

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEURE ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITÉ DE CARTHAGE
ÉCOLE SUPÉRIEURE DE LA STAISTIQUE
ET DE L'ANALYSE DE L'INFORMATION



Enquête sur le temps consacré aux réseaux sociaux

Réalisé par :

Sayf CHAGTMI

Aziz ABIDI

Mohamed BEN MAKHLOUF

27 décembre 2017

Remerciements

Nous tenons à exprimer notre sincère gratitude à Mme Héra Ouaili Mallek pour sa disponibilité pour répondre à toutes nos questions concernant ce projet. Nous remercions aussi toute personne ayant collaboré avec nous pour établir cette enquête.

Table des matières

Introduction	1
1 Collectes des données	2
1.1 Définition de la population et la variable d'intérêt choisie :	2
1.2 Etapes de l'enquête	2
1.2.1 Le questionnaire	2
2 Traitement des données	4
2.1 Généralités	4
2.2 Quelques statistiques des données collectées	5
3 Réalisation des sondages	13
3.1 Sondage aléatoire simple	13
3.1.1 Plan aléatoire simple sans remise	13
3.1.2 Plan simple avec remise	14
3.2 Probabilités inégales	16
3.2.1 Plan de sondage avec remise	16
3.2.2 Plan de sondage sans remise	18
3.3 Stratification	22
3.3.1 Principe de stratification	22
3.3.2 Découpage de la population	22
3.3.3 Propriétés de chaque strate	23
3.3.4 Réalisation des sondages et traitement des résultats	23
3.4 Sondage par grappes et à deux degrés	27
3.4.1 Plan de sondage par grappes	27
3.4.2 Plan de sondage à deux degrés	28
4 Synthèse	30
4.1 Sondages aléatoires PESR, PEAR, PISR et PIAR	30
4.2 Stratification	31
4.3 Sondage par grappes et à deux degrés	32
4.4 Comparaison de tous les types de sondages utilisés	33
5 Redressement	34

Table des figures

1	Les derniers chiffres pour l'utilisation de l'internet, les médias sociaux et les mobiles dans le monde	1
---	---	---

Introduction

Les réseaux sociaux font partie intégrante de notre quotidien et nous ne pouvons plus nous en passer ! Nous utilisons tous ces plateformes pour différentes raisons. Mais savons-nous exactement combien de temps nous passons sur les réseaux sociaux ?

Chaque année, We Are Social réalise un document plus que complet pour bien cerner l'usage d'Internet et des réseaux sociaux dans 30 pays du monde, des USA à la Russie en passant par la France. Des statistiques à connaître lorsqu'on s'intéresse aux médias sociaux.



FIGURE 1 – Les derniers chiffres pour l'utilisation de l'internet, les médias sociaux et les mobiles dans le monde

Pour plus de visibilité concernant ce sujet, et dans le cadre de ce projet académique de sondage, nous avons décidé de mener une enquête le temps consacré aux réseaux sociaux par jour sur une population de 100 individus de notre entourage.

Nous avons donc appliqué les différentes méthodes de sondage, ainsi que les théories étudiées en cours sur un exemple réel. Pour cela, nous avons collecté des informations formant une base de 100 individus qui représentent notre population. Puis, nous avons procédé à des différents types de sondage auprès de cette population.

Nous avons donc traité les résultats de ces méthodes de sondage et comparé leurs précisions. Pour notre travail nous avons utilisé « Rstudio » en faisant appel à différentes bibliothèques telles que ggplot2, sampling, scales, MASS et IpSolve.

Partie 1

Collectes des données

1.1 Définition de la population et la variable d'intérêt choisie :

- **Population cible** : Les personnes qui utilisent les réseaux sociaux
- **Population source** : Etudiants de l'ESSAI, l'entourage, amis sur Facebook. . .
- **Variable d'intérêt** : Nombre d'heures consacrés aux réseaux sociaux par jour

1.2 Etapes de l'enquête

1.2.1 Le questionnaire

L'enquête s'est déroulé par partage sur les réseaux sociaux principalement le Facebook. Notre questionnaire comporte alors 10 questions qui sont :

1. **Vous êtes étudiants :**
 - O Oui
 - O Non
2. **Sexe :**
 - O Homme
 - O Femme
3. **Vous êtes :**
 - O Marié(e)
 - O En couple
 - O Un cœur à prendre
4. **Êtes-vous membre d'un ou plusieurs réseaux sociaux :**
 - O Oui
 - O Non
5. **Quelle plateforme utilisez-vous le plus :**
 - O Facebook
 - O Twitter
 - O Instagram/Snapchat
 - Autres
6. **Combien de contacts (amis) avez-vous sur les sites de réseaux sociaux :**
 - O Moins de 100

- ☐ Entre 100 et 500
 - ☐ Entre 500 et 1000
 - ☐ Plus de 1000
7. **Vous accédez aux réseaux sociaux le plus, via :**
- ☐ Ordinateur bureautique
 - ☐ Ordinateur portable/Tablette
 - ☐ Smartphone
8. **Combien d'heures, les jours de la semaine , vous connectez-vous sur les réseaux sociaux en moyenne :**
- ☐ Entre 0 et 1 heure
 - ☐ Entre 1 et 2 heures
 - ☐ Entre 2 et 3 heures
 - ☐ Entre 3 et 4 heures
 - ☐ Plus de 4 heures
9. **À quelle heure êtes-vous le plus actif sur les médias sociaux :**
- ☐ Le matin avant 8 :00
 - ☐ De 8 :00 à 17 :00
 - ☐ À partir de 17 :00
10. **Sur les médias sociaux, vous êtes plus intéressé par :**
- ☐ Voyages/Mode
 - ☐ Nourriture
 - ☐ Télévision/Films
 - ☐ Santé/Sport
 - ☐ Art/Musique
 - Autres

Partie 2

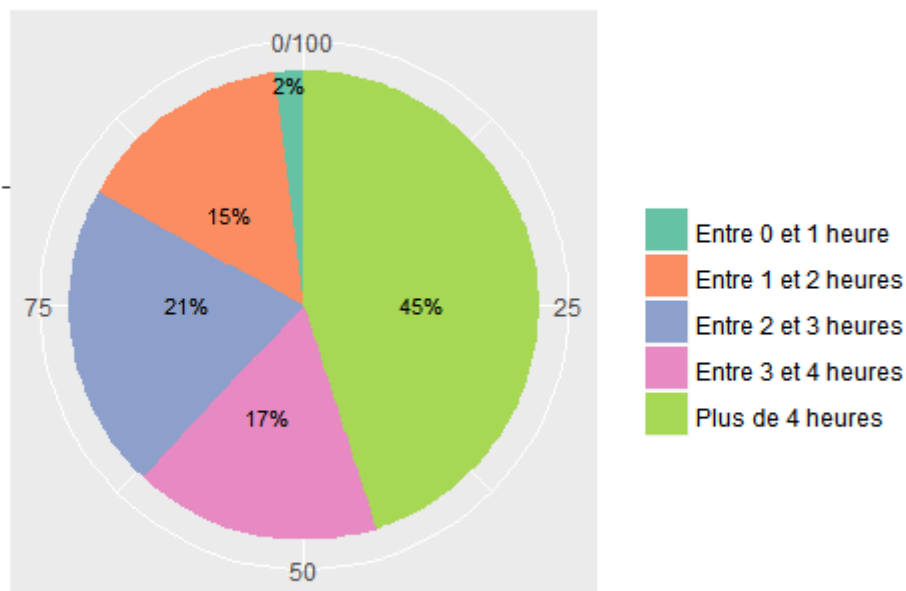
Traitement des données

2.1 Généralités

Notre base de données est composée de 10 variables dont 8 sont qualitatives et 2 sont quantitatives.

La variable d'intérêt est « nb_heures », correspondant à la question : « Combien d'heures, les jours de la semaine, vous connectez-vous sur les réseaux sociaux en moyenne ? ». Cette variable contient 5 modalités : « Entre 0 et 1 heure », « Entre 1 et 2 heures », « Entre 2 et 3 heures », « Entre 3 et 4 heures » et « Plus de 4 heures ». Pour la suite de notre travail nous avons pris le centre de chaque classe à savoir 0.5, 1.5, 2.5, 3.5 et 4.5. La répartition de la population selon notre variable d'intérêt est comme suit :

**La repartition de la population selon la variable d'interet :
Nombre des heures**



On remarque que 45% de la population ont passés plus de 4 heures sur les réseaux sociaux, 17% entre 3 et 4 heures, 21% entre 2 et 3 heures, 15% entre 1 et 2 heures et seulement 2% ont passés moins qu'une heure sur les réseaux sociaux.

- **Moyenne de notre population : 3.8**

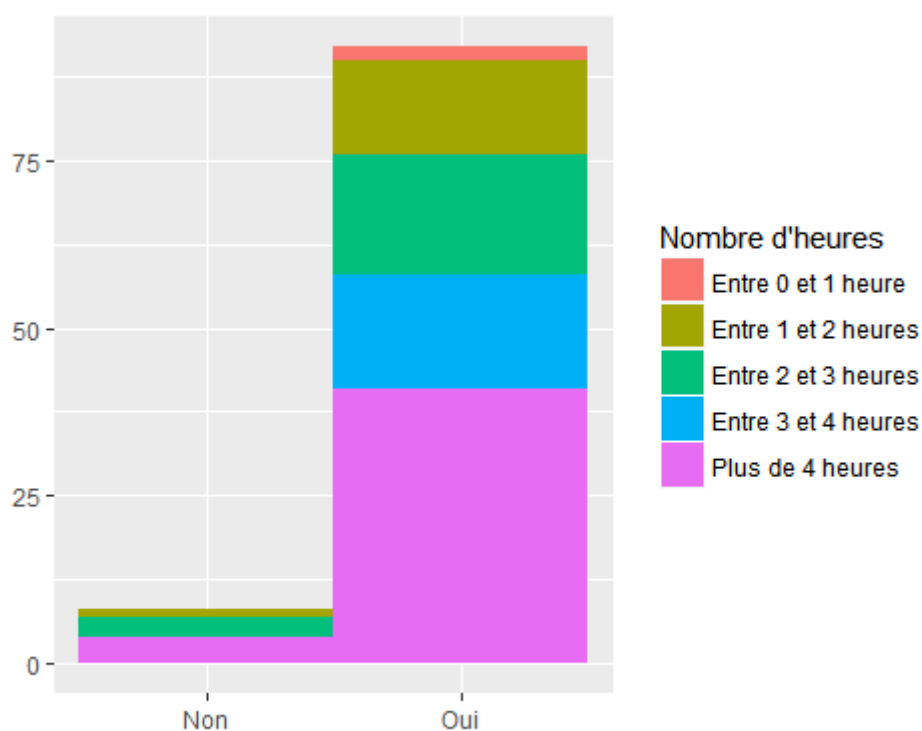
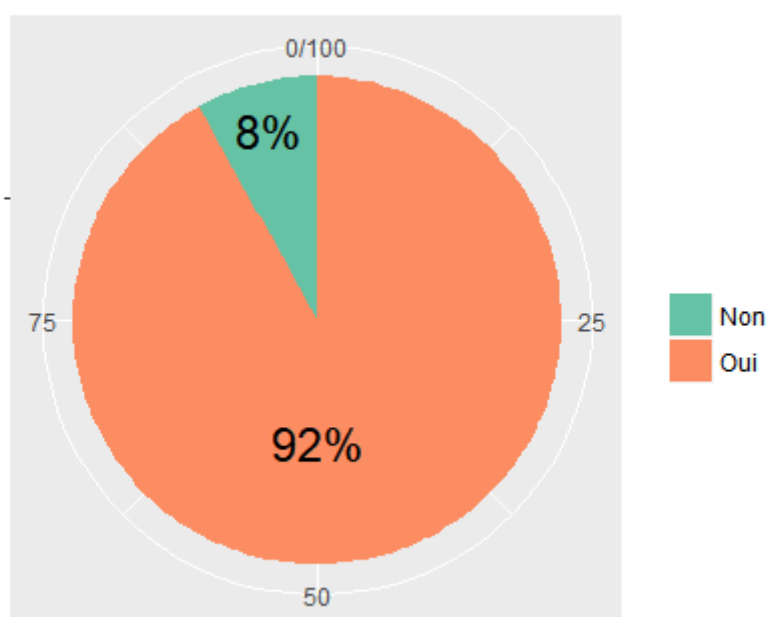
2.2 Quelques statistiques des données collectées

Dans cette partie nous avons représenté quelques statistiques des données collectées à l'aide de camamberts avec le pourcentage de chaque modalité et d'autres graphes dans lesquels on a croisé la variable d'intérêt « nb_heures » avec chacune des variables auxiliaires.

Pour chaque question les statistiques sont représentées comme suit :

Q1 : Vous êtes étudiants

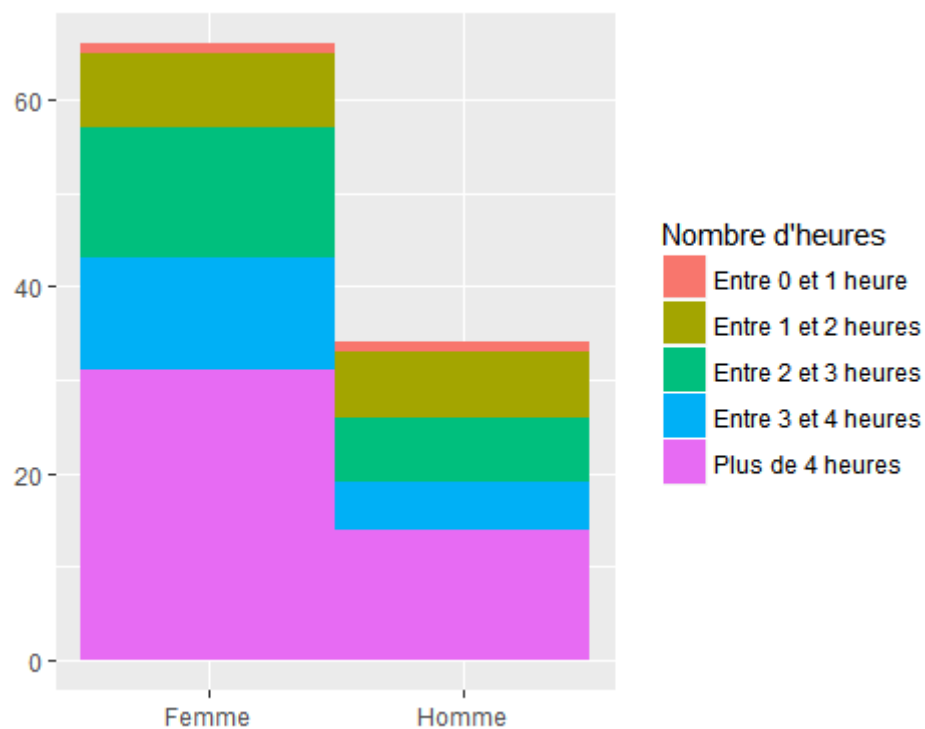
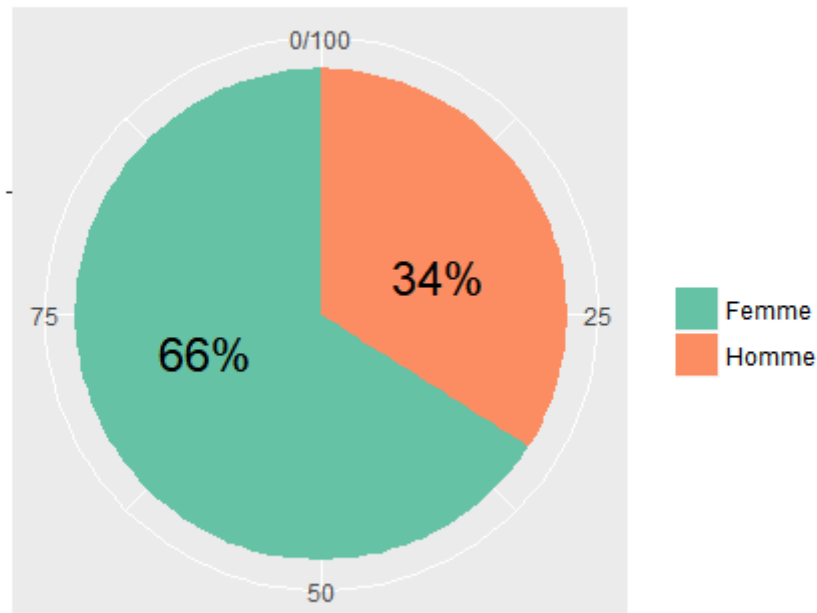
La repartition de la population selon la variable d'inter
Etudiant



On remarque d'après ce graphe que le nombre d'heures passé sur les réseaux sociaux dépend de la variable étudiant.

Q2 : Sexe

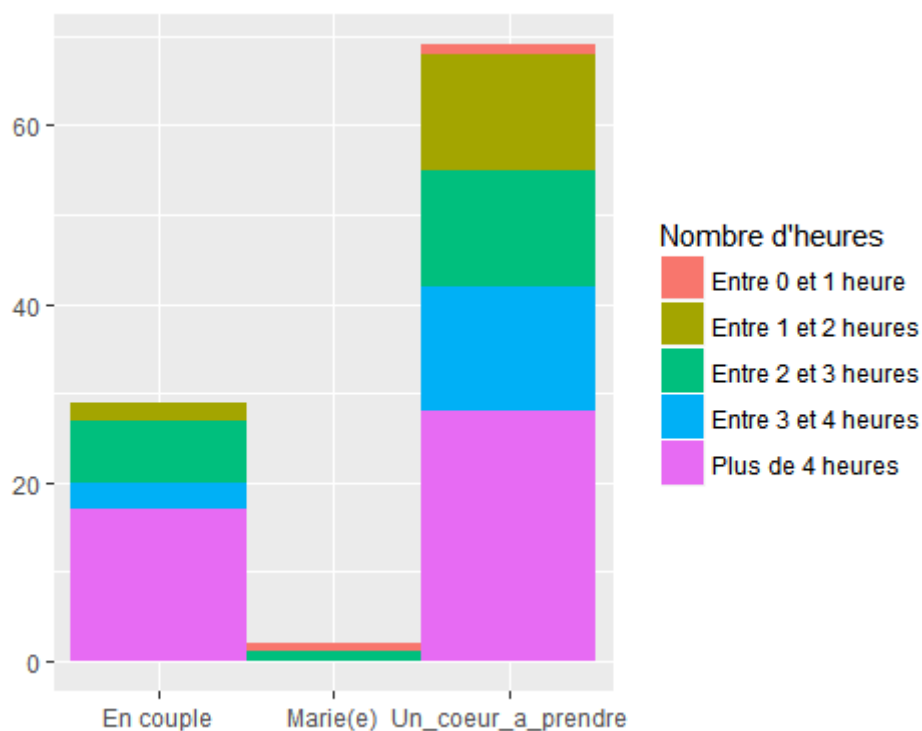
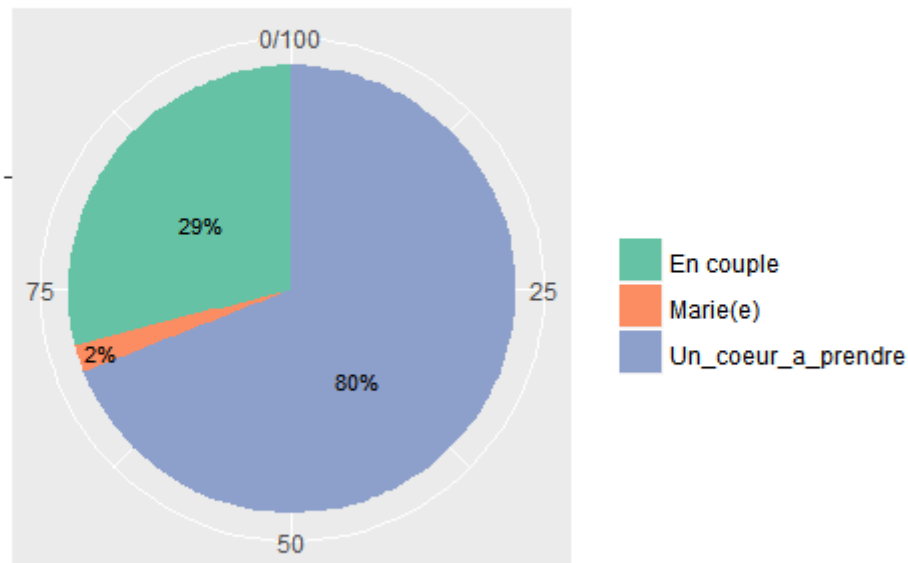
La repartition de la population selon le :
Sexe



66% de notre échantillon sont des femmes et 34% sont des hommes.

Q3 : Vous êtes

La repartition de la population selon :
l'Etat Civil



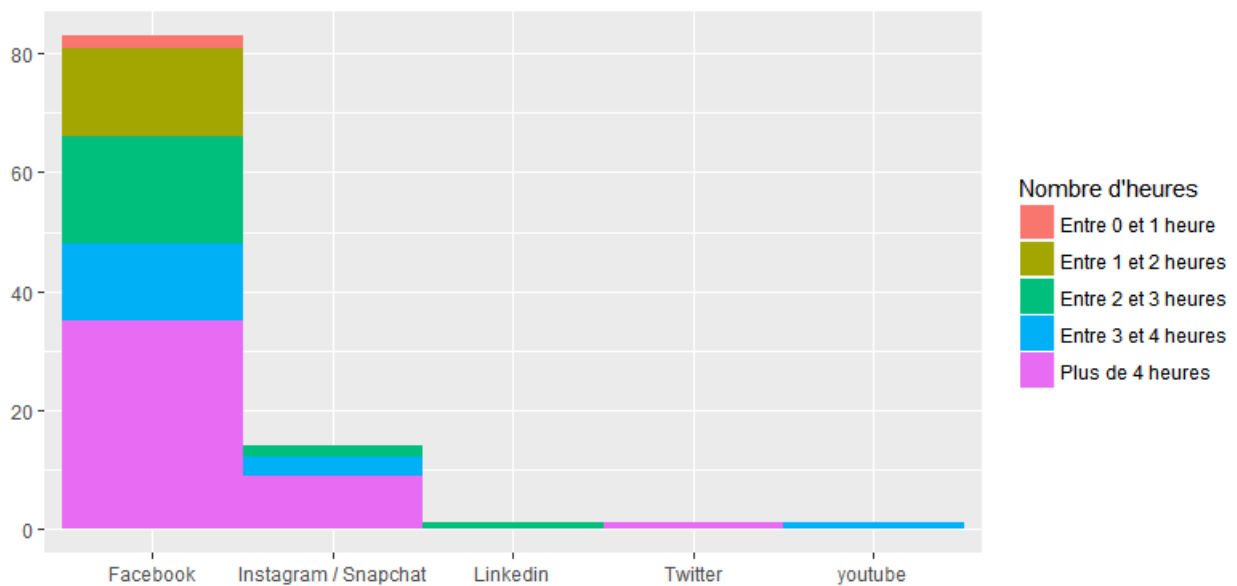
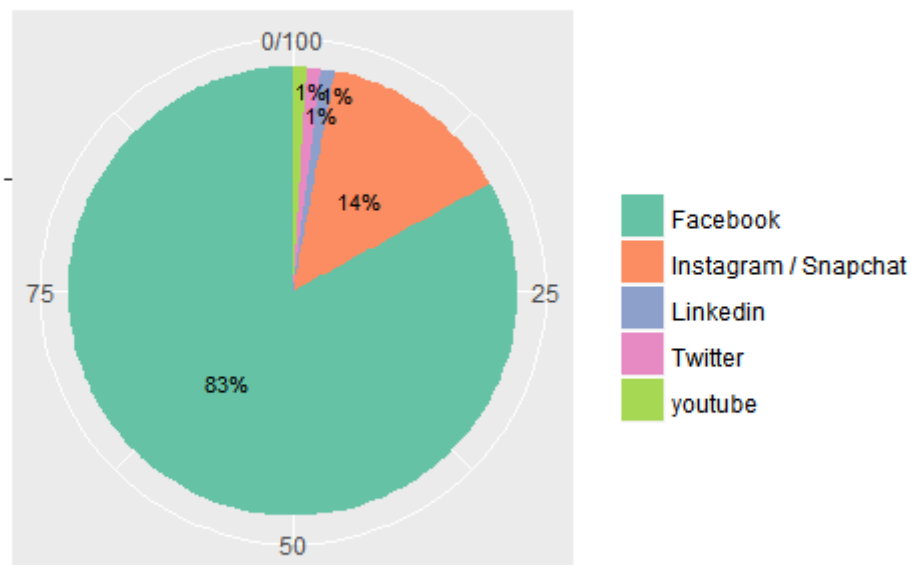
La majorité des utilisateurs des réseaux sociaux dans notre population sont dans la situation "Un cœur à prendre" et ils les utilisent pour plus que 4 heures.

Q4 : Êtes-vous membre d'un ou plusieurs réseaux sociaux

Cette question est une question filtre : notre population n'est composée que des membres des réseaux sociaux.

Q5 : Quelle plateforme utilisez-vous le plus

La repartition de la population selon :
la Plateforme



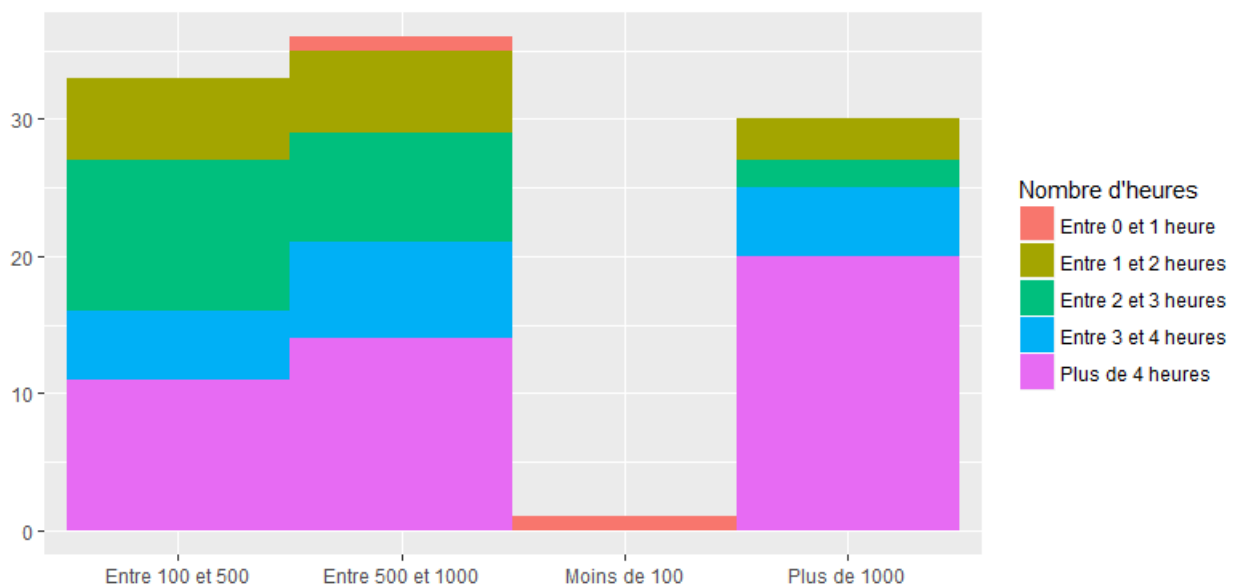
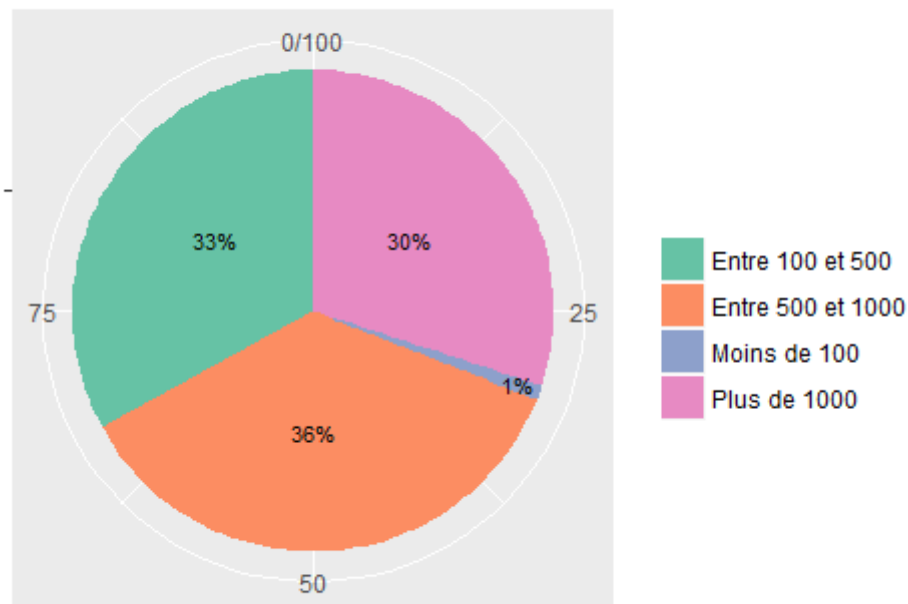
"Facebook" est le réseaux le plus utilisé par les individus : 83% des indivdus utilisent ce réseaux.

En 2ème place, 14% des utilisateurs des réseaux sociaux utilisent "Instagram".

L'utilisation des autres réseaux sociaux est négligeable devant les deux premiers.

Q6 : Combien de contacts (amis) avez-vous sur les sites de réseaux sociaux

La repartition de la population selon :
amis

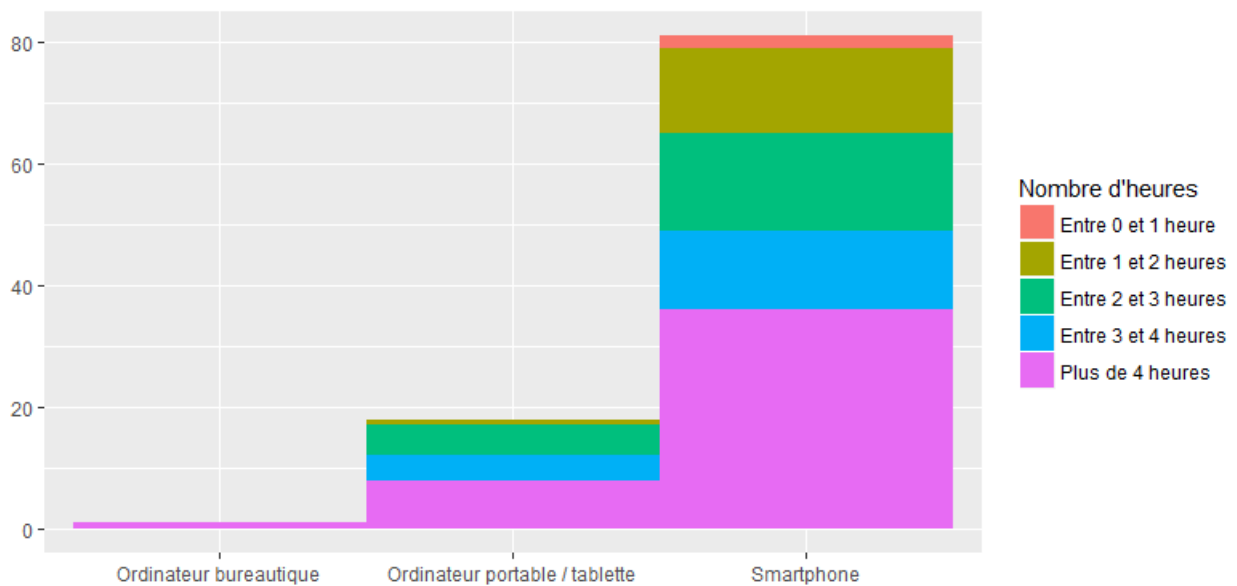
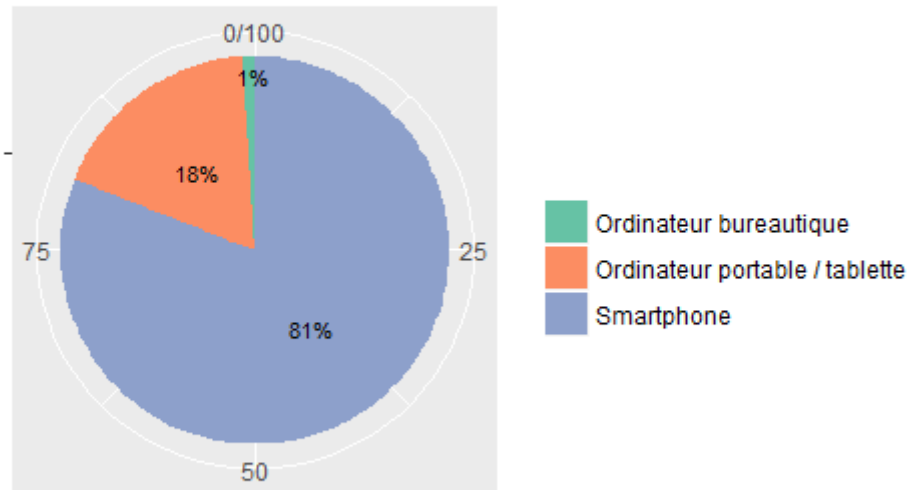


On remarque 99% des individus ont plus que 100 amis sur les réseaux sociaux .

On remarque encore que, le nombre d'heures passées sur les réseaux sociaux est plus ou moins indépendant de nombre d'amis sur ces réseaux.

Q7 : Vous accédez aux réseaux sociaux le plus, via

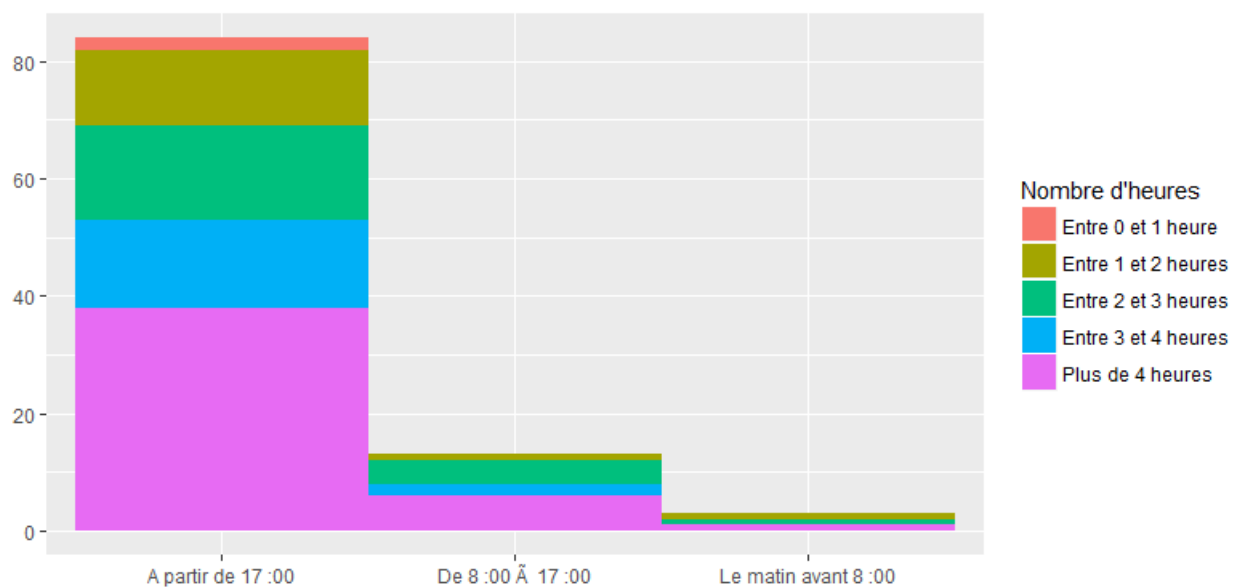
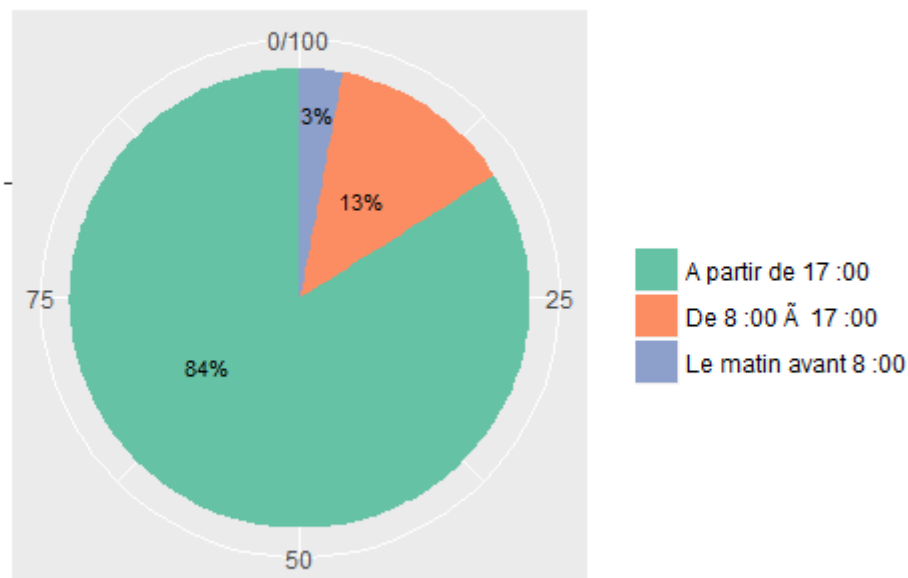
La repartition de la population selon :
l'appareil



La majorité de notre population, 81%, accède aux réseaux sociaux via des smartphones.

Q8 : À quelle heure êtes-vous le plus actif sur les médias sociaux

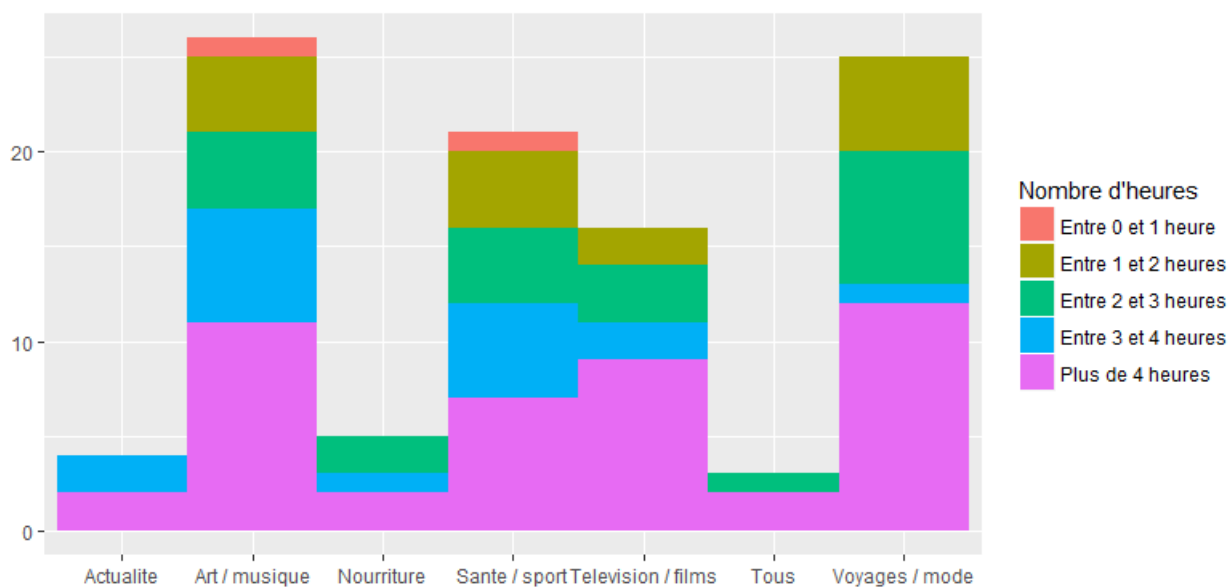
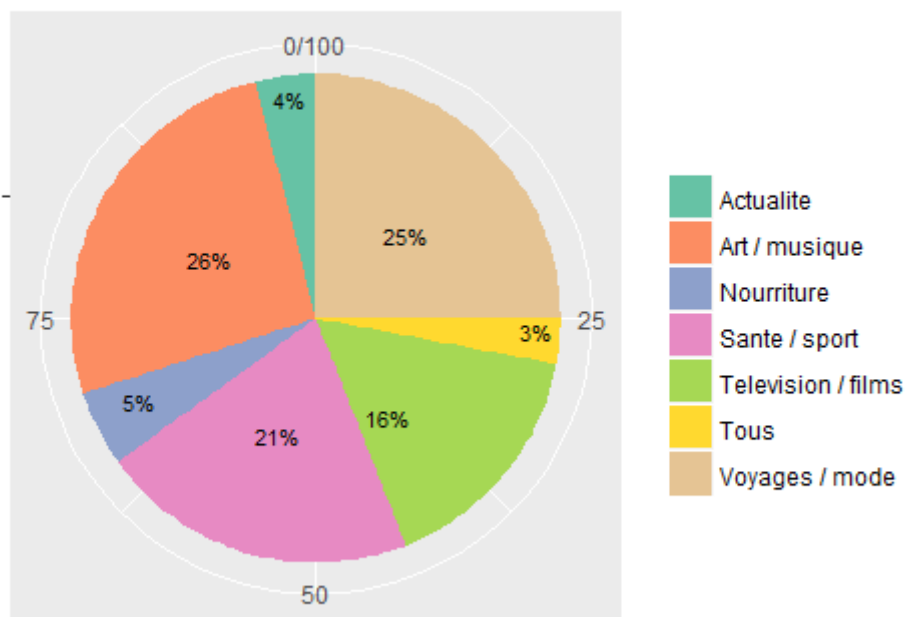
La repartition de la population selon :
le temps



On remarque que le temps favorable pour l'accès aux réseaux sociaux est à partir de 17h et à partir de cette heure presque la moitié des individus les utilisent pour plus de 4 heures.

Q9 : Sur les médias sociaux, vous êtes plus intéressé par

La repartition de la population selon :
l'intérêt



On remarque que les individus sont moins intéressés par l'actualité et la nourriture que les autres types d'activités.

Partie 3

Réalisation des sondages

3.1 Sondage aléatoire simple

Les individus dans le sondage aléatoire simple ont la même probabilité d'être sélectionnés.

3.1.1 Plan aléatoire simple sans remise

Sélection des échantillons

Nous avons tiré un échantillon de 20 individus aléatoirement et sans remise. Pour chaque individu nous avons calculé $\bar{y}_{\text{chapeau_pesr}}$, la variance, la variance estimée et les bornes des intervalles de confiance ainsi que leurs longueurs.

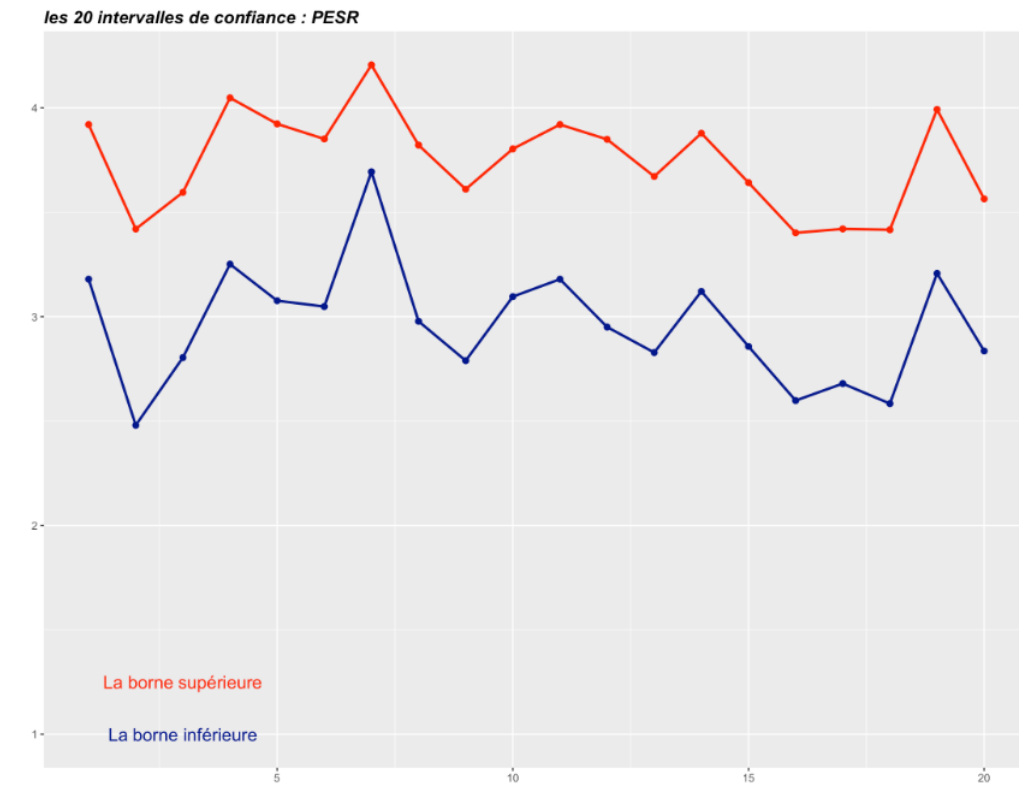
Résultats obtenus

Les résultats obtenus sont regroupés dans le tableau suivant :

	$\bar{y}_{\text{chapeau_pesr}}$	$\text{var}_{\bar{y}_{\text{chapeau_pesr}}}$	$\text{var}_{\text{estim}_{\bar{y}_{\text{chapeau_pesr}}}}$	$\text{IC}_{\text{binf_pesr}}$	$\text{IC}_{\text{sup_pesr}}$	long_IC_pesr
1	3.55	0.057024	0.0459	3.179574	3.920426	0.7408518
2	2.95	0.057024	0.0739	2.479979	3.420021	0.9400419
3	3.20	0.057024	0.0524	2.804214	3.595786	0.7915724
4	3.65	0.057024	0.0531	3.251579	4.048421	0.7968421
5	3.50	0.057024	0.0600	3.076483	3.923517	0.8470336
6	3.45	0.057024	0.0539	3.048589	3.851411	0.8028222
7	3.95	0.057024	0.0219	3.694131	4.205869	0.5117373
8	3.40	0.057024	0.0596	2.977897	3.822103	0.8442054
9	3.20	0.057024	0.0564	2.789385	3.610615	0.8212295
10	3.45	0.057024	0.0419	3.096083	3.803917	0.7078349
11	3.55	0.057024	0.0459	3.179574	3.920426	0.7408518
12	3.40	0.057024	0.0676	2.950460	3.849540	0.8990800
13	3.25	0.057024	0.0595	2.828252	3.671748	0.8434969
14	3.50	0.057024	0.0480	3.121195	3.878805	0.7576098
15	3.25	0.057024	0.0515	2.857627	3.642373	0.7847451
16	3.00	0.057024	0.0540	2.598217	3.401783	0.8035666
17	3.05	0.057024	0.0459	2.679574	3.420426	0.7408518
18	3.00	0.057024	0.0580	2.583602	3.416398	0.8327967
19	3.60	0.057024	0.0516	3.207247	3.992753	0.7855066
20	3.20	0.057024	0.0444	2.835677	3.564323	0.7286458

\Rightarrow La moyenne des $Ybar_chapeau_{pesr} = 3.355$

Représentation des intervalles de confiance



3.1.2 Plan simple avec remise

sélection des échantillons

Nous avons tiré un échantillon de 20 individus aléatoirement et avec remise. Pour chaque individu, nous avons calculé $ybar_chapeau_pesr$, la variance, la variance estimée et les bornes des intervalles de confiance ainsi que leurs longueurs.

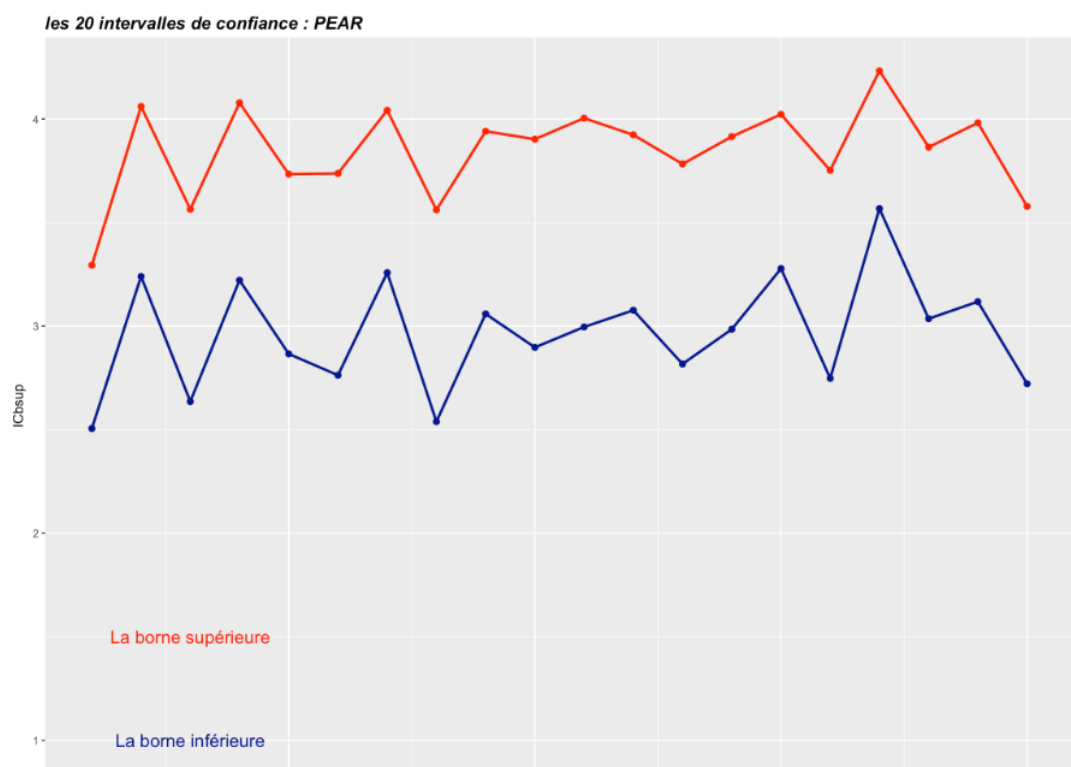
Résultats obtenus

Les résultats obtenus sont regroupés dans le tableau suivant :

	ybar_chapeau_pear	var_ybar_chapeau_pear	var_estim_ybar_chapeau_pear	ICbinf_PEAR	ICbsup_PEAR	long_IC_PEAR
X1	2.90	0.072	0.052000	2.505727	3.294273	0.7885453
X2	3.65	0.072	0.056375	3.239476	4.060524	0.8210475
X3	3.10	0.072	0.072000	2.636061	3.563939	0.9278788
X4	3.65	0.072	0.061375	3.221658	4.078342	0.8566842
X5	3.30	0.072	0.063000	2.866024	3.733976	0.8679511
X6	3.25	0.072	0.079375	2.762879	3.737121	0.9742420
X7	3.65	0.072	0.051375	3.258104	4.041896	0.7837921
X8	3.05	0.072	0.087375	2.538920	3.561080	1.0221593
X9	3.50	0.072	0.065000	3.059190	3.940810	0.8816205
X10	3.40	0.072	0.084500	2.897399	3.902601	1.0052020
X11	3.50	0.072	0.085000	2.995914	4.004086	1.0081716
X12	3.50	0.072	0.060000	3.076483	3.923517	0.8470336
X13	3.30	0.072	0.078000	2.817117	3.782883	0.9657668
X14	3.45	0.072	0.072375	2.984854	3.915146	0.9302920
X15	3.65	0.072	0.046375	3.277662	4.022338	0.7446753
X16	3.25	0.072	0.084375	2.747771	3.752229	1.0044582
X17	3.90	0.072	0.037000	3.567420	4.232580	0.6651596
X18	3.45	0.072	0.057375	3.035851	3.864149	0.8282975
X19	3.55	0.072	0.062375	3.118182	3.981818	0.8636351
X20	3.15	0.072	0.061375	2.721658	3.578342	0.8566842

\Rightarrow La moyenne des ybar_chapeau_pear = 3.4075

Représentation des intervalles de confiance



3.2 Probabilités inégales

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

Les probabilités d'inclusion des individus de notre population deviennent :

```
> pic
[1] 0.20147750 0.33579584 0.20147750 0.08059100 0.08059100 0.08059100 0.08059100
[8] 0.33579584 0.33579584 0.20147750 0.08059100 0.33579584 0.20147750 0.08059100
[15] 0.20147750 0.20147750 0.20147750 0.20147750 0.33579584 0.20147750 0.08059100
[22] 0.08059100 0.20147750 0.20147750 0.33579584 0.33579584 0.20147750 0.08059100
[29] 0.08059100 0.08059100 0.20147750 0.08059100 0.33579584 0.33579584 0.08059100
[36] 0.33579584 0.08059100 0.20147750 0.20147750 0.20147750 0.20147750 0.20147750
[43] 0.33579584 0.08059100 0.20147750 0.20147750 0.33579584 0.08059100 0.20147750
[50] 0.08059100 0.08059100 0.20147750 0.33579584 0.20147750 0.33579584 0.20147750
[57] 0.20147750 0.20147750 0.33579584 0.08059100 0.33579584 0.08059100 0.33579584
[64] 0.08059100 0.08059100 0.20147750 0.08059100 0.33579584 0.08059100 0.20147750
[71] 0.33579584 0.08059100 0.33579584 0.33579584 0.08059100 0.08059100 0.20147750
[78] 0.08059100 0.33579584 0.20147750 0.20147750 0.20147750 0.33579584 0.33579584
[85] 0.20147750 0.33579584 0.33579584 0.33579584 0.33579584 0.33579584 0.33579584
[92] 0.08059100 0.20147750 0.01343183 0.08059100 0.20147750 0.20147750 0.08059100
[99] 0.08059100 0.08059100
```

D'après ce résultat, on remarque qu'il n'existe pas un i_0 tel que $p_i(i_0) > 1$, ce qui implique qu'aucun individu sera sélectionné systématiquement.

3.2.1 Plan de sondage avec remise

Dans ce type de sondage, nous avons utilisé l'estimateur de Hansen & Hurwitz.

Sélection des échantillons

Afin de sélectionner les 20 individus, il faut d'abord calculer p_{ki} et v_{ki} de l'individu i de la manière suivante :

$$p_{ki} = \frac{X_i}{\sum_{i \in P} X_i} \quad ; \quad v_{ki} = \sum_{j=1}^i p_{kj} \quad \text{avec } v_{k0} = 0$$

Algorithme de tirage

Algorithme de tirage :

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

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

Ce tableau présente les 20 individus tirés dans chaque échantillon :

	ech1 [±]	ech2 [±]	ech3 [±]	ech4 [±]	ech5 [±]	ech6 [±]	ech7 [±]	ech8 [±]	ech9 [±]	ech10 [±]	ech11 [±]	ech12 [±]	ech13 [±]	ech14 [±]	ech15 [±]	ech16 [±]	ech17 [±]	ech18 [±]	ech19 [±]	ech20 [±]
1	34	86	49	18	36	73	36	89	55	22	48	81	33	96	19	5	55	70	23	39
2	63	53	56	53	68	77	78	61	33	63	40	86	32	63	43	34	100	96	57	9
3	70	21	95	8	8	32	12	41	42	34	59	10	24	82	68	55	36	73	17	90
4	3	24	4	48	43	84	84	33	89	19	63	75	47	84	66	8	8	84	9	43
5	9	17	70	61	53	49	70	1	19	26	88	86	100	64	32	80	7	26	85	58
6	39	90	89	10	65	53	89	81	38	47	34	13	76	4	51	55	45	42	84	64
7	19	77	12	63	82	52	9	42	56	23	53	14	34	53	83	98	52	1	26	59
8	77	59	80	19	16	20	92	68	36	59	97	8	71	79	47	12	55	63	14	49
9	70	52	61	93	79	46	96	41	2	91	11	78	93	17	87	51	68	15	54	41
10	2	41	92	89	57	87	26	73	38	90	41	63	34	53	47	70	23	82	55	90
11	25	97	47	71	66	53	73	88	39	53	36	66	25	68	53	55	71	26	59	80
12	13	68	89	61	56	41	70	35	29	87	86	24	25	78	12	26	73	66	71	13
13	37	34	78	89	89	9	84	54	25	15	93	76	73	24	19	58	91	18	70	39
14	54	2	9	23	68	43	22	23	49	85	42	100	73	86	79	90	59	9	88	25
15	96	59	9	95	41	77	8	64	19	89	85	34	49	84	33	33	86	68	63	53
16	10	61	68	79	85	67	15	83	87	71	34	47	18	69	80	2	33	79	31	38
17	31	54	73	86	96	13	74	73	80	46	33	20	8	63	88	3	16	19	63	46
18	84	48	16	67	85	3	25	8	88	1	61	33	83	62	67	13	74	26	82	89
19	2	3	75	71	52	68	18	74	66	2	82	83	62	96	88	74	23	34	77	50
20	90	47	15	67	3	59	83	47	38	38	90	43	47	3	25	97	26	35	60	47

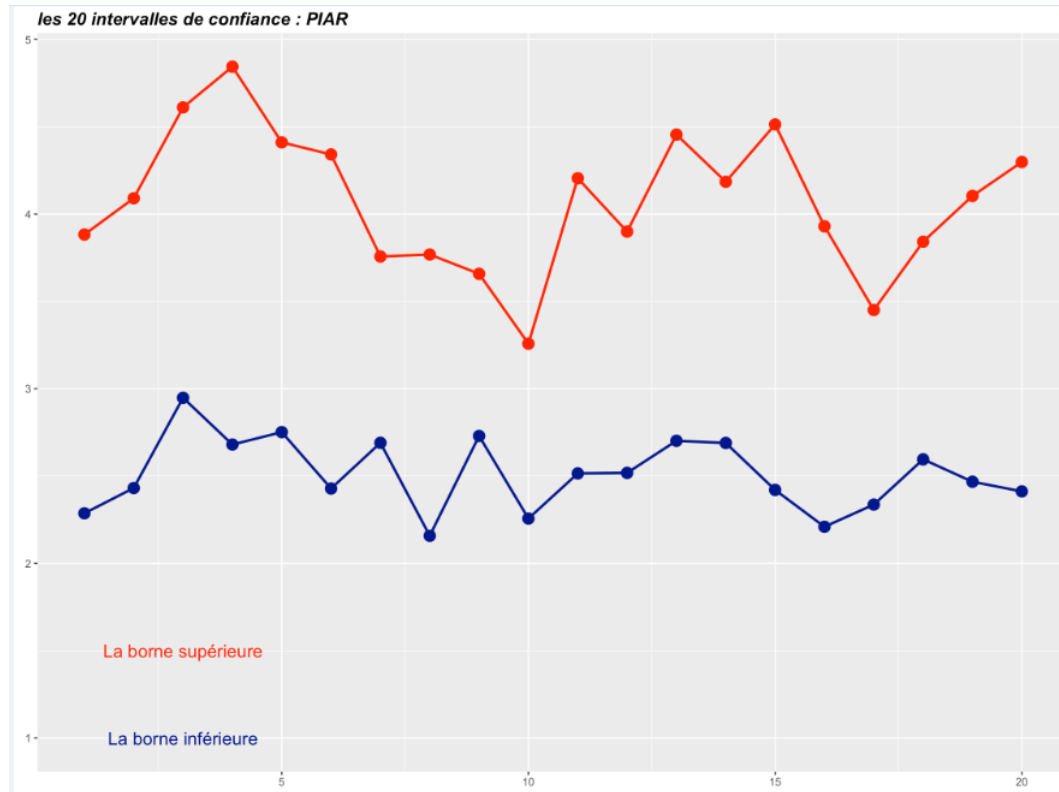
Résultats obtenus

Les résultats obtenus sont regroupés dans ce tableau :

	ybar_chapeau_piar	var_ybar_chapeau_piar	var_estim_ybar_chapeau_piar	IC binf_PiAR	ICbsup_PiAR	long_IC_ PiAR
1	3.084712	0.2445907	0.21310649	2.286545	3.882878	1.5963324
2	3.260910	0.2445907	0.23021884	2.431316	4.090504	1.6591873
3	3.779578	0.2445907	0.23125836	2.948114	4.611043	1.6629290
4	3.762207	0.2445907	0.39153060	2.680330	4.844084	2.1637538
5	3.581045	0.2445907	0.23024704	2.751401	4.410689	1.6592889
6	3.384993	0.2445907	0.30604827	2.428482	4.341504	1.9130220
7	3.223685	0.2445907	0.09519974	2.690211	3.757159	1.0669471
8	2.963110	0.2445907	0.21694205	2.157793	3.768427	1.6106340
9	3.193905	0.2445907	0.07210538	2.729626	3.658184	0.9285575
10	2.757132	0.2445907	0.08381593	2.256569	3.257694	1.0011249
11	3.360177	0.2445907	0.23898360	2.514939	4.205415	1.6904761
12	3.208795	0.2445907	0.15963628	2.517982	3.899608	1.3816269
13	3.578563	0.2445907	0.25710695	2.701862	4.455265	1.7534036
14	3.437108	0.2445907	0.18707931	2.689270	4.184947	1.4956772
15	3.466888	0.2445907	0.36632122	2.420420	4.513357	2.0929364
16	3.069822	0.2445907	0.24760801	2.209467	3.930176	1.7207086
17	2.893623	0.2445907	0.10382998	2.336494	3.450753	1.1142596
18	3.218722	0.2445907	0.12985425	2.595671	3.841772	1.2461005
19	3.285727	0.2445907	0.22405369	2.467316	4.104137	1.6368204
20	3.355213	0.2445907	0.29752727	2.412112	4.298315	1.8862028

==> La moyenne des Ybar_chapeau_piar = 3.293296

Représentation des intervalles de confiance



3.2.2 Plan de sondage sans remise

sélection des échantillons

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 Age de la manière suivante :

- 1- 1. On calcule pk_{ig} et vk_{ig} :

$$pk_{ig} = \frac{x_{ig}}{\sum_{j=1}^5 x_{jg}} \quad \text{et} \quad vk_{ig} = \sum_{j=1}^i pk_{jg}$$

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

Les 20 échantillons tirés sont comme suit :

	ech1	ech2	ech3	ech4	ech5	ech6	ech7	ech8	ech9	ech10	ech11	ech12	ech13	ech14	ech15	ech16	ech17	ech18	ech19	ech20
1	3	4	5	2	5	4	2	1	2	3	1	4	3	2	2	5	3	1	3	2
2	9	10	7	9	8	6	10	8	8	7	9	8	9	6	9	8	9	10	6	8
3	15	12	15	13	12	15	14	15	12	12	15	13	13	15	14	12	12	14	12	13
4	17	20	16	19	17	18	19	18	17	17	17	19	17	19	19	19	19	17	18	19
5	24	24	21	23	25	25	25	25	25	25	24	23	23	24	23	23	25	25	23	25
6	29	27	26	26	30	26	26	29	26	30	26	26	27	26	26	28	26	27	29	27
7	31	31	32	34	31	33	34	31	34	33	34	34	33	31	32	32	34	33	33	35
8	40	36	39	38	39	38	39	40	36	38	40	39	37	36	39	39	36	38	40	39
9	41	45	45	41	42	42	43	43	43	44	42	41	41	43	42	45	43	43	45	45
10	49	48	48	46	49	49	46	49	47	49	49	46	47	46	47	46	49	46	48	47
11	54	55	52	53	53	55	53	54	55	52	52	53	53	53	55	52	55	55	52	53
12	57	57	58	57	56	59	56	56	58	58	56	56	58	57	57	56	57	57	57	57
13	63	65	65	61	62	63	63	63	61	64	62	64	65	65	63	62	61	61	61	63
14	66	69	68	68	68	68	67	66	70	70	70	70	68	68	66	66	70	66	67	68
15	74	75	74	72	73	71	75	71	71	71	74	75	74	74	73	71	71	72	72	71
16	77	77	79	79	77	80	76	77	80	79	78	79	77	80	79	78	80	79	78	79
17	84	81	81	85	83	81	85	82	83	82	84	85	84	85	84	81	83	81	83	83
18	87	86	88	88	88	90	88	87	87	86	89	90	87	86	88	90	88	89	87	90
19	92	92	91	91	93	91	91	93	95	95	93	93	91	91	91	91	91	91	95	91
20	96	96	96	100	96	100	97	96	97	100	99	98	97	96	97	97	96	96	96	98

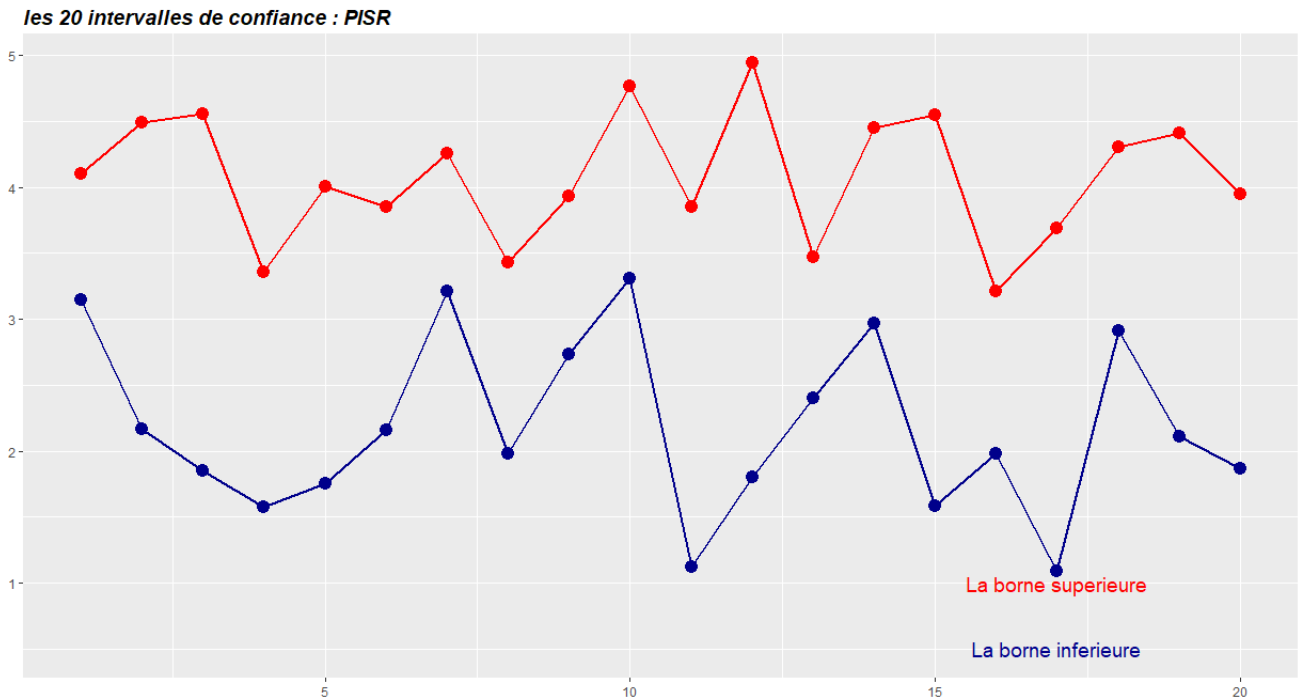
Résultats obtenus

Les résultats obtenus sont regroupés dans ce tableau :

	y_bar_chapeau_pisr	var_est_ybar_chapeau	lcbinf_pisr	ICbsup_pisr	long_IC_pisr
1	3.624700	0.07667074	3.145949	4.103451	0.9575023
2	3.328367	0.44990585	2.168640	4.488094	2.3194542
3	3.202533	0.60954153	1.852649	4.552418	2.6997692
4	2.468100	0.26611385	1.576174	3.360026	1.7838516
5	2.878400	0.42227587	1.754848	4.001952	2.2471037
6	3.006533	0.23996661	2.159559	3.853508	1.6939492
7	3.733333	0.09137556	3.210684	4.255982	1.0452978
8	2.707533	0.17619383	1.981778	3.433289	1.4515110
9	3.333200	0.12163821	2.730182	3.936218	1.2060352
10	4.037233	0.17823979	3.307276	4.767190	1.4599142
11	2.488633	0.62084290	1.126292	3.850974	2.7246822
12	3.373100	0.82625600	1.801463	4.944737	3.1432745
13	2.938767	0.09564428	2.404049	3.473484	1.0694352
14	3.710433	0.18531798	2.966123	4.454743	1.4886197
15	3.066667	0.73330544	1.586068	4.547266	2.9611980
16	2.594267	0.12597515	1.980593	3.207940	1.2273472
17	2.389267	0.56241710	1.092612	3.685921	2.5933089
18	3.609800	0.16137235	2.915240	4.304360	1.3891193
19	3.260767	0.44467387	2.107803	4.413731	2.3059283
20	2.908467	0.36204669	1.868122	3.948811	2.0806895

\Rightarrow La moyenne des $y_bar_{chapeau_pisr} = 3.133005$

Représentation des intervalles de confiance

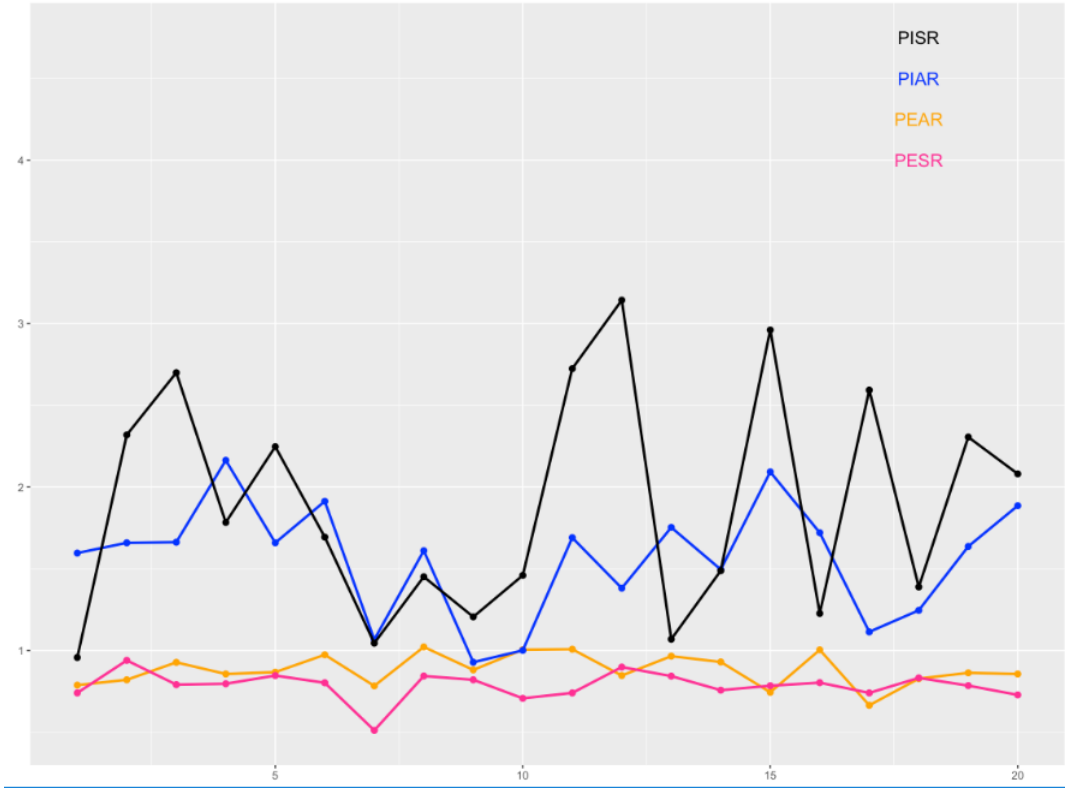


Synthèse

D'après les résultats obtenus après les sondages PESR, PEAR, PIAR et PISR et à partir de ce graphique on peut conclure que le sondage simple sans remise est le plus précis.

On remarque également que le plan simple avec remise est assez précis par rapport aux deux types de sondages à probabilités inégales.

Ceci peut être expliqué par la faible corrélation entre la variable d'intérêt et la variable auxiliaire puisque les probabilités d'inclusion dépendent de cette variable lors du sondage à probabilités inégales.



3.3 Stratification

3.3.1 Principe de 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 souspopulations 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.

3.3.2 Découpage de la population

Le découpage se fait de sorte que les strates soient les plus homogènes possibles par rapport aux variables auxiliaires qui sont fortement corrélées à la variable d'intérêt.

Dans notre cas, et suite à l'observation des graphes croisés de la partie descriptive, il nous a paru que la variable « amis » est la plus corrélée que toute autre variable auxiliaire.

Nous avons, par conséquent, choisi de découper notre population selon cette variable.

Cette variable « amis » correspond dans le questionnaire à la question 6 « Combien de contacts (amis) avez-vous sur les sites de réseaux sociaux : » .

Nous avons eu comme réponses à cette question :

Moins de 100 : 1
Entre 100 et 500 : 33
Entre 500 et 1000 : 36
Plus de 1000 : 30

Par commodité de ne pas travailler avec strates avec un ou deux individus, nous avons distribué les individus ayant comme réponse « Moins de 100 » sur des strates ayant des propriétés proches de point de vue variable d'intérêt « nb_heure ».

Nous avons donc inclus les deux individus ayant comme réponses « Moins de 100 » dans la strate « Entre 500 et 1000 » pour des raisons de ressemblance en nombre d'heures « Entre 0 et 1 heure ». Nous avons désormais, 3 strates définies comme suit :

- * Strate 1 : "Entre 100 et 500" contenant 33 individus
- * Strate 2 : "Entre 500 et 1000" contenant 37 individus
- * Strate 3 : "Plus de 1000" contenant 30 individus

On peut donc visualiser les valeurs de la variable d'intérêt des individus de chaque strate(respectivement strate 1, strate 2, strate 3) :

```
> strates
[[1]]
[1] 1.5 3.5 1.5 2.5 4.5 1.5 4.5 2.5 4.5 2.5 4.5 4.5 3.5 4.5 2.5 2.5 4.5 3.5 4.5 1.5 3.5 4.5 3.5
[24] 2.5 1.5 2.5 2.5 2.5 1.5 4.5 2.5 4.5 2.5

[[2]]
[1] 2.5 1.5 1.5 2.5 4.5 3.5 2.5 4.5 4.5 1.5 3.5 4.5 3.5 4.5 2.5 2.5 3.5 2.5 4.5 1.5 4.5 4.5 3.5
[24] 4.5 4.5 0.5 3.5 3.5 2.5 2.5 1.5 4.5 4.5 4.5 0.5 4.5 1.5

[[3]]
[1] 4.5 1.5 4.5 3.5 4.5 4.5 4.5 3.5 4.5 4.5 4.5 4.5 4.5 4.5 4.5 1.5 2.5 4.5 3.5 4.5 4.5 2.5
[24] 4.5 4.5 1.5 4.5 4.5 3.5 3.5
```

3.3.3 Propriétés de chaque strate

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

	Strate.1	Strate.2	Strate.3
Moyenne	3.136364	3.175676	3.9000000
Variance	1.288125	1.652973	0.9968276
Variance corrigée	1.301136	1.669670	1.0068966

Avant de procéder à chaque type de stratification, nous avons calculé la variance inter strates et la variance intra strates de la variable d'intérêt (y).

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

3.3.4 Réalisation des sondages et traitement des résultats

Les strates ont la même taille

- Sélection de l'échantillon

Pour choisir le nombre d'individus à tirer de chaque strate, nous calculons l'arrondi de la division du nombre d'individus à tirer dans l'échantillon par le nombre des strates (3).

Nous avons donc prélevé aléatoirement et sans remise, exceptionnellement, 21 individus, 7 individus de chaque strate.

On a refait l'expérience 20 fois pour obtenir 20 échantillons.

- Résultats obtenus

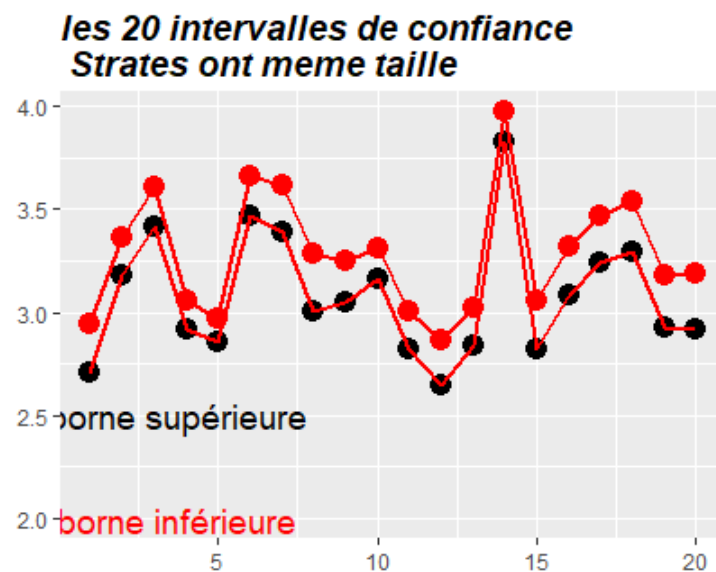
Nous avons calculé la moyenne estimée (\bar{y}_{hat}), la variance ainsi que la variance estimée de \bar{y}_{hat} correspondant à chaque échantillon. Nous avons par suite utilisé cette variance estimée pour le calcul de l'intervalle de confiance.

Tous ces résultats ont été regroupés dans le même tableau :

	ybar_hat	var_ybar_hat_strat	var_estim_ybar_hat_strat	ic_inf_strat	ic_sup_strat	long_ic_strat
1	2.827143	0.05394418	0.06074063	2.708091	2.946194	0.2381033
2	3.271429	0.05394418	0.04710798	3.179097	3.363760	0.1846633
3	3.514286	0.05394418	0.04892393	3.418395	3.610177	0.1917818
4	2.990000	0.05394418	0.03648089	2.918497	3.061503	0.1430051
5	2.911429	0.05394418	0.02759926	2.857334	2.965523	0.1081891
6	3.564286	0.05394418	0.05032505	3.465649	3.662923	0.1972742
7	3.502857	0.05394418	0.05958219	3.386076	3.619638	0.2335622
8	3.147143	0.05394418	0.07120260	3.007586	3.286700	0.2791142
9	3.148571	0.05394418	0.05114768	3.048322	3.248821	0.2004989
10	3.238571	0.05394418	0.03871391	3.162692	3.314451	0.1517585
11	2.915714	0.05394418	0.04758108	2.822455	3.008973	0.1865178
12	2.755714	0.05394418	0.05709647	2.643805	2.867623	0.2238182
13	2.927143	0.05394418	0.04684119	2.835334	3.018952	0.1836175
14	3.901429	0.05394418	0.03715473	3.828605	3.974252	0.1456465
15	2.938571	0.05394418	0.06028237	2.820418	3.056725	0.2363069
16	3.204286	0.05394418	0.05956067	3.087547	3.321025	0.2334778
17	3.352857	0.05394418	0.05646679	3.242182	3.463532	0.2213498
18	3.418571	0.05394418	0.06287421	3.295338	3.541805	0.2464669
19	3.054286	0.05394418	0.06524638	2.926403	3.182169	0.2557658
20	3.051429	0.05394418	0.06985306	2.914517	3.188341	0.2738240

==> La moyenne des $\bar{Y}_{\text{strat_estimé}}$ est : 3.181786

- Représentation des intervalles de confiance



Les strates sont à allocation proportionnelles

- Sélection de l'échantillon

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) :

$n_h/N_h = n/N = f$ Par suite : $n_1 = 7$; $n_2 = 7$; $n_3 = 6$

Nous avons donc tiré de chaque strate le nombre d'individus correspondant suivant un tirage aléatoire sans remise.

On a refait l'expérience 20 fois pour obtenir 20 échantillons.

- Résultats obtenus

Nous avons calculé la moyenne estimée (\bar{y}_{hat}), la variance ainsi que la variance estimée de \bar{y}_{hat} pour chaque échantillon tiré. Nous avons par suite utilisé cette variance est. Tous ces résultats sont regroupés dans le tableau :

	↑ ybar_hat ↓	var_ybar_hat_prop ↓	var_estim_ybar_hat_strat ↓	ic_inf_prop ↓	ic_sup_prop ↓	long_ic_prop ↓
1	3.877857	0.05394658	0.02729417	3.830666	3.925049	0.09438324
2	2.643571	0.05394658	0.05458001	2.549203	2.737940	0.18873767
3	3.416667	0.05394658	0.05286726	3.325259	3.508074	0.18281500
4	3.011429	0.05394658	0.04746815	2.929356	3.093501	0.16414487
5	3.440714	0.05394658	0.06204086	3.333446	3.547983	0.21453730
6	3.011429	0.05394658	0.07943921	2.874078	3.148779	0.27470078
7	2.780000	0.05394658	0.06096945	2.674584	2.885416	0.21083236
8	3.714048	0.05394658	0.02307739	3.674147	3.753948	0.07980162
9	3.358571	0.05394658	0.04491058	3.280921	3.436222	0.15530077
10	3.072381	0.05394658	0.05771278	2.972596	3.172166	0.19957080
11	3.522381	0.05394658	0.05418916	3.428688	3.616074	0.18738610
12	3.366905	0.05394658	0.05256941	3.276012	3.457797	0.18178503
13	3.210952	0.05394658	0.05717221	3.112102	3.309803	0.19770151
14	3.202619	0.05394658	0.05708055	3.103927	3.301311	0.19738453
15	3.077381	0.05394658	0.06774103	2.960257	3.194505	0.23424847
16	2.942619	0.05394658	0.07529298	2.812437	3.072801	0.26036311
17	3.302143	0.05394658	0.06129733	3.196160	3.408126	0.21196615
18	3.223810	0.05394658	0.05361229	3.131114	3.316505	0.18539129
19	3.540238	0.05394658	0.04745426	3.458190	3.622287	0.16409684
20	3.479286	0.05394658	0.05181625	3.389695	3.568876	0.17918059

==> La moyenne des $\bar{Y}_{bar_strat_estimé}$ est : 3.25975

- Représentation des intervalles de confiance

Les strates sont à allocation optimale

- Sélection de l'échantillon

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.

On détermine alors n_h suivant la formule suivante :

$$\frac{n_h}{n} = \frac{N_h \sqrt{\sigma_{hc}^2}}{\sum_{h=1}^M N_h \sqrt{\sigma_{hc}^2}}$$

Nous avons obtenu : $n_1=6$; $n_2=8$; $n_3=6$

Nous avons donc tiré, 20 fois, de chaque strate le nombre d'individus correspondant suivant un tirage aléatoire sans remise.

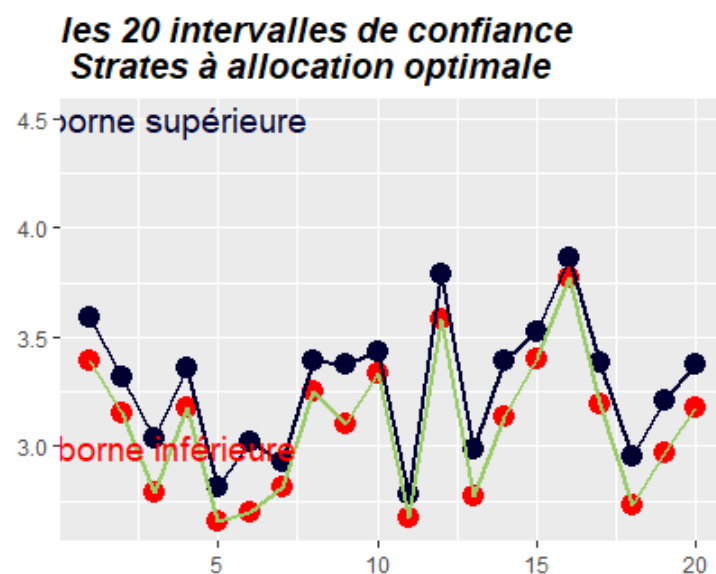
- Résultats obtenus

Nous avons calculé la moyenne estimée (\bar{y}_{hat}), la variance ainsi que la variance estimée de \bar{y}_{hat} pour chacune des 20 chantillons tirés. Nous avons par suite utilisé cette variance. Voici le tableau regroupant tous ces résultats :

	\bar{y}_{hat}	$\text{var}_{\bar{y}_{\text{hat}} \text{ opt}}$	$\text{var}_{\text{estim}} \bar{y}_{\text{hat}} \text{ opt}$	$\text{ic}_{\text{inf}} \text{ strat}$	$\text{ic}_{\text{sup}} \text{ strat}$	long_{ic}
1	3.490000	0.05394658	0.05781682	3.390035	3.589965	0.19993056
2	3.235000	0.05394658	0.04941000	3.149570	3.320430	0.17085978
3	2.911667	0.05394658	0.07223318	2.786775	3.036558	0.24978234
4	3.266667	0.05394658	0.05124273	3.178068	3.355265	0.17719735
5	2.736667	0.05394658	0.04672227	2.655884	2.817449	0.16156562
6	2.858333	0.05394658	0.09360409	2.696492	3.020175	0.32368295
7	2.871667	0.05394658	0.03249818	2.815477	2.927856	0.11237871
8	3.320000	0.05394658	0.04041000	3.250131	3.389869	0.13973778
9	3.240000	0.05394658	0.07830818	3.104605	3.375395	0.27078969
10	3.385000	0.05394658	0.02798182	3.336619	3.433381	0.09676113
11	2.730000	0.05394658	0.03134455	2.675805	2.784195	0.10838944
12	3.685000	0.05394658	0.05779636	3.585070	3.784930	0.19985983
13	2.880000	0.05394658	0.06136773	2.773895	2.986105	0.21220960
14	3.263333	0.05394658	0.07294091	3.137219	3.389448	0.25222966
15	3.461667	0.05394658	0.03767318	3.396530	3.526804	0.13027386
16	3.818333	0.05394658	0.02613682	3.773143	3.863524	0.09038112
17	3.288333	0.05394658	0.05418000	3.194656	3.382011	0.18735444
18	2.845000	0.05394658	0.06564273	2.731504	2.958496	0.22699255
19	3.090000	0.05394658	0.06858000	2.971425	3.208575	0.23714964
20	3.280000	0.05394658	0.05752227	3.180544	3.379456	0.19891202

==> La moyenne des $\bar{Y}_{\text{bar_strat_estimé}}$ est : 3.182833

- Représentation des intervalles de confiance



3.4 Sondage par grappes et à deux degrés

3.4.1 Plan de sondage par grappes

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.

Sélection des échantillons

Nous avons partitionné la population selon la variable auxiliaire «Nombre d'amis» qui présente 4 modalités :

- Grappe 1 : «Moins de 100»
- Grappe 2 : «Entre 100 et 500»
- Grappe 3 : «Entre 500 et 1000»
- Grappe 4 : «Plus de 1000»

	grappe1	grappe2	grappe3	grappe4
moyenne	1	2	3	4
taille de grappe	1	33	36	30

On sélectionne un échantillon de m grappes avec m=2 parmi M=4 grappes.

Résultats obtenus :

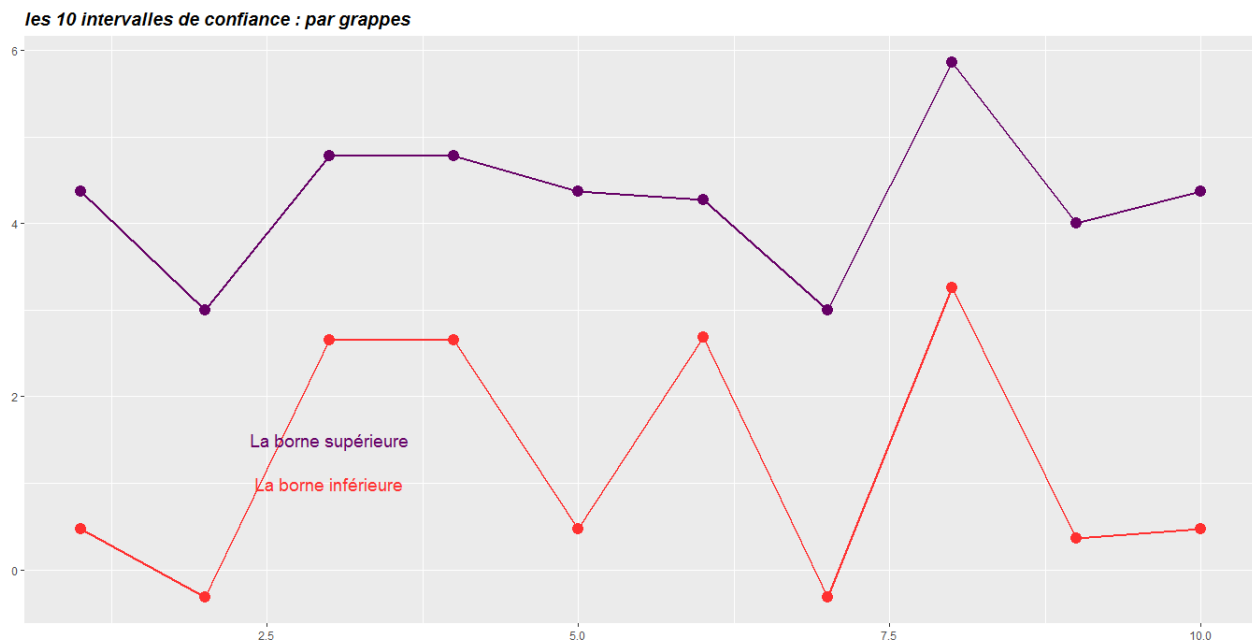
Le tableau récapitulatif des 10 \bar{Y}_{grap} , var_{grap} , les numéros des grappes sélectionnées ainsi que les intervalles de confiance et leurs longueurs.

	$\bar{y}_{grapp\hat{e}}$	$var_{ygrap\hat{e}}$	$icinf_{grapp\hat{e}}$	$icsup_{grapp\hat{e}}$	$longIC_{grap\hat{e}}$	$num_1ere_grap_selct\hat{e}$	$num_2ere_grap_selct\hat{e}$
1	2.42	0.9908833	0.4689548	4.371045	3.902090	4	1
2	1.34	0.7136833	-0.3158037	2.995804	3.311607	1	2
3	3.72	0.2932167	2.6586701	4.781330	2.122660	4	2
4	3.72	0.2932167	2.6586701	4.781330	2.122660	4	2
5	2.42	0.9908833	0.4689548	4.371045	3.902090	1	4
6	3.48	0.1644167	2.6852528	4.274747	1.589494	2	3
7	1.34	0.7136833	-0.3158037	2.995804	3.311607	2	1
8	4.56	0.4416167	3.2574968	5.862503	2.605006	4	3
9	2.18	0.8620833	0.3601705	3.999829	3.639659	3	1
10	2.42	0.9908833	0.4689548	4.371045	3.902090	4	1

\Rightarrow La moyenne des $\bar{Y}_{grap}estim = 2.76$

Représentation des intervalles de confiance

Les 10 intervalles de confiances pour le sondage par grappes :



3.4.2 Plan de sondage 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.

Sélection des échantillons :

Nous avons sélectionné un échantillon de m unités primaires avec $m=2$. Dans chaque unité primaire h nous avons sélectionné un échantillon de taille n_h $N_h/5$.

Résultats obtenus :

. Le tableau récapitulatif des 10 $\bar{Y}_{2\text{degré}}$, $\text{var}_{2\text{degré}}$, les numéros des grappes sélectionnées, les nombres des individus sélectionnés dans chaque grappe ainsi que les intervalles de confiance et leurs longueurs est comme suit :

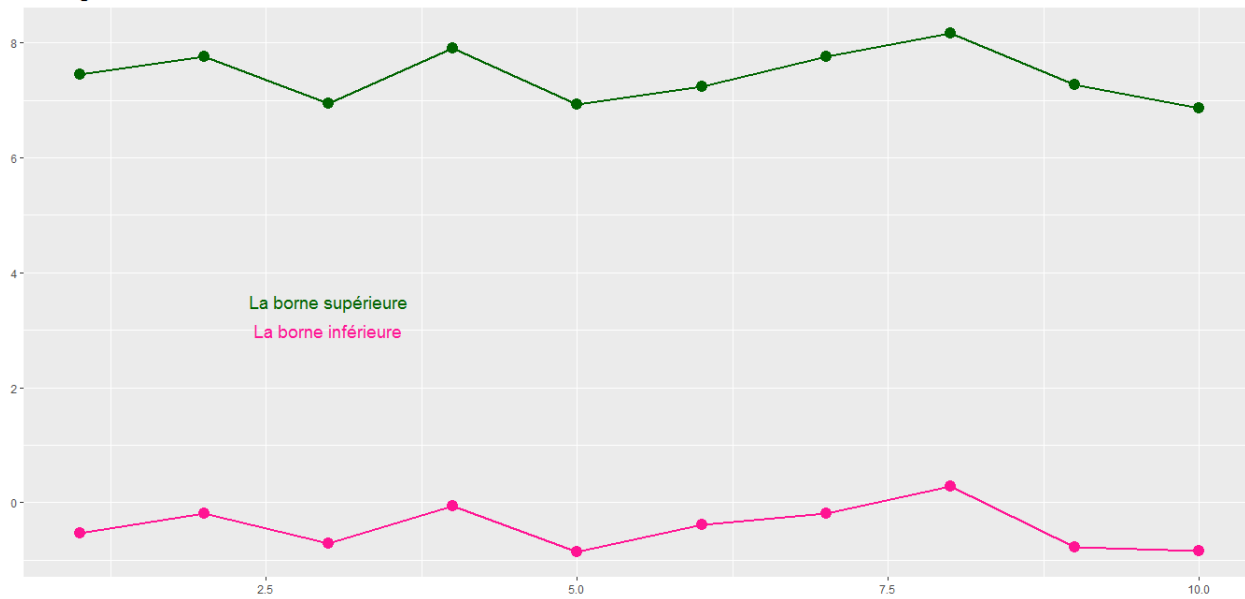
	$\bar{y}_{\text{grappe}2}$	$\text{var}_{\text{y}_{\text{grap}}2\text{dg}}$	$\text{icinf}_{2\text{dg}}$	$\text{icsup}_{2\text{dg}}$	$\text{longIC}_{2\text{dg}}$	$\text{num}_{1\text{ere_grap_selc}}_t$	$\text{num}_{2\text{ere_grap_selc}}$	$\text{nbre_ind_select_G1}$	n
1	3.462000	4.141392	-0.52668029	7.450680	7.977361	4	2	5	
2	3.790000	4.113492	-0.18522211	7.765222	7.950444	1	4	6	
3	3.116000	3.820217	-0.71489366	6.946894	7.661787	2	3	5	
4	3.924667	4.140911	-0.06378235	7.913116	7.976898	2	1	5	
5	3.036667	3.933184	-0.85045566	6.923789	7.774245	1	3	6	
6	3.426667	3.788117	-0.38809831	7.241432	7.629530	1	3	6	
7	3.783333	4.113492	-0.19188877	7.758555	7.950444	4	1	5	
8	4.222000	4.038511	0.28317458	8.160825	7.877651	2	1	5	
9	3.250000	4.213925	-0.77345803	7.273458	8.046916	4	2	5	
10	3.012000	3.857284	-0.83743392	6.861434	7.698868	2	3	5	

\Rightarrow La moyenne des $\bar{Y}_{2\text{deg_estimé}} = 2.158$

Représentation des intervalles de confiance

Les 10 intervalles de confiances pour le sondage à deux degrés :

les 10 intervalles de confiance
à 2 degré

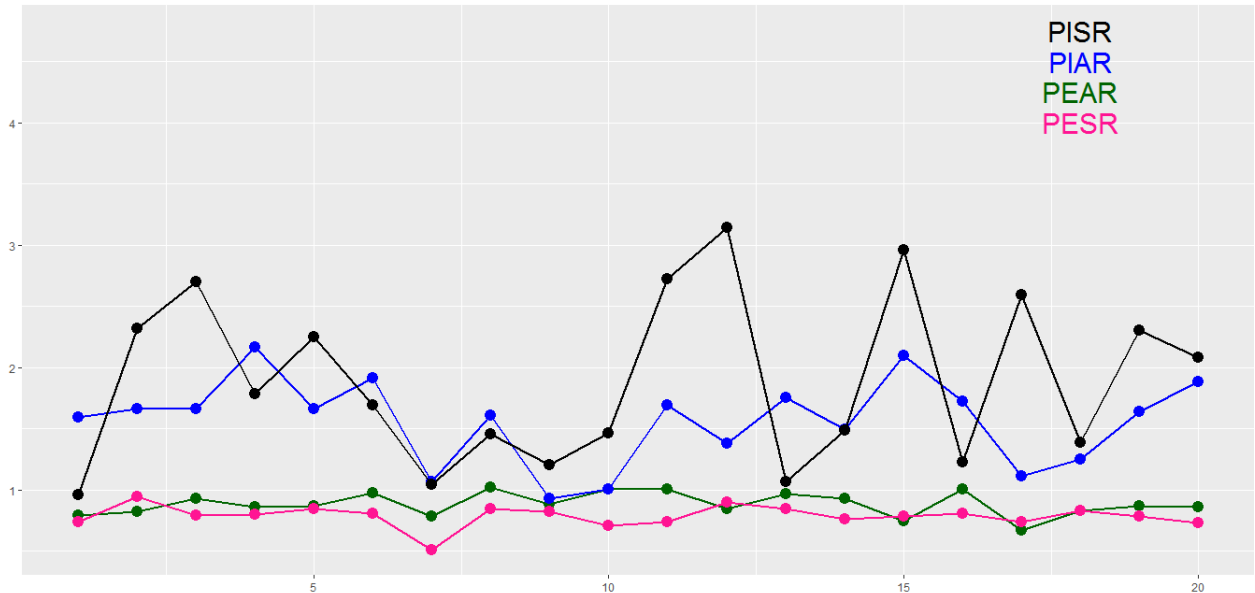


Partie 4

Synthèse

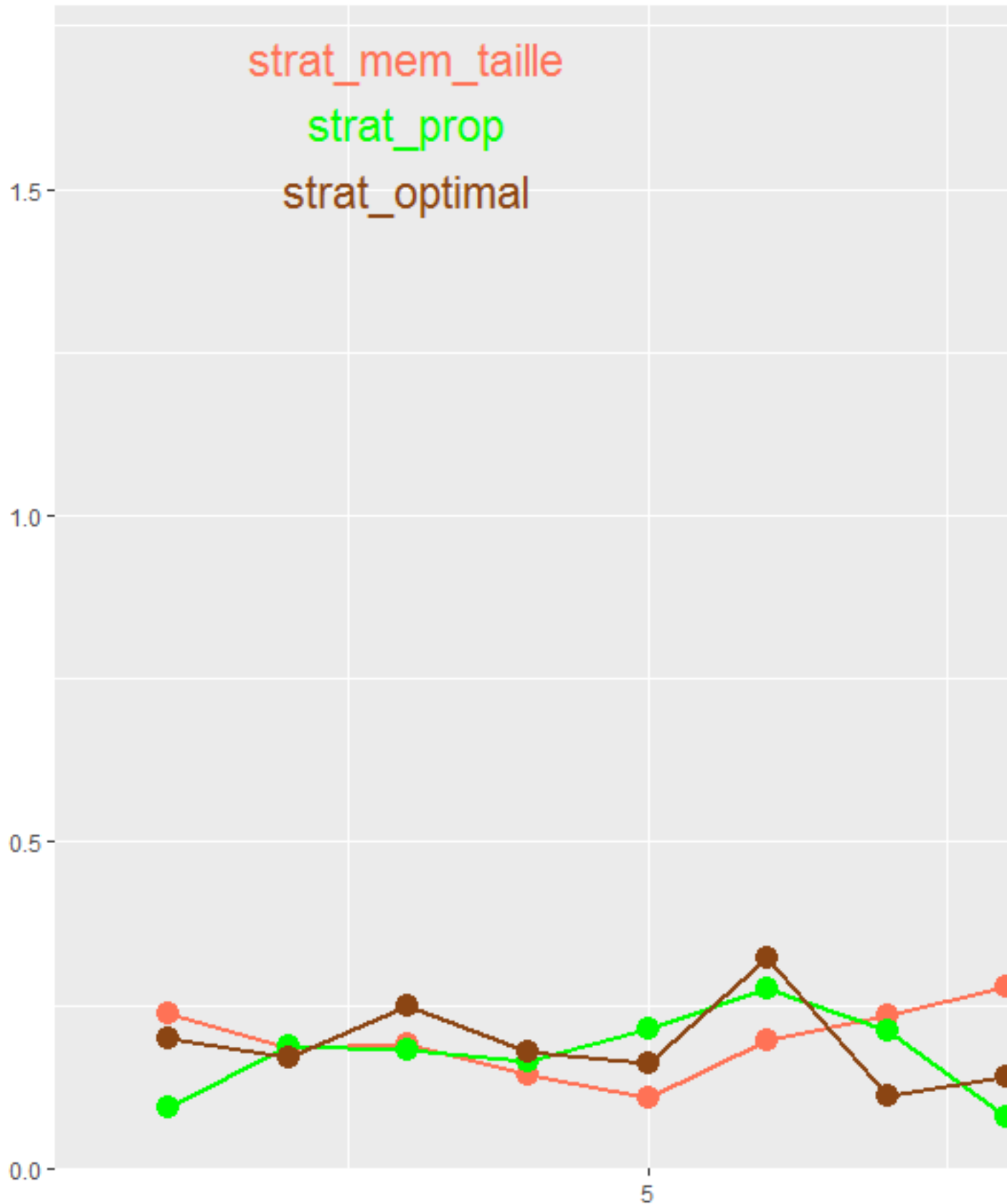
4.1 Sondages aléatoires PESR, PEAR, PISR et PIAR

Synthèse des logueurs des intervalles de confiance
PEAR PESR PIAR PISR



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.

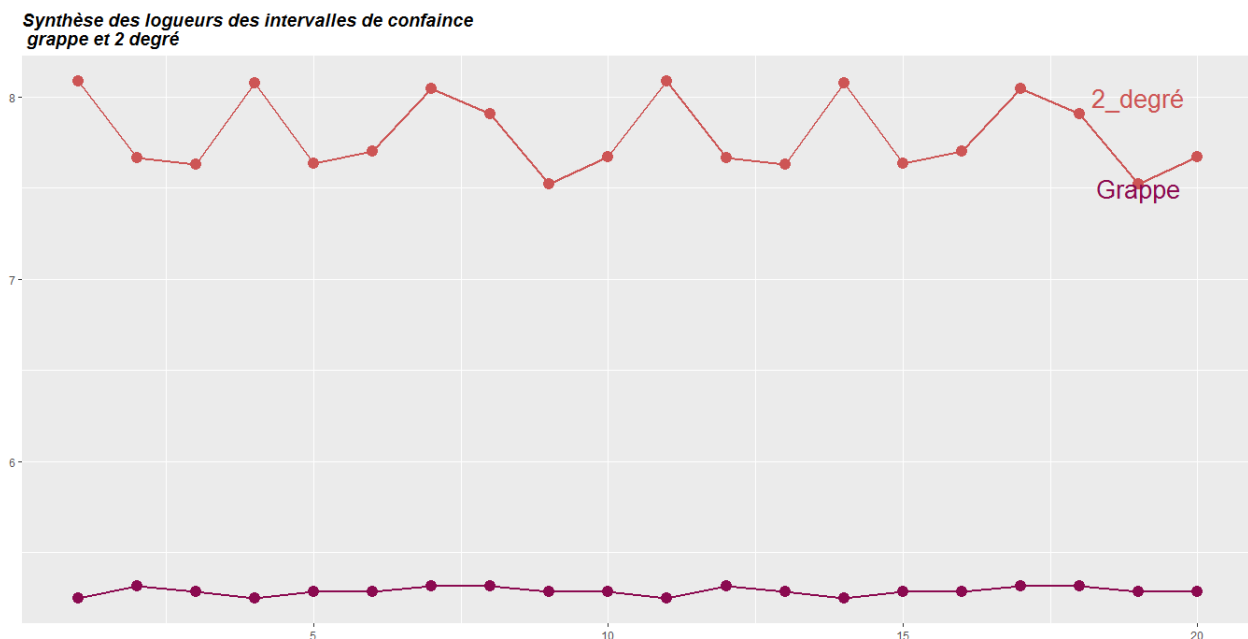
4.2 Stratification

Synthèse des logueurs des intervalles de confiance
strat_egal prop opt

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.

4.3 Sondage par grappes et à deux degrés

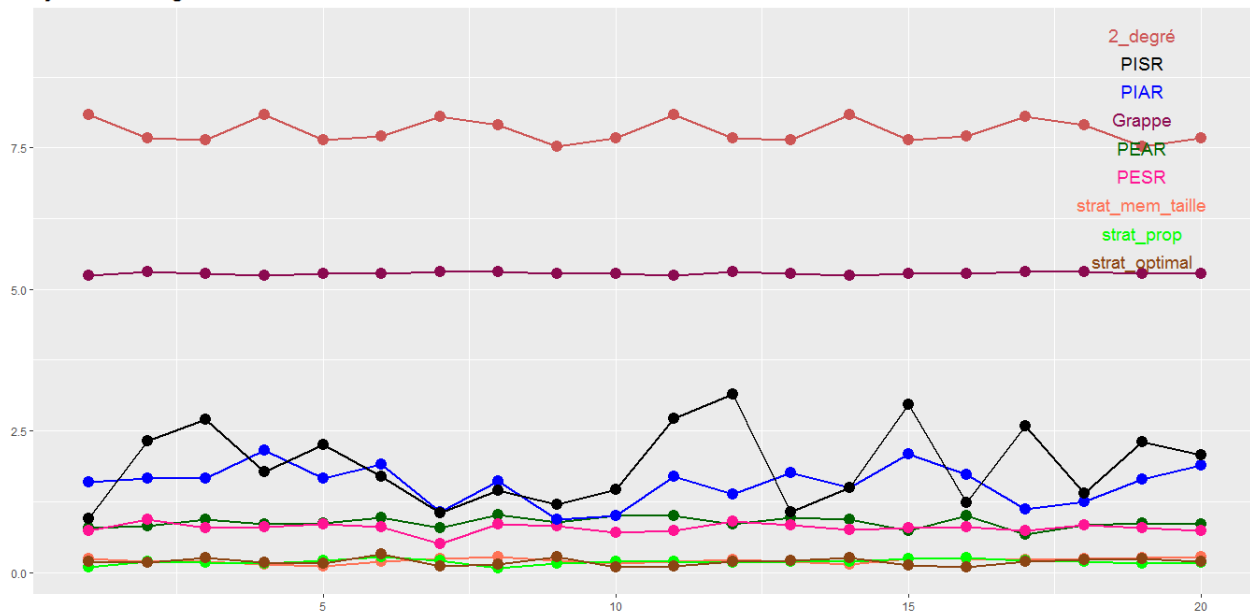


On remarque que le sondage par grappes est beaucoup plus précis que celui à deux degrés (grande différence entre les longueurs des intervalles de confiance). En effet, dans le deuxième on ne sélectionne qu'un échantillon de chaque grappe tirée. Ceci explique la perte en précision qui est, bien entendu, compensée par un gain en termes de coût.

4.4 Comparaison de tous les types de sondages utilisés

sondage	PEAR	PIAR	PISR	PESR	Strat_optima	Strat_prop	Stra_m_tail	Grappe	2degres
Moyenne des longueurs des IC	0.88	1.563	1.892	0.786	0.18732	0.188216	0.20673	5.284504	7.794783

Synthèse des logueurs des intervalles de confiance



D'après la dernière figure et le tableau comparatif des moyennes des longueurs des intervalles de confiance, on remarque que le sondage à deux degrés est le moins précis. Il est fréquemment utilisé pour une raison de coût mais pas nécessairement de précision.

On remarque, par ailleurs, que les sondages PESR, PEAR, PISR et PIAR sont plus précis que celui à deux degrés mais moins précis que les stratifiés.

Quant aux 3 types de sondage stratifié, sont les meilleurs en termes de précision (ils correspondent aux longueurs minimales d'intervalles de confiance) car ils minimisent la variance pour augmenter la précision.

Partie 5

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' Horvitz-Thompson 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] 1.5 3.5 4.5 4.5 4.5 4.5 4.5 4.5 4.5 4.5 0.5 4.5 1.5 1.5 3.5 2.5 4.5
[18] 4.5 4.5 1.5 . . . . .
```

La moyenne de l'échantillon avant redressement est 3.5

La proportion des hommes dans la population est égale à 0.34 alors que dans l'échantillon est 0.6

Donc le poids des femmes dans notre échantillon est $0.34/0.6 = 0.5666667$

L'estimateur de la moyenne des hommes après redressement est la moyenne de l'échantillon multiplié par le poids des hommes est égale à 3.85

Code R

```
#####
#Importer les packages necessaires#
#####

library(ggplot2)
library(scales)
library(xlsx)

#####
#Importer et preparer notre base de donnees#
#####
set.seed(5)
data_100=read.table(file=file.choose(),header=TRUE,sep=";",dec="." )
attach(data_100)
names(data_100) = c('etudiant','sexe','situation','membre','plateforme',
                    , 'amis','appareil','nb_heures','temps','inter t')

y = matrix(100,1)

for(i in 1:100){
  if (data_100[i,8] == "Entre_0_et_1_heure")
    y[i]=0.5
  if (data_100[i,8] == "Entre_1_et_2_heures" )
    y[i]=1.5
  if (data_100[i,8] == "Entre_2_et_3_heures" )
    y[i]=2.5
  if (data_100[i,8] == "Entre_3_et_4_heures" )
    y[i]=3.5
  if (data_100[i,8] == "Plus_de_4_heures" )
    y[i]=4.5
}

a = matrix(100,1)

for(i in 1:100){
  if (data_100[i,6] == "Moins_de_100")
    a[i] = 50
  if (data_100[i,6] == "Entre_100_et_500")
    a[i] = 300
  if (data_100[i,6] == "Entre_500_et_1000")
    a[i] = 750
}
```

```

    if (data_100[i,6] == "Plus_de_1000")
      a[i] = 1250
}

data_100 = cbind(data_100,y,a)

#####
#Calculer les pourcentages de la rpartition selon la variable d'
  inter t#
#####

ylabels = c(sum(y == 4.5),sum(y == 3.5),sum(y == 2.5),sum(y == 1.5),sum
  (y == 0.5))
ylabels = sprintf("%d%s",ylabels,"%")

#####
#R partition selon la Vairable d'inter t#
#####

py = ggplot(data_100,aes(x="",fill=factor(nb_heures)))+coord_polar(
  theta="y")+ geom_bar(width = 1) +labs(x="",y="",fill="nb_heures") +
  scale_fill_brewer("",palette="Set2")+ggtitle("La_rpartition_de_la_
  population_selon_la_variable_d'interet:_\n_Nombre_des_heures") +
  annotate("text",label= ylabels,x=c(1,1,1,1,1.44),y=c(25,55,75,89,99)
  ,size=3)

py

mean(y)
#####
#R partition des tudians #
#####

elabels = c(sum(etudiant == "Oui"),sum(etudiant == "Non"))
elabels = sprintf("%d%s",elabels,"%")

pe = ggplot(data_100,aes(x="",fill=factor(etudiant)))+coord_polar(theta
  ="y")+ geom_bar(width = 1) +labs(x="",y="",fill="etudiant") +scale_
  fill_brewer("",palette="Set2")+ggtitle("La_rpartition_de_la_
  population_selon_la_variable_d'interet:_\n_Etudiant") +annotate("
  text",label= elabels,x=c(1,1.3),y=c(50,96),size=6)

pe

be = ggplot(data_100,aes(x=factor(etudiant),fill=factor(nb_heures)))+
  geom_bar(width = 1) +labs(x="",y="",fill="Nombre_d'heures")
be

#####
#R partiton selon le Sexe#
#####

```



```

slabels = c(sum(sexe == "Homme"),sum(sexe == "Femme"))
slabels = sprintf("%d%s",slabels,"%")

ps = ggplot(data_100,aes(x="",fill=factor(sexe)))+coord_polar(theta="y")
  + geom_bar(width = 1) +labs(x="",y="",fill="Sexe") +scale_fill_
  brewer("",palette="Set2")+ggtitle("La répartition de la population
  selon le : \n Sexe") +annotate("text",label= slabels,x=c(1,1),y=c(20
  ,70),size=6)

ps

bs = ggplot(data_100,aes(x=factor(sexe),fill=factor(nb_heures)))+geom_
  bar(width = 1) +labs(x="",y="",fill="Nombre d'heures")
bs

#####
#R partition selon l' tat civil#
#####

silabels = c(sum(situation == "Un_coeur_a_prendre"),sum(situation == "
  Marie(e)"),sum(situation == "En_couple"))
silabels = sprintf("%d%s",silabels,"%")

psi = ggplot(data_100,aes(x="",fill=factor(situation)))+coord_polar(
  theta="y")+ geom_bar(width = 1) +labs(x="",y="",fill="Etat_Civil")
  +scale_fill_brewer("",palette="Set2")+ggtitle("La répartition de la
  population selon : \n l'Etat_Civil") +annotate("text",label=
  silabels,x=c(1,1.4,1),y=c(40,70,85),size=3)

psi

bsi = ggplot(data_100,aes(x=factor(situation),fill=factor(nb_heures)))
  +geom_bar(width = 1) +labs(x="",y="",fill="Nombre d'heures")
bsi

#####
#R partion selon la plateforme#
#####

plabels = c(sum(situation == "youtube"),sum(situation == "Twitter"),sum
  (situation == "Linkedin"),sum(plateforme == "Instagram_/Snapchat"),
  sum(plateforme == "Facebook"))
plabels = sprintf("%d%s",plabels,"%")

pp = ggplot(data_100,aes(x="",fill=factor(plateforme)))+coord_polar(
  theta="y")+ geom_bar(width = 1) +labs(x="",y="",fill="Plateforme")
  +scale_fill_brewer("",palette="Set2")+ggtitle("La répartition de la
  population selon : \n la Plateforme") +annotate("text",label=
  plabels,x=c(1.4,1.3,1.4,1,1),y=c(1.5,2.5,3.5,10,60),size=3)

pp

```

```

bp = ggplot(data_100,aes(x=factor(plateforme),fill=factor(nb_heures)))
    +geom_bar(width = 1)  +labs(x="",y="",fill="Nombre_d'heures")
bp

#####
#R partition selon les contacts#
#####

clabels = c(sum(amis == "Plus_de_1000"),sum(amis == "Moins_de_100"),sum
    (amis == "Entre_500_et_1000"),sum(amis == "Entre_100_et_500"))
clabels = sprintf("%d%s",clabels,"%")

pc = ggplot(data_100,aes(x="",fill=factor(amis)))+coord_polar(theta="y"
    )+ geom_bar(width = 1)  +labs(x="",y="",fill="Les_amis") +scale_fill
    _brewer("",palette="Set2")+ggtitle("La_repartition_de_la_population_
    selon:_\n_amis") +annotate("text",label= clabels,x=c(1,1.4,1,1),y=c
    (15,30.5,50,85),size=3)
pc

bc = ggplot(data_100,aes(x=factor(amis),fill=factor(nb_heures)))+geom_
    bar(width = 1)  +labs(x="",y="",fill="Nombre_d'heures")
bc

#####
#R partition selon l'appareil#
#####

aplabels = c(sum(appareil == "Smartphone"),sum(appareil == "Ordinateur_
    portable_/_tablette"),sum(appareil == "Ordinateur_bureautique"))
aplabels = sprintf("%d%s",aplabels,"%")

pap = ggplot(data_100,aes(x="",fill=factor(appareil)))+coord_polar(
    theta="y")+ geom_bar(width = 1)  +labs(x="",y="",fill="Les_appareils
    ") +scale_fill_brewer("",palette="Set2")+ggtitle("La_repartition_de_
    la_population_selon:_\n_l'appareil") +annotate("text",label=
    aplabels,x=c(1,1,1.4),y=c(40,90,99.9),size=3)
pap

bap = ggplot(data_100,aes(x=factor(appareil),fill=factor(nb_heures)))+
    geom_bar(width = 1)  +labs(x="",y="",fill="Nombre_d'heures")
bap

#####
#R partition selon le temps #
#####

tlabels = c(sum(temps == "Le_matin_avant_8_00"),sum(temps == "De_8_00
    _ _17_00"),sum(temps == "A_partir_de_17_00"))
tlabels = sprintf("%d%s",tlabels,"%")

```

```

pt = ggplot(data_100,aes(x="",fill=factor(temps)))+coord_polar(theta="y")
+ geom_bar(width = 1) +labs(x="",y="",fill="Le_temps") +scale_
fill_brewer("",palette="Set2")+ggtitle("La_repartition_de_la_
population_selon:_\n_le_temps") +annotate("text",label= tlabels,x=c
(1.3,1,1),y=c(1.5,9,65),size=3)
pt

bt = ggplot(data_100,aes(x=factor(temps),fill=factor(nb_heures)))+geom
_bar(width = 1) +labs(x="",y="",fill="Nombre_d'heures")
bt

#####
#R partition selon l'inter t#
#####

ilabels = c(sum(inter t == "Voyages_/mode"),sum(inter t == "Tous"),
sum(inter t == "Television_/films"),sum(inter t == "Sante_/sport
"),sum(inter t == "Nourriture"),sum(inter t == "Art_/musique"),
sum(inter t == "Actualite"))
ilabels = sprintf("%d%s",ilabels,"%")

pi = ggplot(data_100,aes(x="",fill=factor(inter t)))+coord_polar(theta
="y")+ geom_bar(width = 1) +labs(x="",y="",fill="L'es_amis'inter t
") +scale_fill_brewer("",palette="Set2")+ggtitle("La_repartition_de_
la_population_selon:_\n_l'inter t") +annotate("text",label=
ilabels,x=c(1,1.4,1,1,1.3,1,1.4),y=c(13,26,40,55,68,83,98),size=3)
pi

bi = ggplot(data_100,aes(x=factor(inter t),fill=factor(nb_heures)))+
geom_bar(width = 1) +labs(x="",y="",fill="Nombre_d'heures")
bi

#####
#SAS PESR et PEAR#
#####

set.seed(5)

##### PESR

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 degre de liberte=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 inferieure","La borne
  superieure"), 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_ybar_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 degre de liberte=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],ICbinf,ICbsup,
  longueur)
p1_pear=ggplot(res_pear,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_pear=annotate("text", label =c("La borne inferieure","La borne
  superieure"), 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_ybar_chapeau_pear","ICbinf_PEAR","ICbsup_PEAR","long_IC_PEAR"
  )
View(res_pear)
rownames(res_pear)=c(1:20)
mean(res_pear$ybar_chapeau_pear)

##### proba inegales

x=matrix(100,1)

for (i in 1:100) {x[i] = a[i]}
View(x)
data_100=cbind(data,100)
View(data_100)

summary(data_100)
library(sampling)

pic = inclusionprobabilities(x,20)#calculer les probabilites 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(j,vk[j-1],u,vk[j]))
  }
}
}

#####PIAR
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
ICbinf=1:20
ICbsup=1:20
longueur=1:20
for (i in 1:20) {
  ICbinf[i]=res1[21,i]-1.729*sqrt(res1[23,i])
  ICbsup[i]=res1[21,i]+1.729*sqrt(res1[23,i])
  mu[i]=res1[21,i]
  longueur[i]=ICbsup[i]-ICbinf[i]}
longueur
mean(longueur)
mu1=mean(mu)
View(res1)
echan=res1[1:20,]
colnames(echan)=c("ech1","ech2","ech3","ech4","ech5","ech6","ech7","ech
8","ech9","ech10","ech11□","ech12","ech13","ech14","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],ICbinf,ICbsup,
  longueur)
g1_piar=ggplot(ress,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)
g2_piar=annotate("text", label =c("La_borne_inferieure","La_borne_
  superieure"), 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","IC_binf_PiAR","ICbsup_PiAR","long_IC_
  """"""PIAR")
View(ress)

##### pisr
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 estimee
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 repete 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 inferieure","La borne superieure"),
  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","Icbinf_pisr",
  "","ICbsup_pisr","long_IC_
  """"""""""""""""""""pisr")
View(tt)

#####
#####Straification#####
#####

#####
#Definition des strates#
#####

N=100
n=20
h=3
Nh=as.vector(table(data_100$amis))
Nh
strate1=rep(0,Nh[1]) #Entre 100 et 500
j=1
for (i in 1:100) {
  if(data_100$amis[i]=="Entre_100_et_500") {strate1[j]=y[i]; j=j+1;}
}
strate1
strate2=rep(0,Nh[2]) #Entre 500 et 1000
j=1
for (i in 1:100) {
  if(data_100$amis[i]=="Entre_500_et_1000" || data_100$amis[i]=="Moins_
    de_100") {strate2[j]=y[i]; j=j+1;}
}
strate2
strate3=rep(0,Nh[3]) #Plus de 1000
j=1
for (i in 1:100) {
  if(data_100$amis[i]=="Plus_de_1000") {strate3[j]=y[i]; j=j+1;}
}
strate3

strates=list(strate1,strate2,strate3)
strates

#####
#Propri t s de chaque strate#
#####

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]}
sigma2_hc
sigma2_h
ybar_strates
setNames(sigma2_hc,sigma2_h,ybar_strates)
t=data.frame("Strate_1" = c(ybar_strates[1],sigma2_h[1],sigma2_hc[1]),"
  Strate_2" = c(ybar_strates[2],sigma2_h[2],sigma2_hc[2]),"Strate_3" =
  c(ybar_strates[3],sigma2_h[3],sigma2_hc[3]))

row.names(t) = c("Moyenne","Variance","Variance_corrig e")
t

#####
#Les strates ont la m me taille#
#####

nh=c(rep(round(n/h),h)) #Calcul des nh
nh

#####
#Tirage de 20 echantillons et calcul demand #
#####
### Calcul des nh
nh=c(rep(round(n/h),h))
nh

#####
#Tirage de 20 echantillons et calcul demand #
#####

res_strat=matrix(0,20,7)
for (k in 1:20){

  #ybar_ech et var_ech
  ech1=sample(strate1,nh[1],replace=FALSE)
  ech1
  ybar_hat1=mean(ech1)
  s2_c_h1=var(ech1)
  ech2=sample(strate2,nh[2],replace=FALSE)
  ech2
  ybar_hat2=mean(ech2)
  s2_c_h2=var(ech2)
  ech3=sample(strate3,nh[3],replace=FALSE)
  ech3
  ybar_hat3=mean(ech3)
  s2_c_h3=var(ech3)

```

```

# ybar estim e pour chaque strate
ybar_hat_strates=c(ybar_hat1,ybar_hat2,ybar_hat3)
ybar_hat_strates

# ybar estim e
ybar_hat=(1/N)*sum(Nh*ybar_hat_strates)
ybar_hat
mean(y)
# variances corrig es de l'echantillon dans chaque strate
s2_hc=c(s2_c_h1,s2_c_h2,s2_c_h3)

# Variance totale - Variance intra strates - variance inter strates

fh=NULL
for(i in 1:h){
  fh[i]=nh[i]/Nh[i]
}
var_ybar_hat_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 estim e
var_estim_ybar_hat_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_hat-1.96*var_estim_ybar_hat_strat
ic_sup_strat=ybar_hat+1.96*var_estim_ybar_hat_strat
long_ic=ic_sup_strat-ic_inf_strat

res_strat[k,]=c(ybar_hat,var_ybar_hat_strat,
               var_estim_ybar_hat_strat,var_intra_y,
               ic_inf_strat,ic_sup_strat,long_ic)
}

#####
# Resultats du sondage stratifi regroup s dans un tableau #
#####

res_strat
colnames(res_strat)=c("ybar_hat","var_ybar_hat_strat","var_estim_ybar_
  hat_strat",

```

```

        "var_intra_strat","ic_inf_strat","ic_sup_strat"
        ,"long_ic_strat")
### La moyenne de ybar estim es
ybar_strat=mean(res_strat[,1])
ybar_strat
### Moyenne des longueurs des intervalles de confiance
long_strat=mean(res_strat[,7])
long_strat
### Repr sentation de l'intervalles de confiance pour chacune des 20
    exp riences sur le m me graphe
res_s=NULL
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="red",cex=4)+
    geom_point(aes(y,res_strat[,5]),colour="black",cex=4)+geom_path(
    colour="red",size=1)+geom_path(aes(y,res_strat[,5]),colour="red",
    size=1)
g2_s=annotate("text", label =c("La_borne_inf_rieure","La_borne_
    sup_rieure"), x =c(3,3), y =c(2,2.5),size =5, colour = c("red","
    black"))
g3_s=labs(x="",y="")
g1_s+g2_s+g3_s+ggtitle("les_20_intervalles_de_confiance_\n_Strates_ont_
    meme_taille")+theme(plot.title = element_text(color="black", size=14
    , face="bold.italic"))
res_strat=res_strat[,-4]
View(res_strat)
#####
#Les strates sont allocation proportionnelle#
#####

nh=round(Nh*0.2)
nh
nh=nh[c(-3)]
### Tirage de 20 echantillons et calcul demand
res=matrix(0,20,7)
res_prop=matrix(0,20,7)
for (k in 1:20){
    ###
    #ybar_ech et var_ech
    ech1=sample(strate1,nh[1],replace=FALSE)
    ech1
    ybar_hat1=mean(ech1)
    s2_c_h1=var(ech1)
    ech2=sample(strate2,nh[2],replace=FALSE)
    ech2
    ybar_hat2=mean(ech2)
    s2_c_h2=var(ech2)

    ech3=sample(strate3,nh[3],replace=FALSE)
    ech3
    ybar_hat3=mean(ech3)

```

```

s2_c_h3=var(ech3)
# ybar estim e pour chaque strate
ybar_hat_strates=c(ybar_hat1,ybar_hat2,ybar_hat3)
ybar_hat_strates

# ybar estim e
ybar_hat=(1/N)*sum(Nh*ybar_hat_strates)
ybar_hat
mean(y)
# variances corrig es de l'echantillon dans chaque strate
s2_hc=c(s2_c_h1,s2_c_h2,s2_c_h3)

# Variance totale - Variance intra strates - variance inter strates

fh=NULL
for(i in 1:h){
  fh[i]=nh[i]/Nh[i]
}
var_ybar_hat_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 estim e
var_estim_ybar_hat_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_hat-1.729*var_estim_ybar_hat_strat
ic_sup_strat=ybar_hat+1.729*var_estim_ybar_hat_strat

long_ic=ic_sup_strat-ic_inf_strat

res_prop[k,]=c(ybar_hat,var_ybar_hat_strat,
               var_estim_ybar_hat_strat,var_intra_y,
               ic_inf_strat,ic_sup_strat,long_ic)
}
### Resultats du sondage stratifi regroup s dans un tableau
res_prop
colnames(res_prop)=c("ybar_hat","var_ybar_hat_prop","var_estim_ybar_hat
  _strat",
                    "var_intra_strat","ic_inf_prop","ic_sup_prop"
                    ,"long_ic_prop")

```

```

### La moyenne de ybar estim es
ybar_prop=mean(res_prop[,1])
ybar_prop
### Moyenne des longueurs des intervalles de confiance
long_prop=mean(res_prop[,7])
### Repr sentation de l'intervalles de confiance pour chacune des 20
  exp riences sur le m me graphe
res_prop1=NULL
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="red",
  cex=4)+geom_point(aes(y,res_prop[,5]),colour="red",cex=4)+geom_path(
  colour="red",size=1)+geom_path(aes(y,res_prop[,5]),colour="red",size
  =1)
g2_prop=annotate("text", label =c("La borne inf rieure","La borne
  sup rieure"), x =c(3,3), y =c(3,4.5),size =5, colour = c("red","
  black"))
g3_prop=labs(x="",y="")
g1_prop+g2_prop+g3_prop+ggtitle("les 20 intervalles de confiance\n
  Strates allocation proportionnelle")+theme(plot.title = element_
  text(color="black", size=14, face="bold.italic"))
res_prop=res_prop[,-4]
View(res_prop)

#####
#Les strates sont allocation optimale#
#####

nh=round(((n*Nh*(sigma2_hc)^0.5))/sum((Nh*(sigma2_hc)^0.5)) )
nh
nh=nh[c(-3)]
### Tirage de 20 echantillons et calcul demand
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_hat1=mean(ech1)
  s2_c_h1=var(ech1)
  ech2=sample(strate2,nh[2],replace=FALSE)
  ech2
  ybar_hat2=mean(ech2)
  s2_c_h2=var(ech2)

  ech3=sample(strate3,nh[3],replace=FALSE)
  ech3
  ybar_hat3=mean(ech3)
  s2_c_h3=0
  # ybar estim e pour chaque strate
  ybar_hat_strates=c(ybar_hat1,ybar_hat2,ybar_hat3)

```

```

ybar_hat_strates

# ybar estim e
ybar_hat=(1/N)*sum(Nh*ybar_hat_strates)
ybar_hat
mean(y)
# variances corrig es de l'echantillon dans chaque strate
s2_hc=c(s2_c_h1,s2_c_h2,s2_c_h3)

# Variance totale de ybar_hat - Variance intra strates - variance
  inter strates

fh=NULL
for(i in 1:h){
  fh[i]=nh[i]/Nh[i]
}
var_ybar_hat_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 estim e
var_estim_ybar_hat_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_hat-1.729*var_estim_ybar_hat_opt
ic_sup_opt=ybar_hat+1.729*var_estim_ybar_hat_opt

long_ic_opt=ic_sup_opt-ic_inf_opt

res_opt[k,]=c(ybar_hat,var_ybar_hat_strat,
              var_estim_ybar_hat_opt,var_intra_y,
              ic_inf_opt,ic_sup_opt,long_ic_opt)
}
### Resultats du sondage stratifi regroup s dans un tableau
colnames(res_opt)=c("ybar_hat","var_ybar_hat_opt","var_estim_ybar_hat_
  opt",
                    "var_intra_strat","ic_inf_strat","ic_sup_strat"
                    ,"long_ic")
### La moyenne de ybar estim es
ybar_opt=mean(res_opt[,1])
ybar_opt

```

```

### Moyenne des longueurs des intervalles de confiance
long_opt=mean(res_opt[,7])
### Repr sentation de l'intervalles de confiance pour chacune des 20
  exp riences sur le m me 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="#000033",
cex=4)+geom_point(aes(y,res_opt[,5],colour="red",cex=4)+geom_path(
colour="#000033",size=1)+geom_path(aes(y,res_opt[,5],colour="#99CC6
6",size=1)
g2_opt=annotate("text", label =c("La_borne_inf_rieure","La_borne_
sup_rieure"), x =c(3,3), y =c(3,4.5),size =5, colour = c("red","#00
0033"))
g3_opt=labs(x="",y="")
g1_opt+g2_opt+g3_opt+ggtitle("les_20_intervalles_de_confiance_\n_
Strates_ _allocation_optimale")+theme(plot.title = element_text(
color="black", size=14, face="bold.italic"))
res_opt=res_opt[,-4]
View(res_opt)
ybar_strat
long_strat
ybar_prop
long_prop
ybar_opt
long_opt

#####
###Grappe#####
#####

grappe=matrix(0,100,1)
for (i in 1:100) {
  (if(data_100[i,10]=="Art_/musique") (grappe[i]=1))
  (if(data_100[i,10]=="Tous") (grappe[i]=1))
  (if(data_100[i,10]=="Sante_/sport") (grappe[i]=2))
  (if(data_100[i,10]=="Nourriture") (grappe[i]=2))
  (if (data_100[i,10]=="Television_/films" ) (grappe[i]=3))
  (if (data_100[i,10]== "Actualite") (grappe[i]=3))
  (if (data_100[i,10]=="Voyages_/mode") (grappe[i]=4))
}
data=cbind(data_100,grappe)
View(data)
#par grappe :
p=10
gr=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 #proba d'inclusion pi1h
##### la d finition des grappes
for(j in 1:4){

  gr[j]=list(y[grappe==j])

}

grappe
##### l'echantillonnage
N=100
Nh=as.vector(table(grappe))
sg1=matrix(0,10,2) #la matrice qui va contenir les numero des grappes
  selectionn es
for(i in 1:p){
  sg=sample(1:4,m,replace=F)
  sg1[i,2]=sg[2]
  sg1[i,1]=sg[1]
  sg1
  Nhh=Nh[sg] #le nbre des individus dans chaque grappe selectionn e
  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 estim s
  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])))
d=rbind(d,Nh)
View(d)
colnames(d)=c("grappe1","grappe2","grappe3","grappe4")
rownames(d)=c("moyenne","taille_de_grappe")

d=d[-2,]
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 estim s
var_ygrap=as.numeric(var_ygrap)

```

```

icinf_grappe=NULL
icsup_grappe=NULL
intervalle_grappe=NULL
longIC_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]))
  longIC_grap[i]=list(icsup_grappe[i]-icinf_grappe[i])
}
mean(unlist(longIC_grap))
mugrap=mean(y_bar_grappe) #la moyenne des y_bar_grappe
mugrap #la moyenne des y_bar_grappe
longIC_grap=unlist(longIC_grap)
res_grappe=NULL #le tableau recapitulatif
res_grappe=data.frame(y=1:10,y_bar_grappe,var_ygrap,icinf_grappe,icsup_
  grappe,longIC_grap)
g1_grappe=ggplot(res_grappe,aes(y,icsup_grappe))+geom_point(colour="#66
  0066",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_inf_rieuse","La_borne_
  sup_rieuse"), x =c(3,3), y =c(1,1.5),size =5, colour = c("firebrick
  1","#660066"))
g3_grappe=labs(x="",y="")
g1_grappe+g2_grappe+g3_grappe+ggtitle("les_10_intervals_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","longIC_grap","num_1ere_grap_selct","num_2ere_grap_selct")
View(res_grappe)

#####
## 2 degr s#####
#####
nh1=round(Nh/5) #le nombre des individus qui peuvent etre selectionn s
  dans chaque grappe
grappe_2dg=NULL
##### les m unit s primaires
for(i in 1:4){
  grappe_2dg[i]=list(y[grappe==i])
}
grappe_2dg
sg2dg=matrix(0,10,2) ##### la matrice des numeros des grappes
  selectionn es
nb=matrix(0,10,2) ##### la matrice des nombres des indivus
  selectionn s dans chaque grappe
y_grappe_D2=NULL

```

```

y_bar_grappe2=NULL
y_bar_h2dg=NULL
th2=NULL
var_ygrap2=NULL
ng1=NULL
m=2
t=sum(y)
sg_2dg=NULL

var_y_bar_grap_2dg=NULL
for(i in 1:p){
  sg2=sample(1:4,m,replace=F)
  sg2dg[i,2]=sg2[2]
  sg2dg[i,1]=sg2[1]
  sg2dg
  Nh_2dg=Nh[sg2]
  ygrap2=grappe_2dg[sg2]
  y_grappe_D2[i]=list(unlist(ygrap2))
  ng1[i]=list(length(unlist(ygrap2)))
  sg_2dg[1]=list(sample(unlist(ygrap2[1]),nh1[sg2][1],replace=F))
  sg_2dg[2]=list(sample(unlist(ygrap2[2]),nh1[sg2][2],replace=F))
  nb[i,1]=nh1[sg2][1]
  nb[i,2]=nh1[sg2][2]
  y_bar_h2dg[1]=mean(unlist(sg_2dg[1]))
  y_bar_h2dg[2]=mean(unlist(sg_2dg[2]))
  q=sum(unlist(sg_2dg[1]))
  a=sum(unlist(sg_2dg[2]))
  q=c(q,a)
  y_bar_grappe2[i]=list((sum((M/m)*(Nh_2dg/nh1[sg2])*q))/N) ##### les
    y bar estim s
  th2[1]=sum(unlist(sg_2dg[1]))
  th2[2]=sum(unlist(sg_2dg[2]))
  sigma1=(Nh_2dg[1]/(Nh_2dg[1]-1))*var(unlist(ygrap2[1]))
  sigma2=(Nh_2dg[2]/(Nh_2dg[2]-1))*var(unlist(ygrap2[2]))
  sigma=c(sigma1,sigma2)
  x=((M^2)*(M-m))/(M*m*(M-1))
  u=(sum(th2-(t/M)))^2
  v=(M/m)*((Nh_2dg^2)/nh1[sg2])*(Nh_2dg-nh1[sg2])
  var_y_bar_grap_2dg[i]=((x*u)+sum(v*sigma))/(N^2) # la variance de l'
    estimateur
}

res1=cbind(y_bar_grappe2,var_y_bar_grap_2dg,sg2dg,nb)
colnames(res1)=c("y_bar_grap_estim ", "var_y_bar_grap_2dg", "grappe1", "
  grappe2", "nh1", "nh2")
y_bar_grappe2=as.numeric(y_bar_grappe2)
var_y_bar_grap_2dg=as.numeric(var_y_bar_grap_2dg)
icinf_2dg=NULL
icsup_2dg=NULL
intervalle_2dg=NULL
longIC_2dg=NULL

```

```

for(i in 1:p){
  icinf_2dg[i]=(y_bar_grappe2[i])-1.96*sqrt(var_y_bar_grap_2dg[i])
  icsup_2dg[i]=(y_bar_grappe2[i])+1.96*sqrt(var_y_bar_grap_2dg[i])
  intervalle_2dg[i]=list(c(icinf_2dg[i],icsup_2dg[i]))
  longIC_2dg[i]=list(icsup_2dg[i]-icinf_2dg[i])
}
mean(unlist(longIC_2dg))
mu_2dg=mean(y_bar_grappe2)
mu_2dg
longIC_2dg=unlist(longIC_2dg)
res_2dg=NULL #le tableau recapitulatif
res_2dg=data.frame(y=1:10,y_bar_grappe2,var_y_bar_grap_2dg,icinf_2dg,
  icsup_2dg,longIC_2dg)
g1_2dg=ggplot(res_2dg,aes(y,icsup_2dg))+geom_point(colour="darkgreen",
  cex=4)+geom_point(aes(y,icinf_2dg),colour="deeppink",cex=4)+geom_
  path(colour="darkgreen",size=1)+geom_path(aes(y,icinf_2dg),colour="
  deeppink",size=1)
g2_2dg=annotate("text", label =c("La borne inf r ieure","La borne
  sup r ieure"), x =c(3,3), y =c(3,3.5),size =5, colour = c("deeppink"
  ,"darkgreen"))
g3_2dg=labs(x="",y="")
g1_2dg+g2_2dg+g3_2dg+ggtitle("les 10 intervalles de confiance \n
  2
  degr ") +theme(plot.title = element_text(color="black", size=14,
  face="bold.italic"))
res_2dg=res_2dg[,-1]
res_2dg=cbind(res_2dg,sg2dg,nb)

colnames(res_2dg)=c("y_bar_grappe2","var_y_bar_grap_2dg","icinf_2dg","
  icsup_2dg","longIC_2dg","num_1ere_grap_selc_t","num_2ere_grap_selc_t"
  ,"nbre_ind_select_G1","nbre_ind_select_G2")
View(res_2dg)
#####
##synthese#####
#####

##longeur IC
long_PEAR=res_pear[,6]
long_PESR=res_pesr[,6]
long_PIAR=ress[,6]
long_strat_taille=res_strat[,6]
long_strat_prop=res_prop[,6]
long_strat_opt=res_opt[,6]
long_grappe=res_grappe[,5]
long_2deg=res_2dg[,5]
long_PISR=tt[,5]

tab=data.frame(y=1:20,long_grappe,long_2deg)

#####synthese PEAR PESR PIAR PISR
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_PEAR),colour="blue",cex=4)+
geom_path(aes(y,long_PEAR),colour="blue",size=1)+
geom_point(aes(y,long_PISR),colour="black",cex=4)+
geom_path(aes(y,long_PISR),colour="black",size=1)
b=annotate("text", label =c("PESR","PEAR","PIAR","PISR"), x =c(18,18,18
,18), y =c(4,4.25,4.5,4.75),size =8, colour = c("deeppink","
darkgreen","blue","black"))
c=labs(x="",y="")
a+b+c+ggtitle("Synth se des logueurs des intervalles de confiance \n
PEAR_PESR_PEAR_PISR")+theme(plot.title = element_text(color="black",
size=14, face="bold.italic"))
##### synthese stratification
a=ggplot(tab,aes(y,long_strat_taille))+geom_point(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="chocolate4",cex=4)+
geom_path(aes(y,long_strat_opt),colour="chocolate4",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("
chocolate4","green","coral1"))
c=labs(x="",y="")
a+b+c+ggtitle("Synth se des logueurs des intervalles de confiance \n
strat_egal_prop_opt")+theme(plot.title = element_text(color="black",
size=14, face="bold.italic"))
##### synthese grappe et 2degr
library(ggplot2)
a=ggplot(tab,aes(y,long_grappe))+geom_point(colour="deeppink4",cex=4)+
geom_path(aes(y,long_grappe),colour="deeppink4",size=1)+
geom_point(aes(y,long_2degr),colour="indianred3",cex=4)+
geom_path(aes(y,long_2degr),colour="indianred3",size=1)
b=annotate("text", label =c("Grappe","2_degr "), x =c(19,19), y =c(7.5
,8),size =7, colour = c("deeppink4","indianred3"))
c=labs(x="",y="")
a+b+c+ggtitle("Synth se des logueurs des intervalles de confiance \n
grappe et 2_degr ")+theme(plot.title = element_text(color="black",
size=14, face="bold.italic"))

##### Comparaison de tous les types de sondages utilis s
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_PEAR),colour="blue",cex=4)+
geom_path(aes(y,long_PEAR),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="chocolate4", cex=4)+
geom_path(aes(y, long_strat_opt), colour="chocolate4", 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", "PIAR", "PISR", "2_degr "), x =c(19,19
  ,19,19,19,19,19), y =c(5.5,6,6.5,7,7.5,8,8.5,9,9.5), size =5,
  colour = c("chocolate4", "green", "coral1", "deeppink", "darkgreen", "
  deeppink4", "blue", "black", "indianred3"))
c=labs(x="", y="")
a+b+c+ggtitle("Synth se des logueurs des intervalles de confiance")+
  theme(plot.title = element_text(color="black", size=14, face="bold.
    italic"))
#####
# Redressement #
#####

library(MASS)
library(lpSolve)
library(sampling)
n=20
N=100
# on tire par un sondage aleatoire simple un echantillon de taille n=20
set.seed(1)
s=srswor(n,N)
echantillon=y[s==1]
moyenne=mean(echantillon)
moyenne # moyenne de lechantillon = 3.5
##Alors l'estimateur de la moyenne avant redressement est:
moyenne
##Or on connait que la proportion des femmes dans la population est
P=sum(sexe=="Femme")/N

P # la proprtion des hommes dans la pop = 0.34

p_ech=sum(sexe[s==1]=="Homme")/n
p_ech #la proportion des hommes dans l' chantillon =0.6
##L'estimateur de la moyenne apr s redressement est:
moyenne.red=(P/p_ech)*moyenne
moyenne.red # est egale a 3.85
# tous les individus donc ont le meme poids (P/p_ech) alors que sans
  redressement leur podis est(N/n)

```


Bibliographie