

ÉTUDE DE CAS  
SONDAGE STRATIFIÉ

ALLOCATION

SAS  
AA  
FR  
PROPORTIONNELLE  
SONDAGE  
VAR3  
ZONE  
VARIANCE  
STRATE  
VARI  
VAR2  
PRECISION  
NEYMAN

IUT DE  
CARCASSONNE

IUT STID 2

Groupe  
DADI Abel  
OUATTARA Seydou  
VIGAN Jéros

## SOMMAIRE

<b>CHAPITRE I : ÉTUDE DE L'UNIVERS .....</b>	<b>3</b>
I - 1 - ÉTUDE DES DIFFÉRENTES VARIABLES QUANTITATIVES PAR ZONE .....	6
I - 1 - a - Étude des différentes variables quantitatives par ZONE AA .....	6
I - 1 - b - Étude des différentes variables quantitatives par ZONE BB .....	7
I - 1 - c - Étude des différentes variables quantitatives par ZONE CC .....	8
I - 1 - d - Étude des différentes variables quantitatives par ZONE DD .....	9
<b>CHAPITRE II : ÉTUDE D'UN SONDAGE STRATIFIÉ .....</b>	<b>11</b>
II - 2 - ÉTUDE DE SONDAGE STRATIFIÉ PAR ZONE .....	12
II - 2 - e - ÉTUDE DE SONDAGE STRATIFIÉ PAR ZONE AA .....	12
II - 2 - f - ÉTUDE DE SONDAGE STRATIFIÉ PAR ZONE BB .....	13
II - 2 - g - ÉTUDE DE SONDAGE STRATIFIÉ PAR ZONE CC .....	13
II - 2 - h - ÉTUDE DE SONDAGE STRATIFIÉ PAR ZONE DD .....	14
II - 3 - ESTIMATION DE LA PRÉCISION RELATIVE POUR L'ENSEMBLE DES ZONES .....	14
II - 3 - i - Allocation proportionnelle .....	15
II - 3 - j - Allocation optimale .....	15
II - 4 - ESTIMATION DE LA PRÉCISION RELATIVE PAR ZONES .....	16
<b>CHAPITRE III : ÉTUDE COMPLÉMENTAIRE .....</b>	<b>17</b>
III - 5 - NOMBRE D'UNITÉS POUR PRÉCISION RELATIVE GLOBALE ÉGALE 2% .....	17
III - 5 - k - Allocation proportionnelle .....	17
III - 5 - l - Allocation optimale .....	17
III - 5 - m - Résultat des estimations .....	17
III - 6 - NOMBRE D'UNITÉS PAR STRATE POUR PRÉCISION RELATIVE GLOBALE ÉGALE 2% .....	18
III - 6 - n - Allocation proportionnelle par Strate .....	18
III - 6 - o - Allocation optimale par Strate .....	18
III - 6 - p - Résultats des Estimation de nombre d'unités par Strate si $k = 2\%$ .....	19
<b>CHAPITRE IV : STRATIFICATION DE LA VARIABLE QUANTITATIVE VAR 1 21</b>	
IV - 7 - ESTIMATION DE LA PRÉCISION RELATIVE DE LA VARIABLE VAR1 .....	21
IV - 7 - q - Allocation proportionnelle .....	22
IV - 7 - r - Allocation optimale .....	22
IV - 7 - s - Estimation de nombre d'unités si $k = 2\%$ .....	22
IV - 7 - t - Résultats des Estimation de nombre d'unités par Strate si $k = 2\%$ .....	23
<b>CHAPITRE V : STRATIFICATION DE LA VARIABLE QUANTITATIVE VAR 2..25</b>	
V - 8 - ESTIMATION DE LA PRÉCISION RELATIVE DE LA VARIABLE VAR2 .....	25
V - 8 - u - Allocation proportionnelle .....	26
V - 8 - v - Allocation optimale .....	26
V - 8 - w - Estimation de nombre d'unités si $k = 2\%$ .....	26
V - 8 - x - Résultats des Estimation de nombre d'unités par Strate si $k = 2\%$ .....	27

## Liste des figures

FIGURE N°1 : HISTOGRAMME DES VARIABLES QUANTITATIVES .....	4
FIGURE N°2 : BOXPLOT DES VARIABLES QUANTITATIVES .....	4
FIGURE N°3 : BOXPLOT DE LA VARIABLE VAR 1 / ZONE.....	5
FIGURE N°4 : BOXPLOT DE LA VARIABLE VAR 2 / ZONE.....	5
FIGURE N°5 : BOXPLOT DE LA VARIABLE VAR 3 / ZONE.....	5
FIGURE N°6 : DIAGRAMME EN BARRE DE LA ZONE .....	6
FIGURE N°7 : BOXPLOT DES VARIABLES QUANTITATIVES / ZONE AA .....	7
FIGURE N°8 : BOXPLOT DES VARIABLES QUANTITATIVES / ZONE BB.....	8
FIGURE N°9 : BOXPLOT DES VARIABLES QUANTITATIVES / ZONE CC .....	9
FIGURE N°10 : BOXPLOT DES VARIABLES QUANTITATIVES / ZONE DD .....	10

## CHAPITRE I : Étude de l'univers

Cette étude de cas porte sur un tableau relatif à la France qui est découpée en 4 zones (AA, BB, CC et DD). Le fichier concerne 24 000 unités de production et comporte 5 variables :

- 2 variables qualitatives (la variable PAYS qui prend la modalité FR, la variable ZONE qui prend les modalités AA, BB, CC ou DD)
- 3 variables quantitatives (VAR1, VAR2 et VAR3)

Pour étude de la distribution des variables. Nous allons utiliser deux programmes : **R** et **SAS** en d'une sureté dans les résultats.

L'étude de l'univers qu'est la population porte sur les indicateurs concernant la distribution de chaque variable : somme, moyenne, écart-type corrigé, coefficient de variation.

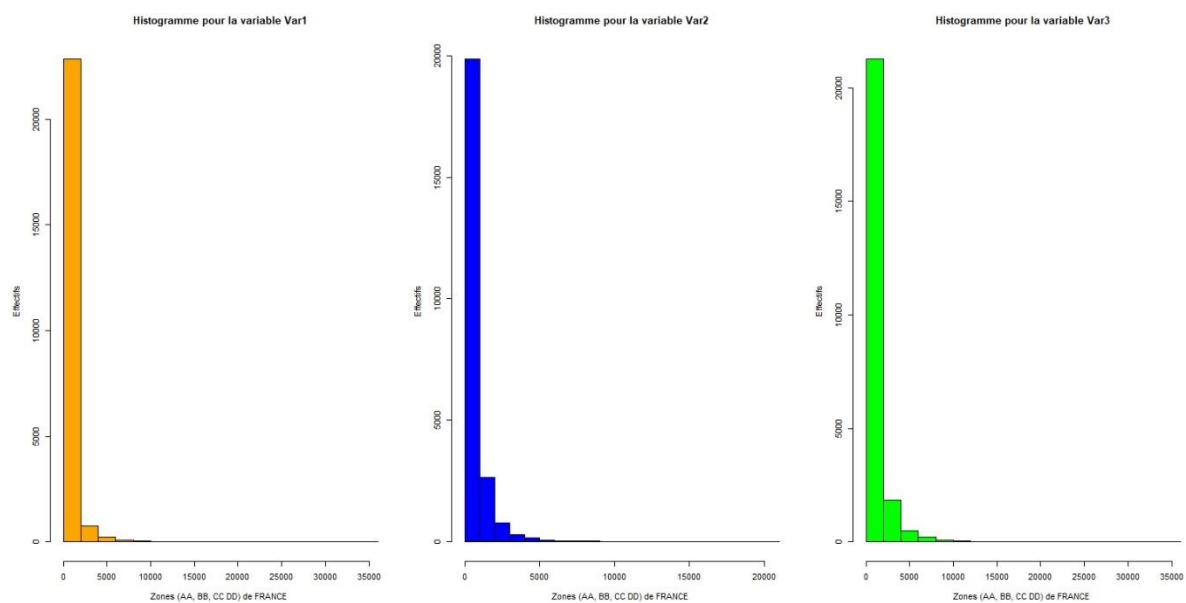
Le tableau ci-dessous résume les résultats par variable.

	Effectifs	Sommes des valeurs	Moyenne des valeurs	Écart type	Coefficient de variation
<b>Var1</b>	24000	6926406	288.6003	1040.740	3.606166
<b>Var2</b>	24000	13490715	562.1131	1053.103	1.873471
<b>Var3</b>	24000	20417121	850.7134	1519.404	1.786035

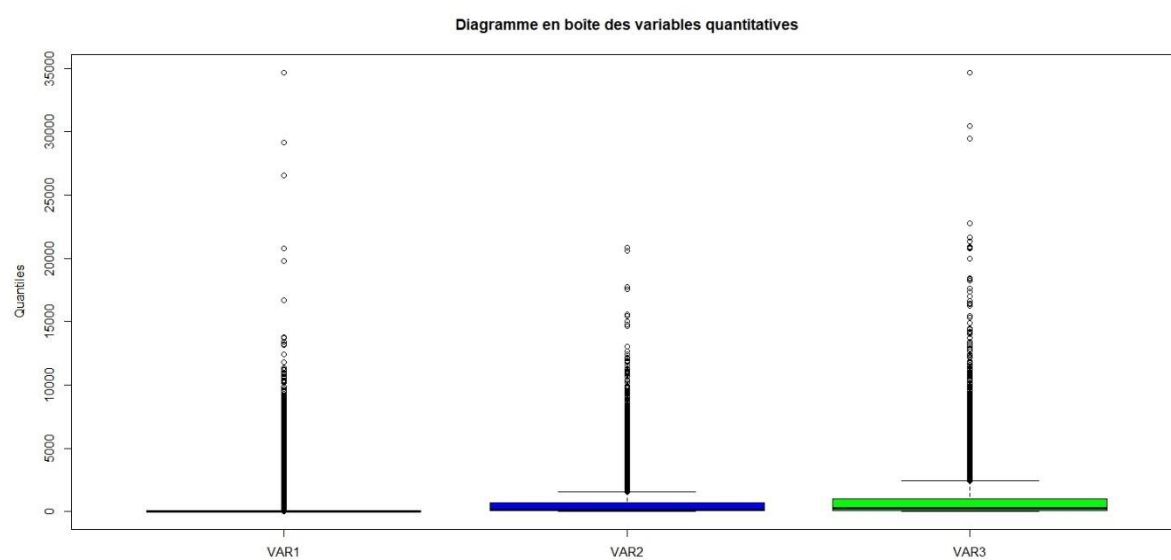
### Commentaire

Le coefficient de variation permet de comparer les distributions entre elles donc on admet que la variable **Var1 de 3,6%** est plus disperse que les deux autres variables, et la variable **Var2 de 1,8%** a plus de variabilité que la **Var3 de 1,78%**.

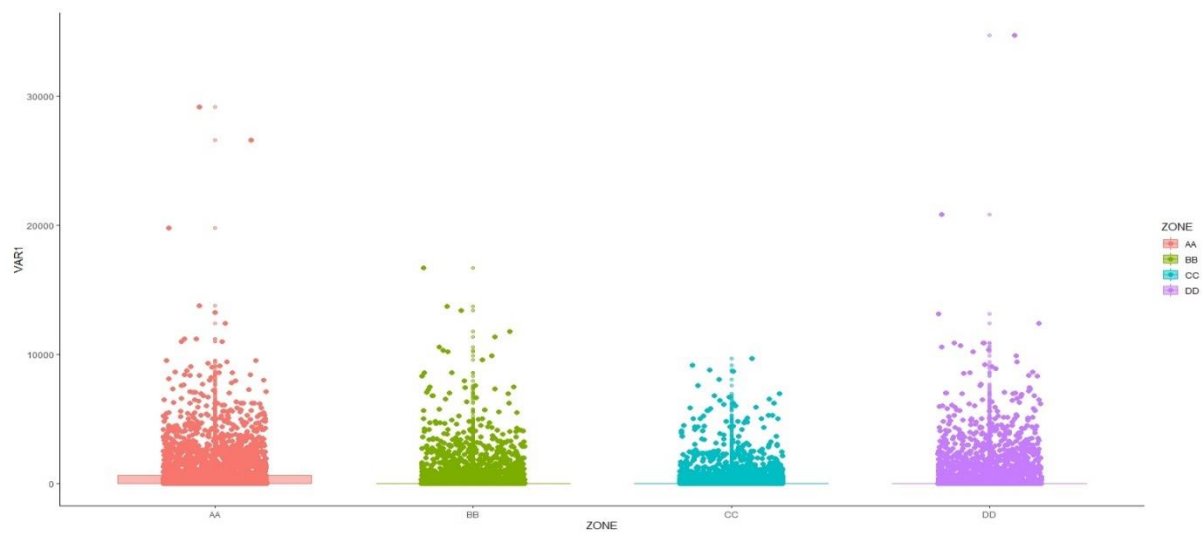
En termes de prévention en sondage par rapport à une taille d'échantillon à déterminer, toute chose étant égale par ailleurs, la variable **Var3** sera **plus précise** donc **meilleure**, la variable **Var2** sera **moins précise** et la variable **Var1** sera **très moins précise donc mauvaise**.



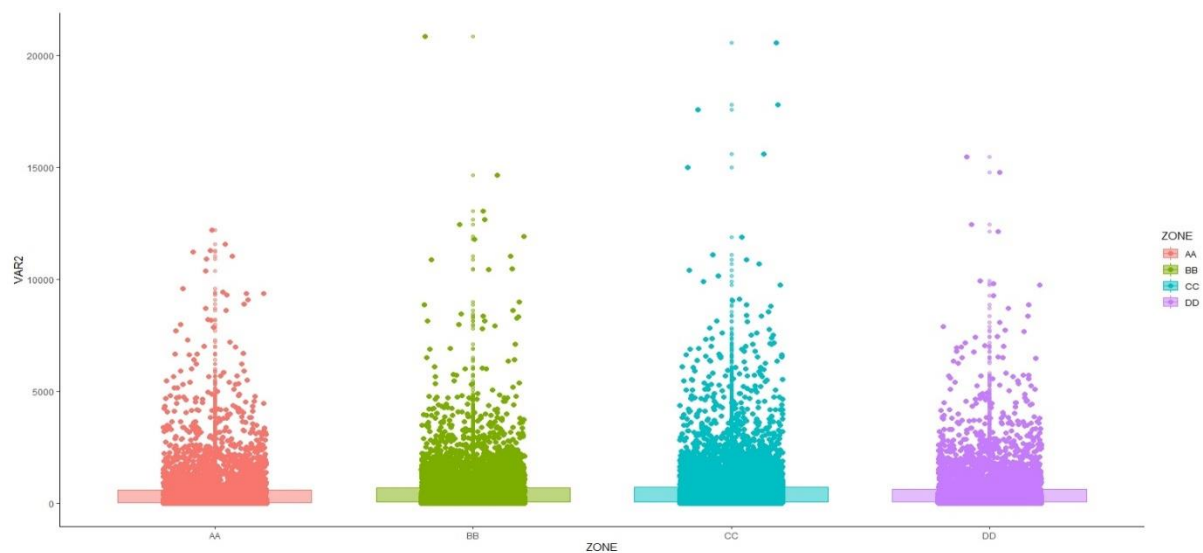
**Figure n°1 : Histogramme des variables quantitatives**



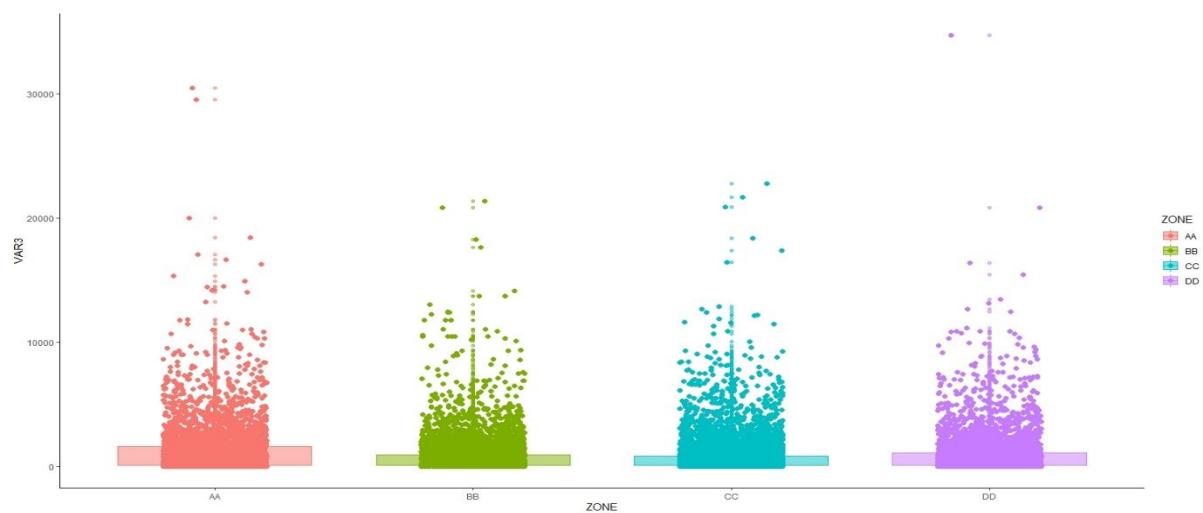
**Figure n°2 : Boxplot des variables quantitatives**



**Figure n°3 :** Boxplot de la variable VAR 1 / ZONE

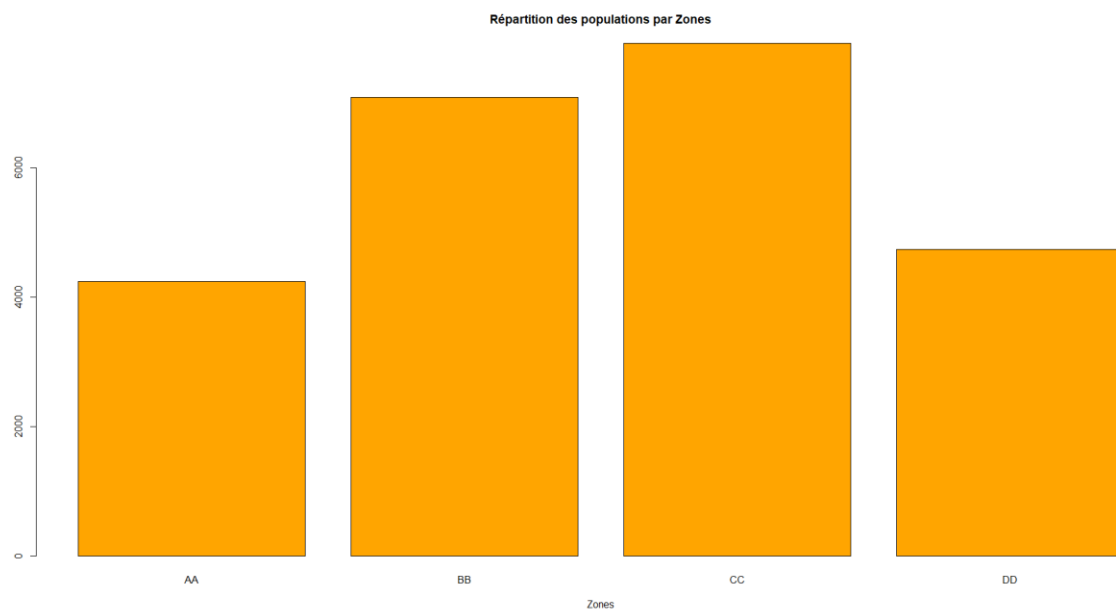


**Figure n°4 :** Boxplot de la variable VAR 2 / ZONE



**Figure n°5 :** Boxplot de la variable VAR 3 / ZONE

## I - 1 - ÉTUDE DES DIFFÉRENTES VARIABLES QUANTITATIVES PAR ZONE



**Figure n°6 :** Diagramme en barre de la zone

### Commentaire

Graphiquement la zone CC est plus représentative par rapport aux autres zones.

### I - 1 - a - Étude des différentes variables quantitatives par ZONE AA

L'étude de l'univers pour la zone AA porte sur les indicateurs concernant la distribution de chaque variable : somme, moyenne, écart-type corrigé, coefficient de variation.

Le tableau ci-dessous résume les résultats par variable.

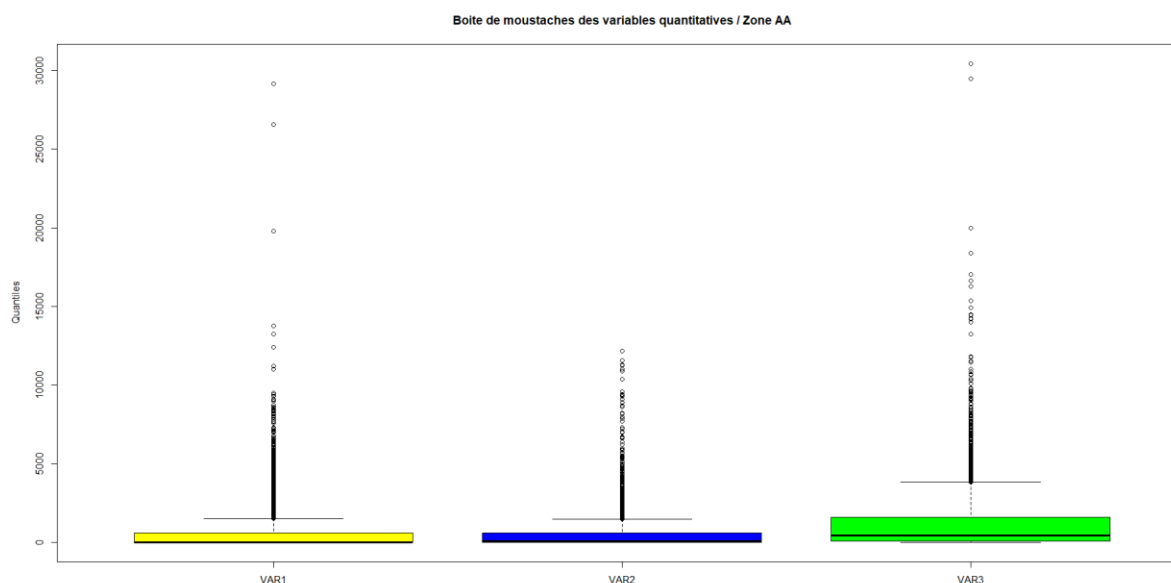
	Nh	Somme	Moyenne	Écart type	Coefficient de Variation
<b>Var1</b>	4247	2888103	680.0337	1542.372	2.268081
<b>Var2</b>	4247	2319580	546.1691	1111.961	2.035928
<b>Var3</b>	4247	5207683	1226.2027	1992.834	1.625208

### Commentaire

De l'analyse du tableau, on admet que dans la zone AA, la variable **Var1 de 2,27%** est plus dispersée que les deux autres variables, et la variable **Var2 de 2,03%** a plus de variabilité que la **Var3 de 1,62%**.

Dans la zone AA, en termes de prévention en sondage par rapport à une taille d'échantillon à déterminer, toute chose étant égale par ailleurs, la variable **Var3** sera **plus précise** donc

meilleure, la variable **Var2** sera **moins précise** et la variable **Var1** sera **très moins précise** donc **mauvaise**.



**Figure n°7 : Boxplot des variables quantitatives / Zone AA**

### **I - 1 - b - Étude des différentes variables quantitatives par ZONE BB**

L'étude de l'univers pour la zone BB porte sur les indicateurs concernant la distribution de chaque variable : somme, moyenne, écart-type corrigé, coefficient de variation.

Le tableau ci-dessous résume les résultats par variable.

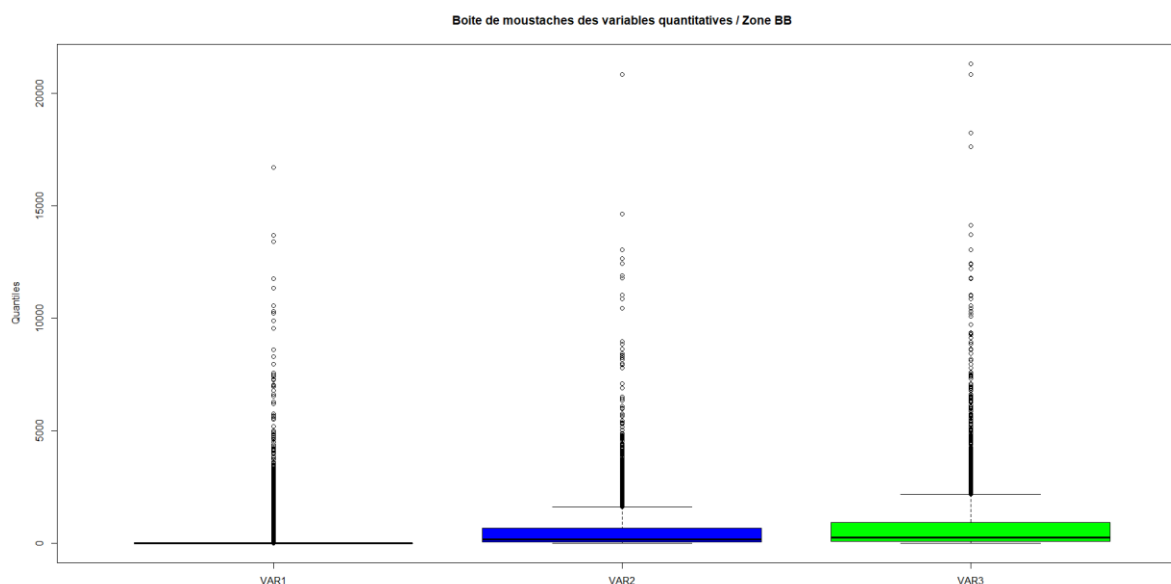
	Nh	Somme	Moyenne	Écart type	Coefficient de Variation
<b>Var1</b>	7092	1323074	186.5587	808.4664	4.333577
<b>Var2</b>	7092	3854650	543.5209	969.8111	1.784313
<b>Var3</b>	7092	5177724	730.0795	1287.5100	1.763520

### **Commentaire**

De l'analyse du tableau, on admet que dans la zone BB, la variable **Var1 de 4,33%** est plus dispersée que les deux autres variables, et la variable **Var2 de 1,78%** a presque la même représentation de variance que la **Var3 de 1,76%**.



Dans la zone BB, en termes de prévention en sondage par rapport à une taille d'échantillon à déterminer, toute chose étant égale par ailleurs, les variables **Var3 et Var2** seront **plus précise** donc **meilleures** par rapport à la variable **Var1** qui sera **très moins précise donc mauvaise**.



**Figure n°8 : Boxplot des variables quantitatives / Zone BB**

### **I - 1 - c - Étude des différentes variables quantitatives par ZONE CC**

L'étude de l'univers pour la zone BB porte sur les indicateurs concernant la distribution de chaque variable : somme, moyenne, écart-type corrigé, coefficient de variation.

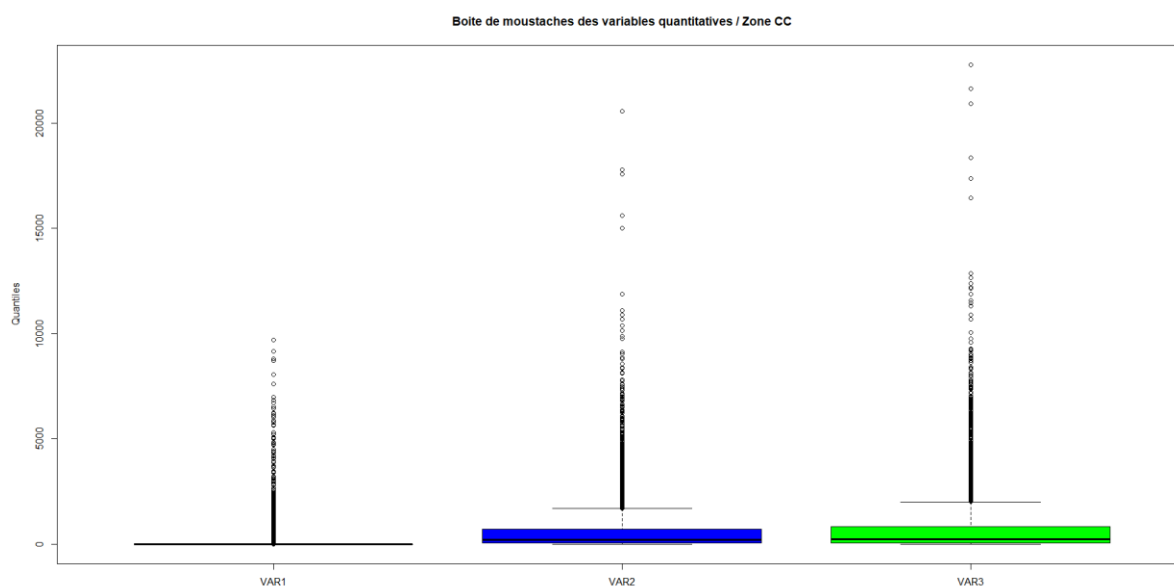
Le tableau ci-dessous résume les résultats par variable.

	Nh	Somme	Moyenne	Écart type	Coefficient de Variation
<b>Var1</b>	7921	943754	119.1458	550.2208	4.618046
<b>Var2</b>	7921	4738865	598.2660	1100.1207	1.838849
<b>Var3</b>	7921	5682619	717.4118	1323.7112	1.845120

### **Commentaire**

De l'analyse du tableau, on admet que dans la zone CC, la variable **Var1 de 4,61%** est surreprésentée que les deux autres variables. Les variables **Var2 et Var3** ont une proportion de représentation égale de **1,8%**.

Dans la zone CC, en termes de prévention en sondage par rapport à une taille d'échantillon à déterminer, toute chose étant égale par ailleurs, les variables **Var3 et Var2** seront **plus précises** donc **meilleures** par rapport à la variable **Var1** qui sera **très moins précise donc mauvaise**.



**Figure n°9** : Boxplot des variables quantitatives / Zone CC

### I - 1 - d - Étude des différentes variables quantitatives par ZONE DD

L'étude de l'univers pour la zone DD porte sur les indicateurs concernant la distribution de chaque variable : somme, moyenne, écart-type corrigé, coefficient de variation.

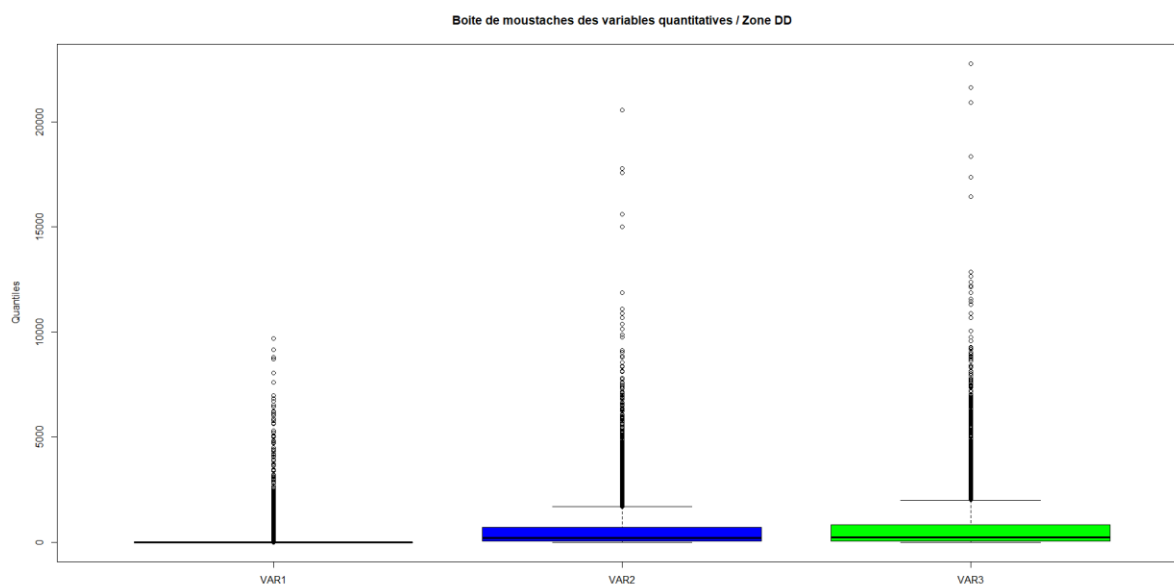
Le tableau ci-dessous résume les résultats par variable.

	Nh	Somme	Moyenne	Écart type	Coefficient de Variation
<b>Var1</b>	4740	1771475	373.7289	1289.048	3.449153
<b>Var2</b>	4740	2577620	543.8017	1036.921	1.906800
<b>Var3</b>	4740	4349095	917.5306	1594.280	1.737577

### Commentaire

De l'analyse du tableau, on admet que dans la zone DD, la variable **Var1 de 3,44%** est surreprésentée que les deux autres variables, et la variable **Var2 de 1,9%** est légèrement plus dispersé que la **Var3 de 1,76%**.

Dans la zone DD, même constat en termes de prévention en sondage par rapport à une taille d'échantillon à déterminer, toute chose étant égale par ailleurs, la variable **Var3** sera **plus précise** donc **meilleure**, la variable **Var2** sera **moins précise** et la variable **Var1** sera **très moins précise** donc **mauvaise**.



**Figure n°10 :** Boxplot des variables quantitatives / Zone DD

## CHAPITRE II : ÉTUDE D'UN SONDAGE STRATIFIÉ

La mise en œuvre d'un plan de sondage sur la variable **Var3** consiste à un découpage par strates suivantes par zone.

STRATE	Conditions
1	< 100
2	De 100 à moins de 500
3	De 500 à moins de 1 000
4	De 1 000 à moins de 2 000
5	De 2 000 à moins de 10 000
6	10 000 et plus

Le découpage de l'univers en 4 régions et 6 strates, dont au total  $4 \times 6 = 24$  strates. L'étude porte sur les indicateurs suivants :

- ✓ L'effectif par strate ( $N_h$ )
- ✓ La somme des valeurs par strate (Total)
- ✓ La moyenne par strate
- ✓ L'écart type corrigé par strate ( $Sh$ )
 

```
shh=function(x){ nn=length(x)-1;yhn=mean(x);eca=((x-yhn)^2);return (sqrt((1/n)*sum(eca)))}
```
- ✓ Le produit  $N_h Sh$
- ✓ L'effectif échantillonné par l'allocation proportionnelle ( $nhAp$ )

$$nhAp = \frac{N_h}{N} * 600$$

- ✓ L'effectif échantillonné par l'allocation optimale ( $nhNey$ )

$$nhNey = \frac{N_h Sh^2}{\sum_{h=1}^H N_h Sh^2} * 600$$

Avec la condition, si l'estimation de  $nhAp = 0$ , nous posons  $nhAp = 2$

Le tableau ci-dessous présente les résultats :

Strate	Nh	Total	Moyenne	Sh	nhAp	nhAp	NhSh <sup>2</sup>	nhNey
<b>AA1</b>	1139	42687	37.47761	29.33289	28.475	28	33410.16	3
<b>AA2</b>	1056	261081	247.23580	113.79350	26.400	26	120165.94	10
<b>AA3</b>	574	414591	722.28397	146.97068	14.350	14	84361.17	7
<b>AA4</b>	643	937863	1458.57387	289.99918	16.075	16	186469.47	15
<b>AA5</b>	806	3135493	3890.18983	1778.18925	20.150	20	1433220.54	<b>116</b>
<b>AA6</b>	29	415968	14343.72414	5085.22784	0.725	1	147471.61	12
<b>BB1</b>	2083	98688	47.37782	27.46534	52.075	52	57210.30	5

Strate	Nh	Total	Moyenne	Sh	nhAp	nhAp	NhSh <sup>2</sup>	nhNey
<b>BB2</b>	2339	558996	238.98931	111.39838	58.475	58	260560.82	21
<b>BB3</b>	1047	763267	729.00382	147.57801	26.175	26	154514.18	13
<b>BB4</b>	1008	1418325	1407.06845	278.04061	25.200	25	280264.93	23
<b>BB5</b>	588	1994704	3392.35374	1493.57071	14.700	15	878219.58	71
<b>BB6</b>	27	343744	12731.25926	3166.51804	0.675	1	85495.99	7
<b>CC1</b>	2448	109545	44.74877	28.19051	61.200	61	69010.36	6
<b>CC2</b>	2630	623705	237.15019	109.70829	65.750	<b>66</b>	288532.80	23
<b>CC3</b>	1150	831723	723.23739	144.24295	28.750	29	165879.39	13
<b>CC4</b>	1019	1421147	1394.64868	277.86066	25.475	25	283140.01	23
<b>CC5</b>	655	2427435	3706.00763	1739.07564	16.375	16	1139094.54	92
<b>CC6</b>	19	269064	14161.26316	4065.63705	0.475	<b>2</b>	77247.10	6
<b>DD1</b>	1323	62001	46.86395	27.59864	33.075	33	36513.00	3
<b>DD2</b>	1452	341527	235.21143	108.92908	36.300	36	158165.03	13
<b>DD3</b>	676	487829	721.64053	143.85679	16.900	17	97247.19	8
<b>DD4</b>	655	934106	1426.11603	278.28128	16.375	16	182274.24	15
<b>DD5</b>	617	2287957	3708.19611	1733.28269	15.425	15	1069435.42	87
<b>DD6</b>	17	235675	13863.23529	6038.32752	0.425	<b>2</b>	102651.57	8
<b>Total</b>	<b>24000</b>	<b>20417121</b>				<b>600</b>		<b>600</b>

### Commentaire

Contrairement à la méthode de l'allocation proportionnelle, la méthode de l'allocation de Neyman donne une répartition qui n'est pas proportionnelle à la population de la strate considérée.

## II - 2 - ÉTUDE DE SONDAGE STRATIFIÉ PAR ZONE

### II - 2 - e - ÉTUDE DE SONDAGE STRATIFIÉ PAR ZONE AA

Strate	Nh	Total	Moyenne	Sh	nhAp	nhAp	NhSh <sup>2</sup>	nhNey
<b>AA1</b>	1139	42687	37.47761	29.33289	28.475	<b>28</b>	33410.16	3
<b>AA2</b>	1056	261081	247.23580	113.79350	26.400	26	120165.94	10
<b>AA3</b>	574	414591	722.28397	146.97068	14.350	14	84361.17	7
<b>AA4</b>	643	937863	1458.57387	289.99918	16.075	16	186469.47	15
<b>AA5</b>	806	3135493	3890.18983	1778.18925	20.150	20	1433220.54	<b>116</b>

<b>AA6</b>	29	415968	14343.72414	5085.22784	0.725	1	147471.61	12
<b>Total</b>	<b>4247</b>	<b>5207683</b>			<b>106.175</b>	<b>105</b>	<b>2005099</b>	<b>163</b>

## II - 2 - f - ÉTUDE DE SONDAGE STRATIFIÉ PAR ZONE BB

<i>Strate</i>	<i>Nh</i>	<i>Total</i>	<i>Moyenne</i>	<i>Sh</i>	<i>nhAp</i>	<i>nhAp</i>	<i>NhSh<sup>2</sup></i>	<i>nhNey</i>
<b>BB1</b>	2083	98688	47.37782	27.46534	52.075	52	57210.30	5
<b>BB2</b>	2339	558996	238.98931	111.39838	58.475	<b>58</b>	260560.82	21
<b>BB3</b>	1047	763267	729.00382	147.57801	26.175	26	154514.18	13
<b>BB4</b>	1008	1418325	1407.06845	278.04061	25.200	25	280264.93	23
<b>BB5</b>	588	1994704	3392.35374	1493.57071	14.700	15	878219.58	<b>71</b>
<b>BB6</b>	27	343744	12731.25926	3166.51804	0.675	1	85495.99	7
<b>Total</b>	<b>7092</b>	<b>5177724</b>			<b>177.3</b>	<b>177</b>	<b>1716266</b>	<b>140</b>

## II - 2 - g - ÉTUDE DE SONDAGE STRATIFIÉ PAR ZONE CC

<b>Strate</b>	<b>Nh</b>	<b>Total</b>	<b>Moyenne</b>	<b>Sh</b>	<b>nhAp</b>	<b>nhAp</b>	<b>NhSh<sup>2</sup></b>	<b>nhNey</b>
<b>CC1</b>	2448	109545	44.74877	28.19051	61.200	61	69010.36	6
<b>CC2</b>	2630	623705	237.15019	109.70829	65.750	<b>66</b>	288532.80	23
<b>CC3</b>	1150	831723	723.23739	144.24295	28.750	29	165879.39	13
<b>CC4</b>	1019	1421147	1394.64868	277.86066	25.475	25	283140.01	23
<b>CC5</b>	655	2427435	3706.00763	1739.07564	16.375	16	1139094.54	<b>92</b>
<b>CC6</b>	19	269064	14161.26316	4065.63705	0.475	<b>2</b>	77247.10	6
<b>Total</b>	<b>7921</b>	<b>5682619</b>			<b>198.025</b>	<b>199</b>	<b>2022904</b>	<b>140</b>

## II - 2 - h - ÉTUDE DE SONDAGE STRATIFIÉ PAR ZONE DD

Strate	Nh	Total	Moyenne	Sh	nhAp	nhAp	NhSh <sup>2</sup>	nhNey
<b>DD1</b>	1323	62001	46.86395	27.59864	33.075	33	36513.00	3
<b>DD2</b>	1452	341527	235.21143	108.92908	36.300	<b>36</b>	158165.03	13
<b>DD3</b>	676	487829	721.64053	143.85679	16.900	17	97247.19	8
<b>DD4</b>	655	934106	1426.11603	278.28128	16.375	16	182274.24	15
<b>DD5</b>	617	2287957	3708.19611	1733.28269	15.425	15	1069435.42	<b>87</b>
<b>DD6</b>	17	235675	13863.23529	6038.32752	0.425	2	102651.57	8
<b>Total</b>	<b>4740</b>	<b>4349095</b>			<b>118.5</b>	<b>119</b>	<b>1646286</b>	<b>134</b>

## II - 3 - ESTIMATION DE LA PRÉCISION RELATIVE POUR L'ENSEMBLE DES ZONES

Formules de calcul

```
#precision relative
precisionRelative=(sqrt(varianceStr)/MoySt)*100;precisionRelative|
```

```
precisionRelative=(sqrt(varianceTot)/T)*100;precisionRelative
```

```
#Estimateur de variance de la moyenne

varianceStratifie=function(nhpetit,grandNh,Sh,N){
  cmp1=(grandNh*grandNh)/(N*N)
  cmp2=1-(nhpetit/grandNh)
  cmp3=(Sh * Sh)/nhpetit
  cmp=cmp1*cmp2*cmp3
  return(sum(cmp))
}

varianceT=function(nhpetit,grandNh,Sh){
  cmp1=(grandNh*grandNh * Sh *Sh)/(nhpetit)
  cmp2=(Sh * Sh * grandNh)
  cmp=cmp1- cmp2
  return(sum(cmp))
}
```

### II - 3 - i - Allocation proportionnelle

Moyenne des Strates	Variance des Strates	Précision. Relative (%)	Précision absolue (95%)	
			Borne inférieure	Borne supérieure
850.7134	682.5666	3.071067	799.506	901.920

### II - 3 - j - Allocation optimale

Moyenne des strates	Variance des Strates	Précision. Relative (%)	Précision absolue (95%)	
			Borne inférieure	Borne supérieure
850.7134	140.5886	1.393772	827.474	873.953

### Commentaire

Pour la variable VAR3, on constate dans le cas de l'allocation proportionnelle, la précision relative est inférieure à 5%, l'échantillon est significativement représentatif au niveau 5% donc moindre de marge d'erreur.

En plus, on remarque que la précision obtenue avec l'allocation de Neyman est bien meilleure que celle obtenue avec l'allocation proportionnelle, et est inférieure à 2%, l'échantillon est significativement représentatif au niveau 2%, très significativement représentatif au niveau 5% donc très moindre de marge d'erreur.

En somme, la variable VAR3 est bien précise, donc meilleure dans les deux cas d'allocation, ce qui confirme nos premières analyses sur le coefficient de variation de ladite variable



## II - 4 - Estimation de la précision relative par zones

	Moyenne	Variance	Précision. Relative. AP (%)	Précision relative OP (%)
<b>Zone AA</b>	1226.2027	1180.4330	2.477844	2.801937
<b>Zone BB</b>	730.0795	384.3740	1.179329	2.685387
<b>Zone CC</b>	717.4118	361.5186	1.150869	2.650311
<b>Zone DD</b>	917.5306	786.5727	1.998318	3.056672

### Commentaire

On constate en général que toutes les précisions relatives sont inférieures à 5%, donc l'échantillonnage par stratification par zone dans les deux cas d'allocation (proportionnelle ou Neyman) sont significativement représentatif au niveau 5%.

Par contre au niveau 2%, l'allocation proportionnelle est beaucoup meilleure que celle de l'allocation de Neyman (dans le cas d'une petite taille de la population (Univers)).

## CHAPITRE III : ÉTUDE COMPLÉMENTAIRE

### III - 5 - Nombre d'unités pour précision relative globale égale 2%

#### III - 5 - k - Allocation proportionnelle

La formule de calcul du nombre d'unités

```
#allocation proportionnelle nAP
calnAP=function(k,grandNh,Sh,N,MoyenneX){
  cmp1= N * sum(grandNh*Sh*Sh)
  cmp2=(k*N*MoyenneX)*(k*N*MoyenneX)
  cmp3=sum(Sh * Sh * grandNh)
  cmp=cmp1/(cmp2 + cmp3)
  return(cmp)
}
```

#### III - 5 - l - Allocation optimale

La formule de calcul du nombre d'unités

```
#allocation de Neymar
calnNey=function(k,grandNh,Sh,N,MoyenneX){
  cmp1= sum(grandNh*Sh) * sum(grandNh*Sh)
  cmp2=(k*N*MoyenneX)*(k*N*MoyenneX)
  cmp3=sum(Sh * Sh * grandNh)
  cmp=cmp1/(cmp2 + cmp3)
  return(cmp)
}
```

#### III - 5 - m - Résultat des estimations

Nombre de Strate	K	nAP	nAP.arrondi	nNey	nNey.arrondi
24	0.02	1368.27	<b>1368</b>	308.8956	<b>309</b>

## Commentaire

On remarque pour une précision relative de 2%, la taille d'échantillon d'enquête dans le cas de l'allocation proportionnelle est très supérieure à celle de l'allocation de Neyman qui est beaucoup moins coûteuse à réaliser dans le cas d'une enquête.

### III - 6 - Nombre d'unités par strate pour précision relative globale égale 2%

#### III - 6 - n - Allocation proportionnelle par Strate

```
#par strate

## allocation proportionnelle

calNhStAP=function(k,grandNh,Sh,T){
  cmp1=(grandNh*grandNh * Sh *Sh)
  cmp2=(Sh * Sh * grandNh)
  cmp3=(k*T)*(k*T)
  cmp=cmp1/(cmp2+ cmp3)
  return(cmp)|
}
```

#### III - 6 - o - Allocation optimale par Strate

```
## allocation optimale

calNhStNey=function(grandNh,Sh,nNey){
  cmp1=(grandNh*Sh)
  cmp2=sum(grandNh*Sh)
  cmp3=cmp1/ cmp2
  cmp= cmp3*nNey
  return(cmp)
}
```

### III - 6 - p - Résultats des Estimation de nombre d'unités par Strate si k = 2%

Strate	K	nAP.par.Strate	nAP.arrondi.Strate	nNey.par.Strate	nNey.arrondi..Strate
<b>AA1</b>	0.02	0.0066943	0	1.396411	1
<b>AA2</b>	0.02	0.0865921	0	5.022454	5
<b>AA3</b>	0.02	0.0426780	0	3.525959	4
<b>AA4</b>	0.02	0.2084614	0	7.793676	8
<b>AA5</b>	0.02	12.1335933	12	59.902869	60
<b>AA6</b>	0.02	0.1298431	0	6.163722	6
<b>BB1</b>	0.02	0.0196289	0	2.391161	2
<b>BB2</b>	0.02	0.4070930	0	10.890397	11
<b>BB3</b>	0.02	0.1431622	0	6.458073	6
<b>BB4</b>	0.02	0.4708533	0	11.713950	12
<b>BB5</b>	0.02	4.5893823	5	36.706055	37
<b>BB6</b>	0.02	0.0437661	0	3.573389	4
<b>CC1</b>	0.02	0.0285611	0	2.884356	3
<b>CC2</b>	0.02	0.4991820	0	12.059513	12
<b>CC3</b>	0.02	0.1649961	0	6.933093	7
<b>CC4</b>	0.02	0.4805611	0	11.834117	12
<b>CC5</b>	0.02	7.6902664	8	47.609582	48
<b>CC6</b>	0.02	0.0357189	0	3.228619	3
<b>DD1</b>	0.02	0.0079955	0	1.526097	2
<b>DD2</b>	0.02	0.1500124	0	6.610664	7

Strate	K	nAP.par.Strate	nAP.arrondi.Strate	nNey.par.Strate	nNey.arrondi..Strate
<b>DD3</b>	0.02	0.0567112	0	4.064542	4
<b>DD4</b>	0.02	0.1991909	0	7.618332	8
<b>DD5</b>	0.02	6.7835785	7	44.698110	45
<b>DD6</b>	0.02	0.0629609	0	4.290424	4

### Commentaire

En comparant les tailles d'échantillon des strates par zone obtenus si la précision relative est de 2%, on se rend compte à quel point l'allocation de Neyman est bien plus efficace par rapport à l'allocation proportionnelle, dans la mesure où, pour avoir un même niveau de précision, un petit nombre d'individus à enquêter est nécessaire.

## CHAPITRE IV : Stratification de la variable quantitative VAR 1

### IV - 7 - Estimation de la précision relative de la variable VAR1

Strate	Nh	Total	Moyenne	Sh	nhAp	nhAp1	NhSh	nhNey
AA1	2438	15277	6.266202	18.855651	61	60.950	45970.08	10
AA2	641	165405	258.042122	115.913154	16	16.025	74300.33	16
AA3	336	237157	705.824405	141.101600	8	8.400	47410.14	10
AA4	331	481413	1454.419940	286.543056	8	8.275	94845.75	20
AA5	491	1829518	3726.105906	1650.678877	12	12.275	810483.33	169
AA6	10	159333	15933.300000	6835.238207	2	0.250	68352.38	14
BB1	6128	12212	1.992820	11.028441	153	153.200	67582.29	14
BB2	383	92437	241.349870	111.470926	10	9.575	42693.36	9
BB3	198	141700	715.656566	147.724064	5	4.950	29249.36	6
BB4	163	234673	1439.711656	296.796845	4	4.075	48377.89	10
BB5	212	744066	3509.745283	1613.375441	5	5.300	342035.59	71
BB6	8	97986	12248.250000	2239.323543	2	0.200	17914.59	4
CC1	6861	18260	2.661420	12.661830	172	171.525	86872.81	18
CC2	602	141614	235.239203	111.722853	15	15.050	67257.16	14
CC3	202	137149	678.955445	136.927962	5	5.050	27659.45	6
CC4	122	164639	1349.500000	270.164757	3	3.050	32960.10	7
CC5	134	482092	3597.701492	1746.927344	3	3.350	234088.26	49
CC6	0	0	0.000000	0.000000	0	0.000	0.00	0
DD1	3905	3985	1.020487	7.811305	98	97.625	30503.15	6
DD2	171	45467	265.888889	119.642332	4	4.275	20458.84	4
DD3	138	98408	713.101449	146.289149	3	3.450	20187.90	4
DD4	194	278438	1435.247423	284.115919	5	4.850	55118.49	12
DD5	321	1189653	3706.084112	1647.398021	8	8.025	528814.76	110
DD6	11	155524	14138.545454	7464.689376	2	0.275	82111.58	17

#### IV - 7 - q - Allocation proportionnelle

Moyenne Strate	Variance moyenne Stratifié	Précision relative
288.6003	297.9287	5.980802

#### IV - 7 - r - Allocation optimale

Moyenne Strate	Variance moyenne Stratifié	Précision relative
288.6003	16.28766	1.398404

#### Commentaire

Pour la variable VAR1, on constate dans le cas de l'allocation proportionnelle, la précision relative est supérieure à 5%, l'échantillon est non significativement représentatif au niveau 5% donc présente beaucoup de marge d'erreur.

En plus, on remarque que la précision obtenue avec l'allocation de Neyman est bien meilleure que celle obtenue avec l'allocation proportionnelle, et est inférieure à 2%, l'échantillon est significativement représentatif au niveau 2%, très significativement représentatif au niveau 5% donc très moindre de marge d'erreur.

#### IV - 7 - s - Estimation de nombre d'unités si $k = 2\%$

Nombre de strate	K	nAP	nAP.arrondi	nNey	nNey.arrondi
24	0.02	4476.612	<b>4477</b>	714.9129	<b>715</b>

#### Commentaire

On remarque pour une précision relative de 2%, la taille d'échantillon d'enquête dans le cas de l'allocation proportionnelle est très supérieure à celle de l'allocation de Neyman qui est beaucoup moins coûteux à réaliser dans le cas d'une enquête.

**IV - 7 - t - Résultats des Estimations de nombre d'unités par Strate si k = 2%**

<b>Strate</b>	<b>K</b>	<b>nAP.par.Strate</b>	<b>nAP.arrondi.Strate</b>	<b>nNey.par.Strate</b>	<b>nNey arrondi Strate</b>
<b>AA1</b>	0.02	0.1101172	0	11.430181	11
<b>AA2</b>	0.02	0.2875483	0	18.474327	18
<b>AA3</b>	0.02	0.1170888	0	11.788243	12
<b>AA4</b>	0.02	0.4681071	0	23.582821	24
<b>AA5</b>	0.02	31.9995527	32	201.521766	202
<b>AA6</b>	0.02	0.2376756	0	16.995405	17
<b>BB1</b>	0.02	0.2379979	0	16.803926	17
<b>BB2</b>	0.02	0.0949592	0	10.615446	11
<b>BB3</b>	0.02	0.0445717	0	7.272677	7
<b>BB4</b>	0.02	0.1218689	0	12.028868	12
<b>BB5</b>	0.02	5.9258994	6	85.045077	85
<b>BB6</b>	0.02	0.0166890	0	4.454354	4
<b>CC1</b>	0.02	0.3932484	0	21.600399	22
<b>CC2</b>	0.02	0.2356303	0	16.723084	17
<b>CC3</b>	0.02	0.0398589	0	6.877354	7
<b>CC4</b>	0.02	0.0565848	0	8.195329	8
<b>CC5</b>	0.02	2.7959276	3	58.204628	58
<b>DD1</b>	0.02	0.0484851	0	7.584422	8
<b>DD2</b>	0.02	0.0218087	0	5.086966	5
<b>DD3</b>	0.02	0.0212344	0	5.019600	5



<b>Strate</b>	<b>K</b>	<b>nAP.par.Strate</b>	<b>nAP.arrondi.Strate</b>	<b>nNey.par.Strate</b>	<b>nNey arrondi Strate</b>
<b>DD4</b>	0.02	0.1581847	0	13.704878	14
<b>DD5</b>	0.02	13.9395917	14	131.486585	131
<b>DD6</b>	0.02	0.3404696	0	20.416547	20

## CHAPITRE V : Stratification de la variable quantitative VAR 2

### V - 8 - Estimation de la précision relative de la variable VAR2

Strate	Nh	Total	Moyenne	Sh	nhAp	nhAp1	NhSh	nhNey
AA1	2091	51182	24.47728	28.88378	52	52.275	60395.978	8
AA2	949	233716	246.27608	115.55179	24	23.725	109658.652	14
AA3	501	356920	712.41517	143.57774	13	12.525	71932.448	9
AA4	409	571813	1398.07579	283.14146	10	10.225	115804.856	15
AA5	290	1027422	3542.83448	1688.64033	7	7.250	489705.696	63
AA6	7	78527	11218.14286	568.70772	0	0.175	3980.954	1
BB1	2618	104926	40.07869	30.11369	65	65.450	78837.646	10
BB2	2286	543241	237.63823	111.80389	57	57.150	255583.696	33
BB3	939	681405	725.67093	147.75503	23	23.475	138741.975	18
BB4	874	1214920	1390.06865	272.09564	22	21.850	237811.592	31
BB5	364	1170053	3214.43132	1386.97885	9	9.100	504860.300	65
BB6	11	140105	12736.81818	2964.91912	0	0.275	32614.110	4
CC1	2754	112173	40.73094	29.53432	69	68.850	81337.511	10
CC2	2629	619983	235.82465	109.46480	66	65.725	287782.971	37
CC3	1107	800571	723.18970	145.20024	28	27.675	160736.669	21
CC4	911	1269651	1393.68935	277.89384	23	22.775	253161.291	33
CC5	509	1784886	3506.65226	1606.20501	13	12.725	817558.349	105
CC6	11	151601	13781.90909	3671.09740	0	0.275	40382.071	5
DD1	1938	72343	37.32869	30.07983	48	48.450	58294.701	8
DD2	1455	336486	231.26186	107.97364	36	36.375	157101.641	20

Strate	Nh	Total	Moyenne	Sh	nhAp	nhAp1	NhSh	nhNey
<b>DD3</b>	567	409475	722.17813	142.95178	14	14.175	81053.657	10
<b>DD4</b>	486	690885	1421.57407	274.07490	12	12.150	133200.404	17
<b>DD5</b>	290	1013653	3495.35517	1632.84579	7	7.250	473525.279	61
<b>DD6</b>	4	54778	13694.50000	1660.60642	0	0.100	6642.426	1

#### V - 8 - u - Allocation proportionnelle

Moyenne Strate	Variance moyenne stratifiée	Précision relative
562.1131	287.0739	3.014208

#### V - 8 - v - Allocation optimale

Moyenne Strate	Variance moyenne stratifiée	Précision relative
562.1131	55.3681	1.32375

#### Commentaire

Pour la variable VAR2, on constate dans le cas de l'allocation proportionnelle, la précision relative est inférieure à 5%, l'échantillon est significativement représentatif au niveau 5% donc moindre de marge d'erreur.

En plus, on remarque que la précision obtenue avec l'allocation de Neyman est bien meilleure que celle obtenue avec l'allocation proportionnelle, et est inférieure à 2%, l'échantillon est significativement représentatif au niveau 2%, très significativement représentatif au niveau 5% donc très moindre de marge d'erreur.

#### V - 8 - w - Estimation de nombre d'unités si $k = 2\%$

Nombre de strates	K	nAP	nAP.arrondi	nNey	nNey.arrondi
24	0.02	1320.836	<b>1321</b>	280.7524	<b>281</b>

## Commentaire

On remarque pour une précision relative de 2%, la taille d'échantillon d'enquête dans le cas de l'allocation proportionnelle est très supérieure à celle de l'allocation de Neyman qui est beaucoup moins coûteuse à réaliser dans le cas d'une enquête.

### V - 8 - x - Résultats des Estimations de nombre d'unités par Strate si $k = 2\%$

Strate	K	nAP.par.Strate	nAP.arrondi.Strate	nNey.par.Strate	nNey.arrondi..Strate
AA1	0.02	0.0501044	0	3.6459665	4
AA2	0.02	0.1651507	0	6.6198411	7
AA3	0.02	0.0710654	0	4.3423967	4
AA4	0.02	0.1841315	0	6.9908733	7
AA5	0.02	3.2571294	3	29.5624087	30
AA6	0.02	0.0002177	0	0.2403211	0
BB1	0.02	0.0853735	0	4.7592477	5
BB2	0.02	0.8969452	1	15.4290010	15
BB3	0.02	0.2643404	0	8.3755345	8
BB4	0.02	0.7761582	1	14.3561399	14
BB5	0.02	3.4678091	3	30.4772574	30
BB6	0.02	0.0145917	0	1.9688390	2
CC1	0.02	0.0908736	0	4.9101588	5
CC2	0.02	1.1371359	1	17.3727973	17
CC3	0.02	0.3547814	0	9.7033037	10
CC4	0.02	0.8795189	1	15.2827660	15
CC5	0.02	9.0186924	9	49.3541207	49
CC6	0.02	0.0223544	0	2.4377729	2
DD1	0.02	0.0466786	0	3.5191173	4
DD2	0.02	0.3389459	0	9.4838654	9
DD3	0.02	0.0902290	0	4.8930232	5
DD4	0.02	0.2435922	0	8.0410026	8
DD5	0.02	3.0476716	3	28.5856341	29
DD6	0.02	0.0006060	0	0.4009880	0

## COMMENTAIRE GÉNÉRAL

À l'issue de cette étude de cas, nous pouvons dire que les meilleures précisions sont obtenues pour la variable VAR3, car les précisions sont inférieures à 5% dans chacune des zones, et dans une moindre mesure, pour la variable 2 pour laquelle la précision est toutefois supérieure à 5%.

Les précisions sont mauvaises pour la variable VAR1. Cela ne nous surprend pas dans la mesure où nous avons trouvé un coefficient de variation très élevé pour cette dernière variable dans chaque zone.

On remarque que les précisions obtenues avec l'allocation de Neyman sont globalement bien meilleures que celles obtenues avec l'allocation proportionnelle.

La précision est ainsi toujours inférieure à 2% pour la variable VAR3, donc très bonne. Elle est toujours inférieure à 5% pour la variable 2, donc bonne.

Elle est bonne pour la variable VAR1 dans les régions AA et DD, mais reste légèrement supérieure à 5% dans les régions BB et CC.

L'allocation de Neyman est bien plus efficace par rapport l'allocation proportionnelle, dans la mesure où, pour avoir un même niveau de précision, un petit nombre d'individus à enquêter est nécessaire donc revient moins coûteux.