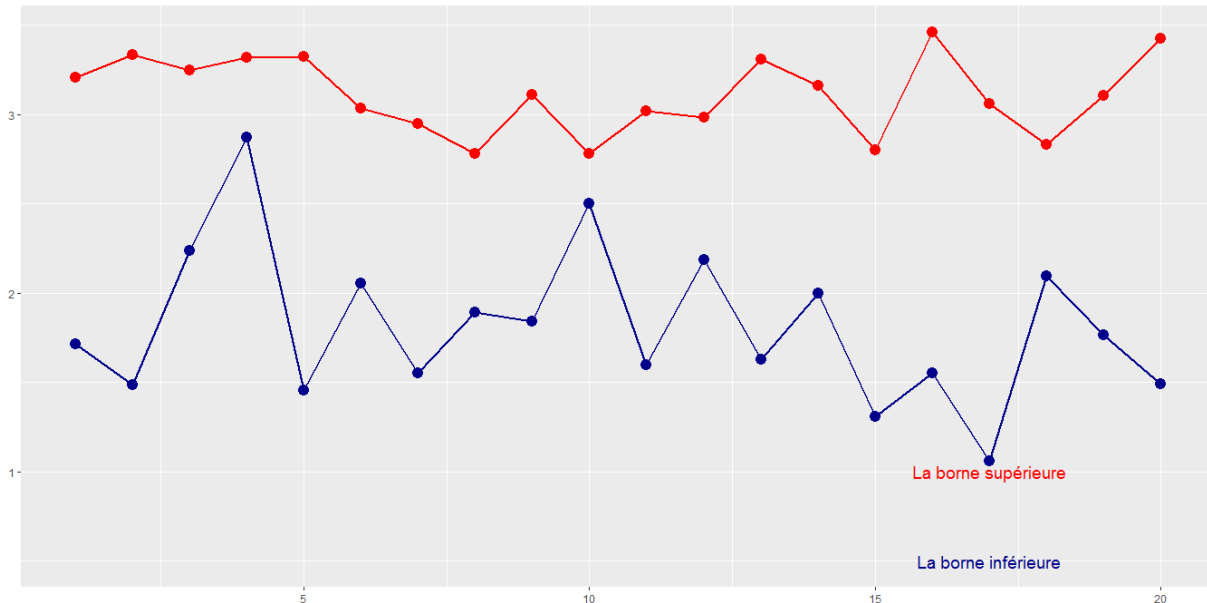
- Représentation des moyennes des échantillons obtenus

	y_bar_chapeau_pisr	var_est_ybar_chapeau	lcbinf_pisr	ICsup_pisr	long_IC_pisr
1	2.45975	0.18570723	1.714659	3.204841	1.4901823
2	2.41000	0.28622464	1.484986	3.335014	1.8500289
3	2.74225	0.08513003	2.237779	3.246721	1.0089424
4	3.09750	0.01677353	2.873573	3.321427	0.4478548
5	2.39025	0.29107178	1.457436	3.323064	1.8656279
6	2.54400	0.08083578	2.052417	3.035583	0.9831659
7	2.25000	0.16392494	1.549969	2.950031	1.4000628
8	2.33675	0.06555192	1.894072	2.779428	0.8853555
9	2.47850	0.13474730	1.843820	3.113180	1.2693606
10	2.64250	0.00643770	2.503773	2.781227	0.2774536
11	2.31025	0.16898888	1.599488	3.021012	1.4215235
12	2.58350	0.05297413	2.185551	2.981449	0.7958971
13	2.46950	0.23555259	1.630351	3.308649	1.6782975
14	2.58000	0.11399651	1.996231	3.163769	1.1675373
15	2.05475	0.18661458	1.307841	2.801659	1.4938183
16	2.50800	0.30477231	1.553485	3.462515	1.9090299
17	2.05775	0.33449050	1.057780	3.057720	1.9999396
18	2.46300	0.04547064	2.094311	2.831689	0.7373786
19	2.43650	0.15043426	1.765893	3.107107	1.3412149
20	2.46000	0.31291294	1.492821	3.427179	1.9343575

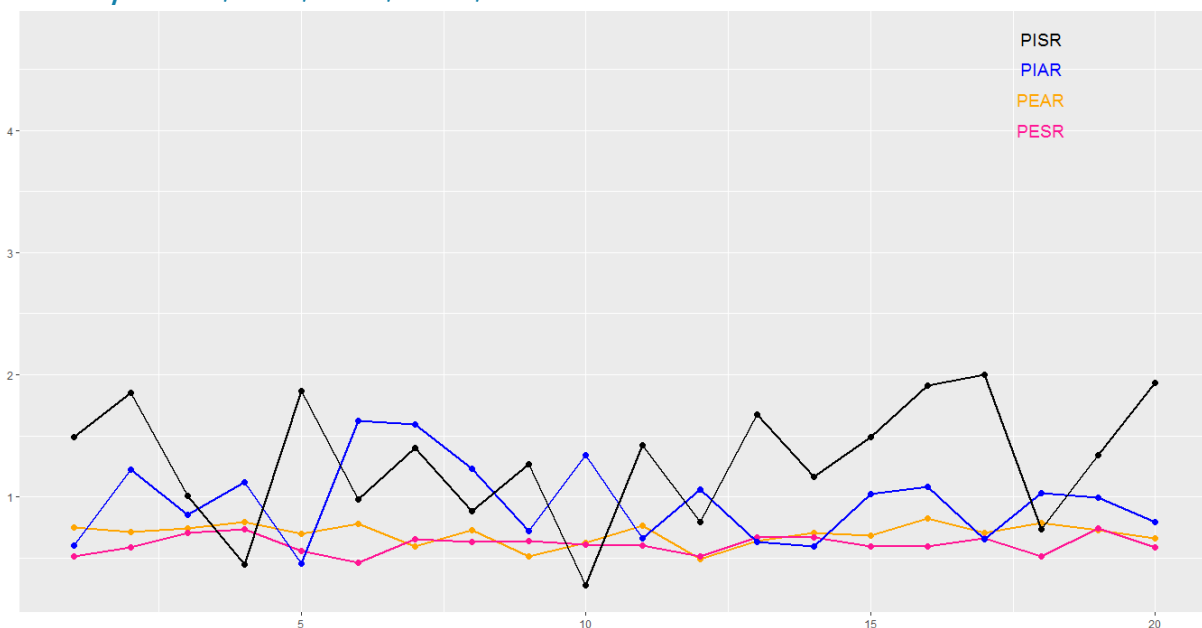
*Moyenne des Y\_bar chapeau pisr= 2.463738*

- Représentation des intervalles de confiance

les 20 intervalles de confiance : PISR



### iii. Synthèse, PISR, PIAR, PEAR, PESR



Les résultats obtenus d'après les sondages PESR, PEAR, PIAR et PISR nous mènent à conclure que le sondage le plus précis est le plan simple sans remise et on trouve aussi que le plan simple avec remise est assez précis par rapport aux deux types de sondage à probabilités inégales.

Ce résultat est expliqué par le fait que la variable auxiliaire  $X$ , dont les probabilités d'inclusion dépendent lors du sondage à probabilités inégales, n'est pas fortement corrélée (0.34) avec la variable d'intérêt.

## 3. Sondage par stratification

La stratification est une méthode de sondage permettant d'introduire de l'information auxiliaire dans le plan de sondage.



On procède alors à un découpage de la population en  $H$  sous populations. Ces sous-populations sont nommées les strates.

Après le découpage, on procède généralement à un sondage aléatoire simple sans remise à l'intérieur de chaque strate. On tire alors dans chaque strate un nombre  $n_h$  prédéterminé d'individus. C'est pour cela que le sondage stratifié est considéré comme un plan à probabilités inégales.

- **Découpage de la population**

Nous avons choisi de découper notre population selon le genre pour avoir le plus de résultat claire.

On peut donc visualiser les valeurs de la variable d'intérêt des individus de chaque strate (strate1 représente les femmes, strate2 représente les hommes) :

```
## [[1]]
##  [1] 2 2 3 3 3 3 3 2 2 2 4 3 2 1 1 3 3 3 1 1 2 1 3 3 4 2 2 4 3 2 3 2 2 4
3
## [36] 2 1 4 1 2 4 4 3 2 3 3 2 4 3 3 4 2 4 2 3 1 2
##
## [[2]]
##  [1] 2 2 1 4 2 2 2 2 3 2 2 4 3 1 2 2 2 2 2 2 1 3 2 2 1 2 2 3 3 3 3 4 1 3
2
## [36] 4 2 2 4 4 2 4 3
```

- **Propriété de chaque strate :**

Ci-dessous, nous avons regroupé dans un même tableau, différentes propriétés de chacune des strates à savoir la moyenne de la variable d'intérêt dans la strate, sa variance et sa variance corrigée :

	strate1	strate2
<b>moyenne</b>	2.5614000	2.418605
<b>variance</b>	0.8845489	0.812392
<b>variance-corrigée</b>	0.8934837	0.820598

Nous avons aussi calculé la variance inter strates et la variance intra strates de la variable d'intérêt.

	var_inter_y	var_intra_y	var_y
<b>1</b>	0.005397634	0.8434195	0.85

On peut donc vérifier que leur somme coïncide avec la variance totale de  $y$ .

### i. States de même taille

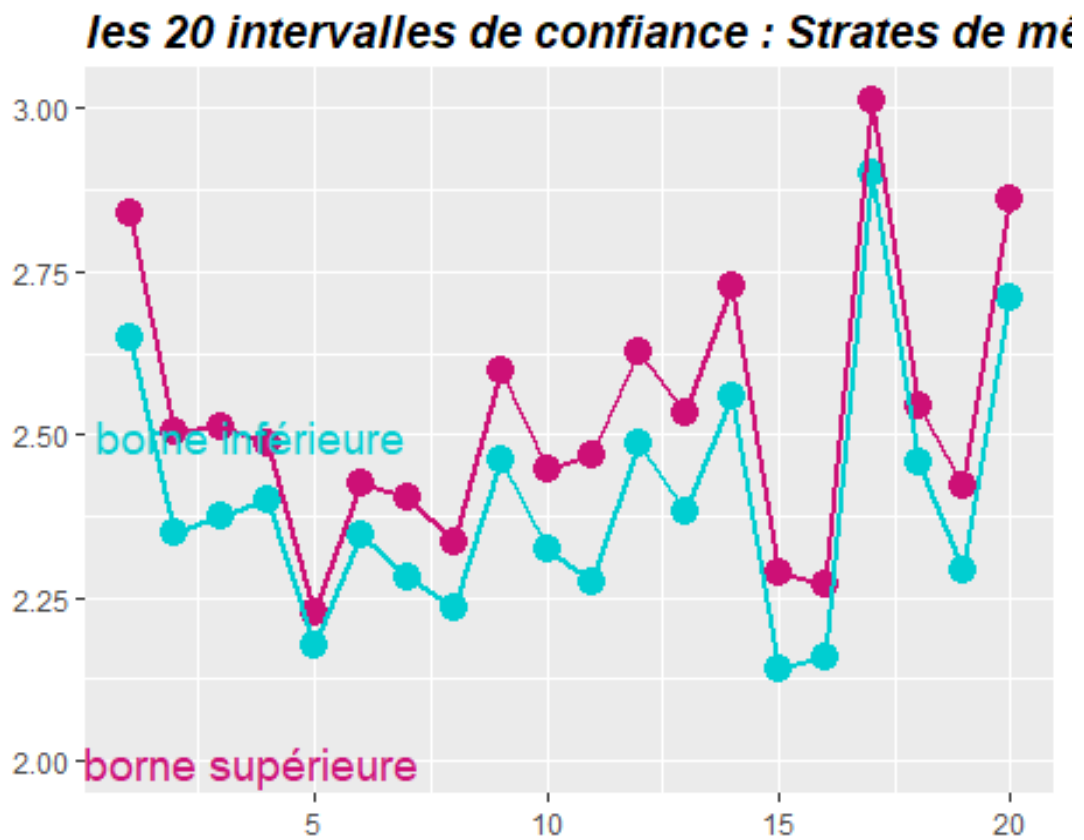
Nous avons calculé la moyenne estimée ( $\bar{y}_{chap}$ ), la variance ainsi que la variance estimée de  $\bar{y}_{chap}$  pour chacun des 20 échantillons tirés. Nous avons par suite utilisé cette variance estimée pour le calcul de l'intervalle de confiance.

Voici le tableau regroupant tous ces résultats :

	$\bar{y}_{chap}$	$var_{\bar{y}_{chap}}_{strat}$	$var_{estim_{\bar{y}_{chap}}}_{strat}$	$ic_{sup}_{strat}$	$long_{ic}$
1	2.743	0.03466235	0.04835367	2.837773	0.18954637
2	2.428	0.03466235	0.03978667	2.505982	0.15596373
3	2.444	0.03466235	0.03457133	2.511760	0.13551963
4	2.444	0.03466235	0.02266467	2.488423	0.08884549
5	2.201	0.03466235	0.01276967	2.226029	0.05005709
6	2.387	0.03466235	0.01998567	2.426172	0.07834381
7	2.343	0.03466235	0.03140167	2.404547	0.12309453
8	2.285	0.03466235	0.02600833	2.335976	0.10195267
9	2.529	0.03466235	0.03462167	2.596858	0.13571693
10	2.386	0.03466235	0.03155933	2.447856	0.12371259
11	2.371	0.03466235	0.04919100	2.467414	0.19282872
12	2.557	0.03466235	0.03560567	2.626787	0.13957421
13	2.457	0.03466235	0.03819233	2.531857	0.14971395
14	2.643	0.03466235	0.04372833	2.728708	0.17141507
15	2.215	0.03466235	0.03721500	2.287941	0.14588280
16	2.215	0.03466235	0.02810833	2.270092	0.11018467
17	2.957	0.03466235	0.02873900	3.013328	0.11265688
18	2.501	0.03466235	0.02124167	2.542634	0.08326733
19	2.358	0.03466235	0.03331000	2.423288	0.13057520
20	2.786	0.03466235	0.03814333	2.860761	0.14952187

La Moyenne de Y Bar estimée  
 $\approx 2.4625$

- Représentation des intervalles de confiances



## ii. Strates à allocation proportionnelles

Dans ce type de sondage, le nombre d'individus à tirer de chaque strate est proportionnel au nombre d'individus de cette strate.

On a, dans ce cas, le taux de sondage dans chaque strate est égale au taux de sondage global (f) :

$$\frac{nh}{Nh} = \frac{n}{N}$$

- Représentation des résultats

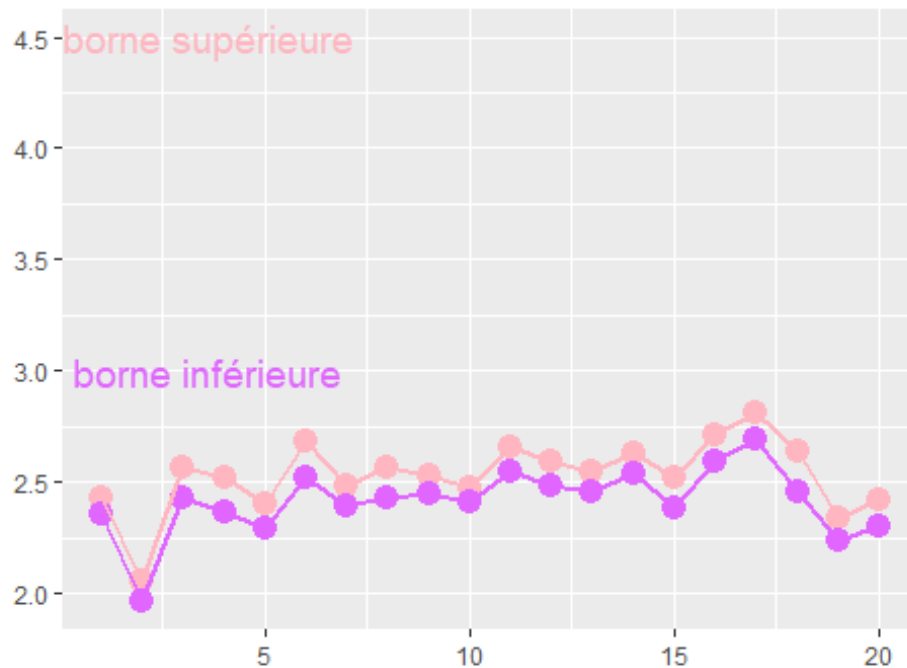
	ybar_chap	var_ybar_chap_prop	var_estim_ybar_chap_strat	ic_inf_prop	ic_sup_prop	long_ic_prop
1	2.497980	0.03407422	0.02937610	2.447189	2.548771	0.10158256
2	2.350606	0.03407422	0.04138097	2.279058	2.422154	0.14309541
3	2.442121	0.03407422	0.03645309	2.379094	2.505149	0.12605477
4	2.601616	0.03407422	0.03300382	2.544553	2.658680	0.11412722
5	2.645354	0.03407422	0.02802240	2.596903	2.693804	0.09690145
6	2.434040	0.03407422	0.03999158	2.364895	2.503186	0.13829087
7	2.290707	0.03407422	0.02574838	2.246188	2.335226	0.08903789
8	2.609697	0.03407422	0.04945474	2.524190	2.695204	0.17101449
9	2.342525	0.03407422	0.02606961	2.297451	2.387600	0.09014871
10	2.697172	0.03407422	0.03165012	2.642449	2.751895	0.10944612
11	2.438081	0.03407422	0.03760540	2.373061	2.503101	0.13003947
12	2.398384	0.03407422	0.03706494	2.334299	2.462469	0.12817055
13	2.761111	0.03407422	0.03170366	2.706295	2.815927	0.10963125
14	2.549798	0.03407422	0.02390267	2.508470	2.591126	0.08265542
15	2.286667	0.03407422	0.03866848	2.219809	2.353524	0.13371562
16	2.302828	0.03407422	0.02211557	2.264590	2.341066	0.07647566
17	2.645354	0.03407422	0.05927482	2.542867	2.747840	0.20497233
18	2.553838	0.03407422	0.04183221	2.481510	2.626166	0.14465578
19	2.689091	0.03407422	0.03775325	2.623816	2.754366	0.13055074
20	2.402424	0.03407422	0.03634093	2.339591	2.465258	0.12566693

La moyenne de  $\bar{Y}$  estimée

$$\equiv 2.473081$$

- Représentation des intervalles de confiances

### les 20 intervalles de confiance : Strates à alloc



### iii. Strates à allocation optimale

Dans ce type de sondage le nombre d'individus tiré de chaque strate ( $n_h$ ) est, non seulement proportionnel au nombre d'individus de cette strate ( $N_h$ ), mais aussi à la variance dans cette strate.

Nous avons partitionné la population selon la variable auxiliaire coût de téléphone actuel qui présente quatre modalités :

[0,200 [ : Grappe 1

[200,500[ : Grappe 2

[500,800[ : Grappe 3

[800,...[ : Grappe 4

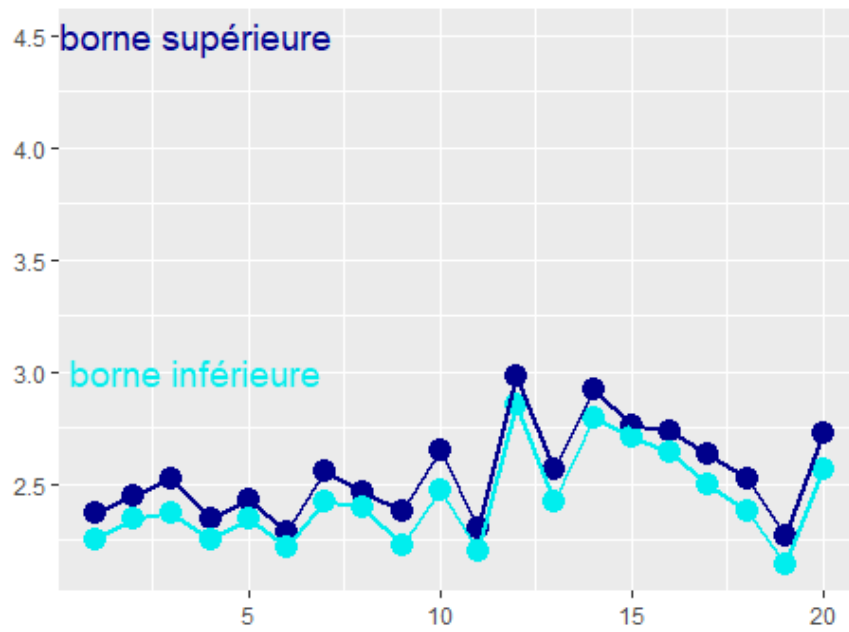
- Représentation des résultats

	ybar_chap	var_ybar_chap_opt	var_estim_ybar_chap_opt	ic_inf_strat	ic_sup_strat	long_ic
1	2.199192	0.03407422	0.02400974	2.157679	2.240705	0.08302569
2	2.489899	0.03407422	0.03007718	2.437896	2.541902	0.10400690
3	2.103636	0.03407422	0.02491221	2.060563	2.146710	0.08614643
4	2.434040	0.03407422	0.02498360	2.390844	2.477237	0.08639328
5	2.493939	0.03407422	0.03590754	2.431855	2.556024	0.12416827
6	2.489899	0.03407422	0.02195496	2.451939	2.527859	0.07592025
7	2.410505	0.03407422	0.04318591	2.335837	2.485173	0.14933688
8	2.549798	0.03407422	0.04367792	2.474279	2.625317	0.15103825
9	2.852626	0.03407422	0.03657801	2.789383	2.915870	0.12648675
10	2.454242	0.03407422	0.04181436	2.381945	2.526539	0.14459406
11	2.693131	0.03407422	0.03885203	2.625956	2.760306	0.13435031
12	2.705253	0.03407422	0.04443254	2.628429	2.782076	0.15364772
13	2.199192	0.03407422	0.02807086	2.150657	2.247726	0.09706902
14	2.653434	0.03407422	0.03143089	2.599090	2.707778	0.10868800
15	2.502020	0.03407422	0.03749832	2.437186	2.566855	0.12966920
16	2.394343	0.03407422	0.01854139	2.362285	2.426401	0.06411613
17	2.314949	0.03407422	0.03100258	2.261346	2.368553	0.10720692
18	2.613737	0.03407422	0.05087983	2.525766	2.701709	0.17594244
19	2.553838	0.03407422	0.03706494	2.489753	2.617924	0.12817055
20	2.147374	0.03407422	0.03587185	2.085351	2.209396	0.12404485

La moyenne des  $y_{\text{bar}} \text{ chapeau} = 2.48203$

- Représentation des intervalles de confiances

### les 20 intervalles de confiance : Strates à alloc



#### iv. Synthèse

	Strates égales	Strates Proportionnelles	Strates Optimales
Y_bar	2.4625	2.473081	2.48203
Longueur de strate	0.1284187	0.1146346	0.1162505

On remarque que le sondage stratifié à même taille est moins précis que les deux autres types de sondage stratifié effectués. En effet, les longueurs de ses intervalles de confiance sont nettement plus importantes dans la plupart des cas.

Par contre, le sondage stratifié à allocations optimales est le plus précis car dans celui-ci on prend en considération, non seulement les nombre d'individus de chaque strate, mais aussi de la variance au sein de chaque strate.

Le sondage stratifié à allocations proportionnelles, légèrement moins précis que celui à allocations optimales car ne prend en considération que les tailles des strates.

#### 4. Sondage par grappes et à deux degrés

Le sondage par grappes consiste à tirer non pas directement un individu, mais des grappes. L'ensemble des individus de la grappe sont tous inclus dans l'étude.

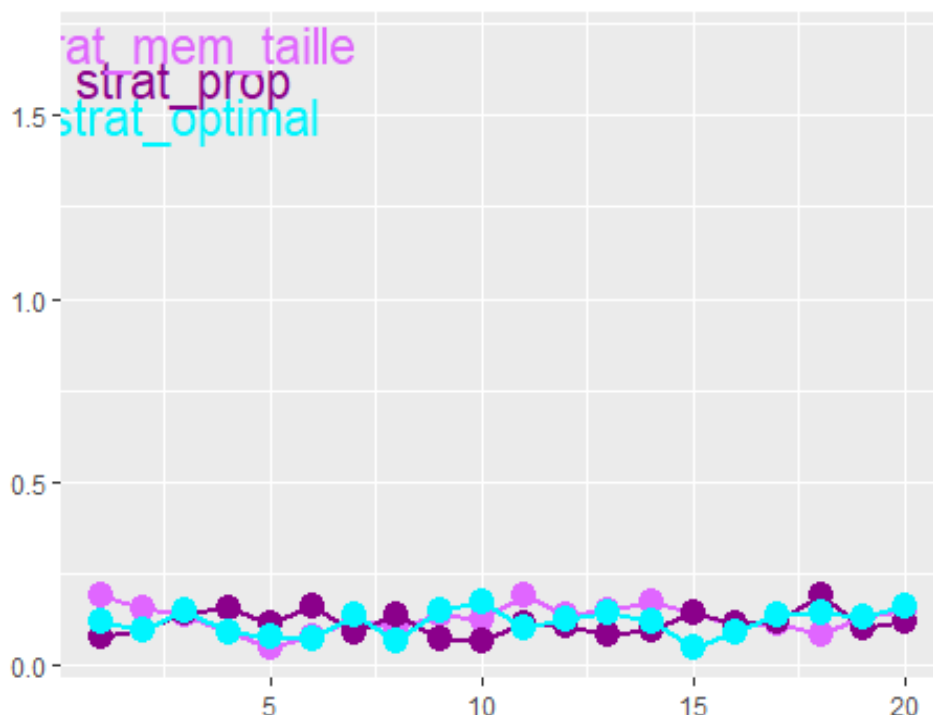
##### i. Plan de sondage par grappes

La probabilité d'inclusion = 0.5

	grappe1	grappe2	grappe3	grappe4
d	1.882353	2.617021	2.5	2.666667
Nh	17.000000	47.000000	6.0	30.000000

- Représentation des resultat dans un tableau

#### Synthèse des logueurs des intervalles de confi

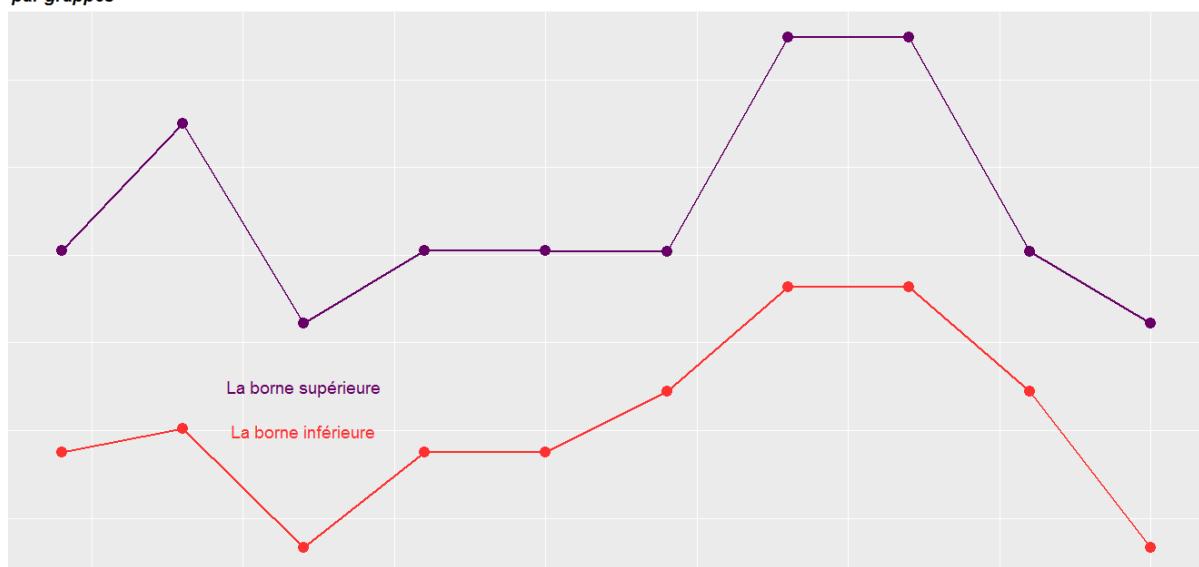


	$\hat{y}_{\text{bar\_grappe}}$	$\hat{\text{var\_ygrap}}$	$\hat{\text{icinf\_grappe}}$	$\hat{\text{icsup\_grappe}}$	$\hat{\text{longIC\_grap}}$	$\hat{\text{num\_1ere\_grap\_selct}}$	$\hat{\text{num\_2ere\_grap\_selct}}$
1	1.90	0.3416667	0.7543357	3.045664	2.291329	3	4
2	2.76	0.7888667	1.0191639	4.500836	3.481672	2	3
3	0.94	0.4248667	-0.3375632	2.217563	2.555126	3	1
4	1.90	0.3416667	0.7543357	3.045664	2.291329	4	3
5	1.90	0.3416667	0.7543357	3.045664	2.291329	3	4
6	2.24	0.1648667	1.4441660	3.035834	1.591668	4	1
7	4.06	0.5288667	2.6346249	5.485375	2.850750	4	2
8	4.06	0.5288667	2.6346249	5.485375	2.850750	2	4
9	2.24	0.1648667	1.4441660	3.035834	1.591668	4	1
10	0.94	0.4248667	-0.3375632	2.217563	2.555126	1	3

La moyenne de  $\hat{Y}_{\text{bar\_grappe}} = 2.294$

- Représentation graphique des intervalles de confiances

les 10 intervalles de confiance  
par grappes

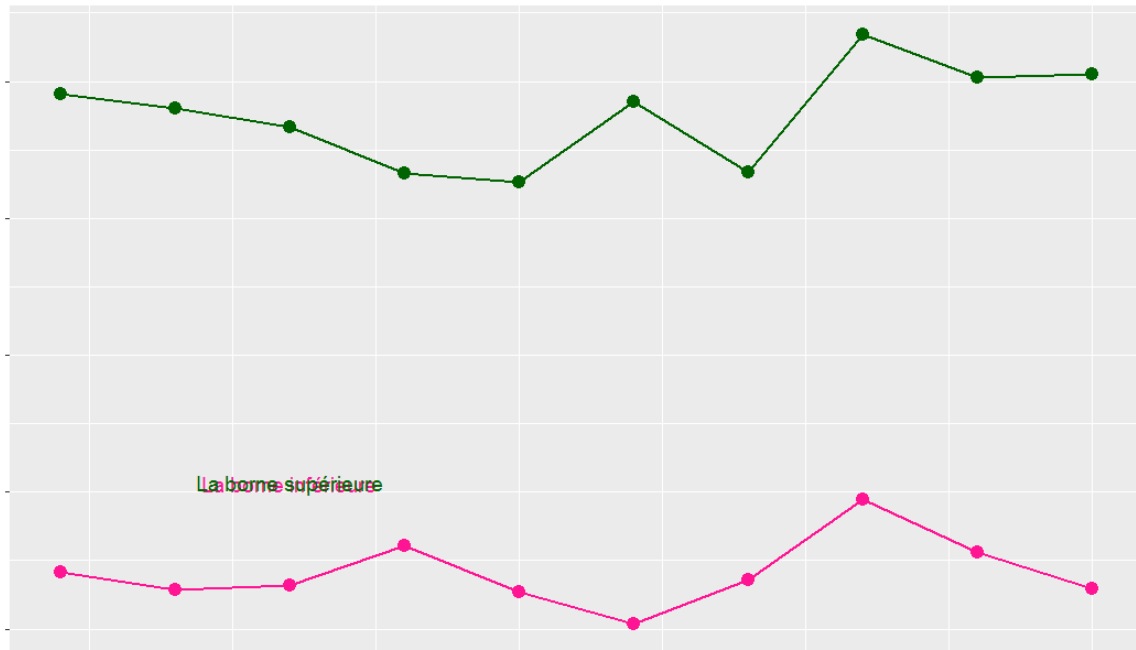


## ii. Plan de sondage par grappe e à deux degrés

Le sondage à deux degrés consiste à tirer d'une manière aléatoire et successive un nombre  $n_h$  d'individus de chaque grappe.



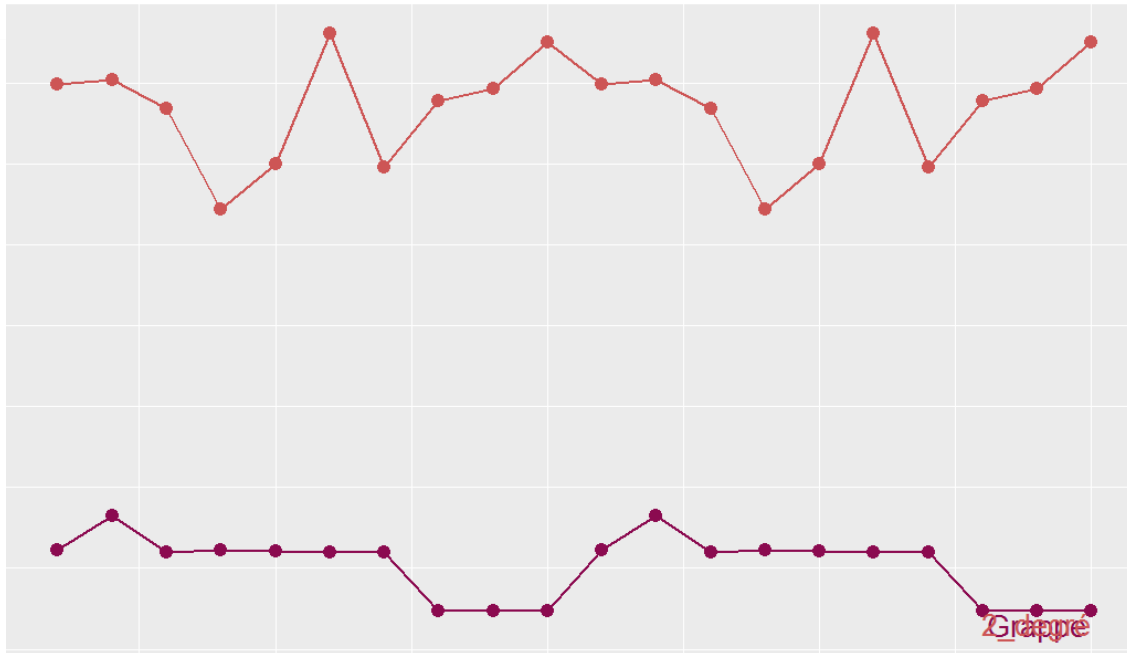
les 10 intervalles de confiance : à 2 degré



La moyenne de  $\bar{Y}_{\text{grappe}}$  = 2.396

### iii. Synthèse

Synthèse des logueurs des intervalles de confiance : grappe et 2 degré



## V. Redressement

Le redressement consiste à améliorer un estimateur en prenant compte l'information auxiliaire.

Une information auxiliaire permet de détecter un déséquilibre et ainsi de le corriger.

Vue la disproportion homme-femme dans notre population, la variable auxiliaire choisie pour le redressement est : «Sexe »

L'estimateur d'HT ne contient pas d'information auxiliaire à part les probabilités d'inclusion, alors nous allons construire des estimateurs qui en tiennent compte.

Supposons que nous avons pris un échantillon de l'échantillon «PESR» pour cela nous avons généré un échantillon selon le plan de sondage PESR.

Nous avons procédé à un redressement post stratifié qui consiste à choisir l'échantillon puis le découpé en strate selon la variable auxiliaire.

L'échantillon tiré

[1] 3 1 2 1 3 2 3 2 4 3 4 2 1 3 3 2 3 3 4 4

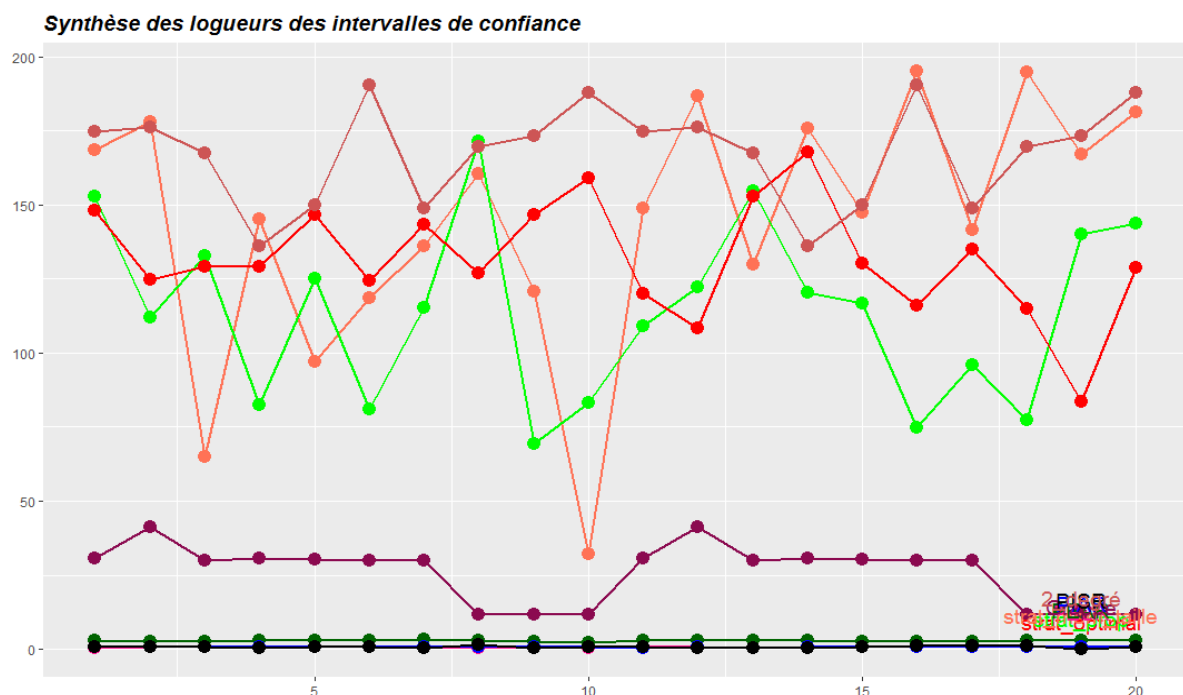
- La moyenne de l'échantillon avant redressement est 2.5

La proportion des femmes dans la population est égale à 0.57 alors que dans l'échantillon est 0.4

Donc le poids des femmes dans notre échantillon est  $\frac{0.57}{0.4} = 1.425$

L'estimateur de la moyenne des femmes après **redressement** est la moyenne de l'échantillon multiplié par le poids des femmes est égale à **2.1375**

## VI. Synthèse



D'après le graphe qui résume les moyennes des longueurs des intervalles de confiance de tous les types de sondage utilisés précédemment, nous remarquons que le sondage à deux degré est le moins précis, nous remarquons par ailleurs que les sondages stratifiés sont plus précis que celui à deux degrés, mais le moins précis dans notre cas que les sondages PESR, PEAR, PISR et PIAR et c'est du faute qu'on a utilisé une variable auxiliaire très faiblement corrélé à notre variable d'intérêt.

## VII. Annexe

```
BD=read.table(file='C:/Users/user/Desktop/sondage/BD111 (1).csv',header = T,sep =
";")
```

```
View(BD)
```

```
colnames(BD)=c('genre','nbr_tél_posse  ','nbre_t  l_cadeau','d  pense','budget_mensuel',
'cout_t  l_actuel')
```

```
View(BD)
```

```
N=100
```

```
mu=mean(y)
```

```
sigma2c=var(y)
```

```
var=((N-1)/N)*sigma2c
```

```
T=sum(y)
```

```
n=20
```

```
tirage=function(y){
```

```
  echantillon=sample(y,n,replace=F)
```

```
  ybarpesr=mean(echantillon)
```

```
  varybarpesr=var*(1-(n/N))/n
```

```
  sc2=var(echantillon)
```

```
  s2=((n-1)/n)*sc2
```

```
  varestimybarpesr=s2*(1-(n/N))/n
```

```
  v=c(echantillon,ybarpesr,varybarpesr,varestimybarpesr)
```

```
  return(v)
```

```
}
```

```
res=replicate(20,tirage(y))
```

```
View(res)
```

```
ICbinf=1:20
```

```

ICbsup=1:20
longueur=1:20
for (i in 1:20) {
  ICbinf[i]=res[21,i]-1.729*sqrt(res[23,i])# 1.729 :c'est le quantile de student avec
alpha=0.05 et degré de liberté=19
  ICbsup[i]=res[21,i]+1.729*sqrt(res[23,i])
  mu[i]=res[21,i]
  longueur[i]=ICbsup[i]-ICbinf[i]
}

mean(longueur)
mu1=mean(mu)
res=data.frame(res)
res=t(res)
res_pesr=NULL
res_pesr=data.frame(y=1:20,res[,21],res[,22],res[,23],ICbinf,ICbsup,longueur)
p1_pesr=ggplot(res_pesr,aes(y,ICbsup))+geom_point(colour="red",cex=2)+geom_point(
aes(y,ICbinf),colour="darkblue",cex=2)+geom_path(colour="red",size=1)+geom_path(
aes(y,ICbinf),colour="darkblue",size=1)

p2_pesr=annotate("text", label =c("La borne inférieure","La borne supérieure"), x
=c(3,3), y =c(1,1.25),size =5, colour =c("darkblue","red"))

p3_pesr=labs(x="",y="")

p1_pesr+p2_pesr+p3_pesr+ggtitle("les 20 intervalles de confiance :
PESR")+theme(plot.title = element_text(color="black",size=14, face="bold.italic"))

res_pesr=res_pesr[-1]

colnames(res_pesr)=c("ybar_chapeau_pesr","var_ybar_chapeau_pesr","var_estim_yb
ar_chapeau_pesr","ICbinf_pesr","ICbsup_pesr","long_IC_pesr")

rownames(res_pesr)=c(1:20)

View(res_pesr)

View(res)

mean(res_pesr$ybar_chapeau_pesr)

```

```
#####pear

tirage=function(y){
  echantillon=sample(y,n,replace=T)
  ybarpear=mean(echantillon)
  varybarpear=sigma2c/n
  sc2=var(echantillon)
  s2=((n-1)/n)*sc2
  varestimybarpear=s2/n
  v=c(echantillon,ybarpear,varybarpear,varestimybarpear)
  return(v)
}
res=replicate(20,tirage(y))
res
ICbinf=1:20
ICbsup=1:20
longueur=1:20
for (i in 1:20) {
  ICbinf[i]=res[21,i]-1.729*sqrt(res[23,i])#1.729:quantile de student avec alpha=0.05%
et degré de liberté=19
  ICbsup[i]=res[21,i]+1.729*sqrt(res[23,i])
  mu[i]=res[21,i]
  longueur[i]=ICbsup[i]-ICbinf[i]
}
longueur
mean(longueur)
mu1=mean(mu)
mu1
```

```

res=data.frame(res)

res=t(res)

res_pear=NULL

res_pear=data.frame(y=1:20,res[,21],res[,22],res[,23],lCbinf,lCbsup,longueur)

p1_pear=ggplot(res_pear,aes(y,lCbsup))+geom_point(colour="red",cex=2)+geom_point(aes(y,lCbinf),colour="darkblue",cex=2)+

geom_path(colour="red",size=1)+geom_path(aes(y,lCbinf),colour="darkblue",size=1)

p2_pear=annotate("text", label =c("La borne inférieure","La borne supérieure"), x
=c(3,3), y =c(1,1.5),size =5, colour =c("darkblue","red"))

p3_pear=labs(x="",y="")

p1_pear+p2_pear+ggtitle("les 20 intervalles de confiance : PEAR")+theme(plot.title =
element_text(color="black",size=14, face="bold.italic"))

res_pear=res_pear[-1]

colnames(res_pear)=c("ybar_chapeau_pear","var_ybar_chapeau_pear","var_estim_y
bar_chapeau_pear","lCbinf_PEAR","lCbsup_PEAR","long_IC_PEAR")

View(res_pear)

rownames(res_pear)=c(1:20)

#####PIAR

library(ggplot2)

library(sampling)

y=matrix(100,1)

y=BD$nbr_tél_possédé

BD=cbind(BD,y)

View(BD)

N=100

mu=mean(BD$y)

mu

x=matrix(100,1)

for (i in 1:100) {

```

```
(if(BD$budget_mensuel[i]=="100") (x[i]=100))
```

```
(if(BD$budget_mensuel[i]=="250") (x[i]=250))
```

```
(if (BD$budget_mensuel[i]=="400") (x[i]=400)))}
```

```
sonda=cbind(BD,x)
```

```
attach(sonda)
```

```
summary(sonda)
```

```
View(sonda)
```

```
library(sampling)
```

```
pic = inclusionprobabilities(x,20)#calculer les probabilités d'inclusion
```

```
sum(pic)
```

```
pic
```

```
pk=rep(0,length(x))
```

```
vk=rep(0,length(x))
```

```
for(i in 1:N){
```

```
  pk[i]=(x[i])/sum(x)
```

```
}
```

```
pk
```

```
sum(pk)
```

```
vk[1]=pk[1]
```

```
for(i in 2:length(x)){
```

```
  vk[i]=vk[i-1]+pk[i]
```

```
}
```

```
vk
```

```
pvk=cbind(pk,vk)
```

```
pvk
```

```
f=function(res){
```



```

xipi=rep(0,20)
xipi1=rep(0,20)
xipi2=rep(0,20)
sumxipi=0
sumxipi1=0
res=matrix(nrow=20,ncol=4)
for(i in 1:20){
  u=runif(1,0,1)
  if(0<u&&u<vk[1]){ res[i,]=c(1,0,u,vk[1])}
  for(j in 2:100){
    if(vk[j-1]<u&&u<vk[j]){
      (res[i,]=c(i,vk[j-1],u,vk[j]))
    }
  }
}
}

```

```

for(i in 1:20){
  xipi[i]=y[res[i,1]]/pk[res[i,1]]
}
sumxipi=sum(xipi)
T_HH=sumxipi/n
ybarpiar=T_HH/N

```

```

for(i in 1:100){
  xipi1[i]=(((y[i]/pk[i])-T)^2)*pk[i]
}

```

```
sumxipi1=sum(xipi1)
```

```
varybarpiar=sum(xipi1)/(n*(N^2))
```

```
for(i in 1:20){
```

```
  xipi2[i]=((y[res[i,1]]/pk[res[i,1]])-(N*ybarpiar))^2
```

```
}
```

```
varestimybarpiar=sum(xipi2)/(n*(n-1)*(N^2))
```

```
v=c(res[,1],ybarpiar,varybarpiar,varestimybarpiar)
```

```
return(v)
```

```
}
```

```
res1=replicate(20,f(res))
```

```
res1
```

```
lCbinf=1:20
```

```
lCbsup=1:20
```

```
longueur=1:20
```

```
for (i in 1:20) {
```

```
  lCbinf[i]=res1[21,i]-1.729*sqrt(res1[23,i])
```

```
  lCbsup[i]=res1[21,i]+1.729*sqrt(res1[23,i])
```

```
  mu[i]=res1[21,i]
```

```
  longueur[i]=lCbsup[i]-lCbinf[i]
```

```
}
```

```
longueur
```

```
mean(longueur)
```

```
mu1=mean(mu)
```

```
View(res1)
```

```
echan=res1[1:20,]
```

```
colnames(echan)=c("ech1","ech2","ech3","ech4","ech5","ech6","ech7","ech8","ech9","ech10",
"ech11","ech12","ech13","e
```

```
ch14","ech15","ech16","ech17","ech18","ech19","ech20")
```

```
View(echan)
```

```
ress=res1[21:23,]
```

```
ress=t(ress)
```

```
colnames(ress)=c("y_bar_piar","var_ybar_piar","var_est")
```

```
ress=data.frame(y=1:20,ress[,1],ress[,2],ress[,3],lCbinf,lCbsup,longueur)
```

```
g1_piar=ggplot(ress,aes(y,lCbsup))+geom_point(colour="red",cex=4)+geom_point(aes(y,lCbinf),colour="darkblue",cex=4)+geom_path(colour="red",size=1)+geom_path(aes(y,lCbinf),colour="darkblue",size=1)
```

```
g2_piar=annotate("text", label =c("La borne inférieure","La borne supérieure"), x =c(3,3), y =c(1,1.5),size =5, colour =c("darkblue","red"))
```

```
g3_piar=labs(x="",y="")
```

```
g1_piar+g2_piar+g3_piar+ggtitle("les 20 intervalles de confiance : PIAR")+theme(plot.title = element_text(color="black",size=14, face="bold.italic"))
```

```
ress=ress[-1]
```

```
colnames(ress)=c("ybar_chapeau_piar","var_ybar_chapeau_piar","var_estim_ybar_chapeau_piar","lCbinf_PiAR","lCbsup_PiAR","long_IC_PiAR")
```

```
View(ress)
```

```
mean(ress[,1])
```

```
#####  
### pirs
```

```
set.seed(5)
```

```
s=sum(x)
```

```
g=function(r){
```

```
  f=function(result){
```

```
    res=matrix(nrow=20,ncol=4)
```

```
    p=1
```

```
    q=5
```

```
  #on tire 1 individu dans chaque groupe
```

```
  for (i in 1:20){
```

```
pk=rep(0,5)
```

```
vk=rep(0,5)
```

```
xn=rep(0,5)
```

```
Trhc=0
```

```
Tgpi=rep(0,20)
```

```
#les 5 variables x dans le groupe i
```

```
j=1
```

```
for(k in p:q){
```

```
  xn[j]=x[k]
```

```
  j=j+1
```

```
}
```

```
#les probabilités d'inclusion dans le groupe i
```

```
j=1
```

```
for(k in p:q){
```

```
  pk[j]=(x[k])/sum(xn)
```

```
  j=j+1
```

```
}
```

```
vk[1]=pk[1]
```

```
for(k in 2:5){
```

```
  vk[k]=vk[k-1]+pk[k]
```

```
}
```

```
#on tire 1 individu de le groupe i
```

```
u=runif(1,0,1)
```

```
if(0<u&&u<vk[1]){ res[i,]=c(p,0,u,vk[1])}
```

```
for(j in 2:5){
```

```
  if(vk[j-1]<u&&u<vk[j]){
```

```

        (res[i,]=c(p+j-1,vk[j-1],u,vk[j]))
    }
}

```

```

    q=q+5
    p=p+5
}

```

```

p=1
for (i in 1:20){
  Tg=rep(0,5)
  for (k in 1:5){
    Tg[k]=x[p+k-1]*(y[res[i,1]]/x[res[i,1]])

  }
  Tgpi[i]=sum(Tg)
  p=p+5

}
result=cbind(res[,1],Tgpi)
return(result)

}

```

#L'estimateur Horvitz Thompson

```
T_rhc=sum((f(result))[,2])
```

#la moyenne

```
y_bar_pisr=T_rhc/100
```

```

p=1
xg=rep(0,20)
xig=rep(0,5)
for (i in 1:20){
  for (k in 1:5){
    xig[k]=x[p+k-1]
  }
  xg[i]=sum(xig)
  p=p+5
}

```

```

s1=rep(0,20)
for (i in 1:20) {
  s1[i]=s*(xg[i]*((y[(f(result))[i,1]]/x[(f(result))[i,1]]^2))
}

```

```

som=sum(s1)
#la variance estimée
var_estim_y_pisr=abs((((1-(n/N))/(n*(n-1)))*(som-(T_rhc^2)))*(n/(N^2)))

```

```

r=c((f(result))[,1],y_bar_pisr,var_estim_y_pisr)
return(r)
}

```

```

g(r)
#on répète le tirage 20 fois
res2=replicate(20,g(r),set.seed(5))
res2
#les intervalles de confiance

```

```

ICbinf=1:20
ICbsup=1:20
longueur=1:20
for (i in 1:20) {
  ICbinf[i]=res2[21,i]-1.729*sqrt(res2[22,i])
  ICbsup[i]=res2[21,i]+1.729*sqrt(res2[22,i])
  mu[i]=res2[21,i]
  longueur[i]=ICbsup[i]-ICbinf[i]
}
tt=data.frame(y=1:20,ICbinf,ICbsup,longueur)

t=ggplot(tt,aes(y,ICbsup))+geom_point(colour="red",cex=4)+geom_point(aes(y,ICbinf),
,colour="darkblue",cex=4)+geom_path(colour="red",size=1)+geom_path(aes(y,ICbinf
),colour="darkblue",size=1)

t1=annotate("text", label =c("La borne inférieure","La borne supérieure"), x =c(17,17),
y =c(0.5,1),size =5, colour = c("darkblue","red"))

t2=labs(x="",y="")

t+t1+t2+ggtitle("les 20 intervalles de confiance : PISR")+theme(plot.title =
element_text(color="black", size=14,face="bold.italic"))

longueur#les longueurs des intervalles de confiance
mean(longueur)#la moyenne des longueurs
mu#y_bar_pisr dans chaque echantillon
mu1=mean(mu)
mu1#la moyenne des y_bar_pisr
ech=res2[1:20,]
colnames(ech)=c("ech1","ech2","ech3","ech4","ech5","ech6","ech7","ech8","ech9","ech10",
,"ech11","ech12","ech13","ech
14","ech15","ech16","ech17","ech18","ech19","ech20")
View(ech)

tt=cbind(res2[21,],res2[22,],tt[,-1])
colnames(tt)=c("y_bar_chapeau_pisr","var_est_ybar_chapeau","lcbinf_pisr","ICbsup_pi
sr","long_IC_pisr")
View(tt)

```

```
#####Synthèse
tab=data.frame(y=1:20,long_PEAR,long_PESR,long_PJAR,long_PISR,long_strat_taille,lo
ng_strat_prop,long_strat_opt,long_g
rappe,long_2deg)

a=ggplot(tab,aes(y,long_PEAR))+geom_point(colour="darkgreen",cex=4)+
geom_path(colour="darkgreen",size=1)+
geom_point(aes(y,long_PESR),colour="deeppink",cex=4)+
geom_path(aes(y,long_PESR),colour="deeppink",size=1)+
geom_point(aes(y,long_PJAR),colour="blue",cex=4)+
geom_path(aes(y,long_PJAR),colour="blue",size=1)+
geom_point(aes(y,long_PISR),colour="black",cex=4)+
geom_path(aes(y,long_PISR),colour="black",size=1)+
geom_point(aes(y,long_strat_taille),colour="coral1",cex=4)+
geom_path(aes(y,long_strat_taille),colour="coral1",size=1)+
geom_point(aes(y,long_strat_prop),colour="green",cex=4)+
geom_path(aes(y,long_strat_prop),colour="green",size=1)+
geom_point(aes(y,long_strat_opt),colour="red",cex=4)+
geom_path(aes(y,long_strat_opt),colour="red",size=1)+
geom_point(aes(y,long_grappe),colour="deeppink4",cex=4)+
geom_path(aes(y,long_grappe),colour="deeppink4",size=1)+
geom_point(aes(y,long_2deg),colour="indianred3",cex=4)+
geom_path(aes(y,long_2deg),colour="indianred3",size=1)

b=annotate("text",
label=c("strat_optimal","strat_prop","strat_mem_taille","PESR","PEAR","Grappe","PJAR"
,"PISR","2_degré"),
x=c(19,19,19,19,19,19,19,19,19),
y=c(9.5,10.5,11.5,12.5,13.5,14.5,15.5,16.5,17.5),size =5, colour =

c("red","green","coral1","deeppink","darkgreen","deeppink4","blue","black","indianred
3"))

c=labs(x="",y="")

a+b+c+ggtitle("Synthèse des logeurs des intervalles de confiance")+theme(plot.title =
element_text(color="black", size=14,face="bold.italic"))
```



#####la stratification #####

```
BD111=read.table(file='C:/Users/USER/Documents/ProjetSondage/BD111.csv',header = T, sep = ";")
View(BD111)
colnames(BD111)=c('genre', 'nbr_tél_posse  ', 'nbre_t  l_cadeau', 'd  pense', 'budget_mensuel', 'cout_t  l_actuel')
attach(BD111)

library(lpSolve)
library(MASS)
library(sampling)
N=100
n=20
y=BD111$nbr_t  l_posse  
set.seed(6)
h=2
Nh=as.vector(table(BD111$genre))
Nh

## [1] 43 57

strate1=c(1:Nh[2]) #femme
j=1

k=1
for (i in 1:100) {
  if(BD111$genre[i]==1) {strate1[k]=y[i]; k=k+1;}
}
strate1

## [1] 2 2 3 3 3 3 3 2 2 2 4 3 2 1 1 3 3 3 1 1 2 1 3 3 4 2 2 4 3 2 3 2 2 4
3
## [36] 2 1 4 1 2 4 4 3 2 3 3 2 4 3 3 4 2 4 2 3 1 2

strate2=rep(0,Nh[1]) #femme
d=1
for (i in 1:100) {
  if(BD111$genre[i]==0) {strate2[d]=y[i]; d=d+1;}
}
strate2

## [1] 2 2 1 4 2 2 2 2 3 2 2 4 3 1 2 2 2 2 2 2 1 3 2 2 1 2 2 3 3 3 3 4 1 3
2
## [36] 4 2 2 4 4 2 4 3

strate1

## [1] 2 2 3 3 3 3 3 2 2 2 4 3 2 1 1 3 3 3 1 1 2 1 3 3 4 2 2 4 3 2 3 2 2 4
3
## [36] 2 1 4 1 2 4 4 3 2 3 3 2 4 3 3 4 2 4 2 3 1 2

strate2
```

```

## [1] 2 2 1 4 2 2 2 2 3 2 2 4 3 1 2 2 2 2 2 2 1 3 2 2 1 2 2 3 3 3 3 4 1 3
2
## [36] 4 2 2 4 4 2 4 3

strates=list(strate1,strate2)
strates

## [[1]]
## [1] 2 2 3 3 3 3 3 2 2 2 4 3 2 1 1 3 3 3 1 1 2 1 3 3 4 2 2 4 3 2 3 2 2 4
3
## [36] 2 1 4 1 2 4 4 3 2 3 3 2 4 3 3 4 2 4 2 3 1 2
##
## [[2]]
## [1] 2 2 1 4 2 2 2 2 3 2 2 4 3 1 2 2 2 2 2 2 1 3 2 2 1 2 2 3 3 3 3 4 1 3
2
## [36] 4 2 2 4 4 2 4 3

#### ybar, variances et variances corrigées des strates
sigma2_hc=NULL
sigma2_h=NULL
ybar_strates=NULL
for (i in 1:h){
  ybar_strates[i]=list(mean(unlist(strates[i])))
  ybar_strates=as.vector(unlist(ybar_strates))
  ybar_strates
  sigma2_hc[i]=var(unlist(strates[i]))
  sigma2_h[i]=((N-1)/N)*sigma2_hc[i]}
ybar_strates

## [1] 2.561404 2.418605

sigma2_h

## [1] 0.8845489 0.8123920

sigma2_hc

## [1] 0.8934837 0.8205980

resultat=data.frame(strate1=c(2.56140, 0.8845489, 0.8934837),strate2=c( 2.4
18605,0.8123920
,0.8205980),
                      row.names = c("moyenne","variance","variance-corrigée")
)
View(resultat)
##### 1/ Les strates ont la même taille#####
### Calcul des nh
nh=c(rep(round(n/h),h))
nh

## [1] 10 10

sum(nh)

## [1] 20

```

```

### Tirage de 20 echantillons et calcul demandé
res_strat=matrix(0,20,7)
res_strat

##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,]    0    0    0    0    0    0    0
## [2,]    0    0    0    0    0    0    0
## [3,]    0    0    0    0    0    0    0
## [4,]    0    0    0    0    0    0    0
## [5,]    0    0    0    0    0    0    0
## [6,]    0    0    0    0    0    0    0
## [7,]    0    0    0    0    0    0    0
## [8,]    0    0    0    0    0    0    0
## [9,]    0    0    0    0    0    0    0
## [10,]   0    0    0    0    0    0    0
## [11,]   0    0    0    0    0    0    0
## [12,]   0    0    0    0    0    0    0
## [13,]   0    0    0    0    0    0    0
## [14,]   0    0    0    0    0    0    0
## [15,]   0    0    0    0    0    0    0
## [16,]   0    0    0    0    0    0    0
## [17,]   0    0    0    0    0    0    0
## [18,]   0    0    0    0    0    0    0
## [19,]   0    0    0    0    0    0    0
## [20,]   0    0    0    0    0    0    0

for (k in 1:20){
  #####ybar_ech et var_ech
  ech1=sample(strate1,nh[1],replace=FALSE)
  ech1
  ybar_chap1=mean(ech1)
  s2_c_h1=var(ech1)
  ech2=sample(strate2,nh[2],replace=FALSE)
  ech2
  ybar_chap2=mean(ech2)
  s2_c_h2=var(ech2)
  # ybar estimée pour chaque strate
  ybar_chap_strates=c(ybar_chap1,ybar_chap2)
  ybar_chap_strates
  # ybar estimée
  ybar_chap=(1/N)*sum(Nh*ybar_chap_strates)
  ybar_chap
  mean(y)
  # variances corrigées de l'échantillon dans chaque strate
  s2_hc=c(s2_c_h1,s2_c_h2)
  # Variance totale - Variance intra strates - variance inter strates
  fh=NULL
  for(i in 1:h){
    fh[i]=nh[i]/Nh[i]
  }
  var_ybar_chap_strat=(1/N^2)*(sum(((Nh^2)/nh)*(1-fh)*sigma2_hc))
  var_inter_y=(1/N)*sum(Nh*(ybar_strates-mean(y))^2)
  var_intra_y=(1/N)*sum(Nh*sigma2_h)
  var_inter_y+var_intra_y
}

```

```

var_y=((N-1)/N)*var(y)
#variance estimée
var_estim_ybar_chap_strat=(1/N^2)*(sum(((Nh^2)/nh)*(1-fh)*s2_hc))
var_y=((N-1)/N)*var(y)
var_y

## les intervalles de confiance
ic_inf_strat=NULL
ic_sup_strat=NULL
interval_strat=NULL
long_ic=NULL
ic_inf_strat=ybar_chap-1.96*var_estim_ybar_chap_strat
ic_sup_strat=ybar_chap+1.96*var_estim_ybar_chap_strat
long_ic=ic_sup_strat-ic_inf_strat
res_strat[k,]=c(ybar_chap,var_ybar_chap_strat,
               var_estim_ybar_chap_strat,var_intra_y,
               ic_inf_strat,ic_sup_strat,long_ic)
}
var_inter_y

## [1] 0.005397634

var_intra_y

## [1] 0.8434195

var_inter_y+var_intra_y

## [1] 0.8488171

var_y

## [1] 0.85

resultat=data.frame(var_inter_y=c(0.005397634),var_intra_y=c(0.8434195),var_y=c(0.85))
View(resultat)
### Resultats du sondage stratifié regroupés dans un tableau
res_strat

##      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,] 2.743 0.03466235 0.04835367 0.8434195 2.648227 2.837773 0.18954637
## [2,] 2.428 0.03466235 0.03978667 0.8434195 2.350018 2.505982 0.15596373
## [3,] 2.444 0.03466235 0.03457133 0.8434195 2.376240 2.511760 0.13551963
## [4,] 2.444 0.03466235 0.02266467 0.8434195 2.399577 2.488423 0.08884549
## [5,] 2.201 0.03466235 0.01276967 0.8434195 2.175971 2.226029 0.05005709
## [6,] 2.387 0.03466235 0.01998567 0.8434195 2.347828 2.426172 0.07834381
## [7,] 2.343 0.03466235 0.03140167 0.8434195 2.281453 2.404547 0.12309453
## [8,] 2.285 0.03466235 0.02600833 0.8434195 2.234024 2.335976 0.10195267
## [9,] 2.529 0.03466235 0.03462167 0.8434195 2.461142 2.596858 0.13571693
## [10,] 2.386 0.03466235 0.03155933 0.8434195 2.324144 2.447856 0.12371259
## [11,] 2.371 0.03466235 0.04919100 0.8434195 2.274586 2.467414 0.19282872
## [12,] 2.557 0.03466235 0.03560567 0.8434195 2.487213 2.626787 0.13957421
## [13,] 2.457 0.03466235 0.03819233 0.8434195 2.382143 2.531857 0.14971395
## [14,] 2.643 0.03466235 0.04372833 0.8434195 2.557292 2.728708 0.17141507

```

```
## [15,] 2.215 0.03466235 0.03721500 0.8434195 2.142059 2.287941 0.14588280
## [16,] 2.215 0.03466235 0.02810833 0.8434195 2.159908 2.270092 0.11018467
## [17,] 2.957 0.03466235 0.02873900 0.8434195 2.900672 3.013328 0.11265688
## [18,] 2.501 0.03466235 0.02124167 0.8434195 2.459366 2.542634 0.08326733
## [19,] 2.358 0.03466235 0.03331000 0.8434195 2.292712 2.423288 0.13057520
## [20,] 2.786 0.03466235 0.03814333 0.8434195 2.711239 2.860761 0.14952187

colnames(res_strat)=c("ybar_chap", "var_ybar_chap_strat", "var_estim_ybar_cha
p_strat",
                      "var_intra_strat", "ic_inf_strat", "ic_sup_strat"
                      , "long_ic")
View(res_strat)
### La moyenne de ybar estimées
ybar_strat=mean(res_strat[,1])
ybar_strat

## [1] 2.4625

### Moyenne des longueurs des intervalles de confiance
long_strat=mean(res_strat[,7])
long_strat

## [1] 0.1284187

### Représentation de l'intervalles de confiance pour chacune des 20 expéri
ences sur le même graphe
res_s=NULL
library(ggplot2)
res_s=data.frame(y=1:20, res_strat[,1], res_strat[,2], res_strat[,3], res_strat
[,4], res_strat[,5], res_strat[,6], res_strat[,7])
g1_s=ggplot(res_s, aes(y, res_strat[,6]))+geom_point(colour="deeppink3", cex=4
)+geom_point(aes(y, res_strat[,5]), colour="darkturquoise", cex=4)+geom_path(c
olour="deeppink3", size=1)+geom_path(aes(y, res_strat[,5]), colour="darkturquo
ise", size=1)
g2_s=annotate("text", label =c("La borne supérieure", "La borne inférieure")
, x =c(3,3), y =c(2,2.5), size =5, colour = c("deeppink3", "darkturquoise"))
g3_s=labs(x="", y="")
g1_s+g2_s+g3_s+ggtitle("les 20 intervalles de confiance : Strates de même t
ailles")+theme(plot.title = element_text(color="black", size=14, face="bold
.italic"))

res_strat=res_strat[, -4]
View(res_strat)

#####
#####

## 2/ Les strates sont à allocation proportionnelle
nh=round(Nh*0.2)
nh

## [1] 9 11

### Tirage de 20 échantillons et calcul demandé
res=matrix(0,20,7)
res_prop=matrix(0,20,7)
res_prop
```

```

##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,]    0    0    0    0    0    0    0
## [2,]    0    0    0    0    0    0    0
## [3,]    0    0    0    0    0    0    0
## [4,]    0    0    0    0    0    0    0
## [5,]    0    0    0    0    0    0    0
## [6,]    0    0    0    0    0    0    0
## [7,]    0    0    0    0    0    0    0
## [8,]    0    0    0    0    0    0    0
## [9,]    0    0    0    0    0    0    0
## [10,]   0    0    0    0    0    0    0
## [11,]   0    0    0    0    0    0    0
## [12,]   0    0    0    0    0    0    0
## [13,]   0    0    0    0    0    0    0
## [14,]   0    0    0    0    0    0    0
## [15,]   0    0    0    0    0    0    0
## [16,]   0    0    0    0    0    0    0
## [17,]   0    0    0    0    0    0    0
## [18,]   0    0    0    0    0    0    0
## [19,]   0    0    0    0    0    0    0
## [20,]   0    0    0    0    0    0    0

for (k in 1:20){
  ####ybar_ech et var_ech
  ech1=sample(strate1,nh[1],replace=FALSE)
  ech1
  ybar_chap1=mean(ech1)
  s2_c_h1=var(ech1)
  ech2=sample(strate2,nh[2],replace=FALSE)
  ech2
  ybar_chap2=mean(ech2)
  s2_c_h2=var(ech2)
  # ybar estimée pour chaque strate
  ybar_chap_strates=c(ybar_chap1,ybar_chap2)
  ybar_chap_strates
  # ybar estimée
  ybar_chap=(1/N)*sum(Nh*ybar_chap_strates)
  ybar_chap
  mean(y)
  # variances corrigées de l'échantillon dans chaque strate
  s2_hc=c(s2_c_h1,s2_c_h2)
  # Variance totale - Variance intra strates - variance inter strates
  fh=NULL
  for(i in 1:h){
    fh[i]=nh[i]/Nh[i]
  }
  var_ybar_chap_strat=(1/N^2)*(sum(((Nh^2)/nh)*(1-fh)*sigma2_hc))
  var_inter_y=(1/N)*sum(Nh*(ybar_strates-mean(y))^2)
  var_intra_y=(1/N)*sum(Nh*sigma2_h)
  var_inter_y+var_intra_y
  var_y=((N-1)/N)*var(y)
  #variance estimée
  var_estim_ybar_chap_strat=(1/N^2)*(sum(((Nh^2)/nh)*(1-fh)*s2_hc))
  var_y=((N-1)/N)*var(y)
}

```

```

## les intervalles de confiance
ic_inf_strat=NULL
ic_sup_strat=NULL
interval_strat=NULL
long_ic=NULL
ic_inf_strat=ybar_chap-1.729*var_estim_ybar_chap_strat
ic_sup_strat=ybar_chap+1.729*var_estim_ybar_chap_strat
long_ic=ic_sup_strat-ic_inf_strat
res_prop[k,]=c(ybar_chap,var_ybar_chap_strat,
               var_estim_ybar_chap_strat,var_intra_y,
               ic_inf_strat,ic_sup_strat,long_ic)
}

### Resultats du sondage stratifié regroupés dans un tableau
res_prop

##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,] 2.394343 0.03407422 0.02330866 0.8434195 2.354043 2.434644
## [2,] 2.012121 0.03407422 0.02833086 0.8434195 1.963137 2.061105
## [3,] 2.497980 0.03407422 0.04155943 0.8434195 2.426124 2.569836
## [4,] 2.442121 0.03407422 0.04528147 0.8434195 2.363830 2.520413
## [5,] 2.346566 0.03407422 0.03213705 0.8434195 2.291001 2.402131
## [6,] 2.601616 0.03407422 0.04659948 0.8434195 2.521046 2.682187
## [7,] 2.438081 0.03407422 0.02542207 0.8434195 2.394126 2.482036
## [8,] 2.502020 0.03407422 0.03891065 0.8434195 2.434744 2.569297
## [9,] 2.489899 0.03407422 0.02124880 0.8434195 2.453160 2.526638
## [10,] 2.446162 0.03407422 0.01940817 0.8434195 2.412605 2.479718
## [11,] 2.601616 0.03407422 0.03229766 0.8434195 2.545774 2.657459
## [12,] 2.541717 0.03407422 0.03094396 0.8434195 2.488215 2.595219
## [13,] 2.502020 0.03407422 0.02460883 0.8434195 2.459472 2.544569
## [14,] 2.585455 0.03407422 0.02691854 0.8434195 2.538912 2.631997
## [15,] 2.454242 0.03407422 0.04110820 0.8434195 2.383166 2.525319
## [16,] 2.653434 0.03407422 0.03213705 0.8434195 2.597869 2.708999
## [17,] 2.753030 0.03407422 0.03515292 0.8434195 2.692251 2.813810
## [18,] 2.549798 0.03407422 0.05444893 0.8434195 2.455656 2.643940
## [19,] 2.290707 0.03407422 0.02910333 0.8434195 2.240387 2.341027
## [20,] 2.358687 0.03407422 0.03408475 0.8434195 2.299754 2.417619
##           [,7]
## [1,] 0.08060136
## [2,] 0.09796812
## [3,] 0.14371253
## [4,] 0.15658332
## [5,] 0.11112991
## [6,] 0.16114100
## [7,] 0.08790951
## [8,] 0.13455302
## [9,] 0.07347834
## [10,] 0.06711344
## [11,] 0.11168532
## [12,] 0.10700421
## [13,] 0.08509733
## [14,] 0.09308431
## [15,] 0.14215216

```

```
## [16,] 0.11112991
## [17,] 0.12155880
## [18,] 0.18828440
## [19,] 0.10063931
## [20,] 0.11786508

colnames(res_prop)=c("ybar_chap", "var_ybar_chap_prop", "var_estim_ybar_chap_
strat",
                    "var_intra_strat", "ic_inf_prop", "ic_sup_prop"
                    , "long_ic_prop")
### La moyenne de ybar estimées
ybar_prop=mean(res_prop[,1])
ybar_prop

## [1] 2.473081

### Moyenne des longueurs des intervalles de confiance
long_prop=mean(res_prop[,7])
### Représentation de l'intervalles de confiance pour chacune des 20 expéri
ences sur le même graphe
res_prop1=NULL
library(ggplot2)
res_prop1=data.frame(y=1:20, res_prop[,1], res_prop[,2], res_prop[,3], res_prop
[,4], res_prop[,5], res_prop[,6], res_prop[,7])
g1_prop=ggplot(res_prop1, aes(y, res_prop[,6]))+geom_point(colour="lightpink"
, cex=4)+geom_point(aes(y, res_prop[,5]), colour="mediumorchid1", cex=4)+geom_p
ath(colour="lightpink", size=1)+geom_path(aes(y, res_prop[,5]), colour="medium
orchid1", size=1)
g2_prop=annotate("text", label =c("La borne inférieure", "La borne supérieur
e"), x =c(3,3), y =c(3,4.5), size =5, colour = c("mediumorchid1", "lightpink"
))
g3_prop=labs(x="", y="")
g1_prop+g2_prop+g3_prop+ggtitle("les 20 intervalles de confiance : Strates
à allocation proportionnelle")+theme(plot.title = element_text(color="black"
, size=14, face="bold.italic"))
```

```
res_prop=res_prop[, -4]
View(res_prop)
```

```
##### 3/ Les strates sont à allocation optimale
nh=round(((n*Nh*(sigma2_hc)^0.5))/sum((Nh*(sigma2_hc)^0.5)) )
nh
```

```
## [1] 9 11
```

```
### Tirage de 20 echantillons et calcul demandé
res_opt=matrix(0,20,7)
for (k in 1:20){
  ### #ybar_ech et var_ech
  ech1=sample(strate1, nh[1], replace=FALSE)
  ech1
  ybar_chap1=mean(ech1)
```



```

s2_c_h1=var(ech1)
ech2=sample(strate2,nh[2],replace=FALSE)
ech2
ybar_chap2=mean(ech2)
s2_c_h2=var(ech2)

# ybar estimée pour chaque strate
ybar_chap_strates=c(ybar_chap1,ybar_chap2)
ybar_chap_strates
# ybar estimée
ybar_chap=(1/N)*sum(Nh*ybar_chap_strates)
ybar_chap
mean(y)
# Variances corrigées de l'échantillon dans chaque strate
s2_hc=c(s2_c_h1,s2_c_h2)
# Variance totale de ybar_chap - Variance intra strates - variance inter
strates
fh=NULL
for(i in 1:h){
  fh[i]=nh[i]/Nh[i]
}
var_ybar_chap_opt=(1/N^2)*(sum(((Nh^2)/nh)*(1-fh)*sigma2_hc))
var_inter_y=(1/N)*sum(Nh*(ybar_strates-mean(y))^2)
var_intra_y=(1/N)*sum(Nh*sigma2_h)
var_inter_y+var_intra_y
var_y=((N-1)/N)*var(y)
#variance estimée
var_estim_ybar_chap_opt=(1/N^2)*(sum(((Nh^2)/nh)*(1-fh)*s2_hc))
var_y=((N-1)/N)*var(y)
## les intervalles de confiance
ic_inf_opt=NULL
ic_sup_opt=NULL
long_ic_opt=NULL
ic_inf_opt=ybar_chap-1.729*var_estim_ybar_chap_opt
ic_sup_opt=ybar_chap+1.729*var_estim_ybar_chap_opt
long_ic_opt=ic_sup_opt-ic_inf_opt
res_opt[k,]=c(ybar_chap,var_ybar_chap_strat,
              var_estim_ybar_chap_opt,var_intra_y,
              ic_inf_opt,ic_sup_opt,long_ic_opt)
}
### Résultats du sondage stratifié regroupés dans un tableau
colnames(res_opt)=c("ybar_chap","var_ybar_chap_opt","var_estim_ybar_chap_opt",
                    "var_intra_strat","ic_inf_strat","ic_sup_strat",
                    "long_ic")
### La moyenne de ybar estimées
ybar_opt=mean(res_opt[,1])
ybar_opt

## [1] 2.48203

### Moyenne des longueurs des intervalles de confiance
long_opt=mean(res_opt[,7])
### Représentation de l'intervalle de confiance pour chacune des 20 expérie

```

nces sur le même graphe

```
res_opt1=NULL
res_opt1=data.frame(y=1:20,res_opt[,1],res_opt[,2],res_opt[,3],res_opt[,4],
res_opt[,5],res_opt[,6],res_opt[,7])
g1_opt=ggplot(res_opt1,aes(y,res_opt[,6]))+geom_point(colour="darkblue",cex
=4)+geom_point(aes(y,res_opt[,5]),colour="cyan2",cex=4)+geom_path(colour="d
arkblue",size=1)+geom_path(aes(y,res_opt[,5]),colour="cyan2",size=1)
g2_opt=annotate("text", label =c("La borne inférieure","La borne supérieure
"), x =c(3,3), y =c(3,4.5),size =5, colour = c("cyan2","darkblue"))
g3_opt=labs(x="",y="")
g1_opt+g2_opt+g3_opt+ggtitle("les 20 intervalles de confiance : Strates à a
llocation optimale")+theme(plot.title = element_text(color="darkblue", size
=14, face="bold.italic"))
```

```
res_opt=res_opt[, -4]
```

```
View(res_opt)
```

```
ybar_strat
```

```
## [1] 2.4625
```

```
long_strat
```

```
## [1] 0.1284187
```

```
ybar_prop
```

```
## [1] 2.473081
```

```
long_prop
```

```
## [1] 0.1146346
```

```
ybar_opt
```

```
## [1] 2.48203
```

```
long_opt
```

```
## [1] 0.1162505
```

```
ybar_strat
```

```
## [1] 2.4625
```

```
ybar_prop
```

```
## [1] 2.473081
```

```
ybar_opt
```

```
## [1] 2.48203
```

```
long_strat
```

```
## [1] 0.1284187
```

```
long_prop
```

```
## [1] 0.1146346
long_opt
## [1] 0.1162505
#### synthese

long_strat_taille=res_strat[,6]
long_strat_prop=res_prop[,6]
long_strat_opt=res_opt[,6]
tab=data.frame(y=1:20, long_strat_taille,long_strat_prop,long_strat_opt)
##### Syntheses stratification
a=ggplot(tab,aes(y,long_strat_taille))+geom_point(colour="mediumorchid1",cex=4)+
  geom_path(aes(y,long_strat_taille),colour="mediumorchid1",size=1)+
  geom_point(aes(y,long_strat_prop),colour="darkmagenta",cex=4)+
  geom_path(aes(y,long_strat_prop),colour="darkmagenta",size=1)+
  geom_point(aes(y,long_strat_opt),colour="turquoise1",cex=4)+
  geom_path(aes(y,long_strat_opt),colour="turquoise1",size=1)

b=annotate("text", label =c("strat_optimal","strat_prop","strat_mem_taille"
), x =c(3,3,3), y =c(1.5,1.6,1.7),size =6, colour = c("turquoise1","darkmagenta",
"mediumorchid1"))
c=labs(x="",y="")
a+b+c+ggtitle("Synthèse des logueurs des intervalles de confiance : strat_e
gal prop opt")+theme(plot.title = element_text(color="black", size=14, face
="bold.italic"))
```

```
#####
```

```
#####GRAPPE#####
```

```
grp=matrix(0,100,1)
for (i in 1:100) {
  (if(BD$cout_tél_actuel[i]=="100") (grp[i]=1))
  (if(BD$cout_tél_actuel[i]=="350") (grp[i]=2))
  (if (BD$cout_tél_actuel[i]=="650") (grp[i]=3))
  (if (BD$cout_tél_actuel[i]=="900") (grp[i]=4))
}
sond=cbind(BD,grp)
View(sond)
#par grappe
p=10
```

```

grappe=NULL
sg=NULL
y_bar_grappe=NULL
var_ygrap=NULL
y_grappe=NULL
ng=NULL
Th=NULL
y_bar_h=NULL
M=4
m=2
p1h=m/M
p1h#proba d'inclusion pi1h
##### la définition des grappes

for(j in 1:4){
  grappe[j]=list(y[grp==j])
}
grappe
#### l'echantillonnage
N=100
Nh=as.vector(table(grp))
sg1=matrix(0,10,2) #la matrice qui va contenir les numero des grappes selctionnées
for(i in 1:p){
  sg=sample(1:4,m,replace=F)
  sg1[i,2]=sg[2]
  sg1[i,1]=sg[1]
  sg1
  Nh=Nh[sg] #le nbre des individus dans chaque grappe selectionnée
  ygrap=grappe[sg]
  y_grappe[i]=list(unlist(ygrap))
}

```

```

ng[i]=list(length(unlist(ygrap)))
y_bar_h[1]=mean(unlist(ygrap[1]))
y_bar_h[2]=mean(unlist(ygrap[2]))
y_bar_grappe[i]=list(((M/(m*N))*sum(Nhh*y_bar_h)) #les y bar estimés
Th[1]=sum(unlist(ygrap[1]))
Th[2]=sum(unlist(ygrap[2]))
var_ygrap[i]=list( (((M-m)*M)/((M-1)*m))*sum(((Th)-((1/M)*sum(y)))^2))/N^2 )
}

d=cbind(mean(unlist(grappe[1])),mean(unlist(grappe[2])),mean(unlist(grappe[3])),mean(
unlist(grappe[4])))
View(d)

colnames(d)=c("grappe1","grappe2","grappe3","grappe4")
d=rbind(d,Nh)
rownames(d)=c("moyenne","taille de grappe")
d=d[-2,]
View(d)

mean(unlist(grappe[4]))
res=cbind(y_bar_grappe,var_ygrap,sg1)
colnames(res)=c("y_bar_grappe","var_ygrap","grappe1","grappe2")
y_bar_grappe=as.numeric(y_bar_grappe) # les y bar estimés
var_ygrap=as.numeric(var_ygrap)

icinf_grappe=NULL
icsup_grappe=NULL
intervalle_grappe=NULL
longlC_grap=NULL
for(i in 1:p){
  icinf_grappe[i]=(y_bar_grappe[i])-1.96*sqrt(var_ygrap[i])
  icsup_grappe[i]=(y_bar_grappe[i])+1.96*sqrt(var_ygrap[i])
  intervalle_grappe[i]=list(c(icinf_grappe[i],icsup_grappe[i]))
  longlC_grap[i]=list(icsup_grappe[i]-icinf_grappe[i])
}

```

```

}

mean(unlist(longlC_grap))

mugrap=mean(y_bar_grappe) #la moyenne des y_bar_grappe

mugrap

longlC_grap=unlist(longlC_grap)

res_grappe=NULL #le tableau recapitulatif

res_grappe=data.frame(y=1:10,y_bar_grappe,var_ygrap,icinf_grappe,icsup_grappe,
longlC_grap)

g1_grappe=ggplot(res_grappe,aes(y,icsup_grappe))+geom_point(colour="#660066",
cex=4)+geom_point(aes(y,icinf_grappe),colour="firebrick1",cex=4)+geom_path(colour="#660066",size=1)+geom_path(aes(y,icinf_grappe),colour="firebrick1",size=1)

g2_grappe=annotate("text", label =c("La borne inférieure","La borne supérieure"), x
=c(3,3), y =c(1,1.5),size =5, colour = c("firebrick1","#660066"))

g3_grappe=labs(x="",y="")

g1_grappe+g2_grappe+g3_grappe+ggtitle("les 10 intervalles de confiance \n par
grappes")+theme(plot.title = element_text(color="black", size=14, face="bold.italic"))

res_grappe=res_grappe[,-1]

res_grappe=cbind(res_grappe,sg1)

colnames(res_grappe)=c("y_bar_grappe","var_ygrap","icinf_grappe","icsup_grappe",
"longlC_grap","num_1ere_grap_selct","num_2ere_grap_selct")

View(res_grappe)

#####Synthèse

tab=data.frame(y=1:20,long_PEAR,long_PESR,long_PJAR,long_PISR,long_strat_taille,lo
ng_strat_prop,long_strat_opt,long_g
rappe)

a=ggplot(tab,aes(y,long_PEAR))+geom_point(colour="darkgreen",cex=4)+
geom_path(colour="darkgreen",size=1)+
geom_point(aes(y,long_PESR),colour="deeppink",cex=4)+
geom_path(aes(y,long_PESR),colour="deeppink",size=1)+
geom_point(aes(y,long_PJAR),colour="blue",cex=4)+

```

```

geom_path(aes(y,long_PJAR),colour="blue",size=1)+
geom_point(aes(y,long_PISR),colour="black",cex=4)+
geom_path(aes(y,long_PISR),colour="black",size=1)+
geom_point(aes(y,long_strat_taille),colour="coral1",cex=4)+
geom_path(aes(y,long_strat_taille),colour="coral1",size=1)+
geom_point(aes(y,long_strat_prop),colour="green",cex=4)+
geom_path(aes(y,long_strat_prop),colour="green",size=1)+
geom_point(aes(y,long_strat_opt),colour="red",cex=4)+
geom_path(aes(y,long_strat_opt),colour="red",size=1)+
geom_point(aes(y,long_grappe),colour="deeppink4",cex=4)+
geom_path(aes(y,long_grappe),colour="deeppink4",size=1)

```

```

b=annotate("text",
label=c("strat_optimal","strat_prop","strat_mem_taille","PESR","PEAR","Grappe","PJAR",
,"PISR","2_degré"),
x=c(19,19,19,19,19,19,19,19),
y=c(9.5,10.5,11.5,12.5,13.5,14.5,15.5,16.5),size =5, colour =
c("red","green","coral1","deeppink","darkgreen","deeppink4","blue","black"))
c=labs(x="",y="")

a+b+c+ggtitle("Synthèse des logueurs des intervalles de confiance")+theme(plot.title =
element_text(color="black", size=14,face="bold.italic"))

```

####Redressement

###redressement par le quotient

```

library(MASS)
library(lpSolve)
library(sampling)
n=20
N=100
s=srswor(n,N)

```

```
echantillon=y[s==1]
moyenne=mean(echantillon)
moyenne
moyenne
P=sum(BD$Genre==1)/N
P
as.numeric(p)
p_ech=sum(BD$Genre[s==1]==0)/n
p_ech
x=P*(1/(p))*(moyenne)
  moyenne.red=x
  moyenne.red

#####resultat
> moyenne #moyenne echantillon
[1] 2.5
```