

# SAS : Données hospitalières

Souraya Ahmed Abderemane, Angelo Lamure-Fontanini, Clément Ruiz, Célian Toureille

Décembre 2023

## 1) *Importez les données et familiarisez-vous avec les données.*

Ligne 14 à 104.

La macro import\_tables se charge de l'importation des fichiers avec en entrée l'année de l'import. Les différentes pages des fichiers Excel sont mises en trois tableaux, un par année, seul les variables utiles sont gardées, les variables sont re-typées et certaines variables sont calculées (le département et la région par exemple).

## 2) *Décrivez les données (analyse, problème de qualité, outliers, ...).*

En prenant les questions du sujet je repère les différentes informations dont nous avons besoin.

- Année de l'établissement (annee = annee du fichier).
- taille\_MCO , taille\_M, Taille\_C et taille\_O
- nombre de lits en Médecine / Chirurgie / Obstétrique
- N° finess
- nombre d'accouchement, nombre de lit d'obstétrique, catégorie d'établissement, niveau de maternité
- établissements pratiquant des actes d'obstétrique par région
- nombre min et max d'accouchement

Variables choisies en rapport avec le sujet :

- CI\_AC1 : Nombre de lit installées en médecine
- CI\_AC6 : Nombre de lit installées en chirurgie
- CI\_AC8 : Nombre de lit installées en obstétrie
- CI\_A3 : Nombre de RSA d'obstétrique (HC)
- CI\_A6 : Nombre de RSA d'obstétrique (ambulatoire)
- CI\_A11 : Nombre d'accouchements
- CI\_A9 : Nb de places installées en obstétrique
- CI\_E6 : Niveau de la maternité
- CI\_RH5 : ETP médicaux (dont gynécologues-obstétriciens)
- P3 : IP - DMS Obstétrique (hors ambulatoire)
- P10 : Taux de césarienne
- P11 : Taux de péridurale (accouchements par voie basse)
- RH1 : Nombre d'accouchements par obstétricien et sage-femm
- RH6 : Nombre de sages-femmes par obstétricien
- RH9 : Turn-over global
- finess : identifiant des établissements de santé
- rs : raison sociale
- cat : catégorie établissement de santé (région écrite sur les 2 premier caractères)

- taille\_MCO : taille de l'établissement MCO (Médecine / Chirurgie / Obstétrique)
- taille\_M : taille Médecine
- taille\_C : taille Chirurgie
- taille\_O : taille Obstétrique
- Valeur de Lit et Places.

Soucis de qualité dans les données :

Toutes les données sont en chaînes de caractères. RH7 n'est pas renseignée. Beaucoup de valeur manquantes dans activité globale. Le label des variables n'est pas écrit en entier et n'est pas entièrement expliqué. Certaines colonnes ont des valeurs manquante écrit avec .z

**3)** *En 2019, existe-il des établissements nouveaux ou non présents par rapport à 2018 et / ou 2017 ?*

Non, aucun nouvel établissement est apparu entre 2017 et 2019.

**4)** *Est-ce que certains établissements ont changé de taille (taille\_MCO , taille\_M, Taille\_C et taille\_O) sur la période 2017-2019 ?*

Non, aucun établissement n'a changé de taille.

*Est-ce que des établissements ont changé d'activité ?*

Non, aucun établissement n'a changé d'activité.

**5)** *A l'aide du nombre de lits, essayez de déterminer quels sont les seuils qui ont été utilisés pour constituer les variables taille\_M, taille\_C et taille\_O (variables catégorielles). Justifiez votre raisonnement.*

Proc means pour avoir, pour chaque variable le minimum et maximum du nombre de lit par catégorie. A l'aide des valeurs obtenue nous en déduisant les seuils d'attribution des tailles.

La procédure MEANS

taille\_MCO=T0

Variable d'analyse : CI_AC1_6_8				
N	Moyenne	Ec-type	Minimum	Maximum
191	0.3036649	4.0538399	0	56.0000000

taille\_MCO=T1

Variable d'analyse : CI_AC1_6_8				
N	Moyenne	Ec-type	Minimum	Maximum
560	33.0517857	28.2400108	0	147.0000000

taille\_MCO=T2

Variable d'analyse : CI_AC1_6_8				
N	Moyenne	Ec-type	Minimum	Maximum
318	100.2547170	52.9968860	0	293.0000000

taille\_MCO=T3

Variable d'analyse : CI_AC1_6_8				
N	Moyenne	Ec-type	Minimum	Maximum
346	240.3381503	110.2067811	34.0000000	616.0000000

taille\_MCO=T4

Variable d'analyse : CI_AC1_6_8				
N	Moyenne	Ec-type	Minimum	Maximum
70	669.2857143	261.1185606	190.0000000	1159.00

taille\_MCO=T5

Variable d'analyse : CI_AC1_6_8				
N	Moyenne	Ec-type	Minimum	Maximum
18	2225.83	2528.15	857.0000000	12064.00

FIGURE 1 – Titre

La procédure MEANS

taille\_M=' '

Variable d'analyse : CI_AC1				
N	Moyenne	Ec-type	Minimum	Maximum
184	0.1902174	2.5802342	0	35.0000000

taille\_M=M0

Variable d'analyse : CI_AC1				
N	Moyenne	Ec-type	Minimum	Maximum
17	0.1176471	0.4850713	0	2.0000000

taille\_M=M1

Variable d'analyse : CI_AC1				
N	Moyenne	Ec-type	Minimum	Maximum
529	17.5406427	15.8155428	0	82.0000000

taille\_M=M2

Variable d'analyse : CI_AC1				
N	Moyenne	Ec-type	Minimum	Maximum
520	63.6423077	49.2723443	0	235.0000000

taille\_M=M3

Variable d'analyse : CI_AC1				
N	Moyenne	Ec-type	Minimum	Maximum
136	176.5441176	84.6285013	4.0000000	354.0000000

taille\_M=M4

Variable d'analyse : CI_AC1				
N	Moyenne	Ec-type	Minimum	Maximum
110	450.0454545	211.9019761	48.0000000	1028.00

taille\_M=M5

Variable d'analyse : CI_AC1				
N	Moyenne	Ec-type	Minimum	Maximum
7	2402.86	2572.84	910.0000000	8162.00

FIGURE 2 – Titre

La procédure MEANS

taille\_C=' '

Variable d'analyse : CI_AC6				
N	Moyenne	Ec-type	Minimum	Maximum
246	0.0853659	1.3389111	0	21.0000000

taille\_C=C0

Variable d'analyse : CI_AC6				
N	Moyenne	Ec-type	Minimum	Maximum
207	0	0	0	0

taille\_C=C1

Variable d'analyse : CI_AC6				
N	Moyenne	Ec-type	Minimum	Maximum
231	4.3852814	14.3960607	0	140.0000000

taille\_C=C2

Variable d'analyse : CI_AC6				
N	Moyenne	Ec-type	Minimum	Maximum
306	36.2385621	20.1984892	0	114.0000000

taille\_C=C3

Variable d'analyse : CI_AC6				
N	Moyenne	Ec-type	Minimum	Maximum
359	71.3064067	33.4512192	0	221.0000000

taille\_C=C4

Variable d'analyse : CI_AC6				
N	Moyenne	Ec-type	Minimum	Maximum
86	127.9767442	56.8976329	21.0000000	342.0000000

taille\_C=C5

Variable d'analyse : CI_AC6				
N	Moyenne	Ec-type	Minimum	Maximum
68	293.4264706	381.9974727	62.0000000	3052.00

FIGURE 3 – Titre

La procédure MEANS				
taille_O=' '				
Variable d'analyse : CI_AC8				
N	Moyenne	Ec-type	Minimum	Maximum
1007	0.3128103	2.8246404	0	45.0000000

  

taille_O=O0				
Variable d'analyse : CI_AC8				
N	Moyenne	Ec-type	Minimum	Maximum
24	1.2083333	4.4033502	0	20.0000000

  

taille_O=O1				
Variable d'analyse : CI_AC8				
N	Moyenne	Ec-type	Minimum	Maximum
16	11.1250000	4.7871355	0	18.0000000

  

taille_O=O2				
Variable d'analyse : CI_AC8				
N	Moyenne	Ec-type	Minimum	Maximum
154	19.4480519	6.4585579	0	50.0000000

  

taille_O=O3				
Variable d'analyse : CI_AC8				
N	Moyenne	Ec-type	Minimum	Maximum
124	29.6370968	8.9117942	14.0000000	63.0000000

  

taille_O=O4				
Variable d'analyse : CI_AC8				
N	Moyenne	Ec-type	Minimum	Maximum
158	55.5632911	19.5931333	22.0000000	151.0000000

  

taille_O=O5				
Variable d'analyse : CI_AC8				
N	Moyenne	Ec-type	Minimum	Maximum
20	156.1500000	168.7671317	65.0000000	850.0000000

FIGURE 4 – Titre

6) Indiquez dans un tableau croisé par catégorie d'établissement (cat) et par année, le nombre d'établissement, le nombre de lits en MCO en détaillant par Médecine / Chirurgie / Obstétrique.

		Nb lit MCO		Nb lit M		Nb lit C		Nb lit O	
		N	Sum	N	Sum	N	Sum	N	Sum
Catégorie	Année								
CH	2017	569	87684.00	569	63514.00	569	15543.00	569	8627.00
	2018	569	86308.00	569	63232.00	569	14699.00	569	8377.00
	2019	569	85134.00	569	62832.00	569	14173.00	569	8129.00
CHR	2017	71	67634.00	71	44322.00	71	18429.00	71	4883.00
	2018	71	67527.00	71	44561.00	71	18113.00	71	4853.00
	2019	71	67877.00	71	45254.00	71	17766.00	71	4857.00
CLCC	2017	26	2836.00	26	1983.00	26	853.00	26	0.00
	2018	26	2783.00	26	1940.00	26	843.00	26	0.00
	2019	26	2737.00	26	1907.00	26	830.00	26	0.00
CLI	2017	690	49864.00	690	15350.00	690	29952.00	690	4562.00
	2018	690	48242.00	690	15474.00	690	28474.00	690	4294.00
	2019	690	47361.00	690	15628.00	690	27574.00	690	4159.00
EBNL	2017	147	12502.00	147	7576.00	147	3904.00	147	1022.00
	2018	147	12462.00	147	7639.00	147	3841.00	147	982.00
	2019	147	12682.00	147	7977.00	147	3718.00	147	987.00

FIGURE 5 – Titre

7) Les deux premiers caractères du N° finesse correspondent au département. Indiquez le nombre d'établissement par catégorie (cat) par région.

		N
cat	region	
CH	Hauts-de-France	51
	Auvergne-Rhône-Alpe	84
	Bourgogne-Franche-Co	41
	Bretagne	30
	Centre-Val de Loire	26
	Grand Est	54
	Normandie	33
	Nouvelle-Aquitaine	61
	Occitanie	59
	Pays de la Loire	29
	Provence-Alpes-Côte	46
	Île-de-France	37
CHR	Hauts-de-France	3
	Auvergne-Rhône-Alpe	4
	Bourgogne-Franche-Co	2
	Bretagne	2
	Centre-Val de Loire	2
	Grand Est	4
	Normandie	2
	Nouvelle-Aquitaine	3
	Occitanie	3
	Pays de la Loire	2
	Provence-Alpes-Côte	7
	Île-de-France	34
CLCC	Hauts-de-France	1
	Auvergne-Rhône-Alpe	2
	Bourgogne-Franche-Co	1
	Bretagne	1
	Grand Est	3
	Normandie	4
	Nouvelle-Aquitaine	1
	Occitanie	2
	Pays de la Loire	3
	Provence-Alpes-Côte	3
CLI	Île-de-France	5
	Hauts-de-France	53
	Auvergne-Rhône-Alpe	56
	Bourgogne-Franche-Co	17
	Bretagne	56
	Centre-Val de Loire	22
	Grand Est	39
	Normandie	25
	Nouvelle-Aquitaine	64
	Occitanie	84
	Pays de la Loire	43
	Provence-Alpes-Côte	72
EBNL	Île-de-France	114
	Hauts-de-France	12
	Auvergne-Rhône-Alpe	12
	Bourgogne-Franche-Co	2
	Bretagne	13
	Centre-Val de Loire	2
	Grand Est	33
	Normandie	3
	Nouvelle-Aquitaine	12
	Occitanie	6
	Pays de la Loire	5
	Provence-Alpes-Côte	15
	Île-de-France	32

FIGURE 6 – Titre



8) Indiquez dans un tableau le nombre d'accouchement, le nombre de lit d'obstétrique par catégorie d'établissement (cat) et par niveau de maternité. Quels sont les 5 établissements avec la plus forte activité d'obstétrique ? Ce tableau n'a de sens que pour les établissements pratiquant des actes d'obstétrique.

			Nombre d'accouchements	Nombre de lit installées en obstétrique
			Sum	Sum
Catégorie	Niveau de maternité	Année		
CH	0	2017	26.00	0.00
		2018	232.00	0.00
		2019	74.00	0.00
	1	2017	57135.00	1606.00
		2018	56754.00	1587.00
		2019	53199.00	1490.00
	2	2017	226773.00	5210.00
		2018	222577.00	4919.00
		2019	229389.00	4993.00
	3	2017	82639.00	1809.00
		2018	85161.00	1869.00
		2019	81615.00	1642.00
CHR	0	2017	1.00	0.00
		2019	2.00	0.00
	1	2017	1057.00	20.00
		2018	1135.00	30.00
		2019	1298.00	30.00
	2	2017	17276.00	426.00
		2018	17210.00	416.00
		2019	16635.00	415.00
	3	2017	181001.00	4416.00
		2018	181710.00	4386.00
		2019	182986.00	4391.00
CLI	0	2017	1009.00	0.00
		2018	1460.00	0.00
		2019	9.00	0.00
	1	2017	67746.00	1958.00
		2018	63531.00	1839.00
		2019	60366.00	1728.00
	2	2017	103195.00	2532.00
		2018	101257.00	2421.00
		2019	98753.00	2416.00
	3	2018	666.00	0.00
		2019	701.00	15.00

FIGURE 7 – Titre

	finess	rs	total_accouchement
1	750712184	AP-HP	114402
2	690781810	HOSPICES CIVILS DE LYON	32488
3	770021145	GRAND HÔPITAL DE L'EST FRANCILIEN	21971
4	970408589	CHU DE LA REUNION	18947
5	670780055	CHU DE STRASBOURG	18189

FIGURE 8 – Titre

9) Résumez les indicateurs de la question 8 par région. Indiquez également le nombre min et max d'accouchement.

		Nombre d'accouchements			Nombre de lit installées en obstétrique		
		Sum	Min	Max	Sum	Min	Max
region	Année						
Hauts-de-France	2017	68785.00	54.00	5433.00	1741.00	0.00	131.00
	2018	66117.00	1.00	5495.00	1654.00	0.00	125.00
	2019	65575.00	1.00	5500.00	1614.00	0.00	125.00
Auvergne-Rhône-Alpe	2017	89462.00	1.00	10869.00	2126.00	0.00	262.00
	2018	89319.00	1.00	10718.00	2062.00	0.00	258.00
	2019	87933.00	1.00	10901.00	2025.00	0.00	258.00
Bourgogne-Franche-Co	2017	27093.00	2.00	3376.00	733.00	0.00	95.00
	2018	26283.00	1.00	3209.00	691.00	0.00	95.00
	2019	25583.00	1.00	3185.00	672.00	0.00	95.00
Bretagne	2017	32252.00	3.00	3623.00	752.00	0.00	91.00
	2018	31952.00	1.00	3680.00	738.00	0.00	91.00
	2019	31662.00	1.00	3732.00	710.00	0.00	91.00
Centre-Val de Loire	2017	25156.00	1.00	4830.00	693.00	0.00	110.00
	2018	24953.00	135.00	4893.00	640.00	0.00	109.00
	2019	24558.00	1.00	4936.00	603.00	0.00	104.00
Grand Est	2017	54129.00	1.00	6172.00	1477.00	0.00	147.00
	2018	54585.00	1.00	6002.00	1477.00	0.00	151.00
	2019	54702.00	1.00	6015.00	1418.00	0.00	148.00
Normandie	2017	33453.00	1.00	3227.00	844.00	0.00	90.00
	2018	33265.00	1.00	3320.00	828.00	0.00	91.00
	2019	32926.00	51.00	3076.00	785.00	0.00	91.00
Nouvelle-Aquitaine	2017	54087.00	1.00	5360.00	1433.00	0.00	146.00
	2018	53574.00	1.00	5594.00	1357.00	0.00	148.00
	2019	53604.00	1.00	5615.00	1323.00	0.00	148.00
Occitanie	2017	57217.00	1.00	4720.00	1454.00	0.00	121.00
	2018	56455.00	1.00	5058.00	1382.00	0.00	121.00
	2019	56198.00	1.00	5055.00	1364.00	0.00	121.00
Pays de la Loire	2017	39808.00	537.00	4465.00	949.00	15.00	106.00
	2018	39223.00	502.00	4210.00	921.00	15.00	96.00
	2019	38926.00	477.00	4178.00	916.00	15.00	96.00
Provence-Alpes-Côte	2017	63069.00	1.00	5338.00	1555.00	0.00	152.00
	2018	62236.00	1.00	5241.00	1550.00	0.00	161.00
	2019	62115.00	1.00	5391.00	1509.00	0.00	139.00
Île-de-France	2017	216525.00	1.00	37945.00	4397.00	0.00	850.00
	2018	213443.00	365.00	38519.00	4298.00	0.00	844.00
	2019	211628.00	1.00	37938.00	4323.00	0.00	866.00

FIGURE 9 – Titre

**10)** *A l'aide de variables présentes dans le fichier essayez de calculer un score de « qualité » pour les accouchements (obstétrique) afin de classer les établissements. Justifiez le calcul de votre score et proposez une interprétation de votre score. A l'aide de votre score, indiquez quels sont les 5 meilleurs et 5 moins bons établissements.*

On a décidé de donner à notre score une approche du bien-être des personnes traitées et des conditions de travail du personnel soignant. Pour cela, pour le score, on va s'intéresser à trois caractéristiques que l'on a trouvé pertinentes, le nombre de sages-femmes par obstétricien, le nombre de lits en obstétrique et enfin le turn-over global, que l'on notera respectivement a, b et c. On va donc calculer le score de l'hôpital  $i$  tel que :

$$score_i = \frac{a_i - \min(a)}{\max(a) - \min(a)} + \frac{b_i - \min(b)}{\max(b) - \min(b)} + \frac{\max(c) - c_i}{\max(c) - \min(c)}$$

On a donc faire une *normalisation minmax* sur chacune des variables pour avoir pour chacune une valeur entre 0 et 1, et donc avoir un score entre 0 et 3.

Ici, on remarque que l'on a, pour le turn-over global, mis  $\max(c) - c_i$  : ici, plus le turn-over global est élevé, plus le score sera faible. En effet, un turn-over global élevé signifie que le personnel est surmené.

Si a et/ou b et/ou c est manquantes, on va nullifier le score associé à la variable donnée. Par exemple, si a et c sont manquantes, on aura :

$$score_i = 0 + \frac{b_i - \min(b)}{\max(b) - \min(b)} + 0$$

Plus notre score sera élevé, plus le bien-être des personnes et du personnel sera qualitatif.

**11)** *Etudiez la stabilité de votre score sur les années 2017, 2018 et 2019. Par exemple, constatez-vous que les établissements qui ont un score qui n'est pas stable s'accompagne de changement dans les activités et les structures ? De même, une stabilité dans le score doit se traduire par une stabilité dans les activités et les structures. Donnez des exemples concrets.*

Le score est stable pour la quasi totalité des établissements d'une année à l'autre. Nous avons vu à la question 4 qu'aucun établissement a changé d'activité d'une année sur l'autre.

Le score se basant sur 3 variables, nous concluons donc qu'un fort changement dans le score correspondrait à une chute (ou hausse) brutale du bien-être et des conditions de travail des personnes et du personnel.

**12)** *Calculez votre score par région et par année. Précisez comment vous passez d'un score calculé par établissement à un score calculé par région / année. Existe-il des différences entre les régions ? Quelles sont les limites de votre score ? Indiquez comment pourriez-vous l'améliorer ?*

Oui il existe des différences entre régions, mais qui reste assez faibles.

		score
		Mean
region	annee	
Hauts-de-France	2017	0.99
	2018	0.97
Auvergne-Rhône-Alpe	2017	0.92
	2018	0.90
Bourgogne-Franche-Co	2017	0.97
	2018	0.98
Bretagne	2017	0.99
	2018	0.98
Centre-Val de Loire	2017	0.96
	2018	0.96
Grand Est	2017	0.96
	2018	0.97
Normandie	2017	0.94
	2018	0.94
Nouvelle-Aquitaine	2017	0.96
	2018	0.92
Occitanie	2017	0.97
	2018	0.96
Pays de la Loire	2017	0.99
	2018	0.91
Provence-Alpes-Côte	2017	0.87
	2018	0.96
Île-de-France	2017	1.01
	2018	1.01

FIGURE 10 – Titre

Les limites de notre score réside en quelques points :

- Nous n'avons que 3 variables pour calculer notre score : on aurait besoin de plus de variables liées aux conditions de travail et au bien-être pour avoir un score plus pertinent. Par exemple, la variable RH7 (nombre moyen d'heures travailler) n'est jamais renseignée.
- Il y a beaucoup de données manquantes, ce qui fausse les calculs du score.