



**Formation : *Master 1 Informatique***

# **RAPPORT DE PROJET**

**Analyse des structures et des activités de soins des  
établissements hospitaliers publics et privés**

**Lansana CISSE  
Falconne KPAMEGAN  
Awa KARAMOKO  
Cynthia RAMALIARISOA**

**Année Universitaire : 2023-2024**

# Tables des matières

<b>Liste des figures.....</b>	<b>2</b>
<b>Introduction.....</b>	<b>3</b>
<b>I. Contexte de l'étude.....</b>	<b>5</b>
1. Objectifs du traitement de données en SAS.....	5
2. Présentation de la base de données.....	5
<b>II. Réponses aux questions (Analyse des données).....</b>	<b>6</b>
1. Importez les données et familiarisez-vous avec les données.....	6
2. Description des données (analyse, problème de qualité, outliers,...).....	6
a. Transformation de données en numérique.....	6
b. Etudes descriptives des données.....	7
c. Gestion des données aberrantes.....	7
d. Gestion des données manquantes.....	8
e. Identification et suppression des doublons.....	9
3. En 2019, présence ou non de nouveaux établissements, disparition de certains établissements de 2017 ou 2018 ?.....	9
4. Est-ce que certains établissements ont changé de taille (taille_MCO , taille_M, Taille_C et taille_O) sur la période 2017-2019 ? Est-ce que des établissements ont changé d'activité?.....	10
a. Quels sont les établissements ayant changé de taille sur la période 2017-2019 ?.....	10
b. Quels sont les établissements qui ont changé d'activité ?.....	10
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.....	11
6. Indiquez dans un tableau croisé par catégorie d'établissement (cat) et par année, le nombre d'établissements, le nombre de lits en MCO en détaillant par Médecine, Chirurgie, Obstétrique..	13
7. Les deux premiers caractères du N° finess correspondent au département. Indiquez le nombre d'établissements par catégorie (cat) par région.....	14
8. Indiquez dans un tableau le nombre d'accouchement, le nombre de lits 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.....	15
9. Résumez les indicateurs de la question 8 par région. Indiquez également le nombre min et max d'accouchement.....	16
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.....	17
Pour calculer le score de qualité, nous avons utilisé le taux de césarienne et le taux de péridurale et nous avons divisé par 2 pour chaque établissement en supposant que la péridurale et la césarienne ont la même importance.....	17
11. Étudiez 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'accompagnent 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.....	18
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 pouvez-vous l'améliorer ?	18
<b>Conclusion.....</b>	<b>20</b>
<b>Référence.....</b>	<b>21</b>

# Liste des figures

*Figure 1 - Importation des données de 2017*

*Figure 2 - Transformation de données en numérique*

*Figure 3 - Résultat de MEANS sur le fichier activité globale 2017*

*Figure 4 - Détection des valeurs aberrantes par UNIVARIATE*

*Figure 5 - Tableau montrant les données manquantes*

*Figure 6 - imputation des valeurs manquantes par la moyenne*

*Figure 7 - Résultat de l'imputation*

*Figure 8 - Suppression des doublons*

*Figure 9 - Fusion des données et tri par identifiant*

*Figure 10 - Filtrage des lignes où col1, col2 et cols manquantes*

*Figure 11 - Jointure sur les données établissements et comparaison de tailles*

*Figure 12 - Jointure sur les données des activités globales et comparaison d'activités*

*Figure 13 - Chargement et fusion des données des lits et tri par finess*

*Figure 14 - Transformation de Valeur en numérique et visualisation de la distribution*

*Figure 15 - Quantiles sur la variable Valeur*

*Figure 16 - Tests des seuils et classification des valeurs par catégorie*

*Figure 17 - Calcul du nombre d'établissements avec PROC SQL*

*Figure 18 - Résultat tableau croisé*

*Figure 19 - Code pour le nombre d'établissement par catégorie et par région*

*Figure 20 - Tableau résultat*

*Figure 21 - Code question 8*

*Figure 22 - Liste des 5 établissements avec la plus forte activité obstétrique*

*Figure 23 - Fusion des tables, calcul min et max d'accouchements par région et cat*

*Figure 24 - Tableau montrant les résultats*

*Figure 25 - Calcul du score de qualité pour les accouchements*

*Figure 26 - Liste des 5 meilleurs établissements*

*Figure 27 - Liste des 5 moins bons établissements*

*Figure 28 - Calcul du changement de score d'une année à l'autre*

*Figure 29 - Tableau montrant les résultats*

*Figure 30 - Calcul du score par région et par année*

*Figure 31 - Résultat score par région et par année*

## Introduction

Dans le cadre de ce projet en groupe, nous nous pencherons sur l'exploration et l'analyse de données relatives aux établissements hospitaliers entre 2017 et 2019. Ces données fournissent des informations essentielles sur les établissements hospitaliers qu'ils soient publics ou privés, couvrant à la fois leur structure et leurs activités de soins. Ce rapport synthétise l'ensemble des réponses aux questions posées dans le cadre du projet, tout en exposant les raisonnements qui ont guidé notre approche pour y répondre.

Notre démarche repose sur une analyse approfondie des données disponibles, visant à extraire des informations pertinentes et à dégager des tendances significatives. Nous mettrons en lumière, les méthodologies utilisées ainsi que les choix de traitement des données.

# I. Contexte de l'étude

## 1. Objectifs du traitement de données en SAS

Les objectifs de notre analyse des données des établissements hospitaliers publics et privés de 2017 à 2019 reposent principalement sur une analyse descriptive des capacités hospitalières, en particulier en ce qui concerne les lits disponibles dans les services de Médecine, Chirurgie et Obstétrique.

## 2. Présentation de la base de données

Pour répondre aux exigences (questions) de notre analyse, nous allons exploiter trois ensembles de données, spécifiquement les fichiers Hospi\_2019.xlsx, Hospi\_2018.xlsx et Hospi\_2017.xlsx. Chaque fichier comprend quatre feuilles de données distinctes. Nous examinerons minutieusement les intitulés de ces feuilles et détaillerons les variables qu'elles contiennent.

**Feuille 1 - Indicateur** : Cette feuille décrit les différents indicateurs utilisés pour évaluer les établissements hospitaliers. Elle contient les variables suivantes :

- Axe
- Sous axe
- Indicateur (indic)
- Code hospitalière (code\_hospidiag)
- Libellé (libellé)

**Feuille 2 - Établissement** : Cette feuille fournit des données détaillées sur les établissements de santé. Elle inclut les colonnes suivantes :

- Identifiants des établissements de santé (Finess)
- Raison sociale (Rs)
- Catégorie des établissements de santé (Cat)
- Taille de l'établissement Médecine, Chirurgie, Obstétrique (Taille\_MCO)
- Taille Médecine (Taille\_M)
- Taille Chirurgie (Taille\_C)
- Taille Obstétrique (Taille\_O)

**Feuille 3 - Lits et Places** : Cette feuille se concentre sur les capacités en termes de lits et de places des établissements de santé. Elle comprend les colonnes suivantes :

- Identifiants des établissements de santé (Finess)
- Indicateur
- Valeur

**Feuille 4 - Activité Globale** : Cette feuille est dédiée à la présentation des diverses activités menées par chaque établissement de santé. Les informations contenues dans cette feuille pourraient inclure :

- Identifiants des établissements de santé (Finess)
- Autres colonnes contenant chacune les différentes activités existantes

## II. Réponses aux questions (Analyse des données)

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

Cette image illustre l'importation des différents jeux de données dans le logiciel SAS.

```
options validvarname=any;
libname TD6 '/home/u63636495/TD5/';
FILENAME REFFILE '/home/u63636495/TD5/Hospi_2017.xlsx';

/*importation de toute les feuilles de hospi 2017*/

PROC IMPORT DATAFILE=REFFILE
            OUT=TD6.hospi_2017_indicateur
            DBMS=XLSX
            replace;
            SHEET="Indicateur";
RUN;

PROC IMPORT DATAFILE=REFFILE
            OUT=TD6.hospi_2017_etablissement
            DBMS=XLSX
            replace;
            SHEET="Etablissement";
RUN;
```

*Figure 1 - Importation des données de 2017*

### 2. Description des données (analyse, problème de qualité, outliers,...)

Pour effectuer la description des données, nous allons travailler sur les données numériques de notre jeu de données. En effet, les données numériques sont dans les feuilles "activité globales" de chaque année c'est-à-dire de 2017 à 2019.

#### a. Transformation de données en numérique

La capture ci-dessous est le code pour transformer les données dont nous aurons besoin en numérique pour effectuer notre analyse.

```

data test2017;
  set TD6.hospi_2017_activiteglobale;
  array tab(*) A7--RH10;
  array X(74);
  do i = 1 to dim(tab);
    x(i) = input(tab(i), ? comma10.1) + 0;
  end;
run;

```

Figure 2 - Transformation de données en numérique

## b. Etudes descriptives des données

L'étude descriptive des données permet d'avoir une vue d'ensemble des caractéristiques numériques du jeu de données telles que la moyenne, la médiane, l'écart-type, etc.. Pour ce faire, nous avons mis en œuvre la procédure MEANS sur les données numériques de notre jeu de données c'est-à-dire les feuilles "activité globale" des années 2017, 2018 et 2019.

La capture ci-dessous est le résultat de la procédure MEANS sur le fichier "activité globale" de 2017. Nous vous invitons à consulter le code sas, de le compiler et voir les résultantes des autres années.

La procédure MEANS					
Variable	N	Moyenne	Ec-type	Minimum	Maximum
X1	1308	13.4594037	16.7512997	0	99.9000000
X2	1308	10.5763761	17.7808865	0	246.3000000
X3	1308	136.9547401	161.5639084	0	837.2000000
X4	213	2.9361502	4.9923015	0	20.7000000
X5	194	6.2185567	10.0639698	0.1000000	56.7000000
X6	1281	20.2830601	27.2400094	0	93.3000000
X7	1185	70.8368776	40.7944437	0.6000000	645.5000000

Figure 3 - Résultat de MEANS sur le fichier activité globale 2017

## c. Gestion des données aberrantes

Nous avons utilisé la procédure PROC UNIVARIATE pour détecter les valeurs aberrantes dans les variables numériques.

```

/* activité globale 2017 */
PROC UNIVARIATE DATA=test2017;
  VAR _NUMERIC_;
  OUTPUT OUT=Outliers;
RUN;

```

Figure 4 - Détection des valeurs aberrantes par UNIVARIATE



#### d. Gestion des données manquantes

Afin de gérer les données manquantes, nous avons utilisés PROC MEANS avec l'option NMISS N MAX MIN, qui générera un rapport entre nombre de valeurs manquantes NMISS, le nombre total d'observations N, la valeur maximale MAX et la valeur MIN pour chaque variable spécifiée dans la liste de VAR.

Variable	Nbre manquant	N	Maximum	Minimum
X1	195	1308	99.9000000	0
X2	195	1308	246.3000000	0
X3	195	1308	837.2000000	0
X4	1290	213	20.7000000	0
X5	1309	194	56.7000000	0.1000000
X6	222	1281	93.3000000	0
X7	318	1185	645.5000000	0.6000000
X8	666	837	387.9000000	0.3000000
X9	1004	499	189.4000000	0
X10	54	1449	39175.30	0

Figure 5 - Tableau montrant les données manquantes

Après avoir appliqué cette procédure à nos données, nous avons constaté qu'il y a beaucoup de données manquantes. Pour résoudre ce problème, nous avons décidé de faire une imputation des valeurs manquantes par la moyenne.

L'image ci-dessous est le code que nous avons appliqué à nos données:

```
DATA test2017_I;
  SET test2017;
  ARRAY X _NUMERIC_;
  DO i = 1 TO DIM(X);
    IF X[i] = . THEN X[i] = mean(of X[*]);
  END;
  DROP i;
RUN;

PROC MEANS DATA=test2017_I NMISS N MAX MIN;
RUN;
```

Figure 6 - imputation des valeurs manquantes par la moyenne

Après imputation des données manquantes en utilisant la moyenne, on obtient le résultat ci-dessous qui indique qu'il y a 0 valeur manquante.

Variable	Nbre manquant	N	Maximum	Minimum
X1	0	1503	913.2000000	0
X2	0	1503	913.4000000	0
X3	0	1503	913.5666667	0
X4	0	1503	18650.66	0
X5	0	1503	18650.66	0.1000000
X6	0	1503	10498.02	0
X7	0	1503	18211.06	0.6000000

Figure 7 - Résultat de l'imputation

### e. Identification et suppression des doublons

Pour identifier et supprimer les doublons, nous avons utilisé la procédure PROC SORT et l'option NODUPKEY. Cela indique à SAS de supprimer les doublons basés sur la clé de tri spécifiée. Ci-dessous l'image du code que nous avons appliqué à nos données:

```
proc sort data=test2018_I out=test2018_D nodupkey dupout=Doublons2018;
  by _all_;
run;

proc print data=Doublons2018;
run;
```

Figure 8 - Suppression des doublons

Les doublons des données relatives à l'activité globale des hôpitaux de 2017 à 2019 ont bien été supprimés.

### 3. En 2019, présence ou non de nouveaux établissements, disparition de certains établissements de 2017 ou 2018 ?

Pour répondre à cette question, nous avons d'abord combiné les données relatives aux établissements de 2017 à 2019 puis nous avons transposé les données de sorte que les valeurs de la variable année deviennent des noms de colonnes.

```
/* Fusionner les données pour 2019, 2018 et 2017 */
data etablisements_combine;
  set etablisements_2017 etablisements_2018 etablisements_2019;
run;

/* Trier les données par identifiant d'établissement */
proc sort data=etablisements_combine;
  by fitness;
run;

proc transpose data=etablisements_combine out=question3;
  var annee;
  by fitness;
run;
```

Figure 9 - Fusion des données et tri par identifiant

Nous allons ensuite filtrer les lignes où les col1, col2 et col3 sont manquantes. Ces colonnes permettent de faire ressortir les nouveaux établissements. Il en ressort qu'il n'y a pas de nouveaux établissements.

```
data question3b;
  set question3;
  where col1 is null or col2 is null or col3 is null;
run;
```

Figure 10 - Filtrage des lignes où col1, col2 et cols manquantes

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

a. Quels sont les établissements ayant changé de taille sur la période 2017-2019 ?

Pour répondre à cette question nous avons fait une jointure sur les données établissements puis nous avons comparé les tailles entre elles et on a déduit qu'aucun établissement n'a changé de taille.

```
proc sql;
  create table question4a as
  select h17.finess,
         h17.taille_MCO as TailleMCO_2017, h18.taille_MCO as TailleMCO_2018, h19.taille_MCO as TailleMCO_2019,
         h17.taille_M as TailleM_2017, h18.taille_M as TailleM_2018, h19.taille_M as TailleM_2019,
         h17.taille_C as TailleC_2017, h18.taille_C as TailleC_2018, h19.taille_C as TailleC_2019,
         h17.taille_O as TailleO_2017, h18.taille_O as TailleO_2018, h19.taille_O as TailleO_2019
  from TD6.hospi_2017_etablissement as h17
  left join TD6.hospi_2018_etablissement as h18 on h17.finess = h18.finess
  left join TD6.hospi_2019_etablissement as h19 on h17.finess = h19.finess;
quit;

proc print data= question4a;
run;
```

Figure 11 - Jointure sur les données établissements et comparaison de tailles

b. Quels sont les établissements qui ont changé d'activité ?

De même que la question précédente, nous avons procédé à une jointure sur les données relatives aux activités globales, puis nous avons comparé les activités entre elles. Il en découle que **1327** établissements ont changé d'activité.

```
proc sql;
  create table question4b as
  select distinct a.*
  from TD6.hospi_2017_activiteglobale a
  inner join TD6.hospi_2018_activiteglobale b on a.finess = b.finess
  inner join TD6.hospi_2019_activiteglobale c on a.finess = c.finess
  where (
    a.A7 ne b.A7 or
    a.A8 ne b.A8 or
    a.A9 ne b.A9 or
    a.A10 ne b.A10 or
    a.A11 ne b.A11 or
    a.A12 ne b.A12 or
    a.A13 ne b.A13 or
    a.A14 ne b.A14 or
    a.A15 ne b.A15
  )
  or (
    b.A7 ne c.A7 or
    b.A8 ne c.A8 or
    b.A9 ne c.A9 or
    b.A10 ne c.A10 or
    b.A11 ne c.A11 or
    b.A12 ne c.A12 or
    b.A13 ne c.A13 or
    b.A14 ne c.A14 or
    b.A15 ne c.A15
  );
quit;
```

Figure 12 - Jointure sur les données des activités globales et comparaison d'activités

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.

Pour répondre à cette question, nous avons d'abord chargé les données des lits de 2017 à 2019 à l'aide de la table **Indicateur**. Nous avons effectué une fusion de ces données et fait un tri par **finess**.

```
/* Charger les données des lits de 2017, 2018, 2019 */
proc sql;
  create table lit_2017 as
  select *, 2017 as annee
  from TD6.hospi_2017_litetplace
  where indicateur in ('CI_AC1', 'CI_AC6', 'CI_AC8');
quit;
proc sql;
  create table lit_2018 as
  select *, 2018 as annee
  from TD6.hospi_2018_litetplace
  where indicateur in ('CI_AC1', 'CI_AC6', 'CI_AC8');
quit;
proc sql;
  create table lit_2019 as
  select *, 2019 as annee
  from TD6.hospi_2019_litetplace
  where indicateur in ('CI_AC1', 'CI_AC6', 'CI_AC8');
quit;

/* Fusionner les données de lit pour 2019, 2018 et 2017 */
data lit_combine;
  set lit_2017 lit_2018 lit_2019;
run;

/* Trier les données par finess */
proc sort data=lit_combine;
  by finess;
run;
```

Figure 13 - Chargement et fusion des données des lits et tri par finess

Ensuite, nous avons transformé la variable **Valeur** en numérique et fait une visualisation de la distribution de la variable Valeur à l'aide de la procédure PROC UNIVARIATE.

```
/*transformation la variable valeur en numerique */
data lit_combine;
  set lit_combine;
  Valeur_Num = input(Valeur, ?? comma10.);
run;

/*visualiser la distribution*/
proc univariate data=lit_combine;
  histogram Valeur_Num;
run;
```

Figure 14 - Transformation de Valeur en numérique et visualisation de la distribution

Cette distribution nous a permis d'avoir les quantiles de la variable Valeur.

Quantiles (Définition 5)	
Niveau	Quantile
100Max 100%	8568
99%	700
95%	277
90%	164
75% Q3	78
50% Médiane	38
25% Q1	17
10%	4
5%	0
1%	0
0% Min	0

Figure 15 - Quantiles sur la variable Valeur

Les valeurs que nous allons utiliser pour la suite sont: Q1 = 17, Q2 = 38 (médiane), Q3 = 78. Nous avons donc affecté les valeurs de Q1 et Q3 à deux variables dans notre code nommées **seuil1** et **seuil2**.

Enfin, nous avons effectué des tests avec nos 2 seuils pour **taille\_M**, **taille\_C** et **taille\_O** afin de classer les valeurs par catégorie (Cat1, Cat2, Cat3).

```

/*Pour taille_M*/
data lit_combine_categories;
set lit_combine;
seuil1 = 17;
seuil2 = 78;
/* Définir les catégories en fonction des seuils */
if Valeur_Num < seuil1 then taille_M = 'Cat1';
else if Valeur_Num < seuil2 then taille_M = 'Cat2';
else taille_M = 'Cat3';
run;

/*Pour taille_C*/
data lit_combine_categories;
set lit_combine;
seuil1 = 17;
seuil2 = 78;
/* Définir les catégories en fonction des seuils */
if Valeur_Num < seuil1 then taille_C = 'Cat1';
else if Valeur_Num < seuil2 then taille_C = 'Cat2';
else taille_C = 'Cat3';
run;

```

Figure 16 - Tests des seuils et classification des valeurs par catégorie

En résumé, sur la base de notre raisonnement, nous avons utilisé la procédure PROC UNIVARIATE pour obtenir des statistiques détaillées, y compris les quantiles (Q1, Q2, Q3). En utilisant ces valeurs comme seuils que nous avons identifiés à partir des statistiques descriptives, nous avons mis en place un code qui permet de les classer par catégorie.

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

Pour répondre à cette question, nous avons fusionné les données des **établissements** et des **lits** et triés par finess. Puis, nous avons calculé le nombre d'établissements et le nombre de lits en médecine, en chirurgie et en obstétrique.

```
/* Calculer le nombre d'établissements avec PROC SQL */
proc sql;
  create table etab_combine_count as
  select cat, annee, count(distinct finess) as nombre_eta
         sum(CI_AC1) as CI_AC1,
         sum(CI_AC6) as CI_AC6,
         sum(CI_AC8) as CI_AC8
  from etab_lit_combine
  group by cat, annee;
quit;

proc tabulate data= td6.etab_combine_count;
  class cat annee;
  var nombre_etablissements CI_AC1 CI_AC6 CI_AC8;
  table cat,
         annee * (nombre_etablissements * f=comma10.
                  CI_AC1 * sum * f=comma10.
                  CI_AC6 * sum * f=comma10.
                  CI_AC8 * sum * f=comma10.)
  / box='Tableau croisé par catégorie et année';
run;
```

Figure 17 - Calcul du nombre d'établissements avec PROC SQL

A l'aide de la procédure PROC TABULATE, nous avons mis en place le tableau croisé avec les informations demandées pour chaque année. Ci-dessous le résultat que nous avons obtenu.

Tableau croisé par catégorie et année	annee											
	2017				2018				2019			
	nombre_etablissements	CI_AC1	CI_AC6	CI_AC8	nombre_etablissements	CI_AC1	CI_AC6	CI_AC8	nombre_etablissements	CI_AC1	CI_AC6	CI_AC8
cat	Sum	Sum	Sum	Sum	Sum	Sum	Sum	Sum	Sum	Sum	Sum	Sum
CH	569	63,514	15,543	8,627	569	63,232	14,699	8,377	569	62,832	14,173	8,129
CHR	71	44,322	18,429	4,883	71	44,561	18,113	4,853	71	45,254	17,766	4,857
CLCC	26	1,983	853	0	26	1,940	843	0	26	1,907	830	0
CLI	690	15,350	29,952	4,562	690	15,474	28,474	4,294	690	15,628	27,574	4,159
EBNL	147	7,576	3,904	1,022	147	7,639	3,841	982	147	7,977	3,718	987

Figure 18 - Résultat tableau croisé

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

Pour répondre à cette question, nous avons fusionné les établissements et les lits de 2017 à 2019 et fait un tri sur finess. La première étape est de créer une variable région et d'extraire les deux premiers caractères du N° FINESS pour obtenir la région et nous avons utilisé la procédure PROC FREQ pour compter le nombre d'établissements par catégorie et par région.

```
/* Fusionner les tables etablisements_combine et lit_combine */
data question7;
  set etablisements_combine lit_combine;
run;

/* Trier les données par finess */
proc sort data=question7;
  by finess;
run;

/* Étape 1 : Créer une variable de région */
data question7;
  set etab_lit_combine;
  region = substr(finess, 1, 2); /* Extraction des deux premiers caractères du N° FINESS pour obtenir la régi
run;

/* Étape 2 : Utiliser PROC FREQ pour compter le nombre d'établissements par catégorie et par région */
proc freq data=question7;
  tables cat * region / nocol norow nopercnt;
run;
```

Figure 19 - Code pour le nombre d'établissement par catégorie et par région

Ci-dessous nous avons les résultats obtenu:

						Table de cat par region																	
cat(cat)						region																	
	01	02	03	04	05	44	45	46	47	48	49	50	51	52	53	54		93	94	95	97	98	Total
CH	24	30	15	18	15	18	18	15	15	15	18	21	15	21	18	12		12	12	18	33	0	1707
CHR	0	0	0	0	0	3	3	0	0	0	3	0	3	0	0	3		9	15	0	9	0	213
CLCC	0	0	0	0	3	6	0	0	0	0	3	3	3	0	0	3		0	6	0	0	0	78
CLI	12	6	9	6	3	48	24	3	6	6	30	15	15	9	9	18		39	39	27	114	3	2070
EBNL	0	0	0	0	0	6	3	0	3	0	6	0	0	9	0	12		3	3	6	0	0	441
Total	36	36	24	24	21	81	48	18	24	21	60	39	36	39	27	48		63	75	51	156	3	4509

Figure 20 - Tableau résultat

8. Indiquez dans un tableau le nombre d'accouchement, le nombre de lits 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.

Afin de répondre à cette question, nous avons fusionné toutes les données relatives aux établissements, lits et activités globales. Nous avons ensuite converti les colonnes CI\_A11 et CI\_A8 en données numériques afin de faciliter l'analyse de données. Les indices correspondant aux nombre de lits, nombre d'accouchements, nombre de lits en obstétriques et le niveau de maternité sont répertoriés dans le fichier indicateur.

```
data etab_actglob_lit_combine;
set etab_actglob_lit_combine;
CI_A11N = input(CI_A11, ?? comma10.);
CI_AC8N = input(CI_AC8, ?? comma10.);
niveau_de_maternite = input(CI_E6, ?? comma10.);

run;

proc sql;
create table tableau_accouchements as
select cat, finess, niveau_de_maternite,
       sum(CI_A11N) as total_accouchements,
       sum(CI_AC8N) as total_lits_obstetricaux
from etab_actglob_lit_combine
group by cat, finess, niveau_de_maternite
order by total_lits_obstetricaux desc;
quit;
```

Figure 21 - Code question 8

L'image ci-dessous montre le Top 5 des établissements avec la plus forte activité obstétrique.

Top 5 des établissements avec la plus forte activité obstétrique		
Obs.	finess	total_lits_obstetricaux
1	750712184	24
2	690781810	6
3	130784713	3
4	590780193	3
5	570005165	3

Figure 22 - Liste des 5 établissements avec la plus forte activité obstétrique



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

On commence par fusionner le résultat de la question 8 avec le résultat de la question précédente, on procède ensuite au calcul du minimum et du maximum d'accouchements par région et par catégorie.

```
/* Fusionner les tables*/
data region_accouchement_combine;
set question7 tableau_accouchements;
run;

/* Résumé des indicateurs par région avec le nombre min et max d'accouchements */
proc sql;
create table resume_indicateurs_region as
select
    region,
    cat,
    sum(total_accouchements) as total_accouchements,
    min(total_accouchements) as min_accouchements,
    max(total_accouchements) as max_accouchements
from region_accouchement_combine
group by region, cat
order by region, cat;
quit;
```

Figure 23 - Fusion des tables, calcul min et max d'accouchements par région et cat

L'image ci-dessous nous donne le total, le minimum et le maximum d'accouchements par catégorie et par région.

Obs.	region	cat	total_accouchements	min_accouchements	max_accouchements
1			2354885	0	114402
2		CH	.	.	.
3		CHR	.	.	.
4		CLCC	.	.	.
5		CLI	.	.	.

Figure 24 - Tableau montrant les résultats

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.

Pour calculer le score de qualité, nous avons utilisé le taux de césarienne et le taux de péridurale et nous avons divisé par 2 pour chaque établissement en supposant que la péridurale et la césarienne ont la même importance.

```
/* Calcul du score de qualité pour les accouchements */
data score_qualite;
  set etab_actglob_combine;

  P10N = input(P10, ?? comma10.2);
  P11N = input(P11, ?? comma10.2);
  Taux_de_cesariennes = P10N ;
  Taux_de_péridurale = P11N ;

  /* Calcul du score de qualité */
  score_qualite = (Taux_de_cesariennes + Taux_de_péridurale) / 2;
run;
```

Figure 25 - Calcul du score de qualité pour les accouchements

Les cinq meilleurs établissements qui en découlent sont:

#### Les 5 meilleurs bons établissements

Obs.	finess	score_qualite
1	750300840	71.835
2	750300840	70.725
3	750300840	70.335
4	190000257	66.075
5	920300787	66.005

Figure 26 - Liste des 5 meilleurs établissements

Les cinq moins bon établissements ayant les plus faible score qualité sont:

#### Les 5 moins bons établissements

Obs.	finess	score_qualite
7575	620003376	0.500
7576	880007299	0.955
7577	530000025	1.205
7578	970305629	1.730
7579	690041124	2.175

Figure 27 - Liste des 5 moins bons établissements

11. Étudiez 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'accompagnent 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.

Pour répondre à cette question, nous avons fusionné les scores de qualité et les activités globales. Puis nous avons fait un tri sur finess et année. Donc pour calculer le changement de score d'une année à l'autre, nous avons mis ce code ci-dessous en place pour trouver les scores stables.

```
/*calcul du changement de score d'une année à l'autre */
data scores_stabilite;
  set scrqua_actglob_combine;
  by finess annee;
  retain score_precedent;
  score_precedent = score_qualite;
  if first.finess then score_precedent = .;
  else score_change = score_qualite - score_precedent;
  /* Si le score_change est différent de 0, cela indique un changement dans le score d'une année à l'autre */
  score_stable = (score_change = 0);
run;
```

Figure 28 - Calcul du changement de score d'une année à l'autre

La capture ci-dessous nous donne les informations souhaitées.

La procédure MEANS

Variable d'analyse : score_qualite						
score_stable	N obs	N	Moyenne	Ec-type	Minimum	Maximum
0	12010	0	.	.	.	.
1	1517	1517	41.5459954	16.2036122	0	71.8350000

Figure 29 - Tableau montrant les résultats

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 pouvez-vous l'améliorer ?

Pour répondre à cette question, nous avons fusionné et trié les scores de qualité et les régions. À l'aide de PROC SQL, nous avons calculé le score par région et par année.

```

/* Utilisation de proc SQL pour calculer le score par région et par année */
proc sql;
create table score_qualite_region_annee as
select region, annee, mean(score_qualite) as score_qualite_region
from scrqua_reg_combine
group by region, annee;
quit;

```

*Figure 30 - Calcul du score par région et par année*

region	annee	score_qualite_region
01	2017	36.91625
01	2018	35.86125
01	2019	47.08625
02	2017	45.151666667
02	2018	42.803333333
02	2019	35.804285714
03	2017	47.278333333
03	2018	44.408333333
03	2019	43.285

*Figure 31 - Résultat score par région et par année*

On constate qu'il y a non seulement des différences de scores entre les régions mais aussi par année. Les scores dépendent des données disponibles et comme certaines données manquent ou sont incomplètes, le score est biaisé.

Pour avoir de meilleurs scores, nous pouvons améliorer la qualité du jeu de données afin de le rendre plus complet. Cela pourrait grandement contribuer à rendre les scores plus fiables.

## Conclusion

En conclusion de notre analyse approfondie des données hospitalières, nous avons navigué à travers les informations mises à notre disposition, dévoilant ainsi une compréhension accrue de leur organisation. Grâce à cette analyse rigoureuse, nous avons pu répondre aux questions essentielles concernant les établissements hospitaliers. Cette exploitation nous a permis de discerner l'évolution des activités et des tailles des établissements de 2017 à 2019. La classification des établissements selon des critères spécifiques et leur répartition par région ont ajouté une dimension géographique à notre compréhension, ouvrant la voie à des observations plus pertinentes. Enfin, l'évaluation de la qualité des établissements à travers l'attribution des scores a constitué une base solide pour des comparaisons significatives entre établissements et par région sur la période considérée. Ces résultats constituent une ressource pour l'identification d'axes d'améliorations potentiels dans le domaine de la santé.

## Références

### Documentation SAS utilisée

[https://documentation.sas.com/doc/fr/pgmsascdc/9.4\\_3.5/pgmsashome/home.htm](https://documentation.sas.com/doc/fr/pgmsascdc/9.4_3.5/pgmsashome/home.htm)

Recherche sur le Web