

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



Enquête sur le temps consacré au sport

Elaborée par : -SALMA BOUSLAMA
-ANWAR MSEHLI
-KHADIJA MHIRI

Encadrée par: Mme Mallek Hela

REMERCIEMENT

Avec tout le plaisir, nous remercions en premier lieu notre enseignante de sondage Mme Mallek Hela qui a eu toute la gentillesse de nous guider pendant le déroulement du travail.
On tient également à remercier toute personne ayant collaboré avec nous.

TABLE DE MATIERE

Table des matières

INTRODUCTION	3
PARTIE I : COLLECTE DE DONNEES	5
PARTIE II : TRAITEMENT DES DONNEES	7
I-généralité	5
II- Quelques statistiques des données collectées	7
PARTIE III : REALISATION DES SONDAGE	16
I-sondage aléatoire simple	16
1- Plan simple sans remise	16
2- Plan simple avec remise	17
II- Probabilités inégales	20
1- Plan simple sans remise	20
2- Plan simple avec remise	24
1-Plan de sondage sans remise	20
2-Plan de sondage avec remise	24
III -Sondage stratifié	27
IV- Sondage par grappes et à deux degrés	34
1- Plan de sondage par grappes	34
2- Plan de sondage à deux degrés	36
Partie IV : SYNTHESE	38
I-Sondages aléatoires PESR, PEAR, PISR PIAR et PIAR.....	38
II- Stratification	39
III-Sondage par grappes et à deux degrés	40
IV-Comparaison de tous les types de sondages utilisés	41
PARTIE V : REDRESSEMENT	43
PARTIE VI : ANNEXE	44

TABLE DE FIGURE

Figure 1: la répartition de la population selon la variable nombre d'heure par semaine	9
Figure 2: la répartition de la population selon le sexe	10
Figure 3: la répartition de la population selon la variable âge	11
Figure 4: la répartition de la population selon la variable raison	11
Figure 5: la répartition de la population selon la variable type sport	12
Figure 6: la répartition de la population selon la variable occupation	12
Figure 7: l'influence de la variable sexe sur la variable nombre d'heure par semaine	13
Figure 8: l'influence de la variable âge sur la variable nombre d'heure par semaine	13
Figure 9: l'influence de la variable raison sur la variable nombre d'heure par semaine.....	14
Figure 10: l'influence de la variable type_sport sur la variable nombre d'heure par semaine.....	14
Figure 11: l'influence de la variable occupation sur la variable nombre d'heure par semaine.....	15
Figure 12: les 20 intervalles de confiance PESR.....	18
Figure 13: les 20 intervalles de confiance PEAR	20
Figure 14: les 20 intervalles de confiance PISR.....	24
Figure 15: les 20 intervalles de confiance PIAR.....	27
Figure 16: les 20 intervalles de confiance Strates ayant la même taille.....	30
Figure 17: les 20 intervalles de confiance Strates a allocation proportionnelle.....	32
Figure 18: les 20 intervalles de confiance Strates a allocation optimale.....	34
Figure 19: les 10 intervalles de confiance par grappes.....	36
Figure 20: les 10 intervalles de confiance à 2 degrés.....	38
Figure 21: synthèse des longueurs des intervalles de confiance PEAR PESR PISR PIAR	39
Figure 22: synthèse des longueurs des intervalles de confiance strat_egal prop opt.....	40
Figure 23: synthèse des longueurs des intervalles de confiance grappe et à 2 degré.....	41
Figure 24: synthèse des longueurs des intervalles de confiance	42

TABLE DE TABLEAU

Tableau 1:les dix premières observations de notre base de données.....	9
Tableau 2:résultats obtenus dans le plan simple sans remise SAS	17
Tableau 3:résultats obtenus dans le plan simple avec remise SAS.....	19
Tableau 4:sélection des échantillons dans le plan simple sans remise proba inégales.....	22
Tableau 5:résultats obtenus dans le plan simple sans remise proba inégales	23
Tableau 6:sélection des échantillons dans le plan simple avec remise proba inégales.....	25
Tableau 5:résultats obtenus dans le plan simple avec remise proba inégales	26
Tableau 7:résultats obtenus dans des strates de même taille sondage stratifié	29
Tableau 8:résultats obtenus dans des strates a allocation proportionnelle sondage stratifié	31
Tableau 9:résultats obtenus dans des strates a allocation optimale sondage stratifié	33
Tableau 10:sélection des échantillons dans un plan de sondage par grappes	35
Tableau 11:résultats obtenus dans un plan de sondage par grappes.....	35
Tableau 12:résultats obtenus dans un plan de sondage a deux degrés.....	37
Tableau 13:Comparaison de tous les types de sondages utilisés.....	42

INTRODUCTION

A la recherche du bonheur, l'individu moderne se trouve esclave du sacrifice qu'il fait pour le confort au détriment du travail. C'est ainsi qu'on se trouve devant un quotidien stressant et peu créative. Pour en échapper, le sport présente un moyen efficace pour se divertir, changer la monotonie du quotidien ou de garder la bonne santé

C'est pour cela que nous avons mené une enquête sur le temps consacré au sport chez les tunisiens ayant comme population cible les sportifs.

Ainsi, nous avons fait ce projet dans le but d'appliquer la théorie dans la pratique, de nous habituer avec les méthodes de sondage et de traiter un exemple concret.

Le déroulement du travail nous a nécessité d'enquêter 100 individus qui représentent notre population. Puis, nous avons procédé à des différents types de sondage auprès de cette population en utilisant le logiciel "Rstudio" en faisant appel à différentes librairies.

PARTIE I : COLLECTE DE DONNEES

Période d'enquête :

Le 8/11/2017

Enquêteurs :

Bousslama Salma/Mhiri Khadija/Msehli Anwaar

Population cible :

Les sportifs.

Population source :

Les personnes rencontrées à la cité sportive El Manzeh.

La variable d'intérêt :

Nombre d'heure d'activité physique par semaine.

Les variables auxiliaires :

sexe/âge/type d'activité physique/occupation/raison.

Etapas de l'enquête :

Pour cette enquête nous avons effectué un questionnaire sur 100 personnes. Le contact avec l'enquêté était réel : nous avons interrogé sur place les sportifs dans le parcours d'El Manzeh. Nous avons consacré deux sorties pour élaborer notre projet : une matinée et un après-midi, pour s'assurer de la distribution uniforme du temps et de pouvoir rencontrer des personnes ayant différents âges.

Voici donc le questionnaire :

Enquête sur le temps consacré au sport

1) Pratiquez-vous un sport?

☐ Oui

☐ Non

2) Vous êtes :

☐ Homme

☐ Femme

3) Age :

☐ Moins de 15 ans

☐ [15,30[

☐ [30,45[

☐ Plus de 45 ans

4) Combien d'heures consacrez-vous à l'activité sportive par semaine, en moyenne?

.....

5) Pourquoi pratiquez-vous du sport ?

☐ Maladie

☐ Divertissement

☐ Athlète

☐ Surpoids

☐ Autre

6) Quel type de sport pratiquez-vous ?

☐ Sport d'équipe

☐ Sport individuel

7) Quel est votre statut professionnel ?

☐ Vous êtes élève/étudiant

☐ Vous occupez un emploi

☐ Vous n'avez pas d'activité

Merci pour votre attention !

PARTIE II : TRAITEMENT DES DONNEES

I-GENERALITE :

Notre sondage se compose de six variables : quatre variables qualitatives et deux quantitatives.

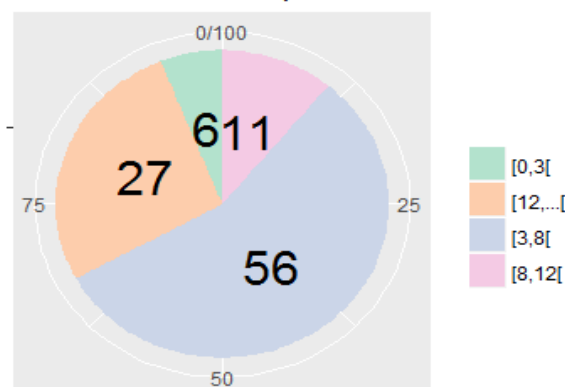
On s'intéresse également au nombre d'activité sportive, la variable d'intérêt, citée dans la quatrième question.

Voici les dix premières observations de notre base de données.

	genre	age	nbre_h_semaine	raison	type_sport	occupation
1	H	[15,30[4	divertissement	individuel	élève/etudiant
2	H	[15,30[3	athlète	individuel	élève/etudiant
3	H	[45,...[17	divertissement	individuel	employer
4	H	[15,30[28	athlète	équipe	élève/etudiant
5	F	[15,30[11	divertissement	individuel	élève/etudiant
6	F	[30,45[11	surpoids	individuel	employer
7	H	[45,...[6	divertissement	individuel	employer
8	H	[15,30[28	surpoids	équipe	employer
9	H	[45,...[5	maladie	individuel	pas d'activité
10	F	[45,...[6	divertissement	équipe	employer

Par suite, notre population se répartie comme suit selon le nombre d'heure de sport :

*la repartition de la population selon la variable c
nombre d'heure par semaine*



PARTIE II : TRAITEMENT DES DONNEES

On observe que : 56% de la population pratique entre 3 et 8 heures de sport par semaine alors que seulement 6% pratique moins de 3 heures. 27% passe plus que 12 heures et 11% entre 8 et 12 heures.

La moyenne de notre population est $\bar{Y}=9.05$

II- Quelques statistiques des données collectées

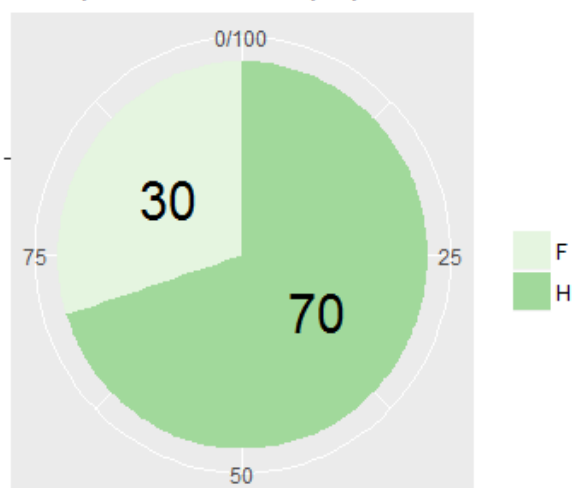
Pour cette partie, on a fait une étude statistique avec le logiciel statistique « R ». En premier lieu, nous avons construit des camemberts pour répartir la population selon chaque variable puis un croisement entre ces dernières et la variable d'intérêt pour plus de précisions.

Question1:Pratiquez-vous un sport?

C'est une question filtre. Notre population n'est constituée que de sportifs.

Question2: Sexe

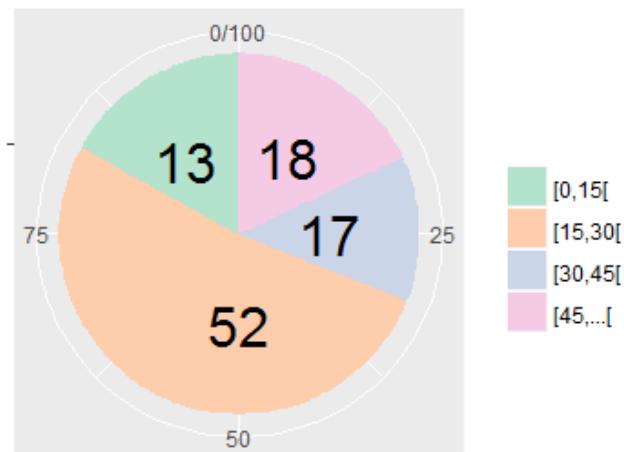
la repartition de la population selon le sexe



PARTIE II : TRAITEMENT DES DONNEES

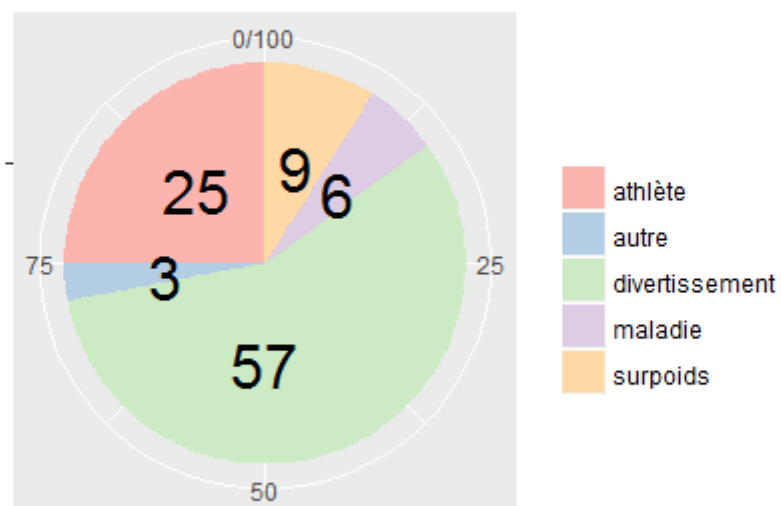
Question3 : Age

la repartition de la population selon la variable age



Question5 : Pourquoi pratiquez-vous du sport ?

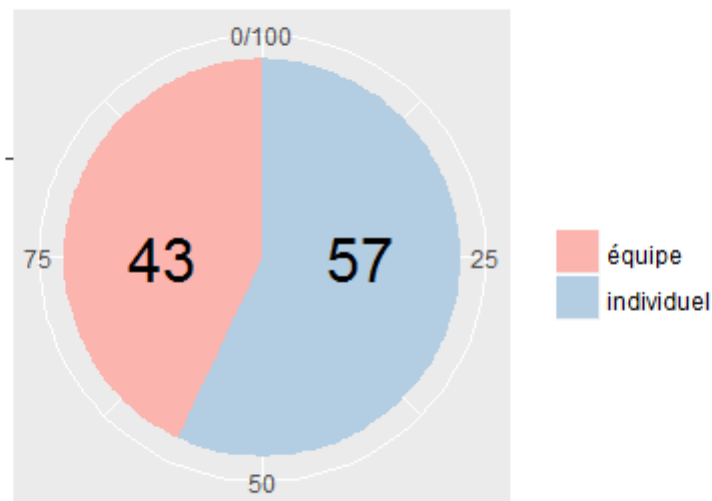
la repartition de la population selon la variable raison



PARTIE II : TRAITEMENT DES DONNEES

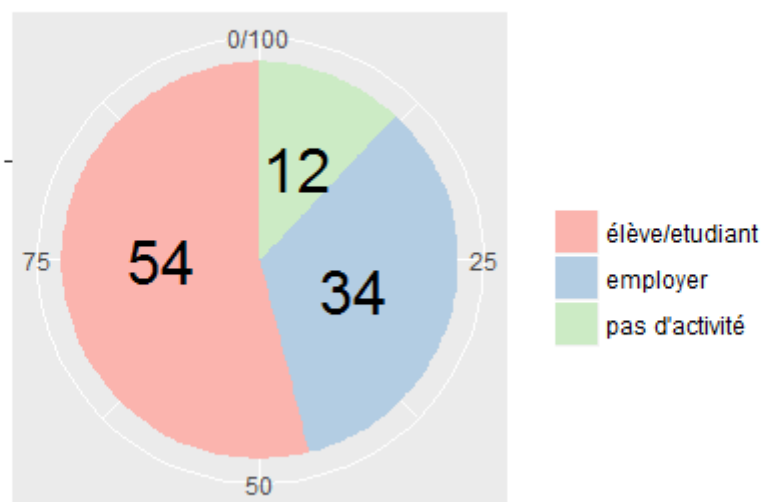
Question6 : Quel type de sport pratiquez-vous ?

la repartition de la population selon la variable type_sport



Question7 : Occupation

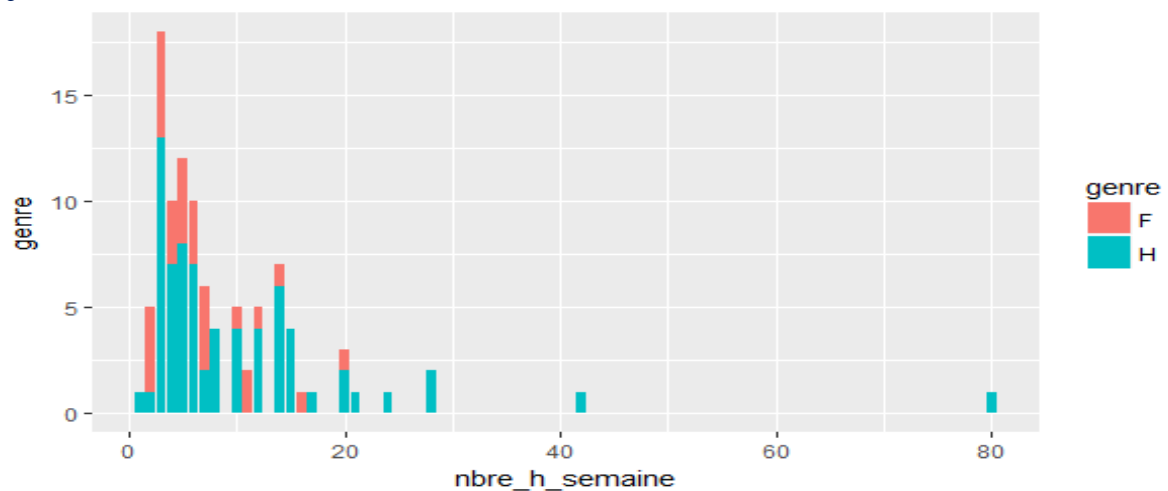
la repartition de la population selon la variable occupation



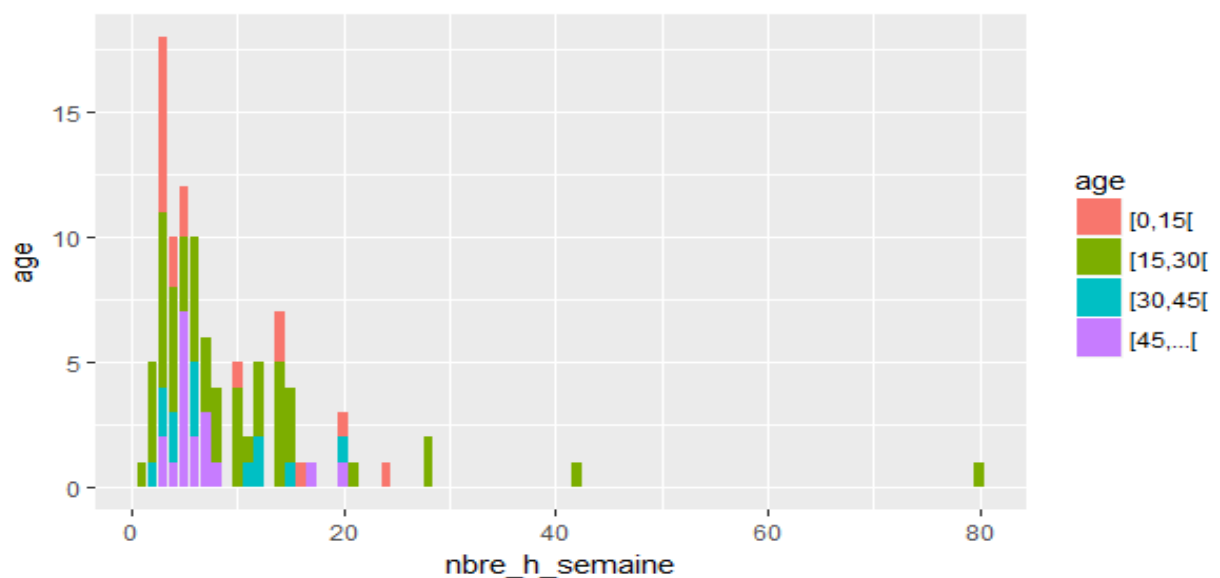
PARTIE II : TRAITEMENT DES DONNEES

Pour mieux comprendre, on a eu recours au package « GGLOT2 ».
Etant donnée le nombre-h-semaine représente la variable d'intérêt, nous voulons savoir l'influence des autres modalités sur cette dernière.

Question2: Sexe

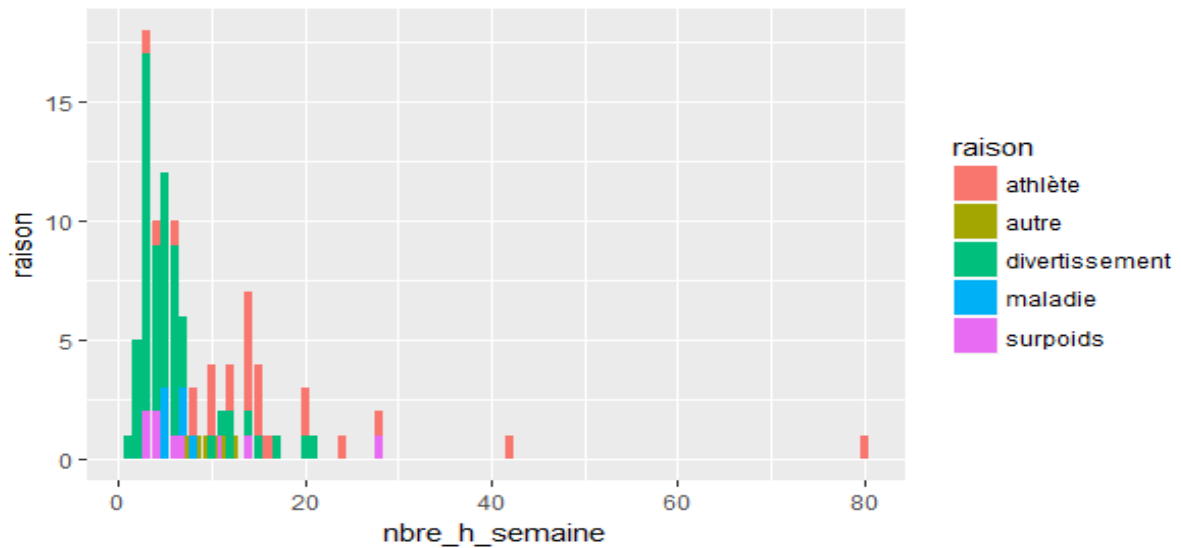


Question3 : Age

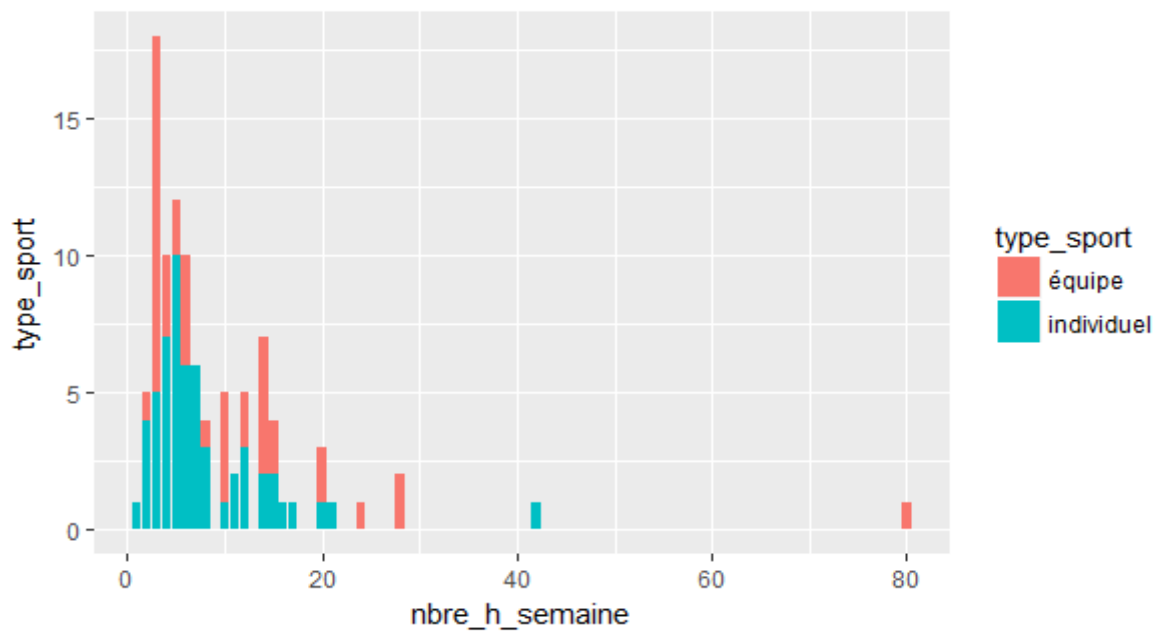


PARTIE II : TRAITEMENT DES DONNEES

Question5 : Pourquoi pratiquez-vous du sport ?

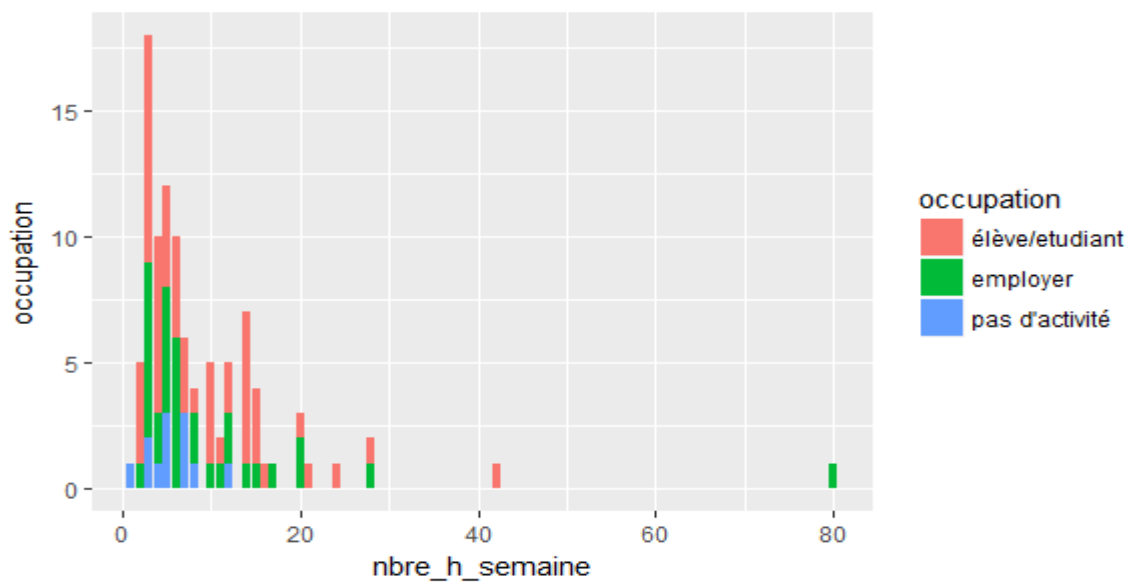


Question6 : Quel type de sport pratiquez-vous ?



PARTIE II : TRAITEMENT DES DONNEES

Question7 : Occupation



ENFIN, cette étude statistique descriptive nous permet de conclure que :

- Les sportifs sont en majorité de sexe masculin et leurs activité est répartie en différentes nombre d'heures alors que les femmes ne dépassent pas les 10h par semaines.
- Selon l'âge, la modalité la plus présente chez les sportifs est de 15ans à 30ans. Ce sont les jeunes qui sont les plus actifs ; de plus le point aberrant (nombre-h=80h) correspond à un jeune.
Les enfants et les vieillards sont moins actifs qui ne dépassent pas en moyenne les 7h d'activité.
Les adultes quant à eux, pratiquent entre 0 et 15h.
- On trouve que la plus part des enquêtés pratiquent le sport pour se divertir en premier lieu, avec un nombre d'heures qui ne dépasse pas les 10h, ou étant donné qu'il soit athlète en second lieu.
- Les sportifs préfèrent les activités individuelles comme le footing ou fréquenter les salles de sport plus que les activités d'équipe.
Ce sont les personnes du premier type qui pratiquent plus de sport.
- On remarque que la modalité « élève/étudiant » pratiquent plus le sport ce qui est conforme avec la réponse de la question 3 qui nous renseigne sur l'âge.
- Le point aberrant correspond à un employé athlète de sexe masculin qui préfère le sport d'équipe.

PARTIE III : REALISATION DES SONDAGES

Pour récapituler, on a choisit une taille de population $N=100$ dont on tire la taille de l'échantillon $n=20$.

Par semaine, les sportifs pratiquent en moyenne un nombre d'heure égale à $\bar{Y}= 9.05$ de variance $=96.876$ et d'écart-type $=9.842$.

I-Sondage aléatoire simple(SAS)

Dans le sondage aléatoire simple, tous les individus ont la même probabilité d'être sélectionnés.

1-Plan simple sans remise

Nous avons tiré aléatoirement et sans remise 20 échantillons. Pour chaque échantillon, nous avons calculé l'estimateur $ybar_hat_pesr$, sa variance, sa variance estimée ainsi que les bornes des intervalles de confiance et leurs longueurs.

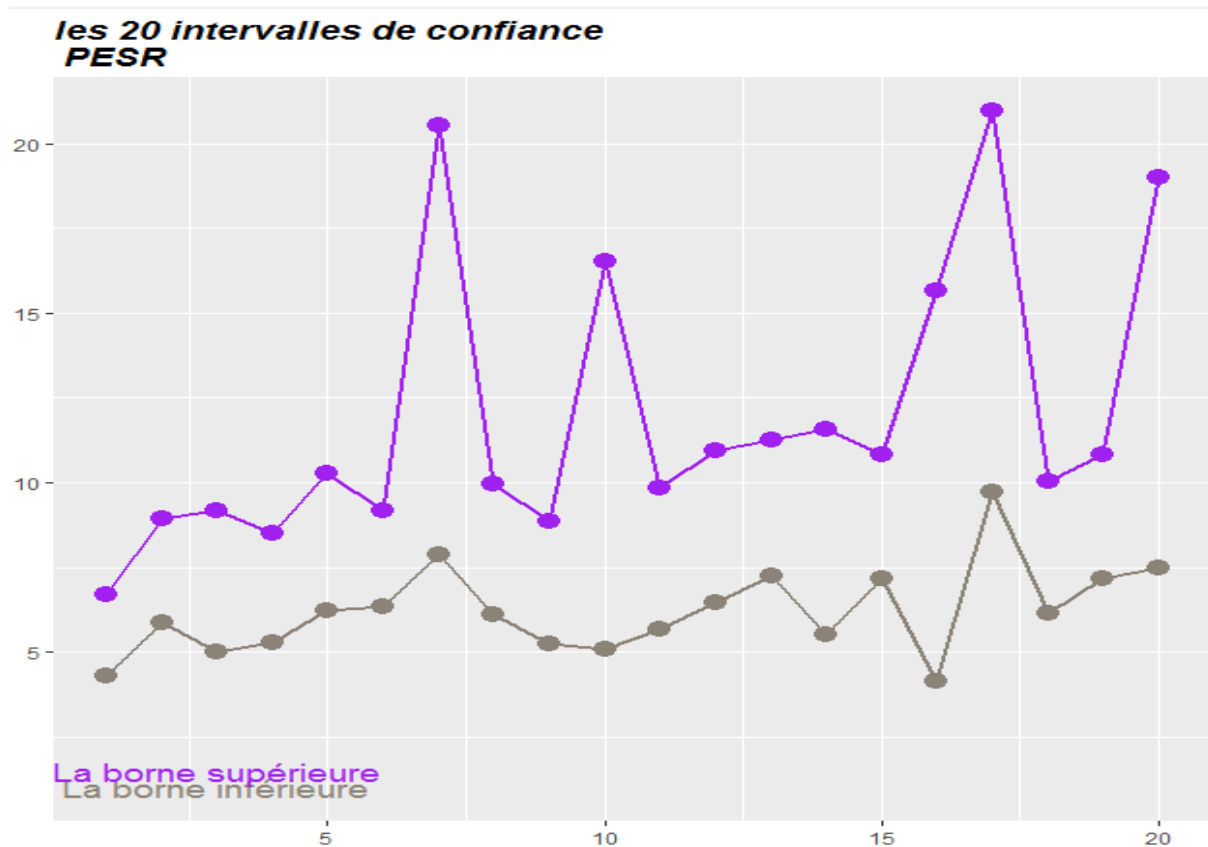
a-Résultats obtenus :

	$ybar_hat_pesr$	$var_ybar_hat_pesr$	$var_estim_ybar_hat_pesr$	$ICbinf_pesr$	$ICbsup_pesr$	$long_IC_pesr$
1	5.50	3.8363	0.4900	4.289700	6.710300	2.420600
2	7.40	3.8363	0.7816	5.871423	8.928577	3.057154
3	7.10	3.8363	1.4516	5.016860	9.183140	4.166280
4	6.90	3.8363	0.8796	5.278423	8.521577	3.243154
5	8.25	3.8363	1.3795	6.219253	10.280747	4.061494
6	7.75	3.8363	0.6755	6.328956	9.171044	2.842089
7	14.20	3.8363	13.4144	7.867421	20.532579	12.665158
8	8.05	3.8363	1.2459	6.120092	9.979908	3.859816
9	7.05	3.8363	1.0979	5.238341	8.861659	3.623317
10	10.80	3.8363	10.9744	5.072232	16.527768	11.455535
11	7.75	3.8363	1.4635	5.658339	9.841661	4.183323
12	8.70	3.8363	1.6844	6.456027	10.943973	4.487946
13	9.25	3.8363	1.3475	7.242945	11.257055	4.014111
14	8.55	3.8363	3.0659	5.522571	11.577429	6.054858
15	9.00	3.8363	1.1240	7.166934	10.833066	3.666132
16	9.90	3.8363	11.0916	4.141729	15.658271	11.516542
17	15.35	3.8363	10.5491	9.734316	20.965684	11.231369
18	8.10	3.8363	1.2836	6.141111	10.058889	3.917778
19	9.00	3.8363	1.1120	7.176745	10.823255	3.646510
20	13.25	3.8363	11.0915	7.491755	19.008245	11.516490

PARTIE III : REALISATION DES SONDAGES

La moyenne des $\bar{y}_{\text{pesr}} = 9.0925$

b-Représentation des intervalles de confiance :



2-Plan simple avec remise

Nous avons tiré aléatoirement et avec remise 20 échantillons. Pour chaque échantillon, nous avons calculé l'estimateur $\bar{y}_{\text{hat_pear}}$, sa variance, sa variance estimée ainsi que les bornes des intervalles de confiance et leurs longueurs.

PARTIE III : REALISATION DES SONDAGES

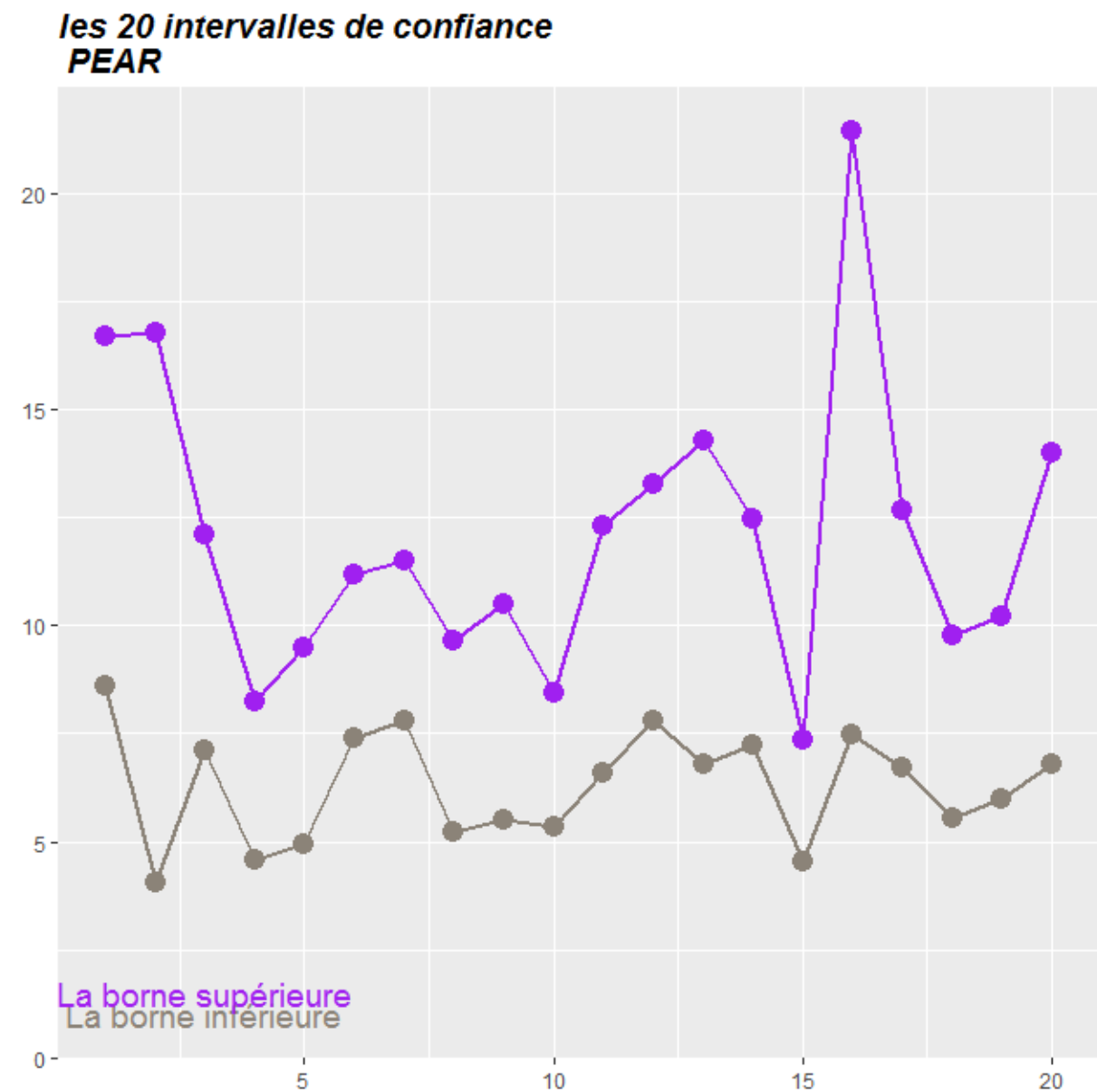
a-Résultats obtenus :

	ybar_hat_pear	var_ybar_hat_pear	var_estim_ybar_hat_pear	ICbinf_PEAR	ICsup_PEAR	long_IC_PEAR
X1	10.05	4.843813	2.197375	7.487009	12.612991	5.125982
X2	6.65	4.843813	1.281375	4.692809	8.607191	3.914381
X3	8.15	4.843813	3.591375	4.873385	11.426615	6.553229
X4	8.25	4.843813	1.184375	6.368347	10.131653	3.763307
X5	14.05	4.843813	13.432375	7.713180	20.386820	12.673641
X6	8.30	4.843813	2.030500	5.836251	10.763749	4.927498
X7	11.95	4.843813	13.667375	5.557988	18.342012	12.784023
X8	10.00	4.843813	1.450000	7.918008	12.081992	4.163983
X9	11.10	4.843813	13.324500	4.788676	17.411324	12.622647
X10	8.45	4.843813	1.942375	6.040308	10.859692	4.819384
X11	8.30	4.843813	1.295500	6.332052	10.267948	3.935897
X12	9.80	4.843813	3.733000	6.459404	13.140596	6.681192
X13	8.15	4.843813	2.311375	5.521366	10.778634	5.257269
X14	8.00	4.843813	2.065000	5.515408	10.484592	4.969183
X15	8.20	4.843813	1.173000	6.327404	10.072596	3.745191
X16	9.70	4.843813	13.840500	3.267632	16.132368	12.864736
X17	8.95	4.843813	1.862375	6.590453	11.309547	4.719093
X18	6.70	4.843813	1.200500	4.805581	8.594419	3.788838
X19	5.60	4.843813	0.427000	4.470181	6.729819	2.259638
X20	8.25	4.843813	3.759375	4.897623	11.602377	6.704753

La moyenne des $\bar{y}_{pear}=8.93$

PARTIE III : REALISATION DES SONDAGES

b-Représentation des intervalles de confiance :



PARTIE III : REALISATION DES SONDAGES

II- Probabilités inégales

Dans le sondage à probabilités inégales, la probabilité d'inclusion π_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 :

```
[1] 0.16727273 0.16727273 0.38545455 0.16727273 0.16727273 0.26909091 0.38545455 0.16727273
[9] 0.38545455 0.38545455 0.26909091 0.05090909 0.26909091 0.16727273 0.16727273 0.16727273
[17] 0.16727273 0.16727273 0.16727273 0.16727273 0.16727273 0.16727273 0.05090909 0.16727273
[25] 0.16727273 0.16727273 0.05090909 0.05090909 0.05090909 0.38545455 0.16727273 0.16727273
[33] 0.16727273 0.38545455 0.16727273 0.16727273 0.16727273 0.16727273 0.16727273 0.16727273
[41] 0.16727273 0.16727273 0.16727273 0.16727273 0.26909091 0.38545455 0.38545455 0.26909091
[49] 0.16727273 0.26909091 0.38545455 0.05090909 0.05090909 0.05090909 0.05090909 0.26909091
[57] 0.26909091 0.38545455 0.05090909 0.05090909 0.38545455 0.26909091 0.38545455 0.16727273
[65] 0.38545455 0.38545455 0.05090909 0.26909091 0.38545455 0.38545455 0.16727273 0.38545455
[73] 0.16727273 0.16727273 0.16727273 0.05090909 0.38545455 0.16727273 0.05090909 0.05090909
[81] 0.16727273 0.16727273 0.05090909 0.05090909 0.16727273 0.26909091 0.16727273 0.26909091
[89] 0.26909091 0.16727273 0.16727273 0.16727273 0.16727273 0.16727273 0.16727273 0.16727273
[97] 0.16727273 0.16727273 0.16727273 0.16727273
```

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

1- Plan de sondage sans remise

a-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 X (l'âge) de la manière suivante :

1. On calcule p_{kig} et v_{kig} :

$$p_{kig} = X_{ig} / \sum X_{ig} \text{ et } v_{kig} = \sum p_{kig}$$

2. On génère une variable aléatoire u suivant la loi uniforme sur $[0,1]$

3. On sélectionne l'individu i vérifiant $v_{ki} - 1 \leq u < v_{ki}$

PARTIE III : REALISATION DES SONDAGES

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	3	4	5	3	1	5	5	3	1	3	3	1	5	1	3	2	1	4	1
2	7	7	7	7	6	7	6	9	10	7	8	6	6	10	7	6	10	10	7	9
3	15	11	15	15	11	15	11	13	11	13	11	13	14	14	14	11	11	13	14	13
4	19	20	16	18	16	18	16	19	19	18	18	20	17	16	17	16	19	17	16	20
5	25	24	22	22	21	21	25	23	25	25	21	21	25	22	21	25	24	22	25	22
6	30	30	29	30	26	29	30	30	26	30	27	29	30	29	26	30	26	26	30	29
7	32	35	34	31	31	32	32	31	34	31	34	32	35	33	32	34	35	33	31	35
8	36	39	36	36	36	38	40	37	36	37	38	36	39	40	40	40	40	38	37	40
9	42	45	45	45	41	45	44	44	44	44	42	43	42	44	45	42	42	42	42	42
10	46	46	50	47	46	46	48	49	46	47	48	48	46	47	50	47	50	49	47	48
11	54	52	51	51	53	55	51	51	51	51	51	55	54	51	51	51	51	55	51	55
12	58	56	56	60	56	58	56	57	56	56	57	57	58	57	57	56	58	60	56	56
13	63	65	62	63	64	61	63	65	63	64	61	62	63	63	65	61	62	65	63	62
14	69	66	66	69	66	66	70	70	69	70	70	66	69	66	68	69	70	67	68	66
15	72	72	73	72	73	75	74	74	73	72	74	71	75	73	75	75	72	74	72	72
16	78	77	79	77	80	78	77	77	78	77	77	79	77	79	77	78	77	77	77	77
17	83	81	85	82	85	81	85	82	82	83	85	85	82	83	85	84	82	81	85	85
18	88	88	89	89	89	89	87	86	88	89	87	88	86	86	86	86	88	87	86	90
19	95	95	92	91	92	92	92	91	92	95	91	94	94	92	95	94	95	94	95	92
20	96	97	100	99	97	100	98	97	98	97	98	100	97	99	100	99	100	98	100	98

PARTIE III : REALISATION DES SONDAGES

b-Résultats obtenus :

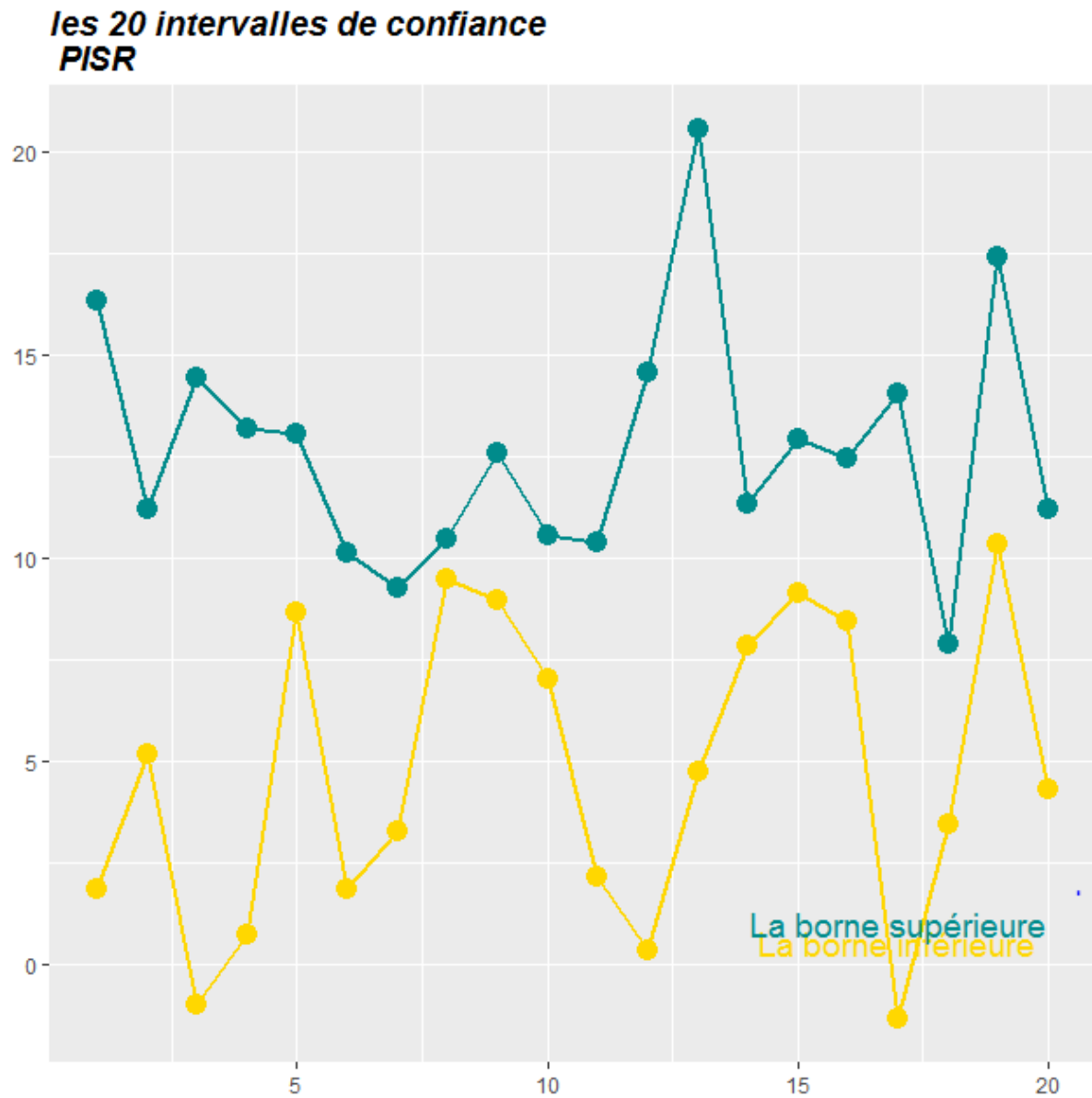
Les résultats obtenus sont regroupés dans ce tableau :

	$\hat{y}_{\text{bar_hat_pisr}}$	$\hat{\text{var_est_ybar_hat}}$	$\hat{\text{Icbinf_pisr}}$	$\hat{\text{ICbsup_pisr}}$	$\hat{\text{long_IC_pisr}}$
1	9.091340	17.5051577	1.8573482	16.325332	14.4679834
2	8.188666	3.0310524	5.1784912	11.198841	6.0203496
3	6.704142	20.0242972	-1.0328769	14.441160	15.4740370
4	6.951051	13.0386997	0.7077809	13.194321	12.4865405
5	10.854769	1.6418622	8.6393118	13.070226	4.4309142
6	5.973877	5.7239652	1.8372773	10.110476	8.2731992
7	6.247909	3.0133964	3.2465144	9.249304	6.0027896
8	9.976490	0.0764028	9.4985761	10.454404	0.9558277
9	10.770631	1.1137983	8.9459029	12.595360	3.6494571
10	8.787733	1.0548258	7.0119689	10.563498	3.5515290
11	6.275752	5.6695412	2.1588650	10.392639	8.2337741
12	7.432991	16.9792563	0.3084918	14.557490	14.2489979
13	12.645359	20.9959333	4.7228525	20.567865	15.8450123
14	9.602221	1.0243927	7.8522608	11.352182	3.4999208
15	11.048398	1.2131812	9.1439994	12.952796	3.8087971
16	10.462458	1.3465467	8.4561130	12.468804	4.0126909
17	6.345457	19.8319625	-1.3543149	14.045228	15.3995431
18	5.644504	1.6436416	3.4278471	7.861162	4.4333145
19	13.880421	4.1819242	10.3446582	17.416183	7.0715247
20	7.749996	3.9569369	4.3106601	11.189331	6.8786712

La moyenne des $\hat{y}_{\text{pisr}}=8.731708$

PARTIE III : REALISATION DES SONDAGES

c-Représentation des intervalles de confiance :



PARTIE III : REALISATION DES SONDAGES

2- Plan de sondage avec remise

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

a-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 :

$$P_{ki} = X_i / \sum X_i \text{ et } V_{ki} = \sum P_{kj} \text{ avec } V_{k0} = 0$$

Algorithme de tirage

1. On génère une variable aléatoire u suivant la loi uniforme sur $[0,1]$

2. On sélectionne l'individu i vérifiant $v_{ki-1} \leq u < v_{ki}$

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

Les 20 individus tirés dans chaque échantillon sont comme suit :

	ech1	ech2	ech3	ech4	ech5	ech6	ech7	ech8	ech9	ech10	ech11	ech12	ech13	ech14	ech15	ech16	ech17	ech18	ech19	ech20
1	69	45	6	45	63	70	57	21	51	38	98	7	97	60	25	47	12	91	34	56
2	14	89	33	31	74	92	18	98	49	81	95	50	9	51	36	14	26	87	8	54
3	57	73	76	51	77	91	65	3	66	97	22	45	96	75	43	68	85	93	9	88
4	66	91	33	26	70	64	74	83	32	93	40	6	15	66	58	95	9	14	81	72
5	15	41	33	16	3	50	68	66	77	77	89	88	44	63	51	13	58	34	96	35
6	43	72	31	69	25	24	32	51	50	56	34	11	83	10	1	8	68	20	2	90
7	58	41	18	98	47	71	15	82	61	15	61	38	9	33	11	47	24	71	70	89
8	9	6	63	83	77	14	70	11	89	48	69	73	1	62	1	26	66	8	6	43
9	86	31	65	85	10	61	27	89	7	44	70	100	100	20	69	69	86	66	30	56
10	43	68	99	31	34	98	10	11	69	52	11	33	89	13	38	56	72	69	19	30
11	20	7	97	10	81	10	46	3	37	24	70	10	62	41	3	7	86	69	99	91
12	69	11	81	9	62	81	23	88	39	7	77	45	39	63	68	20	49	100	4	93
13	58	74	32	96	19	3	74	8	14	69	46	44	62	72	2	66	66	81	9	87
14	18	3	4	48	30	56	78	94	3	47	72	65	26	26	39	46	76	46	60	96
15	36	39	77	32	8	80	68	24	25	21	63	78	84	9	66	89	75	75	48	2
16	11	40	30	45	9	49	56	37	69	96	17	65	95	51	35	50	21	98	63	48
17	33	72	68	53	16	34	65	69	14	49	56	45	77	62	91	5	74	16	8	24
18	32	63	4	73	5	48	66	72	86	34	48	68	21	68	6	30	4	69	42	9
19	31	62	68	10	95	7	80	36	81	57	62	11	7	19	7	31	26	89	57	11
20	38	2	33	68	79	77	89	51	57	97	65	47	75	56	47	70	13	7	87	31

PARTIE III : REALISATION DES SONDAGES

b-Résultats obtenus :

Les résultats obtenus sont regroupés dans ce tableau :

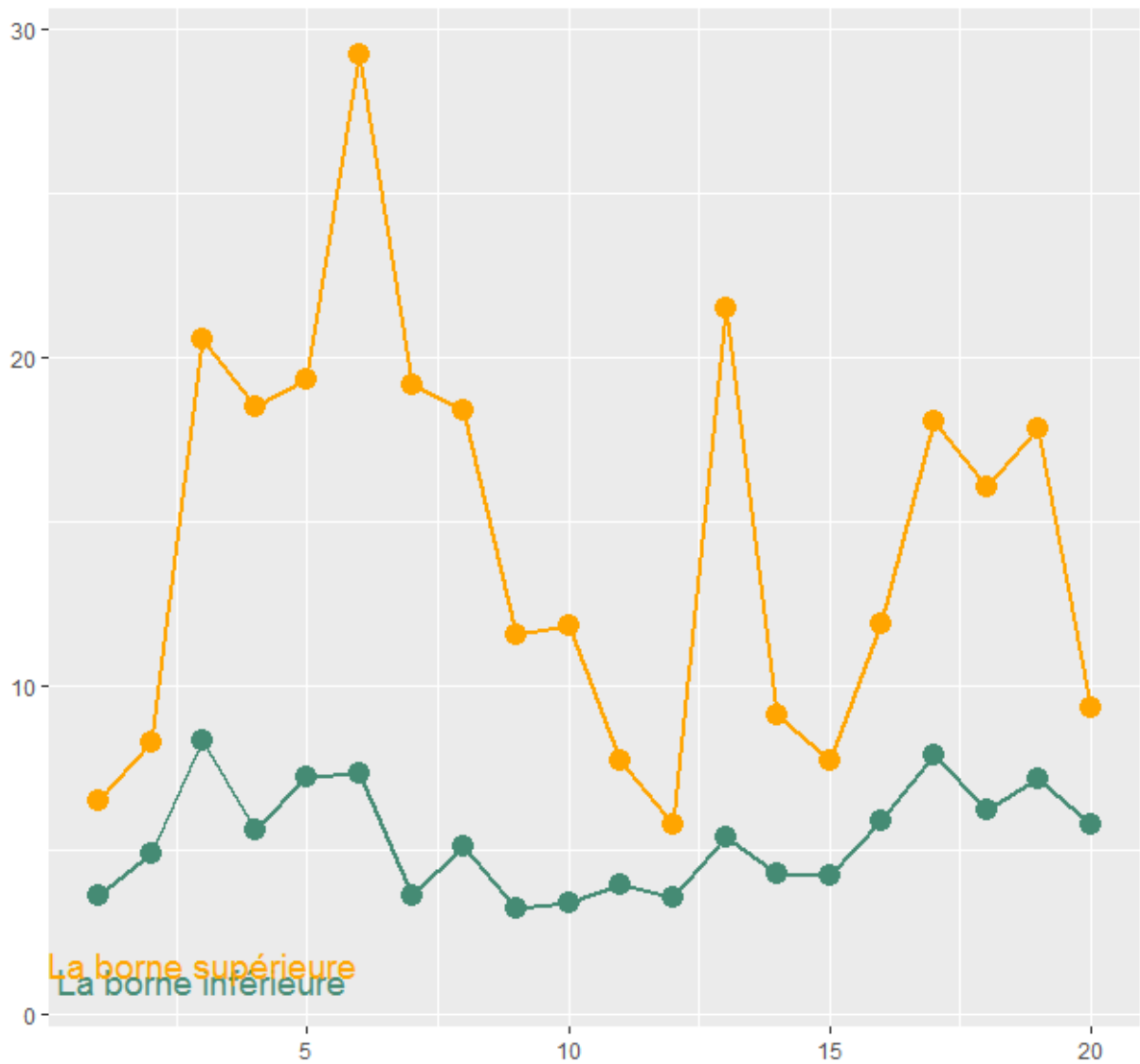
	y	y_bar_piar	var_ybar_piar	var_est_ybar_piar	ICinf	ICsup	longueur
1	1	7.614062	8.418552	3.503123	4.377957	10.850168	6.472211
2	2	5.847355	8.418552	1.123465	4.014725	7.679985	3.665260
3	3	9.972446	8.418552	4.459080	6.321398	13.623495	7.302097
4	4	9.294517	8.418552	2.363257	6.636544	11.952489	5.315945
5	5	9.400690	8.418552	3.079476	6.366565	12.434815	6.068249
6	6	5.597798	8.418552	1.001227	3.867738	7.327859	3.460120
7	7	6.317811	8.418552	2.781301	3.434318	9.201305	5.766987
8	8	10.057432	8.418552	14.130251	3.558082	16.556782	12.998700
9	9	7.680308	8.418552	2.780436	4.797263	10.563353	5.766091
10	10	13.118866	8.418552	11.571166	7.237428	19.000304	11.762877
11	11	7.412098	8.418552	2.127818	4.889999	9.934198	5.044199
12	12	9.984297	8.418552	14.567599	3.385132	16.583462	13.198330
13	13	7.659978	8.418552	2.383979	4.990378	10.329578	5.339200
14	14	13.483493	8.418552	25.352385	4.777779	22.189207	17.411428
15	15	6.514140	8.418552	1.329609	4.520453	8.507827	3.987374
16	16	10.119876	8.418552	21.266580	2.146471	18.093281	15.946810
17	17	8.485041	8.418552	4.552676	4.795874	12.174208	7.378335
18	18	11.938174	8.418552	22.038789	3.821299	20.055049	16.233750
19	19	7.122026	8.418552	2.829511	4.213649	10.030403	5.816754
20	20	6.671904	8.418552	2.911323	3.721780	9.622027	5.900247

La moyenne des $\hat{y}_{piar}=8.714616$

PARTIE III : REALISATION DES SONDAGES

c-Représentation des intervalles de confiance :

**les 20 intervalles de confiance
PIAR**



PARTIE III : REALISATION DES SONDAGES

III -Sondage stratifié :

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.

On a intérêt à découper la population selon une variable auxiliaire qui est la plus corrélée avec 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 « occupation » est la plus corrélée par rapport à toute autre variable auxiliaire.

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

Nous avons eu comme réponses à cette question :

Elève/étudiant (54 réponses)

Employer (34 réponses)

Pas d'activité (12 réponses)

On peut donc visualiser les 3 strates :

> strate1

```
[1] 4 3 28 11 4 6 14 10 8 10 12 6 3 15 15 5 5 4 12 2 7 4 3
5 5 2 7 3 3 3 3 3 7 10 15 10 21 14 4 14 24 42 20 16 14
[47] 4 14 14 6 2 6 4 2
```

> strate2

```
[1] 17 11 6 28 6 6 4 3 14 10 20 3 3 3 3 3 5 5 20 6 5 6 4 8 3 5 15
```

>strate3

```
[1] 5 3 5 5 4 7 7 7 8 3 12 1
```

Propriétés de chacune des strates :

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

<pre>>ybar_strates[1][1] 9.0 92593 >ybar_strates[2] [1] 10.20588 >ybar_strates[3] [1] 3.416667</pre>	<pre>sigma2_hc[1] [1] 58.12334 >sigma2_hc[2] [1] 188.7745 >sigma2_hc[3] [1] 0.8106061</pre>	<pre>>sigma2_h[1] [1] 57.54211 >sigma2_h[2] [1] 186.8868 >sigma2_h[3] [1] 0.8025</pre>
---	---	---

PARTIE III : REALISATION DES SONDAGES

Réalisation des sondages et traitement des résultats

Ce qui diffère entre ces types de sondages stratifiés est le nombre d'individus tiré de chaque strate qui influe directement sur la précision du sondage.

1- Les strates de même taille :

a-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.

b-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.

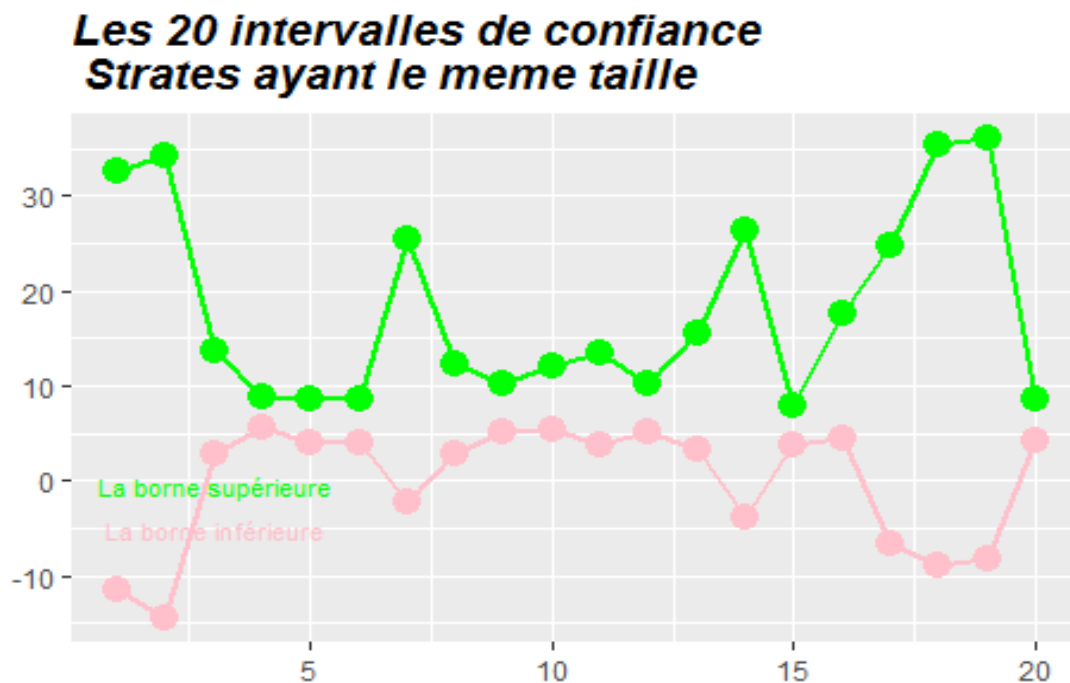
Tous ces résultats ont été regroupés dans le même tableau :

	\bar{y}_{hat}	$\text{var}_{\bar{y}_{\text{hat}} \text{ strat}}$	$\text{var}_{\text{estim}} \bar{y}_{\text{hat}} \text{ strat}$	$\text{ic}_{\text{inf}} \text{ strat}$	$\text{ic}_{\text{sup}} \text{ strat}$	$\text{long}_{\text{ic}} \text{ strat}$
1	8.360000	4.590114	2.4713592	3.516136	13.203864	9.687728
2	11.282857	4.590114	11.1809388	-10.631783	33.197497	43.829280
3	6.340000	4.590114	1.0038367	4.372480	8.307520	3.935040
4	10.200000	4.590114	2.9821837	4.354920	16.045080	11.690160
5	7.822857	4.590114	1.8388980	4.218617	11.427097	7.208480
6	9.814286	4.590114	3.3946000	3.160870	16.467702	13.306832
7	10.334286	4.590114	11.5191469	-12.243242	32.911814	45.155056
8	8.914286	4.590114	1.8672000	5.254574	12.573998	7.319424
9	7.128571	4.590114	0.8782163	5.407267	8.849875	3.442608
10	8.888571	4.590114	1.8614939	5.240043	12.537099	7.297056
11	10.148571	4.590114	13.6917714	-16.687301	36.984443	53.671744
12	5.965714	4.590114	0.5922694	4.804866	7.126562	2.321696
13	6.780000	4.590114	1.2087061	4.410936	9.149064	4.738128
14	8.442857	4.590114	2.2606041	4.012073	12.873641	8.861568
15	9.014286	4.590114	1.9316571	5.228238	12.800334	7.572096
16	5.357143	4.590114	0.9285796	3.537127	7.177159	3.640032
17	12.742857	4.590114	10.8038490	-8.432687	33.918401	42.351088
18	10.534286	4.590114	10.4578163	-9.963034	31.031606	40.994640
19	10.094286	4.590114	1.2094735	7.723718	12.464854	4.741136
20	6.417143	4.590114	1.1110122	4.239559	8.594727	4.355168

La moyenne des $\bar{y}_{\text{strate}} = 9.512143$

PARTIE III : REALISATION DES SONDAGES

c-Représentation des intervalles de confiance



2/ Les strates sont à allocation proportionnelles :

a-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) : $nhNh = n N = f$
Par suite :

$$n_1=11; \quad n_2=7 \quad ; \quad n_3=2$$

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.

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 estimée pour le calcul de l'intervalle de confiance.

PARTIE III : REALISATION DES SONDAGES

b-Résultats obtenus

Tous ces résultats ont été regroupés dans le même tableau :

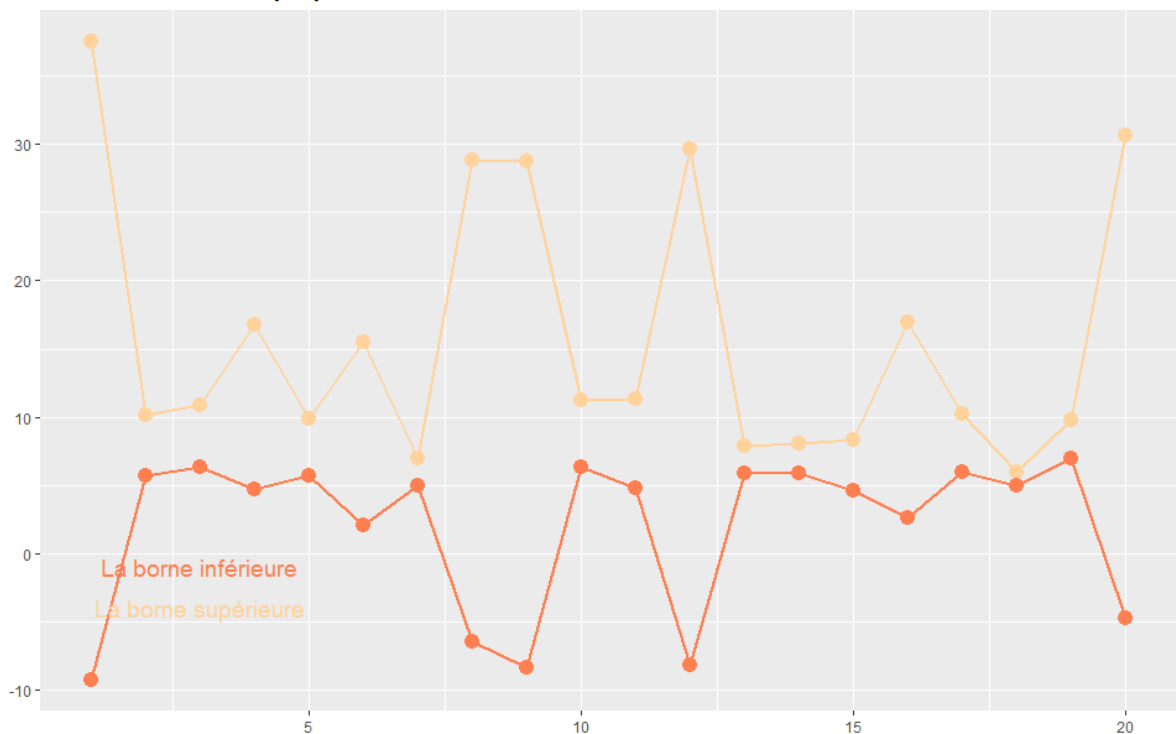
	\bar{y}_{hat}	$\text{var}_{\bar{y}_{\text{hat}}_{\text{prop}}}$	$\text{var}_{\text{estim}_{\bar{y}_{\text{hat}}_{\text{strat}}}$	$\text{ic}_{\text{inf}_{\text{prop}}}$	$\text{ic}_{\text{sup}_{\text{prop}}}$	$\text{long}_{\text{ic}_{\text{prop}}}$
1	7.228052	3.707437	1.5957880	4.468935	9.987169	5.518235
2	9.618442	3.707437	3.1504591	4.171298	15.065585	10.894287
3	12.322338	3.707437	10.3995857	-5.658546	30.303221	35.961767
4	7.656364	3.707437	0.9782693	5.964936	9.347791	3.382855
5	8.014545	3.707437	3.2647529	2.369788	13.659303	11.289515
6	8.174545	3.707437	1.8882095	4.909831	11.439260	6.529428
7	7.120519	3.707437	0.9920564	5.405254	8.835785	3.430531
8	6.321558	3.707437	1.7038154	3.375662	9.267455	5.891794
9	9.092727	3.707437	3.5512181	2.952671	15.232783	12.280112
10	8.452727	3.707437	1.1742176	6.422505	10.482949	4.060444
11	5.962597	3.707437	1.1158935	4.033218	7.891977	3.858760
12	7.950909	3.707437	1.6570712	5.085833	10.815985	5.730152
13	13.007013	3.707437	11.7033624	-7.228101	33.242127	40.470227
14	6.041558	3.707437	0.9407697	4.414968	7.668149	3.253182
15	11.685714	3.707437	11.2769584	-7.812147	31.183575	38.995722
16	10.331948	3.707437	3.7763765	3.802593	16.861303	13.058710
17	8.232468	3.707437	1.5982126	5.469158	10.995777	5.526619
18	9.137403	3.707437	3.1124161	3.756035	14.518770	10.762735
19	8.813247	3.707437	2.1910872	5.024857	12.601636	7.576779
20	7.474545	3.707437	1.0970479	5.577750	9.371341	3.793592

La moyenne des $\bar{y}_{\text{prop}}=8.891$

PARTIE III : REALISATION DES SONDAGES

c-Représentation des intervalles de confiance

Les 20 intervalles de confiance
Strates à allocation proportionnelle



3-Les strates sont à allocation optimale:

a-Sélection de l'échantillon

Dans ce type de sondage le nombre d'individus tiré de chaque strate (n_h) est, non seulement proportionnel 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 : $n_1=9$; $n_2=10$; $n_3=1$

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

b-Résultats obtenus

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

PARTIE III : REALISATION DES SONDAGES

Voici le tableau regroupant tous ces résultats :

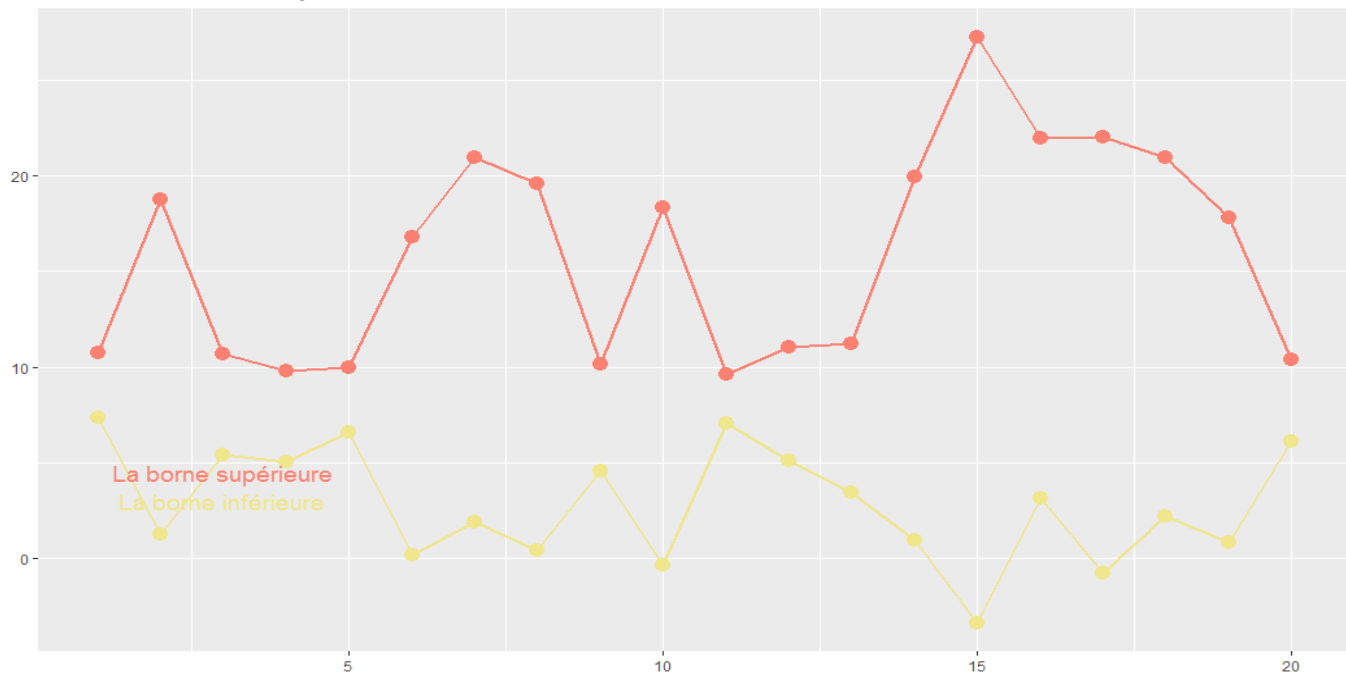
	ybar_hat	var_ybar_hat_opt	var_estim_ybar_hat_opt	ic_inf_strat	ic_sup_strat	long_ic
1	9.054	3.752165	0.9687493	7.3790324	10.728968	3.349935
2	10.000	3.752165	5.0503667	1.2679160	18.732084	17.464168
3	8.068	3.752165	1.5279240	5.4262194	10.709781	5.283561
4	7.446	3.752165	1.3774293	5.0644247	9.827575	4.763151
5	8.298	3.752165	0.9839373	6.5967724	9.999228	3.402455
6	8.486	3.752165	4.7954527	0.1946623	16.777338	16.582675
7	11.452	3.752165	5.4950940	1.9509825	20.953018	19.002035
8	10.044	3.752165	5.5360260	0.4722110	19.615789	19.143578
9	7.372	3.752165	1.6147873	4.5800327	10.163967	5.583935
10	9.010	3.752165	5.3907067	-0.3105318	18.330532	18.641064
11	8.358	3.752165	0.7349173	7.0873279	9.628672	2.541344
12	8.104	3.752165	1.7087827	5.1495148	11.058485	5.908970
13	7.366	3.752165	2.2425760	3.4885861	11.243414	7.754828
14	10.440	3.752165	5.4875200	0.9520779	19.927922	18.975844
15	11.940	3.752165	8.8272900	-3.3223844	27.202384	30.524769
16	12.548	3.752165	5.4388040	3.1443079	21.951692	18.807384
17	10.622	3.752165	6.5758040	-0.7475651	21.991565	22.739130
18	11.596	3.752165	5.4088960	2.2440188	20.947981	18.703962
19	9.320	3.752165	4.9039967	0.8409898	17.799010	16.958020
20	8.294	3.752165	1.2340260	6.1603690	10.427631	4.267262

La moyenne des $\bar{y}_{opt} = 9.3909$

PARTIE III : REALISATION DES SONDAGES

c-Représentation des intervalles de confiance

les 20 intervalles de confiance
Strates à allocation optimale



PARTIE III : REALISATION DES SONDAGES

IV- Sondage par grappes et à deux degrés

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.

a- Sélection des échantillons

Nous avons partitionné la population selon la variable auxiliaire «Age» qui présente 4 modalités :

- [0,15[: Grappe 1
- [15,30[: Grappe 2
- [30,45[: Grappe 3
- [45,...[: Grappe 4

	grappe1	grappe2	grappe3	grappe4
moyenne	8.058824	10.40385	8	6.833333
taille de grappe	17.000000	52.00000	13	18.000000

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

b-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.

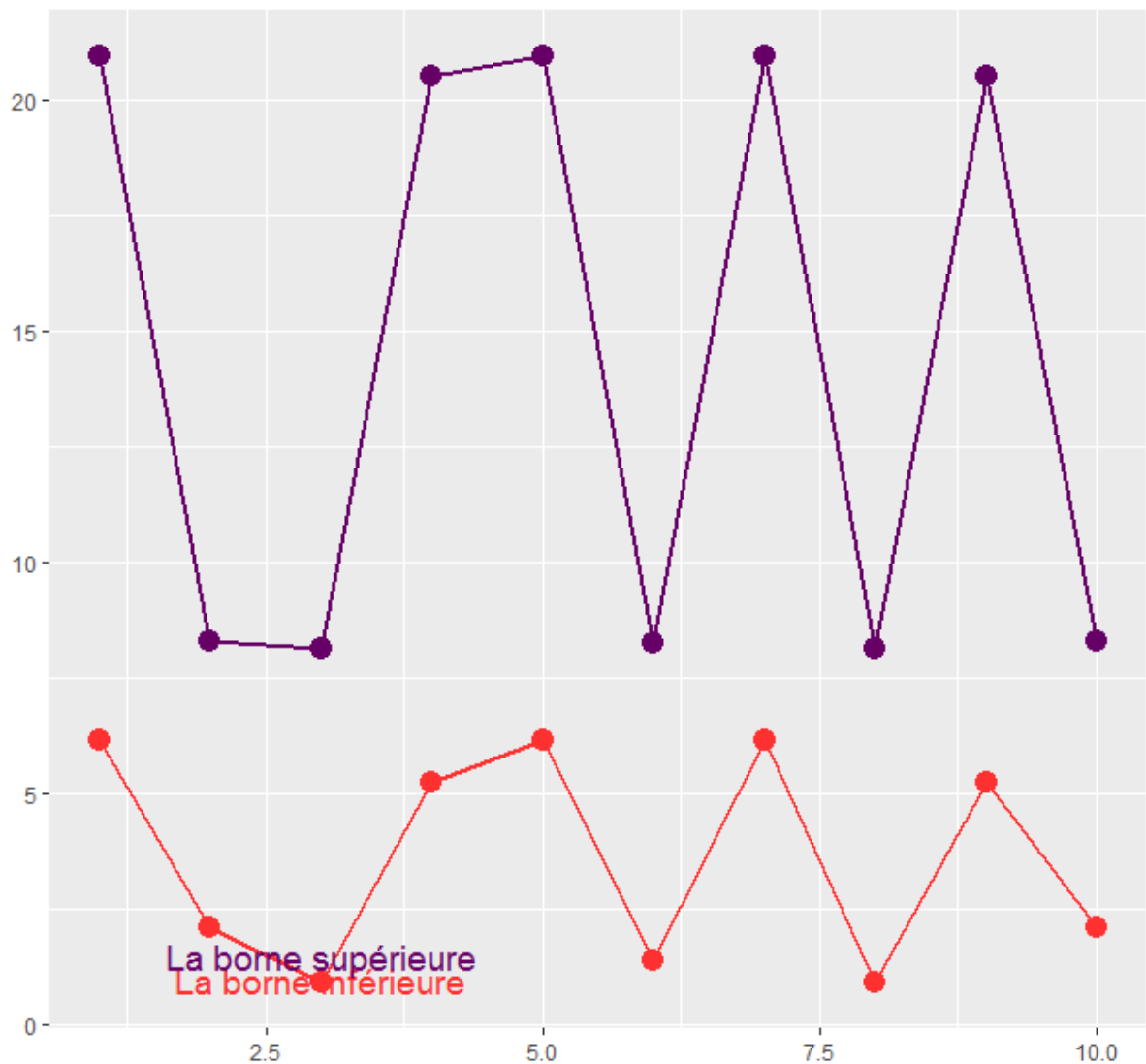
	\bar{y}_{grappe}	$\text{var}_{\text{ygrap}}$	$\text{icinf}_{\text{grappe}}$	$\text{icsup}_{\text{grappe}}$	$\text{longIC}_{\text{grap}}$	$\text{num}_{1\text{ere_grap_selct}}$	$\text{num}_{2\text{ere_grap_selct}}$
1	13.56	14.271083	6.1556909	20.964309	14.808618	1	2
2	5.20	2.483483	2.1112220	8.288778	6.177556	1	4
3	4.54	3.414083	0.9184613	8.161539	7.243077	3	4
4	12.90	15.201683	5.2580901	20.541910	15.283820	3	2
5	13.56	14.271083	6.1556909	20.964309	14.808618	1	2
6	4.82	3.054750	1.3943427	8.245657	6.851315	3	1
7	13.56	14.271083	6.1556909	20.964309	14.808618	1	2
8	4.54	3.414083	0.9184613	8.161539	7.243077	4	3
9	12.90	15.201683	5.2580901	20.541910	15.283820	2	3
10	5.20	2.483483	2.1112220	8.288778	6.177556	4	1

PARTIE III : REALISATION DES SONDAGES

La moyenne des $\bar{Y}_{\text{grappe}} = 9,078$

c-Représentation des intervalles de confiance :

**les 10 intervalles de confiance
par grappes**



PARTIE III : REALISATION DES SONDAGES

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.

a-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 $n_h \sim N_h/5$.

b-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 :

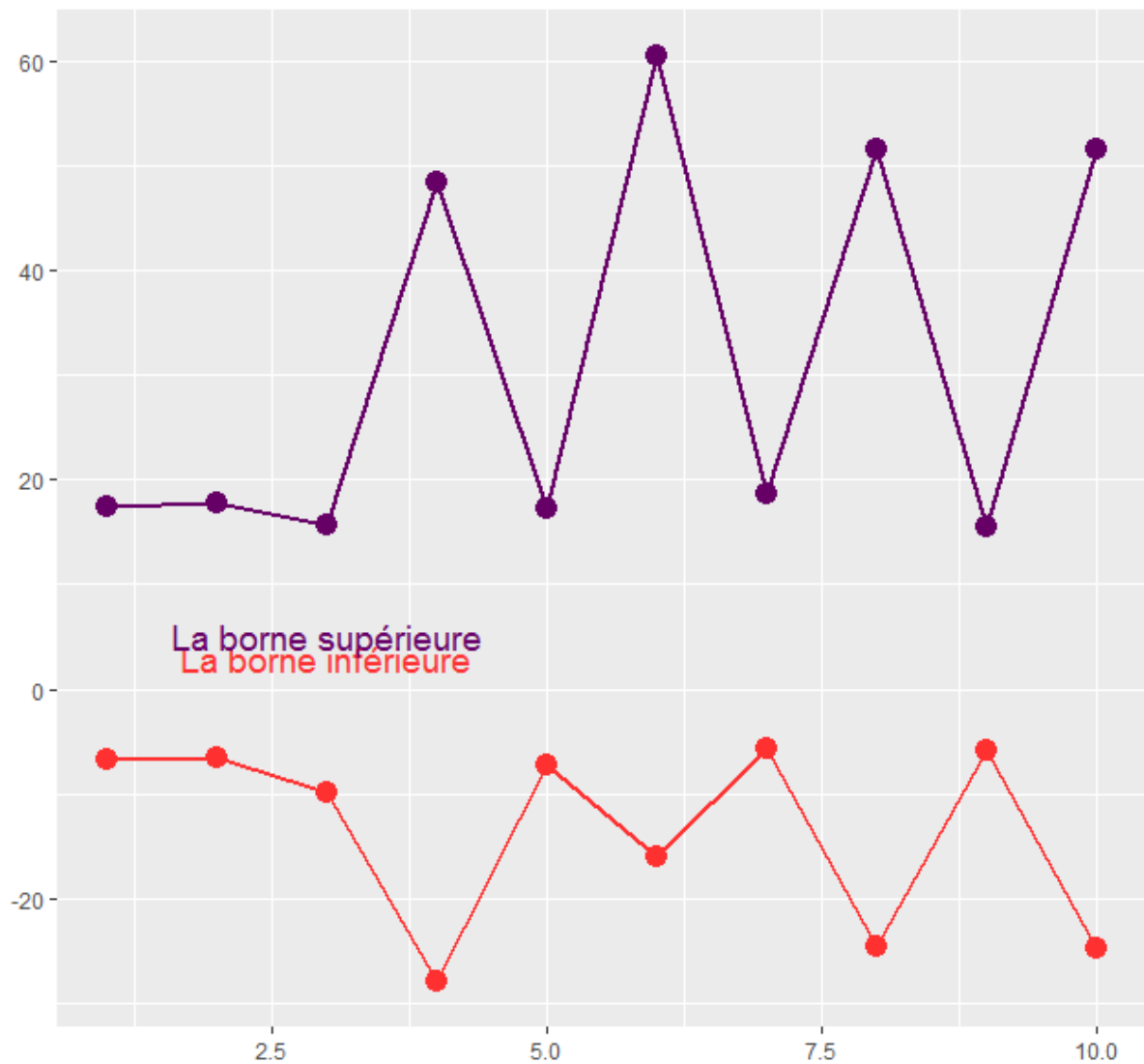
	$\bar{y}_{\text{bar_grappe2}}$	$\text{var_y_bar_grap_2dg}$	icinf_2dg	icsup_2dg	longIC_2dg	$\text{num_1ere_grap_select}$	$\text{num_2ere_grap_select}$	$\text{nbre_ind_select_G1}$	$\text{nbre_ind_select_G2}$
1	5.353333	38.47146	-6.803639	17.51031	24.31394	1	3	3	3
2	5.533333	38.57760	-6.640396	17.70706	24.34746	3	1	3	3
3	2.956667	42.31233	-9.792727	15.70606	25.49879	4	1	4	3
4	10.226667	378.74359	-27.917552	48.37088	76.28844	2	3	10	3
5	4.946667	39.00480	-7.294282	17.18762	24.48190	1	3	3	3
6	22.253333	380.44658	-15.976545	60.48321	76.45976	2	1	10	3
7	6.503333	38.53233	-5.663252	18.66992	24.33317	1	4	3	4
8	13.502667	376.09159	-24.507772	51.51311	76.02088	2	3	10	3
9	4.870000	29.57974	-5.789903	15.52990	21.31981	3	4	3	4
10	13.404000	377.36846	-24.670909	51.47891	76.14982	4	2	4	10

La moyenne des $\bar{Y}_{2\text{deg}}=8,955$

PARTIE III : REALISATION DES SONDAGES

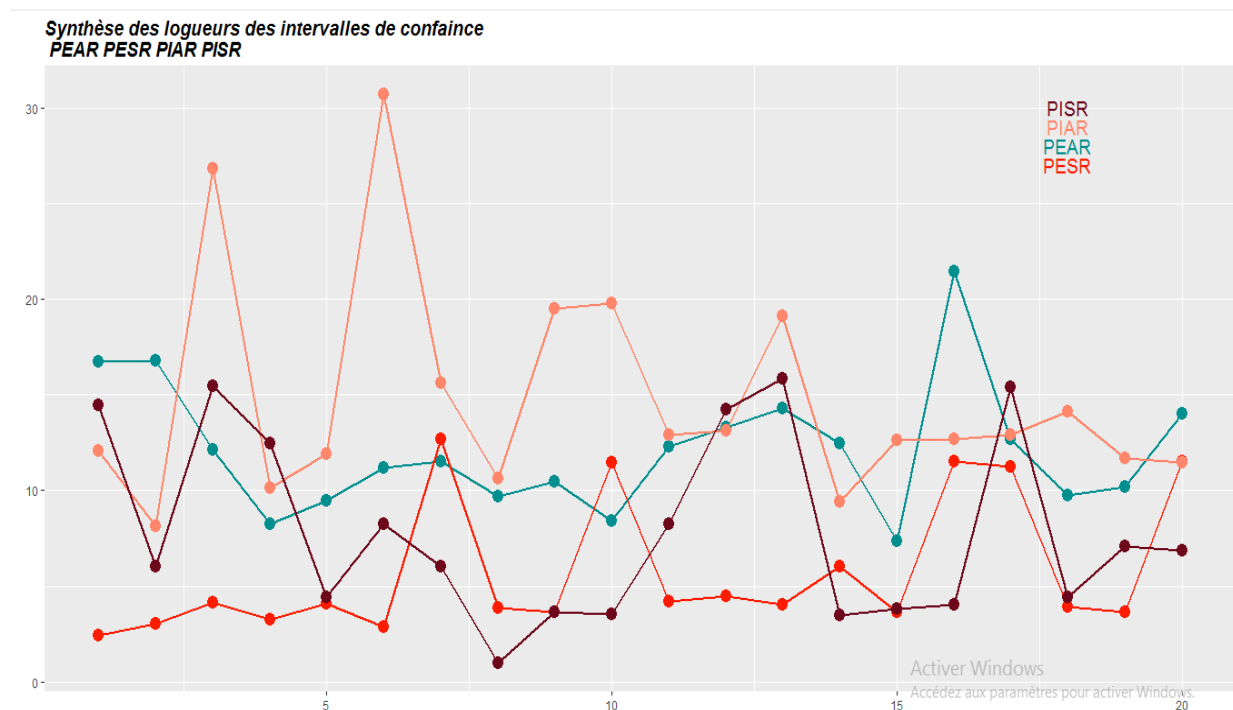
c-Représentation des intervalles de confiance :

***les 10 intervalles de confiance
à 2 degré***



PARTIE IV : SYNTHÈSE

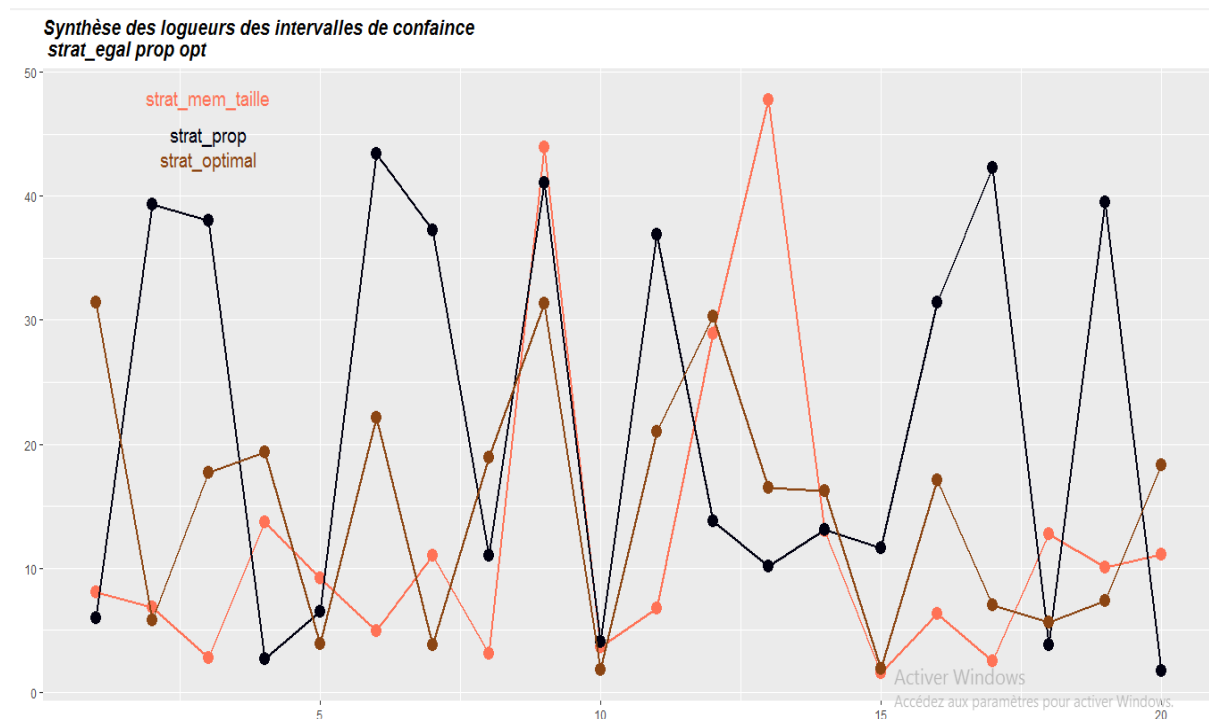
I-Sondages aléatoires PESR, PEAR, PISR PIAR et PIAR



- Ce résultat récapitulatif des sondages à probabilité égale et inégale nous montre que la méthode de sondage aléatoire simple sans remise (PESR) est la plus précise et que celui à probabilité inégale avec remise (PIAR) est le moins précis. En effet, on remarque que la longueur de l'intervalle de confiance pour cette dernière méthode présente plusieurs piques alors que pour le PESR l'alignement de sa longueur est plus stable.
- Ceci est expliqué par le fait que dans la méthode à probabilités inégales on a recourt à utiliser une variables auxiliaire, puisque on manque d'information sur la variable d'intérêt, par suite elle n'est pas aussi fortement corrélée avec la VI que l'on croyait. De plus, il est bien évident que l'échantillon de SAS représente mieux la population alors il lui donne un avantage sur la précision malgré la longueur des procédures effectuer.

PARTIE IV : SYNTHÈSE

II- Stratification

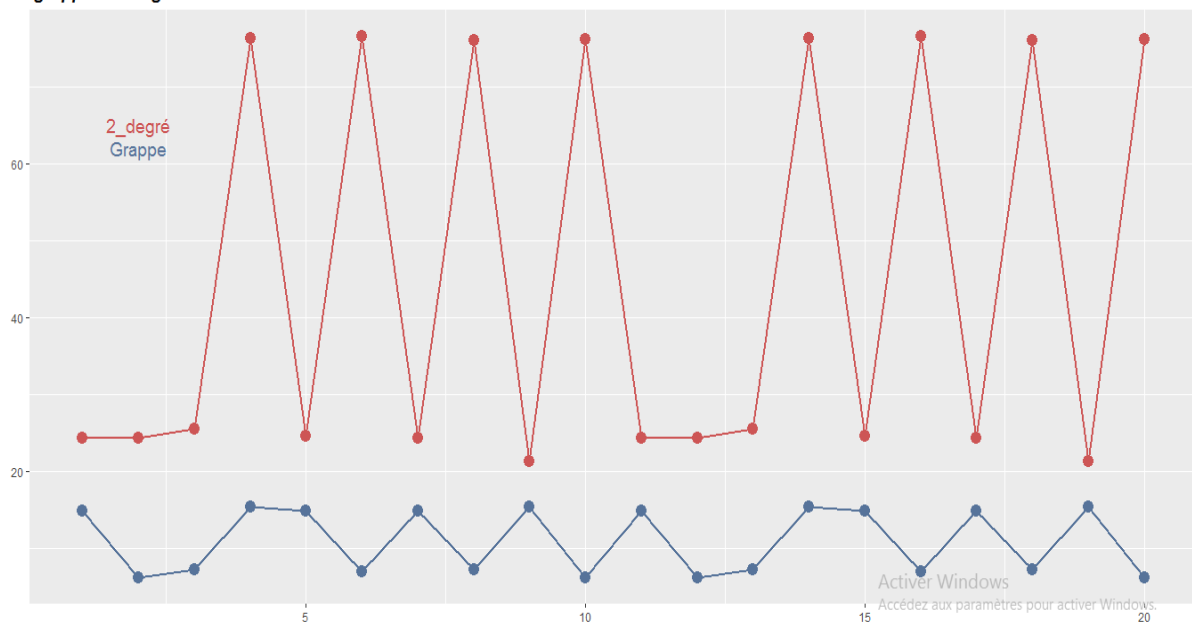


- En cas de stratification, les résultats ne sont pas d'une telle importance car pour les trois méthodes on observe l'instabilité de la longueur de l'intervalle de confiance. La meilleure méthode est la stratification à même taille puis la stratification optimale et en dernière position la stratification à allocation proportionnelle
- La deuxième méthode pour ce plan de sondage est meilleure que la troisième car dans celui-ci on prend en considération, non seulement le nombre d'individus de chaque strate, mais aussi de la variance au sein de chaque strate.

PARTIE IV : SYNTHÈSE

III-Sondage par grappes et à deux degrés

Synthèse des logueurs des intervalles de confiance
grappe et 2 degré



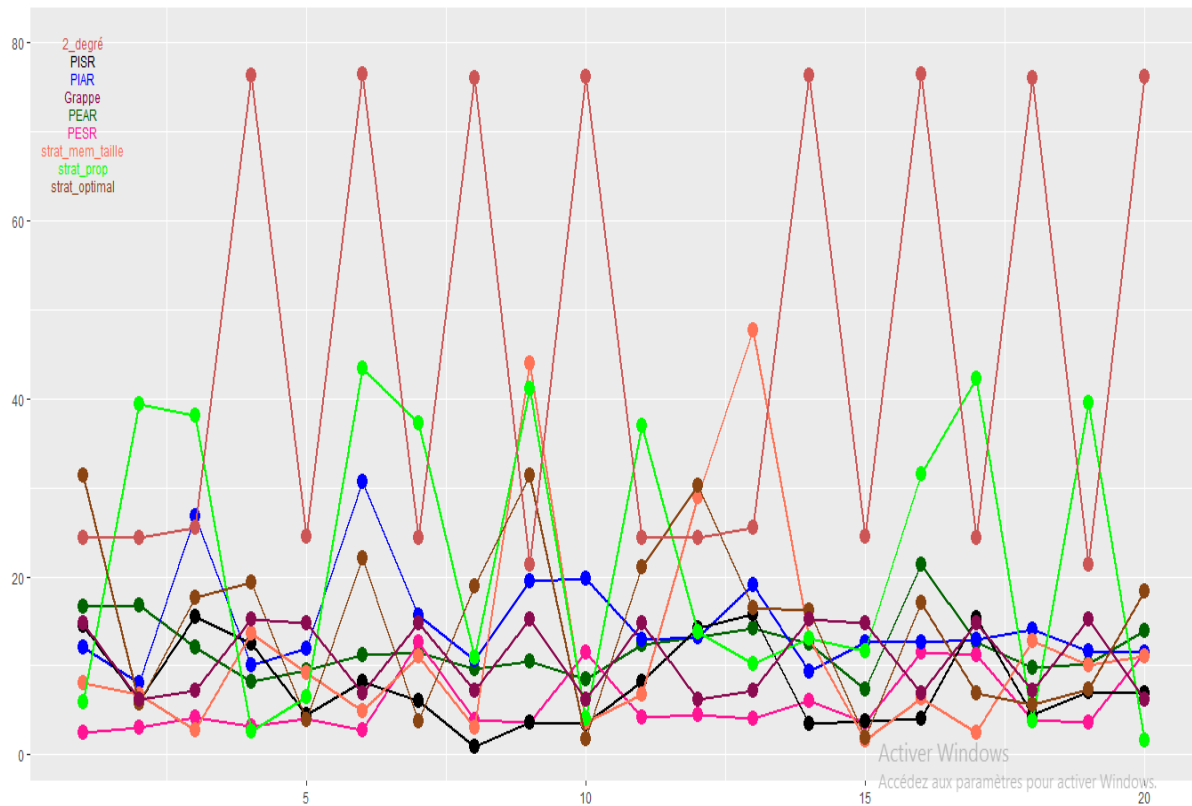
- On peut dire que le sondage par grappes est beaucoup plus précis que celui à deux degrés. Ce qui est expliqué par la grande différence entre les longueurs des intervalles de confiance. En effet, dans le deuxième on ne sélectionne qu'un échantillon de chaque grappe tirée. Ceci explique la perte en précision qui est, bien entendu, compensée par un gain en termes de coût.
- On sait bien aussi que la méthode de grappe réduit le déplacement (on sélectionne m grappes parmi M par un échantillonnage SAS) donc ce qui lui met plus en relief.

PARTIE IV : SYNTHÈSE

IV-Comparaison de tous les types de sondages utilisés

TYPE DE SONDAGE	PESR	PEAR	PISR	PIAR	STRAT MEME TAILL E	STRAT_P ROP	STRAT_O PT	GRAPP E	2_DEGRE
\bar{Y}_{bar}	9.092	8.93	8.731	8.714	9.512	8.891	9.3909	9.078	8.955
MEAN des long_IC	5.781	12.118	7.937	14.766	12.39	21.686	14.873	10.868	44.921

Synthèse des logueurs des intervalles de confiance



PARTIE IV : SYNTHÈSE

- en observant ce dernier output et le tableau récapitulatif en dessus, on peut conclure finalement que le sondage à deux degré est la pire méthode. Il est fréquemment utilisé pour une raison de coût mais en outre il manque de précision.
- En deuxième position, on trouve que le sondage stratifié manque aussi de précision
- Puis en première position, les sondages sans remise à probabilités égales et inégales sont les meilleurs en termes de précision. (ils correspondent aux longueurs minimales d'intervalles de confiance).

Classement des sondage selon la méthode utilisée :

-PESR
-PISR
-GRAPPE
-PEAR
-STRAT_MEME_TAILLE
-PIAR
-STRAT_OPT
-STRAT_PROP
-2_DEGRE

Limite de ce projet:

-en jetant un œil sur la base de données, on remarque qu'elle est constituée plus de variables qualitatives que de variables quantitatives. Ceci rend la vision plus difficile.

De plus, la variable d'intérêt "nombre d'heure d'activité sportive par semaine" doit être **fortement corrélée** avec les variables auxiliaires. Or pour notre projet on ne peut pas trouver une qui expliquera mieux la VI que les variables "l'âge" ou "occupation" ce manque de forte corrélation influencera par la suite sur la variance c'est à dire la précision et les longueurs des intervalles de confiance ce qui défavorise alors le sondage stratifié.

PARTIE V : REDRESSEMENT

Dans la théorie des sondages, le redressement constitue un point très important. Il arrive souvent que l'on possède un certain nombre de variables qualitatives, dites auxiliaires, dont on connaît la répartition sur l'ensemble de la population étudiée. On veut alors se servir de ces variables afin de redresser un sondage et ainsi d'obtenir des répartitions comparables pour la population et l'échantillon.

Ici, on va choisir la variable auxiliaire « genre » pour le redressement.

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

Notre échantillon :

> `echantillon`

[1] 4 3 6 6 8 12 3 5 3 4 8 3 7 10 15 3 4 16 6 2

L'estimateur de la moyenne avant redressement est: 6.4.

la proportion des femme dans la pop = 0.3, alors que la proportion des femme dans l'échantillon=0.25.

Pour les hommes , leur proportion dans la population est de 0.7 alors que dans l'échantillon elle est de 0.75.

Le poids des femmes dans notre échantillon est de($0.3/0.25=1.2$)

L'estimateur de la moyenne après redressement est égale a 5.973333

On remarque tous les individus ont le même poids (P/p_{ech}) alors que sans redressement leur poids est de (N/n) .

PARTIE IV : ANNEXE

#####CHARGEMENT DE LA BASE

```
sport=read.csv('C:/Users/Msehli/Desktop/sondage/projet-sondage.csv',header=TRUE,sep=";")
attach(sport)
View(sport)
y=sport[,3]
y
mean(y)
N=100
n=20
sport1=read.csv('C:/Users/Msehli/Desktop/sondage/sondage2.csv',header=TRUE,sep=";")
sport1
```

#####ETUDE DESCRIPTIVE#####

###selon la var d'interet

```
library(ggplot2)
library(scales)
t1=ggplot(sport1,aes(x="",fill=factor(sport1$nbre_h_semaine)))+coord_polar(theta="y")+
+ geom_bar(width = 1)
t3=labs(x="",y="",fill="intervalle de temps d'activité physique")
c2=scale_fill_brewer("",palette="Pastel2")
k1=annotate("text",label=c(sum(sport1$nbre_h_semaine=="[0,3[",sum(sport1$nbre_h_semaine=="[3,8[",sum(sport1$nbre_h_semaine=="[8,12[",sum(sport1$nbre_h_semaine=="[12,...[")"),x=c(1,1,1,1),y=c(97, 40,5,81),size=8 )
t1+t3+c2+k1+ggtitle("la repartition de la population selon la variable d'interet : \n nombre d'heure par semaine ") +theme(plot.title = element_text(color="black", size=16, face = "italic" ))
```

###selon les variables auxiliaires

##genre

```
t1=ggplot(sport1,aes(x="",fill=factor(sport1$genre)))+coord_polar(theta="y")+
geom_bar(width = 1)
t3=labs(x="",y="",fill="Genre")
c2=scale_fill_brewer("",palette="Pastel")
k1=annotate("text",label=c(sum(sport1$genre=="F"),sum(sport1$genre=="H")),x=c(1,1),y=c(85,35),size=8)
t1+t3+c2+k1+ggtitle("la repartition de la population selon le sexe ") +theme(plot.title = element_text(color="black", size=16, face="italic"))
qplot(sport$genre, data =sport1, geom = "histogram", fill = sport$nbre_h_semaine ,main="hist")
##age
```

```
t1=ggplot(sport1,aes(x="",fill=factor(sport1$age)))+coord_polar(theta="y")+
geom_bar(width = 1)
t3=labs(x="",y="",fill="Genre")
c2=scale_fill_brewer("",palette="Pastel2")
k1=annotate("text",label=c(sum(sport1$age=="[0,15[)",sum(sport1$age=="[15,30[)",sum(sport1$age=="[30,45[)",sum(sport1$age=="[45,...[)"),x=c(1,1,1,1),y=c(25,50,90,9),size=8)
t1+t3+k1+c2+ggtitle("la repartition de la population selon la variable \n age")
)+theme(plot.title = element_text(color="black", size=16, face="italic"))
```

##raison

```
t1=ggplot(sport,aes(x="",fill=factor(sport$raison)))+coord_polar(theta="y")+
geom_bar(width = 1)
t3=labs(x="",y="",fill="Genre")
c2=scale_fill_brewer("",palette="Pastel1")
k1=annotate("text",label=c(sum(sport$raison=="athlète"),sum(sport$raison=="divertissement"),sum(sport$raison=="maladie"),sum(sport$raison=="surpoids"),sum(sport$raison=="autre")),x=c(1,1,1,1,1),y=c(88,50,13,5,73),size=8)
t1+t3+k1+c2+ggtitle("la repartition de la population selon la variable \n raison")
)+theme(plot.title = element_text(color="black", size=16, face="italic"))
```

##type de sport

```
t1=ggplot(sport,aes(x="",fill=factor(sport$type_sport)))+coord_polar(theta="y")+
geom_bar(width = 1)
t3=labs(x="",y="",fill="Genre")
c2=scale_fill_brewer("",palette="Pastel1")
k1=annotate("text",label=c(sum(sport$type_sport=="individuel"),sum(sport$type_sport=="équipe")),x=c(1,1),y=c(25,75),size=8)
t1+t3+k1+c2+ggtitle("la repartition de la population selon la variable \n type_sport")
)+theme(plot.title = element_text(color="black", size=16, face="italic"))
```

##occupation

```
t1=ggplot(sport,aes(x="",fill=factor(sport$occupation)))+coord_polar(theta="y")+
geom_bar(width = 1)
t3=labs(x="",y="",fill="Genre")
c2=scale_fill_brewer("",palette="Pastel1")
k1=annotate("text",label=c(sum(sport$occupation=="élève/etudiant"),sum(sport$occupation=="employeur"),sum(sport$occupation=="pas d'activité")),x=c(1,1,1),y=c(75,30,6),size=8)
t1+t3+k1+c2+ggtitle("la repartition de la population selon la variable \n occupation")
)+theme(plot.title = element_text(color="black", size=16, face="italic"))
```

##genre

```
t1=ggplot(sport1,aes(x="",fill=factor(sport1$genre)))+coord_polar(theta="y")+
geom_bar(width = 1)
t3=labs(x="",y="",fill="Genre")
c2=scale_fill_brewer("",palette="Pastel")
k1=annotate("text",label=c(sum(sport1$genre=="F"),sum(sport1$genre=="H")),x=c(1,1),y=c(85,35),size=8)
```

```

t1+t3+c2+k1+ggtitle("la repartition de la population selon le sexe")+theme(plot.title =
element_text(color="black", size=16, face="italic"))
qplot(sport$genre, data =sport1, geom = "histogram", fill = sport$nbre_h_semaine
,main="hist")
##age
t1=ggplot(sport1,aes(x="",fill=factor(sport1$age)))+coord_polar(theta="y")+
geom_bar(width = 1)
t3=labs(x="",y="",fill="Genre")
c2=scale_fill_brewer("",palette="Pastel2")
k1=annotate("text",label=c(sum(sport1$age=="[0,15[",sum(sport1$age=="[15,30[",su
m(sport1$age=="[30,45[",sum(sport1$age=="[45,...[")"),x=c(1,1,1,1),y=c(25,50,90,9),siz
e=8)
t1+t3+k1+c2+ggtitle("la repartition de la population selon la variable \n age
")+theme(plot.title = element_text(color="black", size=16, face="italic"))
##raison
t1=ggplot(sport,aes(x="",fill=factor(sport$raison)))+coord_polar(theta="y")+
geom_bar(width = 1)
t3=labs(x="",y="",fill="Genre")
c2=scale_fill_brewer("",palette="Pastel1")
k1=annotate("text",label=c(sum(sport$raison=="athlète"),sum(sport$raison=="divertis
sement"),sum(sport$raison=="maladie"),sum(sport$raison=="surpoids"),sum(sport$rai
son=="autre")),x=c(1,1,1,1,1),y=c(88,50,13,5,73),size=8)
t1+t3+k1+c2+ggtitle("la repartition de la population selon la variable \n raison
")+theme(plot.title = element_text(color="black", size=16, face="italic"))
##type de sport
t1=ggplot(sport,aes(x="",fill=factor(sport$type_sport)))+coord_polar(theta="y")+
geom_bar(width = 1)
t3=labs(x="",y="",fill="Genre")
c2=scale_fill_brewer("",palette="Pastel1")
k1=annotate("text",label=c(sum(sport$type_sport=="individuel"),sum(sport$type_sport
=="équipe")),x=c(1,1),y=c(25,75),size=8)
t1+t3+k1+c2+ggtitle("la repartition de la population selon la variable \n type_sport
")+theme(plot.title = element_text(color="black", size=16, face="italic"))
##occupation
t1=ggplot(sport,aes(x="",fill=factor(sport$occupation)))+coord_polar(theta="y")+
geom_bar(width = 1)
t3=labs(x="",y="",fill="Genre")
c2=scale_fill_brewer("",palette="Pastel1")
k1=annotate("text",label=c(sum(sport$occupation=="élève/etudiant"),sum(sport$occu
pation=="employer"),sum(sport$occupation=="pas
d'activité")),x=c(1,1,1),y=c(75,30,6),size=8)
t1+t3+k1+c2+ggtitle("la repartition de la population selon la variable \n occupation
")+theme(plot.title = element_text(color="black", size=16, face="italic"))

```

```
#####ECHANTILLONNAGE #####
```

```
N=100
```

```
n=20
```

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

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

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

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

```
#####PESR#####
```

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

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

```
ybarpesr=mean(ech)
```

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

```
sc2=var(ech)
```

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

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

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

```
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])
```

```
47
```

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

```
longueur[i]=ICbsup[i]-ICbinf[i]
```

```
}
```

```
longueur
```

```
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)
```

```
g1_pesr=ggplot(res_pesr,aes(y,ICbsup))+geom_point(colour="purple",cex=4)+geom_point(aes(y,ICbinf),colour="antiquewhite4",cex=4)+geom_path(colour="purple",size=1)+geom_path(aes(y,ICbinf),colour="antiquewhite4",size=1)
```

```
g2_pesr=annotate("text", label =c("La borne inférieure","La borne supérieure"), x  
=c(3,3), y =c(1,1.5),size =5, colour = c("antiquewhite4","purple"))
```

```
g3_pesr=labs(x="",y="")
```

```
g1_pesr+g2_pesr+g3_pesr+ggtitle("les 20 intervalles de confiance \n
```

```
PESR")+theme(plot.title = element_text(color="black", size=14, face="bold.italic"))
```

```
res_pesr=res_pesr[-1]
```

```

colnames(res_pesr)=c("ybar_hat_pesr","var_ybar_hat_pesr","var_estim_ybar_hat_pesr","I
Cbinf_pesr","ICbsup_pesr","long_IC_pesr")
rownames(res_pesr)=c(1:20)
View(res_pesr)

#####PEAR#####
tirage=function(y){
ech=sample(y,n,replace=T)
ybarpear=mean(ech)
45
varybarpear=sigma2c/n
sc2=var(ech)
s2=((n-1)/n)*sc2
varestimybarpear=s2/n
v=c(ech,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_pesr=NULL
res_pesr=data.frame(y=1:20,res[,21],res[,22],res[,23],ICbinf,ICbsup,longueur)
g1_pear=ggplot(res_pesr,aes(y,ICbsup))+geom_point(colour="purple",cex=4)+geom_poi
nt(aes(y,ICbinf),colour="antiquewhite4",cex=4)+geom_path(colour="purple",size=1)+ge
om_path(aes(y,ICbinf),colour="antiquewhite4",size=1)
g1_pear
g2_pear=annotate("text", label =c("La borne inférieure","La borne supérieure"), x
=c(3,3), y =c(1,1.5),size =5, colour = c("antiquewhite4","purple"))
g3_pear=labs(x="",y="")
g1_pear+g2_pear+g3_pear+ggtitle("les 20 intervalles de confiance \n
PEAR")+theme(plot.title = element_text(color="black", size=14, face="bold.italic"))
res_pesr=res_pesr[-1]
colnames(res_pesr)=c("ybar_hat_pear","var_ybar_hat_pear","var_estim_ybar_hat_pear","
ICbinf_PEAR","ICbsup_PEAR","long_IC_PEAR")

```



```

View(res_pear)
rownames(res_pear)=c(1:20)
y2=mean(res_pear$ybar_hat_pear)
y2

#####PIAR#####
x=matrix(100,1)
for (i in 1:100) {
  (if(sport1$age[i]=="[0,15[") (x[i]=7))
  (if(sport1$age[i]=="[15,30[") (x[i]=23))
  48
  (if (sport1$age[i]=="[30,45[") (x[i]=37))
  (if (sport1$age[i]=="[45,...[") (x[i]=53))
}
sport1=cbind(sport1,x)
attach(sport1)
summary(sport1)
###proba inégales
#piar
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(j,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){
  50
  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","ech8","ech9","ec
h10","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)
colnames(ress)=c("y","y_bar_piar","var_ybar_piar","var_est_ybar_piar","ICinf","ICsup","lo
ngeur")

```

```

View(ress)
y_bar_piar=mean(ress$ress...1.)
y_bar_piar
g1_piar=ggplot(ress,aes(y,ICbsup))+geom_point(colour="orange",cex=4)+geom_point(aes(y,ICbinf),colour="aquamarine4",cex=4)+geom_path(colour="orange",size=1)+geom_path(aes(y,ICbinf),colour="aquamarine4",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("aquamarine4","orange"))
g3_piar=labs(x="",y="")
g1_piar+g2_piar+g3_piar+ggtitle("les 20 intervalles de confiance \n PIAR")+theme(plot.title = element_text(color="black", size=14, face="bold.italic"))
ress=ress[-1]
colnames(ress)=c("ybar_hat_piar","var_ybar_hat_piar","var_estim_ybar_hat_piar","ICbinf_PiAR","ICbsup_PiAR","long_IC_PiAR")
View(ress)

```

```
#####PISR#####
```

```
##traitement de la base
```

```
sport1<-sport
```

```
age1=1:100 #création d'une nouvelle variable qui contiendra les centre de chaque intervalle de la variable age
```

```
sport1<-cbind(sport1,age1)
```

```
View(sport1)
```

```
sport1$age1=as.integer(sport1$age1)
```

```
sport1$age1[sport1$age=='[0,15[']<-7
```

```
sport1$age1[sport1$age=='[15,30[']<-23
```

```
sport1$age1[sport1$age=='[30,45[']<-37
```

```
sport1$age1[sport1$age=='[45,...[']<-53
```

```
str(sport1)
```

```
sport1$age1=as.factor(sport1$age1)
```

```
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){
```

```
  54
```

```
  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))
res2
#les intervalles de confiance
ICbinf=1:20
ICbsup=1:20
55
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="darkcyan",cex=4)+geom_point(aes(y,IC
binf),colour="gold",cex=4)+geom_path(colour="darkcyan",size=1)+geom_path(aes(y,IC
binf),colour="gold",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("gold","darkcyan"))
t2=labs(x="",y="")
t+t1+t2+ggtitle("les 20 intervalles de confiance \n 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","ech1
0","ech11","ech12","ech13","ech14","ech15","ech16","ech17","ech18","ech19","ech20")
View(ech)
tt=cbind(res2[21,],res2[22,],tt[,-1])
colnames(tt)=c("y_bar_hat_pisr","var_est_ybar_hat","Icbinf_pisr","ICbsup_pisr","long_IC_
pisr")
View(tt)
y_bar_pisr=mean(tt$y_bar_hat_pisr)
y_bar_pisr

```

```

####stratification####
##definition des strates
attach(sport)
N=100
n=20
h=3
Nh=as.vector(table(sport$occupation))
Nh
strate1=rep(0,Nh[1]) #élève/etudiant
strate1
j=1
for (i in 1:100) {
if(occupation[i]=="élève/etudiant") {strate1[j]=nbre_h_semaine[i]; j=j+1;}
}
strate1
strate2=rep(0,Nh[2]) #employer
j=1
for (i in 1:100) {
if(occupation[i]=="employer") {strate2[j]=nbre_h_semaine[i]; j=j+1;}
}
strate2
strate3=rep(0,Nh[3]) #pas d'activité
j=1
for (i in 1:100) {
if(occupation[i]=="pas d'activité") {strate3[j]=y[i]; j=j+1;}
}
strate3

strate1
strate2
strate3
strates=list(strate1,strate2,strate3)

#### ybar , variances et variances corrigées des strates
sigma2_hc=NULL
sigma2_h=NULL
ybar_strates=NULL
for (i in 1:3){
ybar_strates[i]=list(mean(unlist(strates[i])))
ybar_strates=as.vector(unlist(ybar_strates))
ybar_strates[i]
sigma2_hc[i]=var(unlist(strates[i]))
sigma2_h[i]=((N-1)/N)*sigma2_hc[i]}
ybar_strates[1]
ybar_strates[2]
ybar_strates[3]
sigma2_hc[1]
sigma2_hc[2]

```

```

sigma2_hc[3]
sigma2_h[1]
sigma2_h[2]
sigma2_h[3]

##### Les strates ont la même taille#####
### Calcul des nh
nh=c(rep(round(n/h),h))
nh
### Tirage de 20 echantillons et calcul demandé
res_strat=matrix(0,20,7)
res_strat
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 estimée pour chaque strate
  ybar_hat_strates=c(ybar_hat1,ybar_hat2,ybar_hat3)
  ybar_hat_strates

  # ybar estimée
  ybar_hat=(1/N)*sum(Nh*ybar_hat_strates)
  ybar_hat
  mean(y)
  # variances corrigées de l'echantillon dans chaque strate
  s2_hc=c(s2_c_h1,s2_c_h2,s2_c_h3)
  s2_hc
  # 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_inter_y+var_intra_y
var_y=((N-1)/N)*var(y)
var_y

#variance estimée
var_estim_ybar_hat_strat=(1/N^2)*(sum(((Nh^2)/nh)*(1-fh)*s2_hc))
var_y=((N-1)/N)*var(y)
var_estim_ybar_hat_strat
## 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)
}
res_strat
### Resultats du sondage stratifié regroupés 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 estimées
ybar_strat=mean(res_strat[,1])
ybar_strat
### Moyenne des longueurs des intervalles de confiance
long_strat=mean(res_strat[,7])
long_strat
### Représentation de l'intervalles de confiance pour chacune des 20 expériences sur
le même graphe
library(ggplot2)
y=sport$nbre_h_semaine
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="green",cex=4)+geom_point(
aes(y,res_strat[,5]),colour="pink",cex=4)+geom_path(colour="green",size=1)+geom_pat
h(aes(y,res_strat[,5]),colour="pink",size=1)
g2_s=annotate("text", label =c("La borne supérieure","La borne inférieure"), x=c(3,3), y
=c(-0.5,-5),size =3, colour = c("green","pink"))

```



```

g3_s=labs(x="",y="")
g1_s+g2_s+g3_s+ggtitle("Les 20 intervalles de confiance \n Strates ayant le meme
taille")+theme(plot.title = element_text(color="black", size=14, face="bold.italic"))
res_strat=res_strat[,-4]
View(res_strat)
y_bar_strate = sum(res_strat[,1]) /20
y_bar_strate

### Les strates sont à allocation proportionnelle ###
nh=round(Nh*0.2)
nh
### Tirage de 20 echantillons et calcul demandé
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 estimée pour chaque strate
  ybar_hat_strates=c(ybar_hat1,ybar_hat2,ybar_hat3)
  ybar_hat_strates

# ybar estimée
  ybar_hat=(1/N)*sum(Nh*ybar_hat_strates)
ybar_hat
mean(y)
  # variances corrigées 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)
62

```

#variance estimée

```

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 stratifié regroupés dans un tableau

```

res_prop
View(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 estimées

```

ybar_prop=mean(res_prop[,1])
ybar_prop

```

Moyenne des longueurs des intervalles de confiance

```

long_prop=mean(res_prop[,7])

```

Représentation de l'intervalles de confiance pour chacune des 20 expériences sur le même 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="burlywood1",cex=4)+geom_point(aes(y,res_prop[,5]),colour="coral",cex=4)+geom_path(colour="burlywood1",size=1)+geom_path(aes(y,res_prop[,5]),colour="coral",size=1)
g2_prop=annotate("text", label =c("La borne inférieure","La borne supérieure"), x =c(3,3), y =c(-1,-4),size =5, colour = c("coral","burlywood1"))
g3_prop=labs(x="",y="")

```

```

g1_prop+g2_prop+g3_prop+ggtitle("Les 20 intervalles de confiance \n Strates à
allocation proportionnelle")+theme(plot.title = element_text(color="black", size=14,
face="bold.italic"))
res_prop=res_prop[,-4]
View(res_prop)
prop = mean (res_prop[,1])
prop

## #Les strates sont à allocation optimale ###
nh=round(((n*Nh*(sigma2_hc)^0.5))/sum((Nh*(sigma2_hc)^0.5)) )
nh
### 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_hat1=mean(ech1)
  s2_c_h1=var(ech1)
  ech2=sample(strate2,nh[2],replace=FALSE)
  ech2
  ybar_hat2=mean(ech2)
  64
  s2_c_h2=var(ech2)

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

  # ybar estimée
  ybar_hat=(1/N)*sum(Nh*ybar_hat_strates)
  ybar_hat
  mean(y)
  # variances corrigées de l'échantillon 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 estimée

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 stratifié regroupés 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 estimées
ybar_opt=mean(res_opt[,1])
ybar_opt
### Moyenne des longueurs des intervalles de confiance
long_opt=mean(res_opt[,7])
### Représentation de l'intervalles de confiance pour chacune des 20 expériences 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="salmon",cex=4)+geom_
point(aes(y,res_opt[,5]),colour="khaki",cex=4)+geom_path(colour="salmon",size=1)+ge
om_path(aes(y,res_opt[,5]),colour="khaki",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("khaki","salmon"))
g3_opt=labs(x="",y="")
g1_opt+g2_opt+g3_opt+ggtitle("les 20 intervalles de confiance \n Strates à allocation
optimale")+theme(plot.title = element_text(color="black", size=14, face="bold.italic"))
res_opt=res_opt[,-4]

```

View(res_opt)

#####GRAPPE#####

```
grp=matrix(0,100,1)
for (i in 1:100) {
  (if(sport2$age[i]=="[0,15[") (grp[i]=1))
  (if(sport2$age[i]=="[15,30[") (grp[i]=2))
  (if (sport2$age[i]=="[30,45[") (grp[i]=3))
  (if (sport2$age[i]=="[45,...[") (grp[i]=4))
}
```

sport2=cbind(sport,grp)

#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 #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

Nhh=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])),me
an(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,]
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
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
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="#660066",cex
=4)+geom_point(aes(y,icinf_grappe),colour="firebrick1",cex=4)+geom_path(colour="#6
60066",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","lon
gIC_grap","num_1ere_grap_selct","num_2ere_grap_selct")
View(res_grappe)

```

```
#####GRAPPE_2DEGRE#####
nh1=round(Nh/5) #le nombre des individus qui peuvent etre selectionnés dans chaque
grappe
grappe_2dg=NULL
# les m unités primaires
for(i in 1:4){
  grappe_2dg[i]=list(y[grp==i])
}
grappe_2dg
sg2dg=matrix(0,10,2) ##### la matrice des numeros des grappes selectionnées
nb=matrix(0,10,2) ### la matrice des nombres des indivdus selectionnés 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
estimés
th2[1]=sum(unlist(sg_2dg[1]))
th2[2]=sum(unlist(sg_2dg[2]))
sigmah1=(Nh_2dg[1]/(Nh_2dg[1]-1))*var(unlist(ygrap2[1]))
sigmah2=(Nh_2dg[2]/(Nh_2dg[2]-1))*var(unlist(ygrap2[2]))
sigma=c(sigmah1,sigmah2)
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_estimé","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,longI
C_2dg)
g1_2dg=ggplot(res_2dg,aes(y,icsup_2dg))+geom_point(colour="darkgreen",cex=4)+geo
m_point(aes(y,icinf_2dg),colour="deeppink",cex=4)+geom_path(colour="darkgreen",siz
e=1)+geom_path(aes(y,icinf_2dg),colour="deeppink",size=1)
g2_2dg=annotate("text", label =c("La borne inférieure","La borne supérieure"), 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 à 2
degré")+theme(plot.title = element_text(color="black", size=14, face="bold.italic"))
res_2dg=res_2dg[,-1]
res_2dg=cbind(res_2dg,sg2dg,nb)
View(res_2dg)
colnames(res_2dg)=c("y_bar_grappe2","var_y_bar_grap_2dg","icinf_2dg","icsup_2dg","lo
ngIC_2dg","num_1ere_grap_selct","num_2ere_grap_selct","nbre_ind_select_G1","nbre_ind
_select_G2")

#####SYNTHESE#####
###longueur###
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_PEAR,long_PESR,long_PJAR,long_PISR,long_strat_taille,long
_strat_prop,long_strat_opt,long_grappe,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="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","PJAR","PISR
","2_degré"), x=c(1,1,1,1,1,1,1,1,1), y=c(64,66,68,70,72,74,76,78,80),size =3, colour =
c("chocolate4","green","coral1","deeppink","darkgreen","deeppink4","blue","black","indi
anred3"))
c=labs(x="",y="")
a+b+c+ggtitle("Synthèse des logueurs des intervalles de confaince")+theme(plot.title =
element_text(color="black", size=14, face="bold.italic"))

```

###synthese PEAR PESR PJAR PISR###

```

a=ggplot(tab,aes(y,long_PEAR))+geom_point(colour="#008E8E",cex=4)+
  geom_path(colour="#008E8E",size=1)+
  geom_point(aes(y,long_PESR),colour="#FE1B00",cex=4)+
  geom_path(aes(y,long_PESR),colour="#FE1B00",size=1)+
  geom_point(aes(y,long_PJAR),colour="#FF866A",cex=4)+
  geom_path(aes(y,long_PJAR),colour="#FF866A",size=1)+
  geom_point(aes(y,long_PISR),colour="#6D071A",cex=4)+
  geom_path(aes(y,long_PISR),colour="#6D071A",size=1)
b=annotate("text", label =c("PESR","PEAR","PJAR","PISR"), x=c(18,18,18,18), y
=c(27,28,29,30),size =5, colour = c("#FE1B00","#008E8E","#FF866A","#6D071A"))
c=labs(x="",y="")
a+b+c+ggtitle("Synthèse des logueurs des intervalles de confaince \n PEAR PESR PJAR
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="#000010",cex=4)+
geom_path(aes(y,long_strat_prop),colour="#000010",size=1)+
geom_point(aes(y,long_strat_opt),colour="chocolate4",cex=4)+
geom_path(aes(y,long_strat_opt),colour="chocolate4",size=1)
75
b=annotate("text", label =c("strat_optimal","strat_prop","strat_mem_taille"), x =c(3,3,3),
y =c(43,45,48),size =5, colour = c("chocolate4","#000010","coral1"))
c=labs(x="",y="")
a+b+c+ggtitle("Synthèse des logueurs des intervalles de confaince \n strat_egal prop
opt")+theme(plot.title = element_text(color="black", size=14, face="bold.italic"))

```

synthese grappe et 2degré###

```

a=ggplot(tab,aes(y,long_grappe))+geom_point(colour="#56739A",cex=4)+
geom_path(aes(y,long_grappe),colour="#56739A",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("Grappe","2_degré"), x =c(2,2), y =c(62,65),size =5, colour =
c("#56739A","indianred3"))
c=labs(x="",y="")
a+b+c+ggtitle("Synthèse des logueurs des intervalles de confaince \n grappe et 2
degré")+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
s=srswor(n,N)
echantillon=y[s==1]
echantillon
moyenne=mean(echantillon)
moyenne# moyenne de lechantillon = 6.4
##Alors l'estimateur de la moyenne avant redressement est:
moyenne
##Or on connait que la proportion des femmes dans la population est
P=sum(genre=="F")/N
P # la proprtion des femme dans la pop = 0.3 #la prop des femme dans l'chantillon=0.25

##Alors que la proportion des femme dans l'échantillon est=0.3
p_ech=sum(genre[s==1]=="F")/n

p_ech
#la proportion des Hommes dans la population est
P=sum(genre=="H")/N
P # la proprtion des homme dans la pop = 0.7
##Alors que la proportion des homme dans l'échantillon est
p_ech=sum(genre[s==1]=="H")/n

```

```
p_ech#la prop des hommes dans l'chantillon=0.75
##L'estimateur de la moyenne après redressement est egale a 5.973333
moyenne.red=(P/p_ech)*moyenne
moyenne.red # est egale a 5.973333
# tous les individus donc ont le meme poids (P/p_ech) alors que sans redressement leur
podis est(N/n)
```

MERCI POUR VOTRE ATTENTION !