



Agence Technique de l'Information sur l'Hospitalisation

SIRET n° 180 092 298 00033 – code APE : 8411 Z

117, boulevard Marius Vivier Merle

69329 LYON CEDEX 03

Téléphone : 04 37 91 33 10

Paris, le 19 Janvier 2009

Référence : DL-VS-139-2009

Analyse de l'Indicateur de Valorisation de l'Activité

La campagne 2009 doit amorcer la mise en place de la tarification à l'activité pour les activités de soins de suite et de rééducation.

Le schéma cible tel qu'envisagé devrait être effectif à l'horizon 2012. Ce schéma cible repose sur 4 compartiments :

- un compartiment similaire à celui qui existe pour les activités MCO relatif au financement des activités dites « missions d'intérêt général et aide à la contractualisation »
- un compartiment spécifique pour ce secteur relatif au financement des plateaux techniques particulièrement coûteux
- un compartiment existant sur l'activité MCO relatif au financement des molécules onéreuses (liste identique à celle applicable aux activités MCO avec ajout de quelques molécules spécifiques au secteur)
- un compartiment financé directement à l'activité sur la base d'une classification et de tarifs.

Pour mettre en place ce schéma cible, des travaux concernant la classification et l'échelle des coûts ont commencé afin de définir une classification médico-économique qui puisse être tarifante. Ces travaux devraient permettre de consolider le schéma cible pour une mise en œuvre effective en 2012.

En attendant cette stabilisation, il est envisagé de tenir compte de la production de l'activité pour l'allocation de ressources. Ce dispositif reposerait sur la mesure de l'activité par le truchement des points IVA. Un indice de modulation¹, qui est la mise en association entre le nombre de point IVA de l'établissement issue du modèle et la valeur financière de l'établissement, servirait ensuite de base pour moduler l'allocation de ressources des établissements.

1 Définition de l'IVA et de l'indice de modulation

1.1 Une description de l'activité revue

Parallèlement aux travaux menés sur la classification SSR pour le modèle cible, l'ATH a conçu un modèle de valorisation basé sur les variables du recueil PMSI-SSR en vigueur. Les recueils hebdomadaires standards (et leurs journées de présence) sont classés selon des critères médicaux dans des groupes de morbidités dominantes (GMD) et affectés d'un nombre de points de base correspondant. A ceux-là sont ajoutés des points supplémentaires au regard des autres variables

¹ Cf Annexe 1 : Méthodologie de la construction de l'indice de modulation.



explicatives du modèle (âge, comorbidités, dépendance physique, dépendance cognitive, nombre d'activités de rééducation-réadaptation-réinsertion, type de prise en charge et type d'hospitalisation). Ainsi, le modèle correspond à un modèle additif qui attribue à chaque journée un nombre de points. L'activité produite par chaque établissement est quantifiée en points, dits points IVA. Pour les établissements DAF et les établissements OQN, la production en points a été calculée sur les données d'activité 2007.

Les IVA permettent de traduire l'activité SSR en points, points plus ou moins importants en fonction de l'activité produite (fonction des critères médicaux, de critères d'âge, de la spécialité, etc.). Pour poursuivre le modèle, il convient de mettre au regard de cette production d'activité une masse financière.

1.2 Associer à la production de l'activité une valeur financière : l'indice de modulation

Dans un premier temps, il a été décidé d'associer à cette production d'activité l'ensemble des financements perçus par les établissements pour leur activité SSR.

Pour un exercice budgétaire donné, les recettes assurance maladie propres à l'activité de soins de suite et de rééducation sont connues.

Pour les établissements « publics » cette activité est financée par dotation globale en dotation annuelle de financement. L'outil de suivi des campagnes budgétaires ARBUST permet de suivre proprement les dotations DAF.

Pour les cliniques privées, la mobilisation des données de l'assurance maladie (SNIIRAM) permet de connaître pour chaque établissement les liquidations associées aux tarifs (avec une possibilité de détail par code prestation, DMT et Mode de Traitement) et les liquidations associées aux dépenses en sus.

Aussi, les données totales financières par établissement peuvent être mises en regard des données d'activité. Cette association permet la construction d'un indice de modulation des recettes des établissements plus en rapport avec leur activité. Toutefois, elle ne tient pas compte des autres compartiments spécifiques.

Compte tenu de la nécessité de prendre en compte des spécificités financières liées à la présence de plateaux techniques, de missions d'intérêt général, et de molécules onéreuses, il est préféré d'utiliser une valeur de point de référence hors compartiment. Ainsi, la valeur de référence du point est calculée en utilisant les seuls financements de l'activité (financements SSR déduction faite des financements des compartiments). Toutefois, pour le calcul du coefficient, la comparaison se fait sur l'intégralité des recettes des établissements. Ainsi, pour un établissement donné, le CA théorique est obtenu par la multiplication du nombre de points IVA de l'établissement par la valeur financière du point à laquelle est ajoutée ensuite les masses financière des compartiments.



2 Analyse de l'indice de modulation

L'analyse a pour objectif de vérifier, a posteriori, au travers d'indicateurs statistiques de la pertinence et de la qualité du modèle de valorisation de l'activité, nommé modèle IVA (Indicateur de Valorisation de l'Activité).

D'un point de vue théorique, l'indice de modulation², qui est la mise en association entre le nombre de point IVA de l'établissement issue du modèle et la valeur financière de l'établissement, ne doit pas être lié de manière statistique aux différentes variables du modèle. C'est ce postulat qui va guider l'axe de l'analyse et servir de critère de qualité du modèle.

L'analyse est constituée de deux séries d'études. La première porte sur les corrélations entre l'indice de modulation et les variables constitutives du modèle. Cette étude a pour objectif de démontrer la pertinence du modèle IVA, dans le sens où aucune variable du modèle n'influe significativement sur l'indice de modulation. La deuxième est basée sur le comportement de l'indice en fonction de catégories d'établissements (autorisations, chiffre d'affaires, etc....). L'objectif est cette fois-ci non pas de valider la pertinence du modèle mais de porter un regard clair sur les remarques des ARH, suite aux simulations 2006 du scénario A (indice de modulation version 1 sur bases 2006).

2.1 Etude des corrélations entre l'indice de modulation et les variables du modèle IVA

Si le modèle IVA était totalement pertinent, les différences liées aux spécificités de chaque structure ne devraient pas se traduire au niveau de l'indice de modulation. Autrement dit, le coût du point IVA ne doit pas être corrélé positivement ou négativement à de quelconques items relatifs à l'activité des structures hospitalières. L'objectif de ce travail est d'effectuer une analyse sur des facteurs choisis pouvant expliquer les variations de la valeur de l'indice. L'analyse factorielle va permettre, au travers de la modélisation statistique, d'établir les relations éventuelles entre les différentes variables du modèle et l'indice de modulation.

Dans un premier temps, l'étude³ porte un regard sur les corrélations linéaires (coefficient de Pearson) et les corrélations non linéaires (coefficient de Spearman) entre les variables du modèle et l'indice de modulation. L'analyse porte également sur l'ensemble des corrélations avec d'autres variables comme le PMJP relatif ou la diversité de l'activité (cf. Annexe 2).

L'étude prend par conséquent en compte les variables suivantes :

Variables constitutives de l'IVA

- Age (Age Moyen des journées de présence)
- Dépendance physique (Pourcentage des journées avec DP > 9)
- Dépendance cognitive (Pourcentage des journées avec DC > 5)
- Nombre activité CdARR (Pourcentage des journées sans activité, entre 0,5 et 4, > 4)
- Comorbidités et actes valorisants (Pourcentage des journées avec CAV)
- Type d'hospitalisation (Pourcentage des journées hors hospitalisation complète)
- Type de prise en charge (Pourcentage des journées avec une prise en charge RF)
- GMD (Pourcentage des journées dans GMD pédiatrie)

Autres variables

- Le PMJP relatif (poids moyen de la journée de présence)
- Diversité activité (nombre de CMC représentant au moins 50% des journées)

² Cf Annexe 1 : Méthodologie de la construction de l'indice de modulation.

³ Cf Annexe 2 : Analyse factorielle



Secteur DAF

	Corrélations linéaires avec INDICE	Corrélations non linéaires avec INDICE
AGE	-0,17	-0,22
DEP PHYS	-0,01	-0,05
DEP CONG	0,09	0,01
CdARR_0	-0,05	-0,1
CdARR_04	0,04	0,08
CdARR_4	0,06	0,16
CAV	0,03	-0,01
TYPE HOSP	0,07	0,17
TYPE PC	0,05	0,08
PED	0,04	0,14
PMJP	-0,074	-0,007
DIV ACTIV	-0,15	-0,17

Secteur OQN

	Corrélations linéaires avec INDICE	Corrélations non linéaires avec INDICE
AGE	0,1	-0,16
DEP PHYS	-0,09	-0,1
DEP CONG	-0,08	-0,09
CdARR_0	-0,01	-0,08
CdARR_04	0	0,07
CdARR_4	0,06	0,1
CAV	0,04	0,03
TYPE HOSP	0,14	0,22
TYPE PC	0,21	0,24
PED	-0,27	-0,02
PMJP	-0,216	-0,147
DIV ACTIV	-0,12	-0,15

Variables un peu plus corrélées que les autres avec l'indice de modulation :

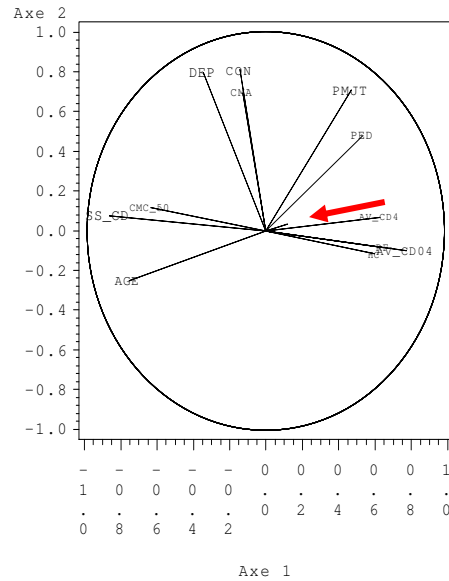
- Sur les deux secteurs
 - L'âge moyen des journées de présence (AGE)
 - Le pourcentage de journées hors hospitalisation complète (TYPE HOSP)
- Spécifiques au secteur DAF
 - Le pourcentage de journées pédiatriques (PED)
- Spécifiques au secteur OQN
 - Le pourcentage de journées avec un type de prise en charge RF (TYPE PC)

Dans l'ensemble les variables du modèle présentent de faibles corrélations linéaires ou non avec l'indice de modulation et des corrélations linéaires ou non importantes avec le PMJP relatif. Par conséquent sur cette analyse descriptive, le modèle IVA apparaît pertinent en prenant bien en compte le poids de l'activité des établissements.

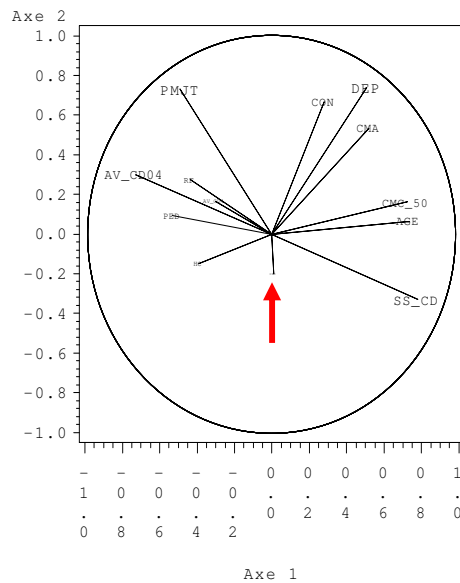
Cette étude des corrélations a été complétée par une analyse factorielle, qui **révèle que, dans l'ensemble, les variables du modèle ne sont pas corrélées avec l'indice de modulation et corrélées avec le PMJP relatif. Par conséquent le modèle est pertinent car il prend bien en compte le poids de l'activité.** Le cercle des corrélations permet de confirmer ces tendances.



Secteur DAF :



Secteur OQN :



La flèche rouge indique la position sur le cercle des corrélations de l'indice de modulation. L'indice de modulation n'est pas fortement représenté sur les deux premiers axes de l'analyse factorielle alors que le PMJP relatif est bien représenté sur les deux axes et les variables du modèle sur un des deux axes. Les cercles montrent par conséquent l'absence de relation entre l'indice de modulation et les variables du modèle.

Afin d'affiner ces résultats, un effet zoom a été mis en place au travers de plusieurs analyses factorielles par rapport à des classes de variables, qui ont été définies au travers d'outils statistiques d'aide à la décision.

Les résultats des analyses factorielles permettent d'établir quelques constats.

1- le degré de diversité des établissements présente une petite corrélation avec l'indice de modulation. Cependant cette variable n'est pas incluse comme telle dans le modèle,



2- Les variables du type de prise en charge et du type d'hospitalisation présentent une absence de corrélation moins évidente sur le secteur OQN.

2.2 Etudes du comportement de l'indice en fonction de catégories d'établissements

Deux points précis ressortent des remarques des ARH. Le premier point repose sur les établissements RF qui seraient en grande majorité sur dotés, c'est-à-dire que leur indice de modulation serait supérieur à 1. Le deuxième point repose sur le fait que les établissements avec une faible diversité d'activité seraient en grande partie sur dotés.

2.2.1 Autorisations d'établissements en vigueur RF/SSM/MIXTE

L'analyse porte sur la relation entre l'indice de modulation et l'autorisation de l'établissement en activité SSM (soins de suite médicalisés), RF (rééducation fonctionnelle) ou MIXTE (SSM et RF). L'autorisation de l'établissement a été renseignée de manière déclarative par les ARH.

La variable déclarative de l'autorisation de l'établissement a d'autant plus de poids dans l'analyse que ce sont les ARH qui ont fait la remarque que, les établissements RF étaient sur dotés. Cette variable étant qualitative, l'outil statistique d'aide à la décision, pour mesurer la relation entre la variable d'activité RF et l'indice de modulation, est l'analyse de la variance.

Les résultats de l'analyse de la variance⁴ indiquent que les tests statistiques confirment l'hypothèse que les établissements RF/Mixte et SSM n'ont pas, en moyenne, des comportements identiques par rapport à l'indice de modulation. En particulier les établissements RF seraient en moyenne sur-dotés alors que les établissements SSM seraient sous dotés. Il convient néanmoins de noter que sur les deux secteurs l'intensité de la liaison reste extrêmement faible, autrement dit la catégorie de l'établissement suivant son type d'activité n'est pas une variable significative pouvant expliquer la variation de l'indice de modulation.

Par ailleurs, ces intensités de liaison sont différentes selon les indices (version 1 sans prise en compte des trois autres compartiments, version 3 avec prise en compte). A titre indicatif l'intensité de liaison avec l'indice de modulation version 3 est de l'ordre de 3% pour le secteur DAF et de 10% pour le secteur OQN alors que pour l'indice de modulation version 1, l'intensité de liaison est de 5% pour le secteur DAF et de 22% pour le secteur OQN. Par conséquent la sur dotation des établissements RF est en partie expliquée par le coût des compartiments MIG, Molécules onéreuses et Plateaux Techniques.

Toutefois les hypothèses nécessaires à la réalisation de ces tests paramétriques ne sont pas toutes vérifiées. Le recours aux tests non paramétriques est nécessaire : test de localisation et test de dispersion.

Les tests non paramétriques amènent également à conclure que les établissements « RF et Mixte » n'ont pas la même distribution et la même localisation que les établissements « SSM » par rapport à l'indice de modulation.

Toutefois, ces résultats sont également liés à la définition des catégories. Ainsi, si on opère la même analyse sur des catégories d'établissements réalisées non pas sur les autorisations mais sur les chiffres d'affaires (CA), les conclusions montrent que les établissements de taille financière importante

⁴ Annexe 3 : Analyse de la Variance entre l'indice de modulation et la catégorie de population RF/SSM/MIXTE



ont un comportement différent par rapport à l'indice de modulation de celui des établissements de faible taille financière.

La comparaison des tests montrent également **que la catégorisation suivant le CA est plus discriminante que la catégorisation RF/SSM/MIXTE**. Le test d'indépendance (test du chi deux) entre les catégories de CA et les catégories d'autorisations est rejeté. Il existe donc un lien entre le CA et les catégories d'autorisation.

Les établissements avec des chiffres d'affaires importants sont des établissements plutôt RF/MIXTE, alors que les établissements avec des chiffres d'affaires faibles sont des établissements plutôt SSM. Sur les deux secteurs, plus les établissements RF et MIXTE comme les établissements SSM ont un CA élevé plus ils ont tendance à être sur-dotés.

Par conséquent la valeur discriminatoire de la catégorisation RF/SSM/MIXTE est à relativiser d'autant plus que les établissements RF et Mixtes sont en général des établissements avec des chiffres d'affaires importants.

Ainsi, une éventuelle explication de la sur dotation des établissements RF et mixtes est peut être dû au fait que ces établissements sont en général les établissements avec les plus grands chiffres d'affaires.

2.2.2 Catégories d'établissements selon la diversité de l'activité (créneau étroit) pour les adultes et une catégorie spécifique pour la pédiatrie.

Les commentaires des ARH ont mis en évidence la présence d'une éventuelle relation entre l'indice de modulation et la diversité de l'activité des établissements. L'objectif est de constater s'il existe un lien entre l'indice de modulation de l'établissement et le degré de diversité des établissements. Les établissements pédiatriques sont isolés, puisque les analyses précédentes entre l'indice de modulation et les variables du modèle IVA avaient permis aussi de soupçonner une relation.

Au préalable, les travaux ont portés sur le secteur DAF, où il s'agit de vérifier si les MECS, les hôpitaux locaux et les pouponnières ne forment pas une catégorie à part, vis-à-vis de l'indice de modulation, afin de ne pas biaiser le reste de l'analyse. L'analyse de la variance est l'outil statistique permettant de mesurer une éventuelle relation entre l'indice de modulation et la catégorie de l'établissement.

L'analyse de la variance⁵ montre que la relation entre la catégorie d'établissement et l'indice de modulation existe, même si elle est relativement faible (l'intensité de la liaison est inférieure à 1%). Etant donné qu'ils sont en moyenne sous dotés (particulièrement les MECS), les établissements de la catégorie « HL MECS POUPONNIERES » sur le secteur DAF et les établissements MECS sur le secteur OQN sont retirés de la suite de l'analyse, afin de ne pas fausser la suite des résultats.

Sur ce nouveau périmètre, les établissements sont classés de la manière suivante sur les deux secteurs :

- Les établissements pédiatriques (plus de 50% de leur journée sont consacrés à la pédiatrie).
- Les établissements avec très peu de diversité (1 ou 2 GMD constituant plus de 50% des journées)
- Les établissements avec peu de diversité (3 GMD constituant plus de 50% des journées)
- Les établissements avec de la diversité (4 à 5 GMD constituant plus de 50% des journées)

⁵ ANNEXE 4 : Analyse de la Variance sur le degré de diversité des établissements



- Les établissements avec beaucoup de diversité (plus de 5 GMD constituant plus de 50% des journées)

Tout d'abord l'analyse de la variance ne permet pas de conclure sur les établissements pédiatriques, dont le nombre est trop faible (23 établissements hors MECS sur le secteur DAF, 2 établissements hors MECS sur le secteur OQN) pour une utilisation statistique pertinente.

Sur le secteur DAF, les tests paramétriques et non paramétriques infirment l'hypothèse que les établissements ont des indices de modulations moyens identiques suivant leur degré de diversité. Il convient de noter que l'intensité de liaison entre le degré de diversité et l'indice de modulation est extrêmement faible.

Sur le secteur OQN, les tests paramétriques et non paramétriques confirment l'hypothèse que les établissements ont des indices de modulations moyens différents suivant leur degré de diversité. Cependant l'intensité de liaison reste trop faible, pour en déduire que le degré de diversité influe fortement les variations de l'indice de modulation.



ANNEXE 1 : **Méthodologie de la construction de l'indice de modulation**

1 Calcul de l'indice

Le principe de l'indice repose sur le rapport entre les ressources actuelles perçues par l'établissement correspondant à leurs activités et des ressources théoriques correspondant à une valorisation des points, issue du modèle IVA par une valeur financière nationale du point.

Les ressources perçues par les établissements au titre de leurs activités correspondent à leur chiffre d'affaire auquel on retranche la masse financière correspondant aux compartiments MIG (seulement sur le secteur DAF), Molécules Onéreuses et Plateaux techniques.

$$VF_i = CA_réel_i - Comp_i$$

VF_i = Valeur financière de l'établissement i au titre de son activité

$CA_réel_i$ = Chiffre d'affaire de l'établissement i

$Comp_i$ = Masse financière de l'établissement i financière correspondant aux compartiments MIG, Molécules Onéreuses et Plateaux techniques

$$VPN = \frac{\sum_i VF_i}{\sum_i IVA_i}$$

VPN = valeur nationale du Point

IVA_i = Nombre de Point IVA de l'établissement i

Tout comme dans le secteur MCO, cette valeur nationale du point est redressée par un coefficient géographique.

$$VPN = \frac{\sum_i VF_i}{\sum_{i,j} IVA_i (1 + \alpha_j)}$$

α_j = Coefficient géographique de la région j .

Ce coefficient est de 5 % pour la Corse, de 7% pour l'île de France, de 25 % pour la Martinique, la Guadeloupe et la Guyane, et de 30% pour la Réunion.

Remarque 1 : L'indice de modulation peut être calculé à partir d'une valeur de point régionale. Dans ce cas il n'est pas nécessaire d'y introduire le coefficient géographique. Lors des simulations du scénario A présentées aux ARH, l'indice de modulation a été calculé en fonction d'une valeur de point régionale. En revanche les analyses sont basées sur le principe d'une valeur de point nationale.

Remarque 2 : Pour le secteur OQN, les dépenses en SUS 2007 sont estimées à partir des données 2006 en appliquant le taux des dépenses en SUS 2006 sur le chiffre d'affaires 2006.



$$\text{taux_sus_2006}_i = \frac{\text{DEP_SUS_2006}_i}{(\text{CA_réel_2006}_i)}$$

$$\text{DEP_SUS_2007}_i = (\text{taux_sus_2006}_i) * \text{CA_réel_2007}_i$$

Avec

DEP_SUS_YYYY_i = dépenses en SUS de l'établissement i de l'année YYYY

CA_réel_YYYY_i = chiffre d'affaires réels de l'établissement i de l'année YYYY

Remarque 3 : Pour les deux secteurs, les montants liés aux compartiments 2007 sont estimés à partir des données 2006 en appliquant le taux des montants des compartiments 2006 sur le chiffre d'affaires 2006.

$$\text{taux_comp_2006}_i = \frac{\text{COMP_2006}_i}{\text{CA_réel_2006}_i}$$

$$\text{COMP_2007}_i = (\text{taux_comp_2006}_i) * \text{CA_réel_2007}_i$$

Avec

COMP_YYYY_i = Montants liés aux compartiments de l'établissement i de l'année YYYY

L'indice I est un rapport entre le financement par l'assurance maladie constaté (« réel ») et le financement théorique calculé selon le modèle IVA.

Si le coefficient est supérieur à 1, l'établissement est théoriquement surdoté dans son financement actuel.

Le chiffre d'affaire théorique (CAT) résulte de la multiplication des points d'activité IVA par la valeur nationale du point IVA, auquel on ajoute la masse financière correspondant aux compartiments MIG, Molécules Onéreuses et Plateaux techniques.

$$\text{CAT}_i = \text{VPN} * \text{IVA}_i + \text{Comp}_i$$

CAT_i = Chiffre d'affaires théoriques de l'établissement i .

Ensuite la valeur de l'indice de l'établissement est le rapport entre sa valeur financière et son chiffre d'affaire théorique

$$I_i = \text{CA_réel}_i / \text{CAT}_i$$

3 Périmètre de l'analyse

Les établissements exclus sont les suivants :

- Etablissements sans valeur financière (1)
- Etablissement avec une valeur de point IVA nulle (2)



Agence Technique de l'Information sur l'Hospitalisation

- Nouveaux Etablissements 2007 (3)
- Etablissements avec un indice de modulation atypique (4)

Le tableau suivant indique la fréquence des établissements exclus et l'importance de leurs masses financières pour les trois premiers cas :

Secteur	Nouveaux Etablissements 2007			Etablissements sans masse financière			Etablissement avec un nombre de Point IVA nul			TOTAL		
	NB ETAB	% ETAB	% masse financière	NB ETAB	% ETAB	% masse financière	NB ETAB	% ETAB	% masse financière	NB ETAB	% ETAB	% masse financière
DAF	7	0,63%	0,19%	8	0,72%	0%	1	0,09%	0,03%	13	1,17%	0,22%
OQN	12	2,63%	1,85%	6	1,32%	0%	0	0%	0%	18	3,95%	1,85%

Les nouveaux établissements sont exclus en raison des remarques 2 et 3 (voir ci-dessus). Etant donné que les montants des dépenses en SUS et les masses financières correspondant aux compartiments MIG, Molécules Onéreuses et Plateaux techniques sont estimés à partir des données 2006, les établissements qui ont ouverts des lits de SSR en 2007 ne peuvent pas être intégrés aux calculs de l'indice de modulation.

Le tableau suivant indique la fréquence des établissements exclus et l'importance de leurs masses financières pour les établissements avec un indice de modulation atypique. Les raisons expliquant ces indices atypiques sont :

- un défaut d'exhaustivité de la base PMSI
- la DAF est allouée pour l'année alors que l'activité PMSI ne correspond pas à une année complète (MECS, établissement qui ferme au cours de l'année...)
- problèmes de découpage budgétaire pour les établissements multi activités....

Secteur	Etablissements avec indice Version 3 atypique		
	NB ETAB	% ETAB	% masse financière
DAF	21	1,89%	1,05%
OQN	8	1,75%	0,59%

Une fois le périmètre établi, l'indice de modulation est recalculé pour ce nouveau champ d'établissements, soit 1076 établissements pour le secteur DAF et 430 établissements pour le secteur OQN.



ANNEXE 2 :

Analyse Factorielle pour analyser la corrélation entre l'indice de modulation et les variables du modèle.

L'objectif de l'analyse factorielle permet, quand on dispose de k variables, de représenter un plan d'espace de dimension l , ($l \in [1 ; k)$) dans un espace de dimension 2, en déformant le moins possible la réalité.

C'est la matrice des variances-covariances (ou celle des corrélations) qui va permettre de réaliser ce résumé pertinent, parce qu'on analyse essentiellement la dispersion des données considérées.

De cette matrice, on va extraire, par procédé mathématique adéquat, les facteurs que l'on recherche, en petit nombre. Ils vont permettre de réaliser les graphiques désirées dans cet espace de petite dimension (le nombre de facteurs retenus), en déformant le moins possible la configuration globale des individus selon l'ensemble des variables initiales (ainsi remplacées par les facteurs).

C'est l'interprétation de ces graphiques qui permettra de comprendre la structure des données analysées. Cette interprétation sera guidée par un certain nombre d'indicateurs numériques et graphiques, appelées aides à l'interprétation, qui sont là pour aider à l'interprétation la plus juste et la plus objective possible.

Etant donné que les variables du modèle sont quantitatives, l'analyse factorielle sera une analyse en composantes principales.

1 Les données

Les données sont toutes formalisées au niveau établissement et porte sur les secteurs DAF et OQN.

Les variables sont les suivantes :

- Age (Age Moyen des journées de présence) - AGE
- Dépendance physique (Pourcentage des journées avec DP > 9) – DEP PHYS
- Dépendance cognitive (Pourcentage des journées avec DC > 5) – DEP CONG
- Nombre activité CdARR (Pourcentage des journées sans activité, entre 0,5 et 4, > 4) – CdARR 0 / CdARR 04 / CdARR 4
- Comorbidités et actes valorisants (Pourcentage des journées avec CAV) - CAV
- Type d'hospitalisation (Pourcentage des journées hors hospitalisation complète) – TYPE HOSP
- Type de prise en charge (Pourcentage des journées avec une prise en charge RF) – TYPE PC
- GMD (Pourcentage des journées dans GMD pédiatrie) - PED
- Le PMJP relatif (poids moyen de la journée de présence) - PMJP
- Diversité activité (nombre de CMC représentant au moins 50% des journées) – DIV ACTIV

Remarque :

La variable DIV ACTIV contient des valeurs manquantes pour certains établissements. En effet, les établissements ayant plus de 50% de leur journée dans la seule CMC numérotée 60 (CMC dit « autres ») n'ont pas de valeur. Les valeurs manquantes sont remplacées par la moyenne de la variable correspondante, ce qui n'a aucune modification sur l'inertie totale pour une ACP centrée ou centrée réduite.

4 Analyse en composantes principales avec l'ensemble des variables

1.1 Résultats préliminaires

Le tableau ci-après présente la moyenne, l'écart type, le minimum et le maximum de chaque variable. Il s'agit pour l'instant d'études univariées.

Secteur DAF

Variable	Moyenne	Écart-type	Minimum	Maximum
PED	0,05	0,19	0	1
DIV ACTIV	2,39	0,87	1	5
AGE	68,93	16,97	1	86
DEP PHYS	46,38	20,53	0	100
DEP CONG	20,55	17,49	0	100
CdARR 0	39,84	27,84	0	100
CdARR 04	57,26	25,54	0	100
CdARR 4	2,90	7,06	0	64
CAV	15,67	12,13	0	98
TYPE PC	40,51	32,95	0	100
TYPE HOSP	4,88	13,65	0	100
PMJP	0,98	0,15	0,24	1,66
IND	1,01	0,37	0,33	2,95

Secteur OQN

Variable	Moyenne	Écart-type	Minimum	Maximum
PED	0,07	0,25	0	1
DIV ACTIV	2,06	1,05	1	7
AGE	65,40	17,30	4	86
DEP PHYS	25,78	18,74	0	99
DEP CONG	9,65	12,63	0	98
CdARR 0	30,52	28,50	0	100
CdARR 04	66,97	27,29	0	100
CdARR 4	2,45	7,44	0	86
CAV	7,70	6,99	0	49
TYPE PC	49,78	37,66	0	100
TYPE HOSP	5,22	16,51	0	100
PMJP	0,99	0,19	0,52	1,86
IND	0,99	0,34	0,30	2,39

La grande majorité des variables du modèle sont assez hétérogène au niveau de la dispersion (écart type > à plus de la moitié de la moyenne). En revanche l'indice de modulation et le PMJP relatif sont plus homogènes, ce qui est logique puisque leur construction est centrée sur la moyenne nationale. Le tableau suivant est la matrice des corrélations. Elle donne les coefficients de corrélation linéaire des variables prises deux à deux. C'est une succession d'analyses bivariées, constituant un premier pas vers l'analyse multivariée.



Secteur DAF

	PED	DIV ACTIV	AGE	DEP PHYS	DEP CONG	CdARR 0	CdARR 04	CdARR 4	CAV	TYPE PC	TYPE HOSP	PMJP	IND
PED	1	- 0,23	- 0,81	0,03	0,25	- 0,26	0,22	0,26	0,11	0,08	0,24	0,51	0,04
DIV ACTIV	- 0,23	1	0,47	0,29	0,09	0,48	-0,43	-0,32	0,13	-0,33	-0,28	-0,13	-0,15
AGE	0,81	0,47	1	0,21	0,15	0,50	-0,42	-0,45	0,02	-0,23	-0,42	-0,35	-0,17
DEP PHYS	0,03	0,29	0,21	1	0,64	0,29	-0,27	-0,16	0,55	-0,14	-0,27	0,45	-0,01
DEP CONG	0,25	0,09	0,15	0,64	1	0,19	-0,20	-0,04	0,43	-0,22	-0,12	0,39	0,09
CdARR 0	- 0,26	0,48	0,50	0,29	0,19	1	-0,97	-0,44	0,07	-0,55	-0,36	-0,39	-0,05
CdARR 04	0,22	- 0,43	- 0,42	- 0,27	- 0,20	- 0,97	1	0,20	-0,06	0,50	0,27	0,32	0,04
CdARR 4	0,26	- 0,32	- 0,45	- 0,16	- 0,04	- 0,44	0,20	1	-0,07	0,38	0,43	0,37	0,06
CAV	0,11	0,13	0,02	0,55	0,43	0,07	-0,06	-0,07	1	-0,06	-0,15	0,36	0,03
TYPE PC	0,08	- 0,33	- 0,23	- 0,14	- 0,22	- 0,55	0,50	0,38	-0,06	1	0,42	0,36	0,05
TYPE HOSP	0,24	- 0,28	- 0,42	- 0,27	- 0,12	- 0,36	0,27	0,43	-0,15	0,42	1	0,14	0,07
PMJP	0,51	- 0,13	- 0,35	0,45	0,39	- 0,39	0,32	0,37	0,36	0,36	0,14	1	- 0,074
IND	0,04	- 0,15	- 0,17	- 0,01	- 0,09	- 0,05	0,04	0,06	0,03	0,05	0,07	- 0,074	1

Secteur OQN

	PED	DIV ACTIV	AGE	DEP PHYS	DEP CONG	CdARR 0	CdARR 04	CdARR 4	CAV	TYPE PC	TYPE HOSP	PMJP	IND
PED	1	- 0,27	- 0,86	0,29	0,03	- 0,12	0,13	0,00	-0,25	-0,18	-0,07	0,60	-0,27
DIV ACTIV	- 0,27	1	0,48	0,45	0,21	0,47	-0,45	-0,14	0,34	-0,30	-0,23	-0,17	-0,12
AGE	0,86	0,48	1	0,48	0,08	0,32	-0,30	-0,12	0,40	0,06	-0,11	-0,47	0,10
DEP PHYS	- 0,29	0,45	0,48	1	0,51	0,13	-0,14	0,01	0,55	0,01	-0,23	0,29	-0,09
DEP CONG	0,03	0,21	0,08	0,51	1	0,10	-0,08	-0,07	0,39	-0,08	-0,17	0,31	-0,08
CdARR 0	- 0,12	0,47	0,32	0,13	0,10	1	-0,97	-0,29	0,18	-0,50	-0,27	-0,43	-0,01
CdARR 04	0,13	- 0,45	- 0,30	- 0,14	- 0,08	- 0,97	1	0,03	-0,17	0,46	0,20	0,37	0,00
CdARR 4	- 0,00	- 0,14	- 0,12	0,01	- 0,07	- 0,29	0,03	1	-0,08	0,23	0,30	0,28	0,06
CAV	0,25	0,34	0,40	0,55	0,39	0,18	-0,17	-0,08	1	-0,08	-0,22	0,06	0,04
TYPE PC	- 0,18	- 0,30	0,06	0,01	- 0,08	- 0,50	0,46	0,23	-0,08	1	0,33	0,32	0,21
TYPE HOSP	- 0,07	- 0,23	- 0,11	- 0,23	- 0,17	- 0,27	0,20	0,30	-0,22	0,33	1	0,01	0,14
PMJP	0,60	- 0,17	- 0,47	0,29	0,31	- 0,43	0,37	0,28	0,06	0,32	0,01	1	- 0,216
IND	- 0,27	- 0,12	0,10	- 0,09	- 0,08	- 0,01	0,00	0,06	0,04	0,21	0,14	- 0,216	1



Plusieurs choses sont à souligner :

- Les variables du modèle sont peu corrélées à l'indice de modulation, ce qui est un indicateur de performance du modèle. Cependant les corrélations sont plus importantes sur le secteur OQN. Il faudrait approfondir, pour les deux secteurs, sur les variables liés à l'âge, à la diversité d'activité, au type d'hospitalisation et au type de prise en charge, où on note quand même une légère corrélation (les statistiques de test rejettent cependant le test d'indépendance) avec l'indice de modulation.
- Les variables du modèle sont plus ou moins corrélées avec le PMJP relatif, à l'exception de la variable DIV ACTIV, de la variable TYPE HOSP et de la variable CAV sur le secteur OQN. Autrement dit toutes les variables influent sur le PMJP relatif, sauf la diversité d'activité des établissements, le type d'hospitalisation et la CAV pour le secteur OQN. Ces variables sont moins corrélées que les autres à l'activité proprement dite.
- Les variables du modèle sont plus ou moins corrélées entre elles. Les plus fortes corrélations ont lieu de manière négative entre l'âge et la pédiatrie et entre l'absence ou non d'activité de rééducation-réadaptation. Ces corrélations sont logiques du point de vue médical. Ces variables sont conservées dans l'analyse car même si elles apportent la même information, elles ont un sens différent qui donnera plus de poids à l'interprétation de l'analyse factorielle.

Cependant le coefficient de Pearson n'est applicable que pour mesurer la relation entre deux variables X et Y ayant une distribution de type gaussien et ne comportant pas de valeurs exceptionnelles. Si ces conditions ne sont pas vérifiées l'emploi de ce coefficient peut aboutir à des conclusions erronées sur la présence ou l'absence d'une relation. C'est pour cette raison sont calculés les coefficients de corrélation de Spearman.

Secteur DAF

	PED	DIV ACTIV	AGE	DEP PHYS	DEP CONG	CdARR 0	CdARR 04	CdARR 4	CAV	TYPE PC	TYPE HOSP	PMJP	IND
PED	1	0,36	0,61	0,20	0,09	- 0,46	0,39	0,62	-0,12	0,43	0,63	0,41	0,14
DIV ACTIV	0,36	1	0,50	0,27	0,22	0,51	-0,47	-0,39	0,24	-0,30	-0,32	-0,16	-0,17
AGE	0,61	0,50	1	0,42	0,20	0,52	-0,46	-0,57	0,30	-0,28	-0,51	-0,15	-0,22
DEP PHYS	0,20	0,27	0,42	1	0,65	0,28	-0,26	-0,19	0,57	-0,12	-0,26	0,46	-0,05
DEP CONG	0,09	0,22	0,20	0,65	1	0,28	-0,28	-0,11	0,44	-0,22	-0,14	0,35	0,01
CdARR 0	0,46	0,51	0,52	0,28	0,28	1	-0,96	-0,61	0,15	-0,51	-0,45	-0,45	-0,10
CdARR 04	0,39	0,47	0,46	0,26	0,28	- 0,96	1	0,48	-0,13	0,48	0,36	0,40	0,08
CdARR 4	0,62	0,39	0,57	0,19	0,11	- 0,61	0,48	1	-0,07	0,53	0,64	0,52	0,16
CAV	0,12	0,24	0,30	0,57	0,44	0,15	-0,13	-0,07	1	0,00	-0,12	0,32	-0,01
TYPE PC	0,43	0,30	0,28	0,12	0,22	- 0,51	0,48	0,53	0,00	1	0,49	0,44	0,08
TYPE HOSP	0,63	0,32	0,51	0,26	0,14	- 0,45	0,36	0,64	-0,12	0,49	1	0,31	0,17
PMJP	0,41	0,16	0,15	0,46	0,35	- 0,45	0,40	0,52	0,32	0,44	0,31	1	- 0,007
IND	0,14	0,17	0,22	0,05	0,01	- 0,10	0,08	0,16	-0,01	0,08	0,17	- 0,007	1



Secteur OQN

	PED	DIV ACTIV	AGE	DEP PHYS	DEP CONG	CdARR 0	CdARR 04	CdARR 4	CAV	TYPE PC	TYPE HOSP	PMJP	IND
PED	1	- 0,43	- 0,64	- 0,29	- 0,24	- 0,43	0,36	0,44	-0,38	0,21	0,37	0,39	-0,02
DIV ACTIV	0,43	1	0,63	0,53	0,43	0,54	-0,53	-0,25	0,50	-0,27	-0,31	-0,13	-0,15
AGE	- 0,64	0,63	1	0,58	0,35	0,52	-0,46	-0,38	0,57	-0,16	-0,38	-0,17	-0,16
DEP PHYS	- 0,29	0,53	0,58	1	0,66	0,23	-0,23	0,01	0,60	0,03	-0,18	0,40	-0,10
DEP CONG	- 0,24	0,43	0,35	0,66	1	0,23	-0,22	-0,07	0,48	-0,08	-0,24	0,31	-0,09
CdARR 0	- 0,43	0,54	0,52	0,23	0,23	1	-0,94	-0,52	0,32	-0,46	-0,37	-0,47	-0,08
CdARR 04	0,36	0,53	0,46	0,23	0,22	- 0,94	1	0,32	-0,31	0,42	0,29	0,39	0,07
CdARR 4	0,44	0,25	0,38	0,01	0,07	- 0,52	0,32	1	-0,10	0,39	0,43	0,50	0,10
CAV	- 0,38	0,50	0,57	0,60	0,48	0,32	-0,31	-0,10	1	-0,06	-0,25	0,08	0,03
TYPE PC	0,21	- 0,27	- 0,16	0,03	- 0,08	- 0,46	0,42	0,39	-0,06	1	0,45	0,38	0,24
TYPE HOSP	0,37	- 0,31	- 0,38	0,18	0,24	- 0,37	0,29	0,43	-0,25	0,45	1	0,14	0,22
PMJP	0,39	- 0,13	- 0,17	0,40	0,31	- 0,47	0,39	0,50	0,08	0,38	0,14	1	-0,15
IND	- 0,02	- 0,15	- 0,16	- 0,10	- 0,09	- 0,08	0,07	0,10	0,03	0,24	0,22	-0,147	1

- Les corrélations entre l'indice de modulation et les variables du modèle sont sensiblement les mêmes que pour les corrélations linéaires, à l'exception des variables liés à l'âge, concernant la rééducation-réadaptation et le type d'hospitalisation. D'une part cela indique que ces variables sont corrélées de manières non linéaires avec le coefficient établissement (coefficient Spearman > Coefficient Pearson). Cela peut poser problème pour la suite de l'analyse, puisque l'ACP se base sur des corrélations linéaires. Il faudra donc être particulièrement attentif sur ces variables, même si les corrélations de Spearman ne sont de toute façon pas très élevées.
- Les corrélations entre le PMJP et les variables du modèle sont sensiblement les mêmes que pour les corrélations linéaires. La variable type d'hospitalisation présente des coefficients de Spearman plus élevés avec le PMJP, ce qui traduit une relation non linéaire mais surtout que la variable TYPE HOSP est bien corrélée avec le PMJP. En revanche la variable âge présente des coefficients de Spearman moins élevés avec le PMJP, ce qui traduit la présence d'individus exceptionnels qui tirent les coefficients de Pearson. Il conviendra d'être plus attentif sur la liaison entre la variable de l'âge des patients et le PMJP
- Les corrélations entre les variables du modèles sont relativement les mêmes que pour les corrélations linéaires. Il faut souligner quand même que les variables liées à l'âge sont nettement moins corrélées, en raison de la présence d'individus exceptionnels

Dans l'ensemble les variables du modèle présentent de faibles corrélations linéaires ou non avec l'indice de modulation et des corrélations linéaires ou non importantes avec le PMJP relatif. Par conséquent sur cette analyse descriptive, le modèle IVA apparaît assez pertinent pour prendre bien en compte le poids de l'activité des établissements.

La probable non-linéarité entre certaines variables du modèle (pédiatrie, rééducation-réadaptation et type d'hospitalisation) et l'indice de modulation doit rester à l'esprit lors de l'analyse factorielle. Cependant les corrélations non linéaires restant faibles, elles ne peuvent remettre en cause le principe de l'analyse factorielle.

Les variables liées à l'âge, la diversité d'activité, le type d'hospitalisation et le type de prise en charge sur le secteur OQN sont quelque peu corrélées avec l'indice de modulation. La suite de l'analyse se portera par conséquent particulièrement sur ces variables.



1.2 Résultats concernant l'analyse factorielle

Continuons l'analyse par celui de la matrice des variances-covariances, matrice de même nature que celle des corrélations, bien que moins parlante. La diagonale de cette matrice fournit les variances des variables considérées. Pour rappel c'est à partir de cette matrice que les axes factoriels sont constitués.

Secteur DAF

	PED	DIV ACTIV	AGE	DEP PHYS	DEP CONG	CdARR 0	CdARR 04	CdARR 4	CAV	TYPE PC	TYPE HOSP	PMJP	IND
PED	0,04	- 0,04	- 2,65	0,13	0,84	- 1,43	1,07	0,36	0,26	0,48	0,63	0,01	0,00
DIV ACTIV	-	0,04	6,89	5,15	1,39	11,55	9,57	1,99	1,42	9,55	3,30	0,02	0,05
AGE	- 2,65	6,89	287,97	73,41	- 44,85	234,24	- 180,08	- 54,25	4,68	-126,67	-97,81	-0,91	- 1,06
DEP PHYS	0,13	5,15	73,41	421,42	228,97	165,55	141,89	23,68	137,37	97,54	74,89	1,41	0,08
DEP CONG	0,84	1,39	44,85	228,97	305,94	93,11	87,55	5,46	92,12	129,53	29,66	1,02	0,56
CdARR 0	1,43	11,55	234,24	165,55	93,11	775,00	688,72	86,75	24,37	508,49	136,42	1,63	0,55
CdARR 04	1,07	9,57	180,08	141,89	87,55	-688,72	652,39	36,88	18,48	420,03	95,00	1,23	0,40
CdARR 4	0,36	- 1,99	- 54,25	- 23,68	- 5,46	- 86,75	36,89	49,90	- 5,82	88,84	41,55	0,39	0,15
CAV	0,26	1,42	4,68	137,37	92,12	24,37	- 18,48	- 5,82	147,02	- 23,30	-25,64	0,65	0,13
TYPE PC	0,48	9,55	126,67	97,54	129,53	-508,49	420,03	88,84	23,31	1	188,45	1,82	0,58
TYPE HOSP	0,63	3,30	97,81	74,89	29,66	-136,42	95,00	41,55	25,64	188,45	186,24	0,30	0,38
PMJP	0,01	- 0,02	- 0,91	1,41	1,02	- 1,63	1,23	0,39	0,65	1,82	0,30	0,02	- 0,00
IND	0,00	- 0,05	- 1,06	- 0,08	0,56	- 0,55	0,40	0,15	0,13	0,58	0,38	- 0,00	0,14

Secteur OQN

	PED	DIV ACTIV	AGE	DEP PHYS	DEP CONG	CdARR 0	CdARR 04	CdARR 4	CAV	TYPE PC	TYPE HOSP	PMJP	IND
PED	0,06	- 0,07	- 3,76	- 1,35	0,08	- 0,88	0,88	- 0,00	- 0,43	- 1,69	- 0,29	0,03	- 0,02
DIV ACTIV	-	0,07	1,10	8,79	2,84	14,03	12,94	1,08	2,49	11,98	4,04	- 0,03	- 0,04
AGE	- 3,76	8,69	299,32	156,50	17,52	156,93	- 141,12	- 15,84	48,45	40,54	- 32,52	- 1,58	0,56
DEP PHYS	1,35	8,79	156,50	351,29	121,53	70,76	71,54	0,81	72,06	9,69	71,93	1,05	- 0,58
DEP CONG	0,08	2,84	17,52	121,53	159,64	34,54	- 28,02	- 6,74	34,45	40,40	36,28	0,77	- 0,32
CdARR 0	- 0,88	14,03	156,93	70,76	34,54	812,49	- 751,03	60,65	36,80	537,91	127,76	- 2,37	- 0,13
CdARR 04	0,88	-	-141,12	71,54	- 28,02	-751,03	744,99	5,35	32,86	473,23	91,02	1,97	- 0,01
CdARR 4	- 0,00	- 1,08	- 15,84	0,81	- 6,74	- 60,65	5,35	55,30	- 3,94	64,68	36,64	0,40	0,15
CAV	- 0,43	2,49	48,45	72,06	34,45	36,80	- 32,86	- 3,93	48,91	- 20,82	- 25,08	0,08	0,10
TYPE PC	- 1,69	- 11,98	40,54	9,69	- 40,40	-537,91	473,23	64,68	- 20,82	1418,24	207,40	2,33	2,64
TYPE HOSP	- 0,29	- 4,04	- 32,52	- 71,93	- 36,28	-127,76	91,02	36,64	- 25,08	207,40	272,66	0,02	0,77
PMJP	0,03	- 0,03	- 1,58	1,05	0,77	- 2,38	1,97	0,40	0,08	2,33	0,02	0,04	- 0,01
IND	- 0,02	- 0,04	0,56	- 0,58	- 0,32	- 0,13	- 0,01	0,15	0,10	2,64	0,77	- 0,01	0,11

Les valeurs propres données ci-dessous sont celles de la matrice de variance-covariance.



Secteur DAF

Axe	Valeur propre	Pourcentage Variance expliquée	Pourcentage Variance expliquée cumulée
1	4,17	0,32	0,32
2	2,63	0,2	0,52
3	1,35	0,1	0,63
4	1,04	0,08	0,71
5	0,98	0,08	0,78
6	0,65	0,05	0,83
7	0,57	0,04	0,88
8	0,54	0,04	0,92
9	0,49	0,04	0,96
10	0,3	0,02	0,98
11	0,19	0,01	0,99
12	0,08	0,01	1
13	0	0	1

Secteur OQN

Axe	Valeur propre	Pourcentage Variance expliquée	Pourcentage Variance expliquée cumulée
1	3,77	0,29	0,29
2	2,21	0,17	0,46
3	2,15	0,17	0,63
4	1,14	0,09	0,71
5	0,93	0,07	0,78
6	0,68	0,05	0,84
7	0,59	0,05	0,88
8	0,53	0,04	0,92
9	0,52	0,04	0,96
10	0,33	0,03	0,99
11	0,1	0,01	1
12	0,05	0	1
13	0	0	1

Chaque ligne correspond à un axe factoriel, dont la valeur propre correspond à la variance de l'axe. La colonne « Pourcentage Variance expliquée » correspond au pourcentage de variance de chaque ligne par rapport au total. La dernière colonne représente le cumul de ces pourcentages.

La somme des valeurs propres est égal à la somme des corrélations entre les variables elles mêmes, c'est-à-dire égal au nombre de variables (le coefficient de corrélation d'une variable avec cette même variable est égale à 1).



Les deux premiers axes factoriels sont retenus pour le reste de l'analyse, ce qui correspond à 52% de la variance expliquée sur le secteur DAF et 46% de la variance expliquée sur le secteur OQN. Les deux premiers axes représentent en effet une part de variance expliquée importante et suffisent pour expliquer les relations entre toutes les variables présentes dans l'analyse.

1.3 Résultats sur les variables

Le résultat sur les variables est la matrice des corrélations variables facteurs. Il s'agit des coefficients de corrélations linéaires entre les variables initiales et les facteurs. Ces corrélations vont permettre de donner un sens aux axes. Afin de renforcer l'analyse des axes, on pourrait calculer la contribution de chaque variable aux différents axes, mais ici on se contentera de la matrice des corrélations.

Secteur DAF

	Axe 1	Axe 2
PED	0,53	0,48
DIV ACTIV	-0,64	0,12
AGE	-0,76	-0,25
DEP PHYS	-0,35	0,8
DEP CONG	-0,15	0,81
CdARR 0	-0,87	0,08
CdARR 04	0,77	-0,1
CdARR 4	0,63	0,07
CAV	-0,13	0,7
TYPE PC	0,65	-0,09
TYPE HOSP	0,6	-0,12
PMJP	0,47	0,71
IND	0,12	0,03

Secteur OQN

	Axe 1	Axe 2
PED	-0,53	0,09
DIV ACTIV	0,72	0,16
AGE	0,73	0,06
DEP PHYS	0,51	0,74
DEP CONG	0,28	0,67
CdARR 0	0,78	-0,33
CdARR 04	-0,73	0,3
CdARR 4	-0,31	0,16
CAV	0,52	0,53
TYPE PC	-0,44	0,28
TYPE HOSP	-0,39	-0,15
PMJP	-0,49	0,73
IND	0,01	-0,2



Le premier axe caractérise une opposition entre deux catégories de variables :

- Une première catégorie comprenant les variables PED, CdARR 04, CdARR 4, TYPE PC et TYPE HOSP où les variables sont fortement corrélées de manière positive pour le secteur DAF et de manière négative pour le secteur OQN avec le premier axe. Ainsi plus le nombre de journées pédiatriques est important, plus l'établissement a des activités de rééducation et de réadaptation.
- Une deuxième catégorie comprenant les variables DIV ACTIV, AGE, CdARR 0 qui sont fortement corrélées de manière négative pour le secteur DAF et de manière positive pour le secteur OQN avec le premier axe. Ainsi plus l'âge moyen des patients est important, moins l'établissement a des activités de rééducation et de réadaptation.

Les Variables des de la catégorie 1 et de la catégorie 2 ont donc une corrélation forte mais inverse. Par exemple plus l'établissement a un nombre de journées pédiatriques, plus l'âge moyen des patients est faible. **L'axe 1 caractérise surtout le lien entre l'âge des patients, les activités de rééducation et de réadaptation et le degré de diversité.** Les variables de dépendances (CONG, PHYS) et de comorbidités (CAV) sont moins contributives sur l'axe 1.

Le deuxième axe caractérise une corrélation fortement positive d'une catégorie de variables.

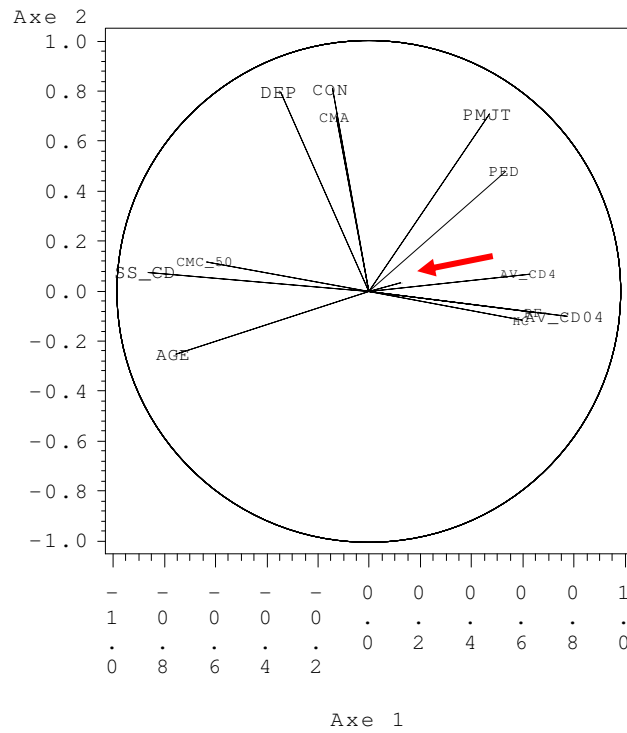
- Cette catégorie comprend les variables de dépendances (CONG, PHYS) et de comorbidités (CAV). Ainsi plus l'établissement aura des patients avec une forte dépendance, plus l'établissement aura des patients avec une forte comorbidité. Les autres variables du modèle sont peu contributives sur cette composante. **L'axe 2 caractérise le lien entre les variables de dépendance et de comorbidités.**

Le premier axe va permettre de caractériser l'indice de modulation et le PMJP par rapport aux variables d'âge et d'activité de rééducation réadaptation. Le deuxième axe va permettre de caractériser l'indice de modulation et le PMJP par rapport aux variables de dépendances et de comorbidités. L'indice de modulation est faiblement corrélé sur les deux axes, alors que le PMJP présente des corrélations importantes surtout sur le deuxième axe. L'indice de modulation a une corrélation plus forte sur l'axe 1 pour le secteur DAF et une corrélation plus forte sur l'axe 2 pour le secteur OQN, mais toutes ces corrélations restent extrêmement faibles.

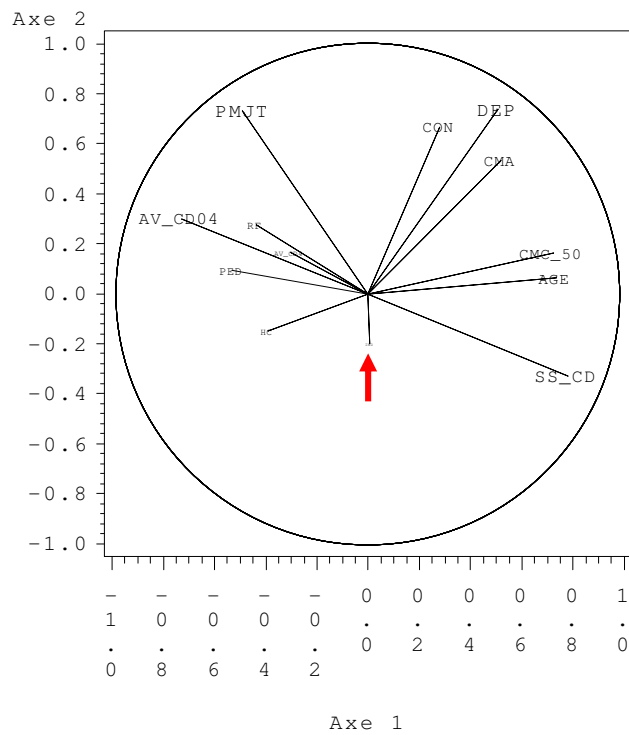
Ces résultats peuvent être confirmés par le cercle des corrélations ci-dessous.



Secteur DAF



Secteur OQN





Les variables d'âge et d'activité de rééducation réadaptation caractérisent l'axe 1 et les variables de dépendance et de comorbidités caractérisent l'axe 2.

L'ensemble des variables du modèle ne sont pas corrélées avec l'indice de modulation et corrélées avec le PMJP relatif. Par conséquent le modèle est pertinent car il prend bien en compte le poids de l'activité.

Les variables du type de prise en charge et du type d'hospitalisation présentent une absence de corrélation moins évidente sur le secteur OQN

Le nombre de variables importantes rend cependant le graphique peu lisible. Afin d'améliorer la lisibilité du graphique, il est intéressant de réaliser les ACP sur des classes de variables.

5 Classification de variables

On va procéder à une classification des variables du modèle IVA. Une fois qu'on aura constitué les classes de variables, on fera une ACP sur les variables de chaque classe, plus les variables PMJP relatif et d'indice de modulation.

La méthode sélectionnée est une classification hiérarchique descendante de variables selon un algorithme itératif opérant une dichotomie sur un groupe de variables et donc initialement sur toutes.

Le choix du nombre de classe sera déterminé quand toutes les secondes valeurs propres de chaque classe seront inférieures à 1. Il existe d'autres critères de sélection du choix du nombre de classes, mais celui qu'on a retenu donnent des classes pertinentes d'un point de vue statistique et médicale.

Le nombre de classes de variables retenues est de 4 pour les deux secteurs DAF et OQN.

Classe 1 : PED, AGE

Cette classe caractérise les variables liées à l'âge.

Classe 2 : DEP PHYS, DEP CONG, CAV

Cette classe caractérise les variables de dépendance et de comorbidités.

Classe 3 : CdARR 0, CdARR 04, TYPE PC et DIV ACTIV

Cette classe caractérise la présence ou non d'activité de réadaptation et rééducation et le degré de diversité de l'établissement.

Classe 4 : CdARR 4, TYPE PC, TYPE HOSP

Cette classe caractérise la présence forte d'activité de rééducation et la présence d'hospitalisation partielle.

Cette classification reste cohérente avec l'ACP initiale, puisque les classes 1, 3 et 4 sont contributives sur l'axe 1 et la classe 2 sur l'axe 2.



On réalise ensuite une ACP sur chacune des classes avec en complément le PMJP relatif et l'indice de modulation, afin d'obtenir un effet zoom

6 Analyse factorielle sur les classes

Les résultats suivants vont concernés la matrice de corrélation entre les variables et les facteurs et les graphiques associés sur les variables.

1.4 Analyse factorielle sur la classe 1

Les valeurs propres données ci-dessous sont celles de la matrice de variance-covariance.

Secteur DAF

Axe	Valeur propre	Pourcentage Variance expliquée	Pourcentage Variance expliquée cumulée
1	2,14	0,54	0,54
2	1,06	0,26	0,8
3	0,63	0,16	0,96
4	0,17	0,04	1

Secteur OQN

Axe	Valeur propre	Pourcentage Variance expliquée	Pourcentage Variance expliquée cumulée
1	2,39	0,6	0,6
2	0,95	0,24	0,83
3	0,56	0,14	0,97
4	0,11	0,03	1

Les deux premiers axes, qui ont un pourcentage de variance expliquée, respectivement pour le secteur DAF et OQN de 80 et 83%, sont retenus dans la suite de l'analyse.

L'interprétation des axes se fait par l'intermédiaire de la matrice des corrélations entre les variables et les facteurs.

Secteur DAF

	Axe 1	Axe 2
PED	0,93	-0,02
AGE	-0,89	-0,17
IND	0,12	0,95
PMJP	0,68	-0,36



Secteur OQN

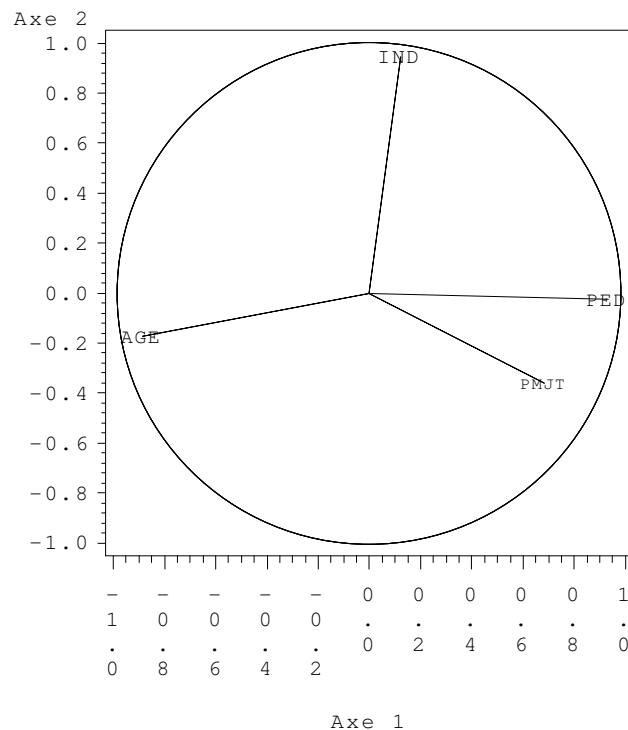
	Axe 1	Axe 2
PED	0,95	0,1
AGE	-0,87	-0,3
IND	-0,36	0,92
PMJP	0,77	-0,03

Le premier axe se caractérise par la forte corrélation des variables liées à l'âge et du PMJP relatif avec l'axe. En revanche l'indice de modulation est peu contributif sur le premier axe, même si la corrélation est plus forte sur le secteur OQN.

Le deuxième axe, sur les deux secteurs, confirme l'absence de corrélation entre les variables liées à l'âge, le PMJP relatif avec l'indice de modulation. Cet axe est l'inverse du premier avec une faible contribution des variables liées à l'âge et du PMJP et une forte contribution de l'indice sur l'axe. Cependant sur le secteur OQN, la variable de l'âge du patient présente une petite corrélation avec l'indice de modulation.

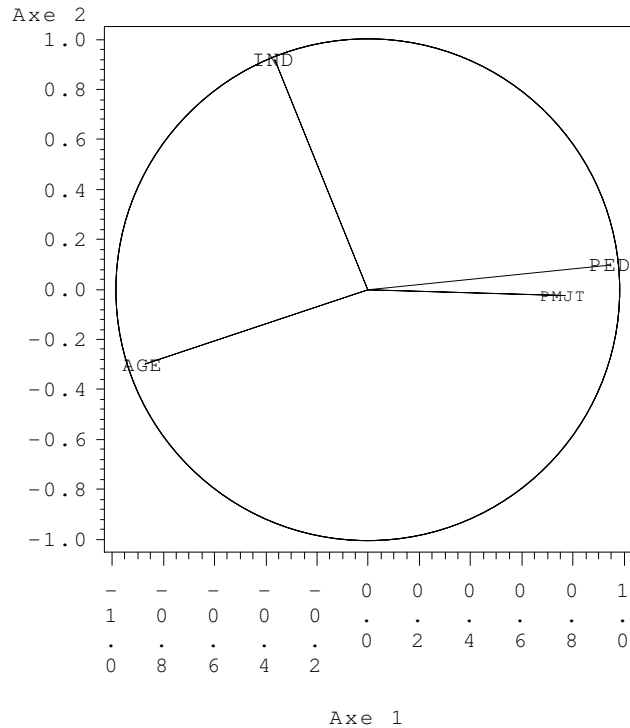
Les graphiques vont aider à renforcer ces interprétations.

Secteur DAF :





Secteur OQN



L'indice de modulation est bien représenté sur l'axe 2 et les autres variables sur l'axe 1.

Les variable liés à l'âge ne sont pas corrélées avec l'indice de modulation et corrélées avec le PMJP relatif. Par conséquent ces variables sont bien prises en compte dans le poids de l'activité.

Cependant cette absence de corrélation est moins flarante sur le secteur OQN.



1.5 Analyse factorielle sur la classe 2 (DEP PHYS, DEP CONG, CAV)

Secteur DAF

Axe	Valeur propre	Pourcentage Variance expliquée	Pourcentage Variance expliquée cumulée
1	2,42	0,48	0,48
2	1,03	0,21	0,69
3	0,64	0,13	0,82
4	0,57	0,11	0,93
5	0,33	0,07	1

Secteur OQN

Axe	Valeur propre	Pourcentage Variance expliquée	Pourcentage Variance expliquée cumulée
1	2,13	0,43	0,43
2	1,17	0,23	0,66
3	0,77	0,15	0,81
4	0,54	0,11	0,92
5	0,39	0,08	1

Les deux premiers axes sont toujours retenus avec respectivement pour le secteur DAF et OQN, de 69 et 66% de part de variance expliquée.

L'interprétation des axes se fait par l'intermédiaire de la matrice des corrélations entre les variables et les facteurs.

Secteur DAF

	Axe1	Axe2
DEP PHYS	0,87	-0,01
DEP CONG	0,8	0,14
CAV	0,75	0,04
IND	0,02	0,98
PMJP	0,68	-0,23



Secteur OQN

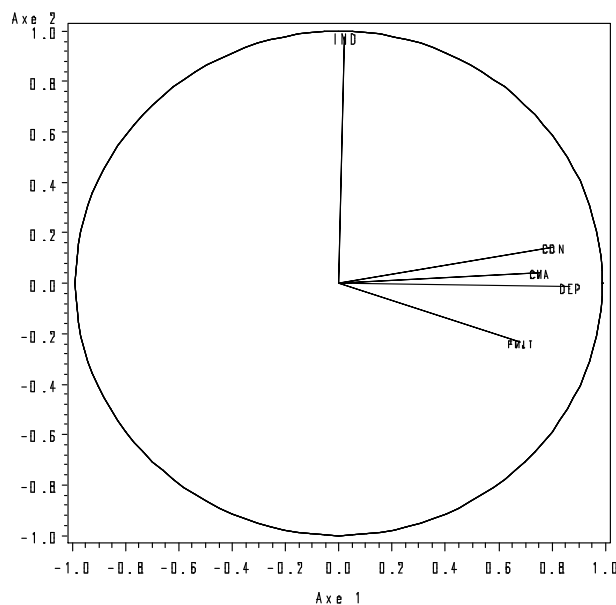
	Axe1	Axe2
DEP PHYS	0,85	0,13
DEP CONG	0,78	0,02
CAV	0,71	0,45
IND	-0,19	0,78
PMJP	0,51	-0,58

Le premier axe se caractérise par la forte corrélation des variables de dépendances, de comorbidités et le PMJP relatif avec l'axe. En revanche l'indice de modulation est peu contributif sur le premier axe, même si la corrélation est plus forte sur le secteur OQN.

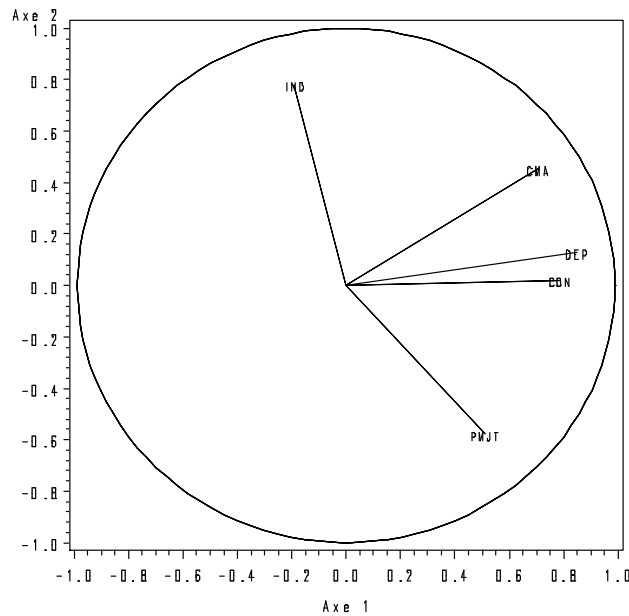
Le deuxième axe, sur les deux secteurs, confirme l'absence de corrélation entre les variables de dépendances, de comorbidités et du PMJP relatif avec l'indice de modulation. Cet axe est l'inverse du premier avec une faible contribution des variables de dépendances, de comorbidités et du PMJP et une forte contribution de l'indice sur l'axe. Cependant sur le secteur OQN, la variable concernant la comorbidité présente une petite corrélation avec l'indice de modulation.

Les graphiques vont aider à renforcer ces interprétations.

Secteur DAF



Secteur OQN



L'indice de modulation est bien représenté sur l'axe 2 et les autres variables sur l'axe 1. Cependant le PMJP sur les deux secteurs et la variable CAV sur le secteur OQN ont une petite contribution sur l'axe 2.

Les variable de dépendances et de comorbidités ne sont pas corrélées avec l'indice de modulation et corrélées avec le PMJP relatif. Par conséquent ces variables sont bien prises en compte dans le poids de l'activité.

Cependant cette absence de corrélation est moins flagrante sur le secteur OQN pour la variable concernant la comorbidité.

1.6 Analyse factorielle sur la classe 3

Secteur DAF

Axe	Valeur propre	Pourcentage Variance expliquée	Pourcentage Variance expliquée cumulée
1	2,89	0,48	0,48
2	1,1	0,18	0,67
3	0,8	0,13	0,8
4	0,63	0,1	0,9
5	0,56	0,09	1
6	0,03	0	1

Secteur OQN

Axe	Valeur propre	Pourcentage Variance expliquée	Pourcentage Variance expliquée cumulée
1	2,87	0,48	0,48
2	1,21	0,2	0,68
3	0,78	0,13	0,81
4	0,62	0,1	0,91
5	0,49	0,08	0,99
6	0,03	0,01	1

Les deux premiers axes sont toujours retenus avec respectivement pour le secteur DAF et OQN, de 67 et 68% de part de variance expliquée.

L'interprétation des axes se fait par l'intermédiaire de la matrice des corrélations entre les variables et les facteurs.

Secteur DAF

	Axe1	Axe2
CdARR 0	-0,94	-0,01
CdARR 04	0,9	0
DIV ACTIV	-0,62	0,37
TYPE PC	0,73	0,08
IND	0,09	-0,86
PMJP	0,52	0,47

Secteur OQN

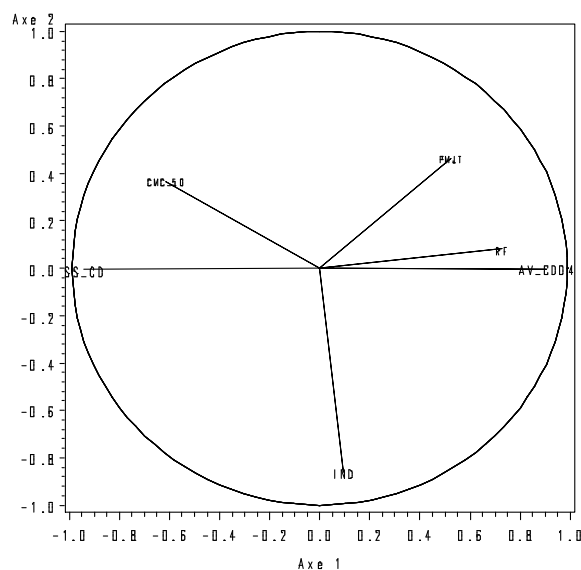
	Axe1	Axe2
CdARR 0	-0,94	0,05
CdARR 04	0,91	-0,05
DIV ACTIV	-0,62	-0,25
TYPE PC	0,68	0,25
IND	0,06	0,9
PMJP	0,56	-0,52

Le premier axe se caractérise par la forte corrélation des variables de rééducation-réadaptation, du type de prise en charge, du degré de diversité et du PMJP relatif avec l'axe. En revanche l'indice de modulation n'est pas contributif sur le premier axe.

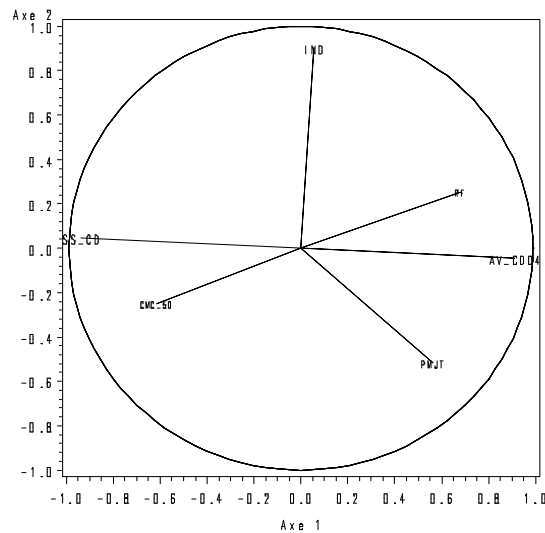
Le deuxième axe, sur les deux secteurs, confirme l'absence de corrélation entre les variables de rééducation-réadaptation, du type de prise en charge, du degré de diversité de l'activité et du PMJP relatif avec l'indice de modulation. Cet axe est l'inverse du premier avec une faible contribution des variables de rééducation-réadaptation, du type de prise en charge, du degré de diversité de l'activité et du PMJP et une forte contribution de l'indice sur l'axe. Cependant le degré de diversité de l'activité sur les deux secteurs et le type de prise en charge sur le secteur OQN présentent une petite corrélation avec l'indice de modulation sur l'axe 2. En effet, Il existe une légère corrélation négative entre la variable DIV ACTIV et l'indice de modulation. Autrement dit plus l'établissement n'est pas diversifié dans plusieurs CMC, plus il a tendance à être surdoté.

Les graphiques vont aider à renforcer ces interprétations

Secteur DAF



Secteur OQN



L'indice de modulation est bien représenté sur l'axe 2 et les autres variables sur l'axe 1. Cependant le PMJP, la variable DIV ACTIV sur les deux secteurs et la variable TYPE PC sur le secteur OQN ont une petite contribution sur l'axe 2.

Les variables de rééducation-réadaptation, du type de prise en charge et du degré de diversité de l'activité ne sont pas corrélées avec l'indice de modulation et corrélées avec le PMJT relatif. Par conséquent ces variables sont bien prises en compte dans le poids de l'activité.



1.7 Analyse factorielle sur la classe 4

Secteur DAF

Axe	Valeur propre	Pourcentage Variance expliquée	Pourcentage Variance expliquée cumulée
1	1,65	0,41	0,41
2	1,06	0,26	0,68
3	0,8	0,2	0,88
4	0,5	0,12	1

Secteur OQN

Axe	Valeur propre	Pourcentage Variance expliquée	Pourcentage Variance expliquée cumulée
1	1,41	0,35	0,35
2	1,25	0,31	0,67
3	0,76	0,19	0,86
4	0,57	0,14	1

Les deux premiers axes sont toujours retenus avec 68 et 67% de part de variance expliquée respectivement sur les deux secteurs.

L'interprétation des axes se fait par l'intermédiaire de la matrice des corrélations entre les variables et les facteurs.

Secteur DAF

	Axe 1	Axe 2
TYPE HOSP	0,72	0,25
CdARR 4	0,85	0,02
IND	0,09	0,9
PMJP	0,63	-0,43

Secteur OQN

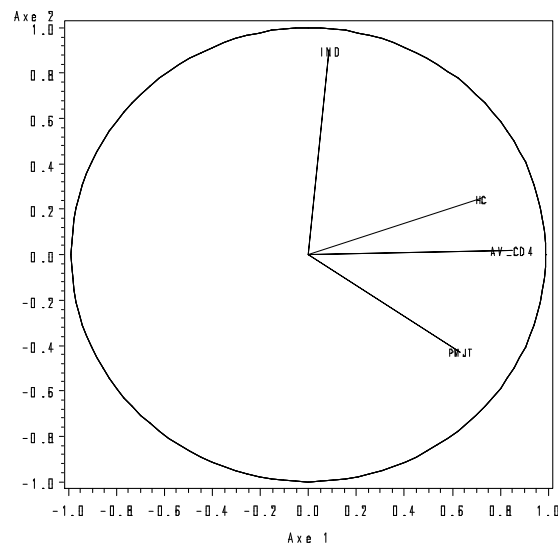
	Axe 1	Axe 2
TYPE HOSP	0,63	0,46
CdARR 4	0,84	0,03
IND	0,04	0,8
PMJP	0,56	-0,63

Le premier axe se caractérise par la forte corrélation des variables du type d'hospitalisation, de forte rééducation-réadaptation et du PMJP relatif avec l'axe. En revanche l'indice de modulation n'est pas contributif sur le premier axe.

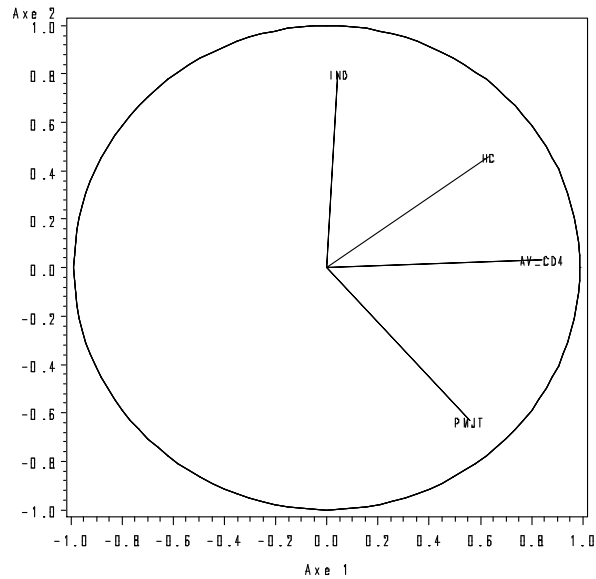
Le deuxième axe, sur les deux secteurs, confirme l'absence de corrélation entre les variables de type d'hospitalisation, de forte rééducation-réadaptation et du PMJP relatif avec l'indice de modulation. Cet axe est l'inverse du premier avec une faible contribution des variables de type d'hospitalisation, de forte rééducation-réadaptation et du PMJP et une forte contribution de l'indice sur l'axe. Cependant le type d'hospitalisation sur les deux secteurs présente une petite corrélation avec l'indice de modulation sur l'axe 2.

Les graphiques vont aider à renforcer ces interprétations

Secteur DAF



Secteur OQN



L'indice de modulation est bien représenté sur l'axe 2 et les autres variables sur l'axe 1. Cependant le PMJP, la variable TYPE HOSP sur les deux secteurs ont une petite contribution sur l'axe 2.

Les variables du type d'hospitalisation et de forte rééducation-réadaptation ne sont pas corrélées avec l'indice de modulation et corrélées avec le PMJP relatif. Par conséquent ces variables sont bien prises en compte dans le poids de l'activité.



ANNEXE 3 :
Analyse de la Variance entre l'indice de modulation et la catégorie de population
RF/SSM/MIXTE

1 Les données

Chaque établissement a été classé par les ARH dans une catégorie d'autorisation (RF, SSM ou MIXTE). Les établissements qui ont eu une réponse NSP (ne se prononce pas) sont retirés de l'analyse. L'enquête auprès des ARH ayant eu lieu en 2006, les établissements n'ayant pas transmis de PMSI en 2006 et les établissements appartenant à une région qui n'a pas répondu à l'enquête sont exclus de l'analyse.

Secteur	Etablissements exclus		
	<i>NB ETAB</i>	<i>% ETAB</i>	<i>% masse financière</i>
DAF	167	15%	11%
OQN	42	9%	6%

Il reste dans le périmètre 909 établissements du secteur DAF et 388 établissements du secteur OQN.

7 Analyse descriptive

Secteur DAF

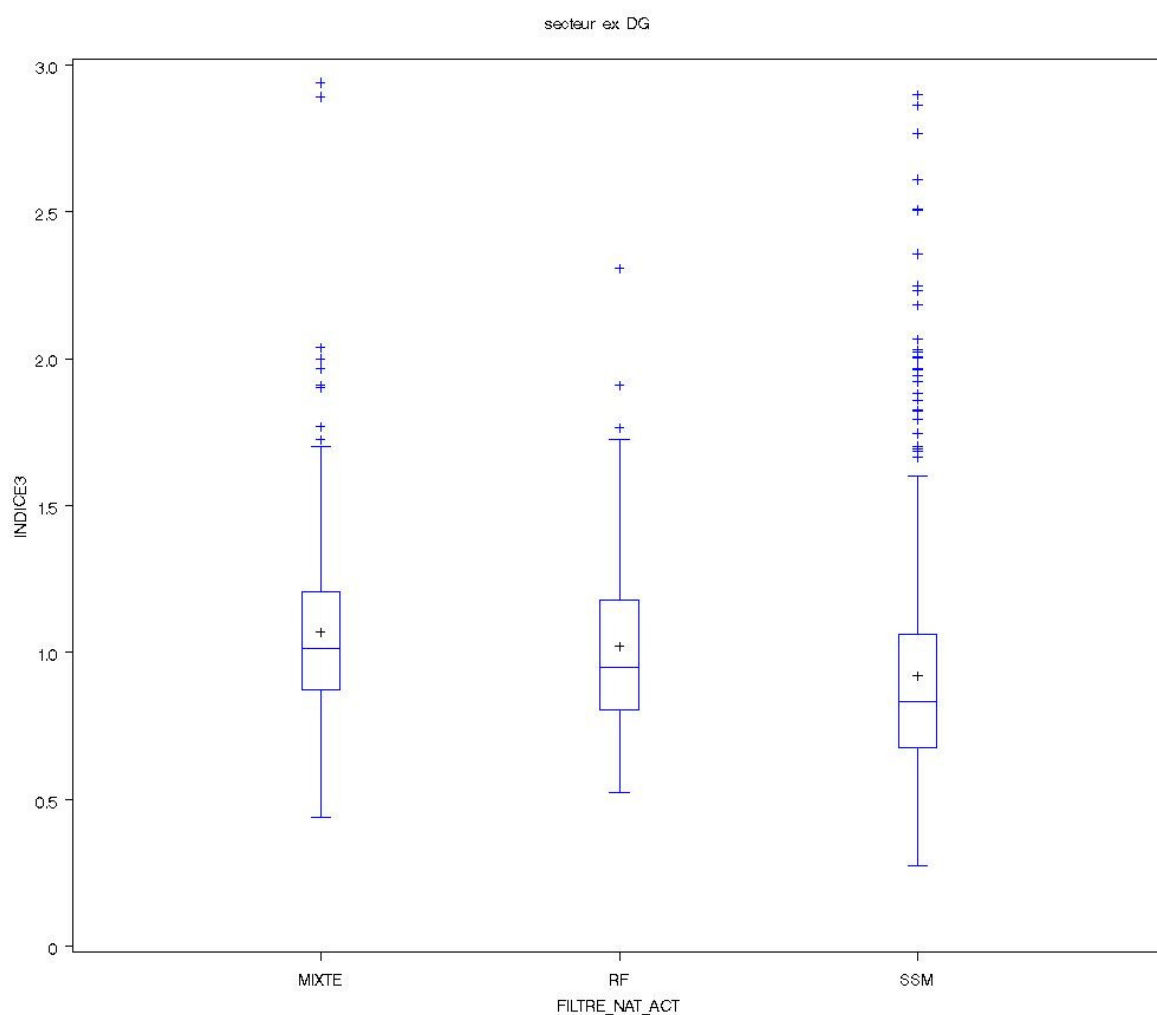
Etant donné que l'analyse de la variance est basée sur une comparaison des moyennes, l'analyse descriptive consiste à étudier les moyennes de l'indice de modulation selon les catégories d'autorisations.

Il s'agit ici plus particulièrement de moyenne pondérée de l'indice de modulation par le chiffre d'affaire théorique, afin de prendre en compte le poids de chaque établissement vis-à-vis de l'importance de sa masse financière, qui en théorie lui sera redistribué. Donc le tableau ci-dessous présente les statistiques sur l'indice moyen de modulation par catégorie d'autorisations.

Le tableau ci-dessous indique que les établissements de RF et MIXTE sont sur dotés. En revanche les établissements de SSM sont en moyenne sous dotés.

	Effectif	Moyenne
MIXTE	192	1,07
RF	100	1,02
SSM	617	0,91

Ces tendances sont confirmées par le tracé de la boîte à moustache



Les catégories « MIXTE » et « RF » sont proches les unes des autres. En revanche la dispersion de la catégorie « SSM » montre que ces établissements sont en moyenne sous dotés, même s'il existe quelques établissements nettement sur dotés.

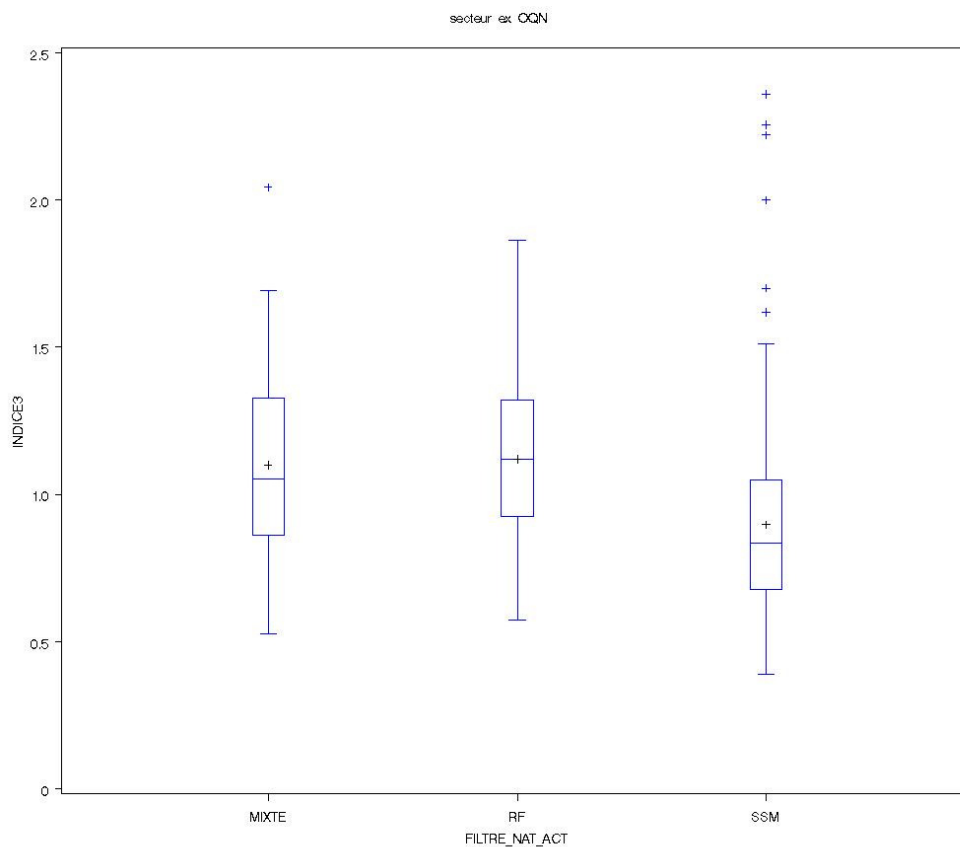


Secteur QQN

	Effectif	Moyenne
MIXTE	55	1,10
RF	80	1,12
SSM	253	0,90

Le tableau pour le secteur QQN indique que les établissements de RF et MIXTE sont en moyenne sur dotés. En revanche les établissements de SSM sont comme sur le secteur DAF en moyenne sous dotés.

Ces tendances sont confirmées par le tracé de la boîte à moustache



2 Analyse de la variance

7.1 Tests Paramétriques

L'étude d'une liaison entre une variable numérique et ordinale se fait par une analyse de la variance. L'hypothèse nécessaire aux tests paramétriques est que la variable numérique doit suivre une loi normale et l'hypothèse souhaitable aux tests paramétriques est l'égalité des variances de chaque modalité de la variable ordinale (test d'homoscédasticité).

Si on considère ces hypothèses vraies, il est alors possible d'appliquer les tests paramétriques, c'est-à-dire de tester si les deux variables sont indépendantes, autrement dit de tester si les moyennes de la variable numérique de chaque modalité de la variable nominale sont égales.

La procédure utilisée est le test ANOVA, qui est une généralisation du test de Student quand la variable nominale a plus de 2 modalités.

L'hypothèse nulle du test est : $H_0 \Rightarrow \text{Moyenne RF} = \text{Moyenne SSM} = \text{Moyenne Mixte}$.

Le test ANOVA, comme pour l'analyse descriptive, se pratique sur les moyennes pondérées, afin de prendre le poids financier de l'établissement.

Secteur DAF

Test ANOVA

Les résultats sont les suivants :

Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	2	3,62	1,81	13,49	<,0001
Error	906	121,59	0,13		
Corrected Total	908	125,21			

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent au moins une des moyennes est différente des autres.

L'intensité de la liaison entre la variable numérique et ordinale est communément mesurée par l'éta carré.

Eta carré = (Somme des carrés interclasse/Somme des carrés totaux) = 0,03

Cet indicateur est comparable au R carré d'une régression. Par conséquent on peut remarquer, que l'intensité existe, mais est très faible (presque 3%).

Le test d'ANOVA ne dit pas quelles moyennes diffèrent significativement. Pour le savoir on recourt au test de Scheffe (test le plus conservateur).

Les moyennes avec la même lettre ne sont pas très différentes.			
Scheffe Groupement	Moyenne	Nb	FILTRE NAT ACT
A	1,07	192	MIXTE
A	1,02	100	RF
B	0,91	617	SSM

Le Principe est simple. Les Modalités de la variable ordinale qui n'ont pas la même lettre ont des moyennes significativement différentes. Par conséquent :

- La Moyenne RF est significativement différente de la Moyenne SSM
- La Moyenne RF est non significativement différente de la Moyenne MIXTE
- La Moyenne SSM est significativement différente de la Moyenne MIXTE

En particulier les établissements RF sont significativement sur dotés en moyenne et les établissements SSM sont significativement sous dotés en moyenne.

Vérification des hypothèses de départ

Test d'homoscédasticité

Le test choisi est celui de Levene qui a pour hypothèse nulle :

$H_0 \Rightarrow$ Variance de RF = Variance de SSM = Variance Mixte

Levene's Test for Homogeneity of Valeur du Point Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
FILTRE_NAT_ACT	2	0,2488	0,1244	0,95	0,3883
Error	906	119	0,1314		

La P-Value étant supérieure à 0.05, l'hypothèse nulle est acceptée. Par conséquent le test d'homoscédasticité est concluant.

Test de normalité de la valeur du point

Le test choisi est celui de Shapiro Wilk, puisque $n \leq 2000$. L'hypothèse nulle est :

$H_0 \Rightarrow$ La valeur du point suit une Loi normale



Tests de normalité				
Test	Statistique		p Value	
Shapiro-Wilk	W	0,88	Pr < W	<0,0001

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent l'indice de modulation ne suit pas une loi normale.

Secteur OQN

Test ANOVA

Les résultats sont les suivants :

Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	2,00	4,08	2,04	22,15	<,0001
Error	385	35,47	0,09		
Corrected Total	387	39,55			

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent au moins une des moyennes est différente des autres.

Eta carré = 0.10

L'intensité de liaison entre la variable numérique et nominale est plus importante sur le secteur OQN (de l'ordre de 10%).



Le test de Scheffe, qui permet de déterminer quelles moyennes diffèrent des autres, donne les résultats suivants.

Les moyennes avec la même lettre ne sont pas très différentes.			
Scheffe Groupement	Moyenne	Nb	FILTRE_NAT_ACT
A	1.12	80	RF
A	1.10	55	MIXTE
B	0.90	253	SSM

Le constat est le même que sur le secteur DAF, mais avec une intensité de relation plus importante. Les établissements RF et MIXTE sont significativement sur dotés en moyenne et les établissements SSM sont significativement sous dotés en moyenne.

Vérification des hypothèses de départ

Test d'homoscédasticité

Levene's Test for Homogeneity of Valeur du Point Variance					
ANOVA of Squared Deviations from Group Means					
Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
FILTRE_NAT_ACT	2	0,0209	0,0105	0,27	0,766
Error	385	15,1042	0,0392		

La P-Value étant supérieure à 0.05, l'hypothèse nulle est acceptée. Par conséquent le test d'homoscédasticité est concluant.

Test de normalité de la valeur du point

Tests de normalité				
Test	Statistique		p Value	
Shapiro-Wilk	W	0,95	Pr < W	<0.0001

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent l'indice de modulation ne suit pas une loi normale.

Sur les deux secteurs, l'indice de modulation ne suit pas une loi normale. Comme l'hypothèse n'est pas vérifiée, il faut passer à l'utilisation des tests non paramétrique.

7.2 Tests non Paramétriques

Le Test ANOVA est un test paramétrique qui suppose que les variables suivent une loi particulière (homoscédasticité, normalité). Dans l'étude de la relation entre l'indice de modulation et la catégorie d'autorisations, si l'hypothèse de l'homoscédasticité est respectée, l'indice de modulation ne suit pas une loi normale.

Un recours possible est l'utilisation de tests non paramétriques. Ces tests ne supposent pas que les variables suivent une loi particulière et se fondent souvent sur les rangs des valeurs de variables plutôt que sur les valeurs elle mêmes (Tests de Wilcoxon, Kruskal-Wallis). De plus ces tests sont peu sensibles aux valeurs aberrantes.

Tests non paramétriques de Kruskal-Wallis :

Le test de Kruskal-Wallis est non paramétrique: il ne fait aucune hypothèse sur la forme des distributions sous-jacentes. Comme de nombreux tests non paramétriques, il travaillera non pas sur les valeurs des observations, mais sur leurs rangs. La statistique du test de Kruskal-Wallis est construite à partir des moyennes des rangs des observations dans les différentes catégories.

L'hypothèse nulle est donc : H_0 : Egalité des moyennes des rangs de chaque modalité ou population

Secteur DAF :

Kruskal-Wallis Test	
Chi-Square	57.12
DF	2
Pr > Chi-Square	<.0001

Secteur OQN :

Kruskal-Wallis Test	
Chi-Square	50.01
DF	2
Pr > Chi-Square	<.0001

Sur les deux secteurs, la P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent les populations n'ont pas des moyennes de rang identiques.

Les tests paramétriques ont indiqués que les établissements « RF » et « MIXTES » ont une moyenne de l'indice de modulation significativement différente des établissements dits « SSM ». Etant donné que les tests suivants nécessitent la présence d'au maximum deux catégories de population, les établissements « RF » et « MIXTE » sont regroupés en une seule catégorie nommée « RF MIXTE ». Les tests suivants vont opposer par conséquent les établissements « RF et MIXTE » aux établissements « SSM ».

Tests non paramétriques de Wilcoxon :

On dispose de deux populations, à savoir

X_1, \dots, X_m (iid) $\sim F(x)$ avec m le nombre d'établissements dans la catégorie A

Y_1, \dots, Y_n (iid) $\sim G(y)$ avec n le nombre d'établissements dans la catégorie B



Le test consiste à confronter l'hypothèse nulle que les deux populations ont des comportements identiques ($F=G$) à la contre hypothèse suivante que les deux populations n'ont pas des comportements identiques ($F \neq G$).

On détermine S_1, \dots, S_n : les rangs des Y_j dans l'ensemble des $m + n$ observations.
On rejette H_0 : si $W_s = S_1 + \dots + S_n$ s'éloigne "trop" de sa valeur moyenne.

La catégorie A représente les établissements dits « RF et MIXTE » alors que la catégorie B représente les établissements dits « SSM ». Les résultats du test sont les suivants

Secteur DAF :

Wilcoxon Scores (Rank Sums) pour la variable INDICE3 Classée par variable FILTRE_NAT_ACT2					
FILTRE_NAT_ACT2	Nb	Somme des scores	Attendue sous H_0	Écart-type sous H_0	Score moyen
RF MIXTE	292	160233	132860	3696.27	548.74
SSM	617	253362	280735	3696.27	410.63

Wilcoxon Two-Sample Test	
Statistic	160233
Normal Approximation	
Z	7.4054
One-Sided Pr > Z	<.0001
Two-Sided Pr > Z	<.0001
t Approximation	
One-Sided Pr > Z	<.0001
Two-Sided Pr > Z	<.0001
Z includes a continuity correction of 0.5.	



Secteur OQN :

Wilcoxon Scores (Rank Sums) pour la variable INDICE3 Classée par variable FILTRE_NAT_ACT2					
FILTRE_NAT_ACT2	Nb	Somme des scores	Attendue sous H0	Écart-type sous H0	Score moyen
RF MIXTE	135	33675	26257	1052.23	249.44
SSM	253	41791	49208	1052.23	165.18

Wilcoxon Two-Sample Test	
Statistic	33675
Normal Approximation	
Z	7.0488
One-Sided Pr > Z	<.0001
Two-Sided Pr > Z	<.0001
t Approximation	
One-Sided Pr > Z	<.0001
Two-Sided Pr > Z	<.0001
Z includes a continuity correction of 0.5.	

Sur les deux secteurs, que le test soit bilatéral ou unilatéral, il n'y a aucune raison d'accepter H0. Au vu de ces résultats les établissements « RF et MIXTE » et « SSM » ont des comportements différents.

Tests non paramétriques de Kolmogorov Smirnov :

Le test de Wilcoxon est un test de localisation. Si F et G ont la même localisation mais diffère dans leur dispersion, le test s'applique toujours mais il n'est pas assez puissant. Un recours possible est le test de Kolmogorov Smirnov.

F_m: fonction de répartition expérimentale associée à X₁, . . . ,X_m.

G_n: fonction de répartition expérimentale associée à Y₁, . . . ,Y_n.

F_m est un estimateur fonctionnel de F qui converge lorsque $m \rightarrow \infty$.

G_n est un estimateur fonctionnel de G qui converge lorsque $n \rightarrow \infty$.

On cherche à voir si $F < G$ au sens stochastique, si et seulement si $F(x) \geq G(x)$ quelque soit x. Comme pour le test de Wilcoxon, la statistique de test est une statistique de rangs et le test ne fonctionne que pour au maximum deux catégories différentes



Secteur DAF :

Kolmogorov-Smirnov Test pour la variable INDICE3 Classé(e) par variable FILTRE_NAT_ACT2			
FILTRE_NAT_ACT2	Nb	EDF au maximum	Écart de la moyenne au maximum
RF MIXTE	292	0.21	-3.28
SSM	617	0.49	2.26
Total	909	0.40	
Maximum Deviation Occurred at Observation 567			
Value of INDICE3 at Maximum = 0.831602			

Kolmogorov-Smirnov Two-Sample Test (Asymptotic)			
KS	0.13	D	0.28
KSa	3.99	Pr > KSa	<.0001

Secteur QQN :

Kolmogorov-Smirnov Test pour la variable INDICE3 Classé(e) par variable FILTRE_NAT_ACT2			
FILTRE_NAT_ACT2	Nb	EDF au maximum	Écart de la moyenne au maximum
RF MIXTE	135	0.19	-2.64
SSM	253	0.54	1.93
Total	388	0.42	
Maximum Deviation Occurred at Observation 321			
Value of INDICE3 at Maximum = 0.857035			

Kolmogorov-Smirnov Two-Sample Test (Asymptotic)			
KS	0.16	D	0.348
KSa	3.27	Pr > KSa	<.0001

Sur les deux secteurs, Le seuil asymptotique du test est inférieur à 5%, par conséquent le test rejette encore une fois H0.

Les tests non paramétriques amènent à conclure que les établissements « RF et Mixte » n'ont pas la même distribution et la même localisation que les établissements « SSM » par rapport à l'indice de modulation.

Cependant pour chaque catégorisation des établissements, il existe un risque à caractère discriminatoire de la catégorisation en question par rapport à l'indice de modulation. A titre de comparaison, la partie suivante concerne la relation entre l'indice de modulation et des catégories d'établissement selon leur Chiffre d'affaires.

3 Analyse de la variance entre l'indice de modulation et des catégories d'établissements selon le chiffre d'affaires

7.3 Les données

Les établissements sont classés en plusieurs catégories selon leur chiffre d'affaires (CA). Les tableaux suivants présentent les catégories, les effectifs et les moyennes des indices de modulation de chaque catégorie sur les deux secteurs.

Secteur DAF :

Catégorie	Effectif	Moyenne
01 CA < 0,7 M euros	76	0,77
02 CA entre 0,7 et 1,4M euros	206	0,83
03 CA entre 1,4 et 2,4M euros	253	0,89
04 CA entre 2,4 et 4,5M euros	245	0,92
05 CA entre 4,5 et 8,5M euros	181	0,98
06 CA > 8,5M euros	115	1,10

Secteur OQN :

Catégorie	Effectif	Moyenne
01 CA < 0,7 M euros	43	0,66
02 CA entre 0,7 et 1,2 M euros	66	0,80
03 CA entre 1,2 et 2,3 M euros	104	0,84
04 CA entre 2,3 et 4 M euros	112	0,98
05 CA entre 4 et 6 M euros	67	1,06
06 CA > 6 M euros	38	1,16

Ces tableaux descriptifs montrent dans un premier temps que les gros établissements ont tendance à être sur-dotés alors que les petits établissements ont tendance à être sous dotés.

7.4 Analyse de la variance

Cette partie va porter sur la comparaison des statistiques de test entre l'analyse de la relation entre l'indice de modulation et la catégorie de l'établissement RF/SSM/MIXTE et l'analyse de la relation entre l'indice de modulation et la catégorie de l'établissement suivant son chiffre d'affaires. Par

conséquent les résultats sur l'analyse de la relation entre l'indice de modulation et la catégorie de l'établissement suivant son chiffre d'affaires ne seront pas forcément détaillés.

Sur les deux secteurs, le test ANOVA rejette l'hypothèse que les moyennes des indices de modulation des catégories de chiffre d'affaires sont significativement identiques. Par conséquent le chiffre d'affaires est aussi discriminant que la catégorie autorisations. Le tableau suivant présente les principaux indicateurs du test ANOVA sur les deux analyses.

Secteur		Indice de Modulation et Catégorisation RF/SSM/MIXTE		Indice de Modulation et Catégorisation suivant le CA	
		DAF	OQN	DAF	OQN
TEST ANOVA	H0	Rejet de H0	Rejet de H0	Rejet de H0	Rejet de H0
	Intensité de Liaison	3%	10%	5%	17%

L'intensité de liaison montre sur les deux secteurs que la catégorisation suivant le chiffre d'affaires est plus discriminante que la catégorisation RF/SSM/MIXTE.

Le test de Scheffe sur les deux secteurs montre aussi que des regroupements peuvent être faits au sein des catégories suivant le chiffre d'affaires. En résumé le test indique que

- les catégories 01 et 02 (établissements au faible chiffre d'affaires) ont des moyennes significativement différentes des catégories 03 et 04 (établissements au chiffre d'affaires moyen) et des catégories 05 et 06 (établissements au fort chiffre d'affaires).
- les catégories 03 et 04 ont des moyennes significativement différentes des catégories 01 et 02 et des catégories 05 et 06.
- les catégories 05 et 06 ont des moyennes significativement différentes des catégories 01 et 02 et des catégories 03 et 04.

Cependant, l'indice de modulation ne suit toujours pas une loi normale, ce qui implique le recours à l'utilisation de tests non paramétriques.

7.5 Analyse non paramétrique

Pour le test de Kruskal-Wallis les six catégories sont conservées pour chacun des secteurs. En revanche pour les tests de Wilcoxon et de Kolmogorov-Smirnov, tests qui nécessitent au maximum deux catégories, les catégories 01 et 02 sont regroupées dans une catégorie dit « établissements à faible CA » et les catégories 05 et 06 sont regroupées dans une catégorie dit « établissements à fort CA ». Les catégories 03 et 04 qui rassemblent les établissements, avec un chiffre d'affaires moyens, sont exclues des tests de Wilcoxon et de Kolmogorov-Smirnov, l'objectif étant de montrer que le comportement des établissements à faible CA s'oppose au comportement des établissements à fort CA.

Le tableau suivant présente les principaux indicateurs des tests non paramétriques sur les deux analyses :

Secteur		Indice de Modulation et Catégorisation RF/SSM/MIXTE		Indice de Modulation et Catégorisation suivant le CA	
		DAF	OQN	DAF	OQN
TEST Kruskal-Wallis	H0	Rejet de H0	Rejet de H0	Rejet de H0	Rejet de H0
TEST Wilcoxon	H0	Rejet de H0	Rejet de H0	Rejet de H0	Rejet de H0
TEST Kolmogorov	H0	Rejet de H0	Rejet de H0	Rejet de H0	Rejet de H0

Les tests montrent le caractère discriminant de chaque catégorisation.

Les tests amènent à conclure que la catégorisation suivant le CA est plus discriminante que la catégorisation RF/SSM/MIXTE.

4 Etude du lien entre le CA et les autorisations

Le paragraphe suivant a pour objectif d'étudier une éventuelle relation entre les catégories de CA et les catégories d'autorisation. Une première solution est de tester l'indépendance entre ces deux types de catégories.

Le test d'indépendance du chi-deux permet de contrôler l'indépendance de deux types de catégories dans une même population donnée.

Test d'indépendance du Chi deux

Soit 2 variables X et Y, les valeurs possibles de X sont réparties en l classes A_1, \dots, A_l , celles de Y en c classes B_1, \dots, B_c . Il s'agit alors de tester si les variables X et Y sont indépendants. La statistique du chi deux est une mesure des écarts entre les effectifs observés et les effectifs théoriques, autrement dit l'écart entre la situation observée et la situation d'indépendance.

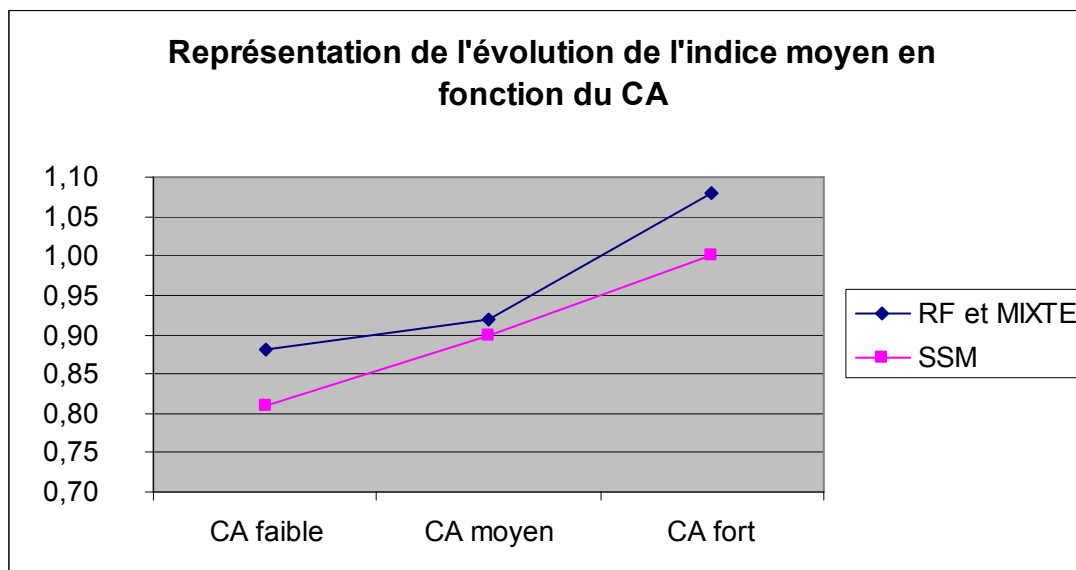
Les résultats du test sont les suivants.

Statistique	DF	Valeur	Proba.
Khi-2	2	291.8896	<.0001

L'hypothèse d'indépendance est largement rejetée. Il existe donc un lien entre le CA et les catégories d'autorisation comme le montre les tableaux et graphiques suivants:

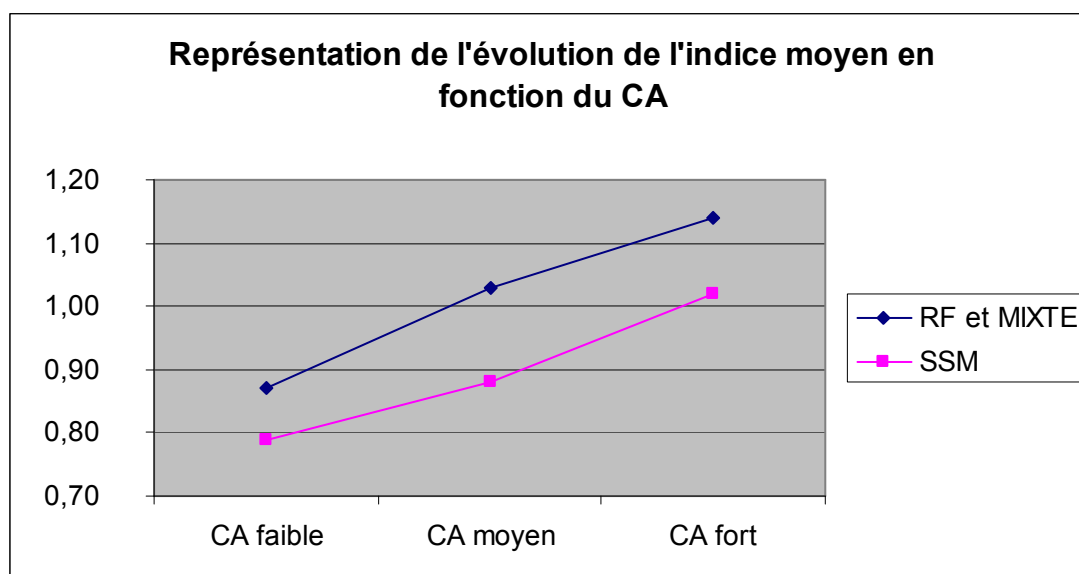
Secteur DAF :

	CA faible		CA moyen		CA fort	
	Effectif	Indice moyen	Effectif	Indice moyen	Effectif	Indice moyen
RF-Mixte	8	0,88	95	0,92	189	1,08
SSM	215	0,81	328	0,9	74	1
	223		423		263	



Secteur OQN :

	CA faible		CA moyen		CA fort	
	effectif	indice moyen	effectif	indice moyen	effectif	indice moyen
RF-Mixte	9	0,87	45	1,03	81	1,14
SSM	80	0,79	151	0,88	22	1,02
	89		196		103	



Les établissements avec des chiffres d'affaires importants sont des établissements plutôt RF/MIXTE, alors que les établissements avec des chiffres d'affaires faibles sont des établissements plutôt SSM. Les établissements avec CA moyen ont une valeur moyenne de l'indice semblable qu'ils soient RF-MIXTE ou SSM. Les courbes montrent assez nettement sur les deux secteurs que plus les établissements RF et MIXTE comme les établissements SSM ont un CA élevé plus ils ont tendance à être sur-dotés.

Une éventuelle explication de la sur dotation des établissements RF et mixtes est peut être dû au fait que ces établissements sont en général les établissements avec les plus grands chiffres d'affaires. Or le modèle a tendance à surdoter les établissements avec des chiffres d'affaires importants.



ANNEXE 4 : **Analyse de la Variance sur le degré de diversité de l'activité**

L'objectif est de voir s'il existe un lien entre l'indice de modulation de l'établissement et la diversité ou non de l'établissement dans un ou quelques GMD. Les établissements pédiatriques seront au préalable isolés.

Sur le secteur DAF, la question sur les établissements MECS, Hôpitaux Locaux et Pouponnières se pose aussi. L'activité de ces établissements a priori atypiques est-elle bien prise en compte dans le modèle IVA ?

L'analyse de la variance permettra dans un premier temps d'identifier si les MECS, Hôpitaux Locaux et Pouponnières sont une catégorie à part sur le secteur DAF. Dans un deuxième temps, elle permettra de dégager un lien entre l'indice de modulation et la diversité des établissements sur les deux secteurs.

1 Les MECS, Hôpitaux Locaux et Pouponnières sur le secteur DAF

1.1 Les données

Les établissements sont classés en trois catégories :

- Les Mecs et Pouponnières (MECS)
- Les Hôpitaux Locaux (HL)
- Les autres établissements

Les effectifs de chaque catégorie sont respectivement les suivants :

	Fréquence	Pourcentage
AUTRES	819	76.12
HL	228	21.19
MECS POUPO	29	2.70

La catégorie MECS et Pouponnières ne comprend que 29 établissements. L'effectif de cette catégorie est faible pour l'analyse de la variance, où en pratique il est commode d'avoir au moins 100 individus pour chaque tranche. C'est pour cette raison que les établissements dit « Hôpitaux Locaux » et « Mecs et Pouponnières » sont regroupés en une seule catégorie.



Les effectifs sont alors :

	Fréquence	Pourcentage
AUTRES	819	76.12
HL, MECS et POUPONNIERES	257	23.88

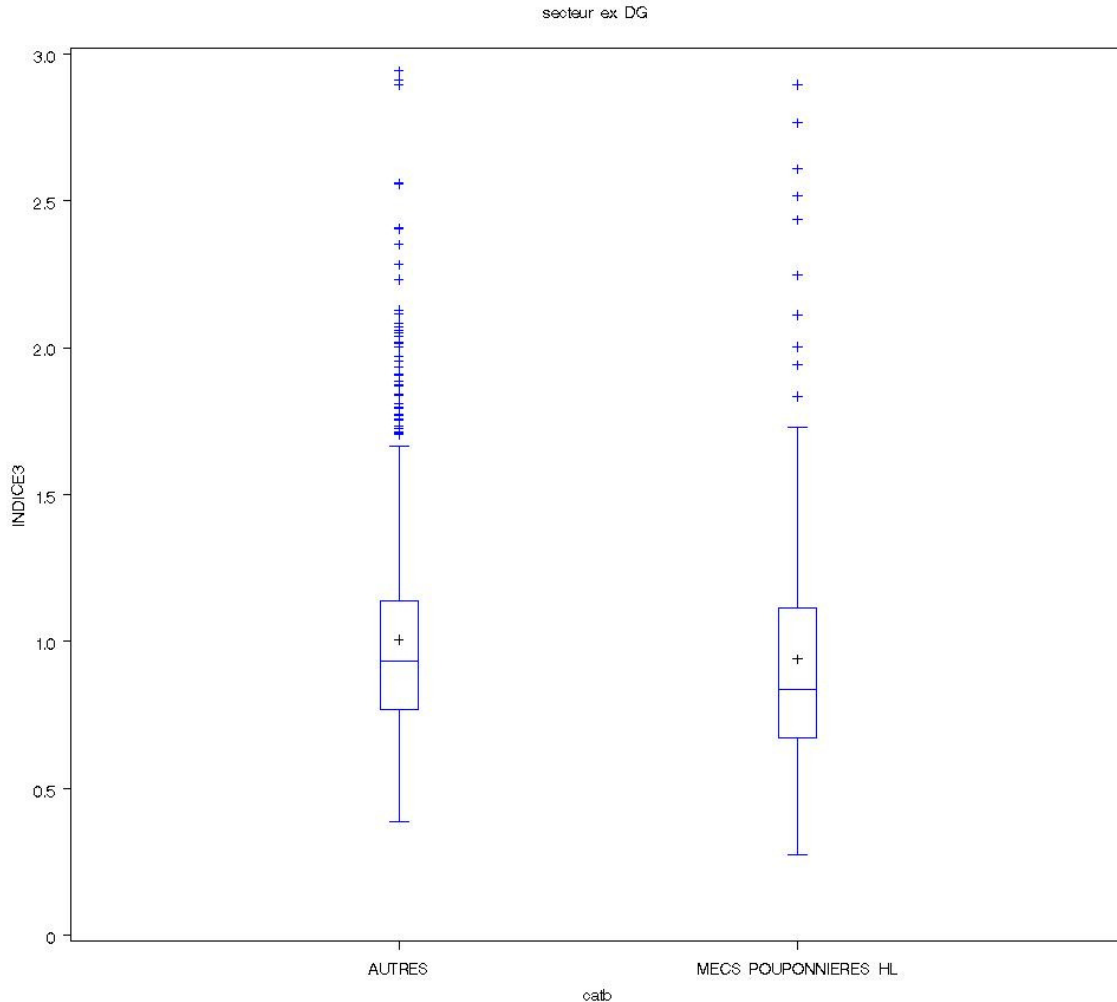
1.2 Analyse descriptive

Etant donné que l'analyse de la variance est basée sur une comparaison des moyennes et des variances, l'analyse descriptive consiste à étudier les moyennes et écarts type empiriques de l'indice de modulation selon les catégories d'établissement sur le secteur DAF.

Il s'agit ici plus particulièrement de moyenne pondérée de l'indice de modulation par le chiffre d'affaire théorique, afin de prendre en compte le poids de chaque établissement vis-à-vis de l'importance de sa masse financière, qui en théorie lui sera redistribué. Donc le tableau ci-dessous présente les statistiques sur l'indice moyen de modulation par catégorie d'établissement.

Analyses Variable : Coefficient		
	Effectif	Moyenne
AUTRES	819	1,01
HL, MECS et POUPONNIERES	257	0,94

Ces tendances sont confirmées par le tracé de la boîte à moustache



La catégorie « autres » présentent plus d'établissements sur dotés que la catégorie « HL, MECS et POUPONNIERES ».

1.3 Analyse Paramétrique

L'étude d'une liaison entre une variable numérique et ordinaire se fait par une analyse de la variance. L'hypothèse nécessaire aux tests paramétriques et que la variable numérique doit suivre une loi normale et l'hypothèse souhaitable aux tests paramétriques est l'égalité des variances de chaque modalité de la variable ordinaire (test d'homoscédasticité).

Si on considère ces hypothèses vraies, il est alors possible d'appliquer les tests paramétriques, c'est-à-dire de tester si les deux variables sont indépendantes, autrement dit de tester si les moyennes de la variable numérique de chaque modalité de la variable nominale sont égales.

La procédure utilisée est le test ANOVA, qui est une généralisation du test de Student quand la variable nominale a plus de 2 modalités.



L'hypothèse nulle du test est : $H_0 \Rightarrow$ Moyenne RF = Moyenne SSM = Moyenne Mixte.

Le test ANOVA, comme pour l'analyse descriptive, se pratique sur les moyennes pondérées, afin de prendre le poids financier de l'établissement.

Test ANOVA

Les résultats sont les suivants :

Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	1	0,8454100	0,8454100	6,05	0,0141
Error	1074	150,1140987	0,1397710		
Corrected Total	1075	150,9595087			

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent au moins une des moyennes est différente des autres.

L'intensité de la liaison entre la variable numérique et ordinale est communément mesurée par l'éta carré.

Eta carré = (Somme des carrés interclasse/Somme des carrés totaux) = 0.005

Cet indicateur est comparable au R carré d'une régression. Par conséquent on peut remarquer, que l'intensité est très faible (moins de 1%).

Le test d'ANOVA ne dit pas quelles moyennes diffèrent significativement. Pour le savoir on recourt au test de Scheffe (le plus conservateur).

Les moyennes avec la même lettre ne sont pas très différentes.		
Scheffe Groupement	Moyenne	Catégorie
A	1,01	AUTRES
B	0,94	HL MECS POUPONNIERES

Le Principe est simple. Les Modalités de la variable ordinale qui n'ont pas la même lettre ont des moyennes significativement différentes. Par conséquent, la Moyenne de la catégorie « Autres » est significativement différente de la Moyenne de la catégorie « HL MECS POUPONNIERES ».

En particulier les Meecs, HL et Pouponnières sont significativement sous dotés en moyenne.

Vérification des hypothèses de départ

Test d'homoscédasticité

Le test choisi est celui de Levene qui a pour hypothèse nulle :



H0=> Variance « Autres » = Variance « HL MECS POUPONNIERES ».

Levene's Test for Homogeneity of Valeur du Point Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
FILTRE_NAT_ACT	1	0,4592	0,4592	3,58	0,0588
Error	1074	137,8	0,1283		

La P-Value étant supérieure à 0.05, l'hypothèse nulle est acceptée. Par conséquent le test d'homoscédasticité est concluant.

Test de normalité de la valeur du point

Le test choisi est celui de Shapiro Wilk, puisque $n \leq 2000$. L'hypothèse nulle est :

H0=> L'indice de modulation suit une Loi normale

Tests de normalité				
Test	Statistique		p Value	
Shapiro-Wilk	W	0.88	Pr < W	<0.0001

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent la valeur du point ne suit pas une loi normale.

Comme l'hypothèse de normalité n'est pas vérifiée, il est nécessaire de passer à une analyse non paramétrique.

1.4 Analyse non paramétrique

L'indice de modulation ne suivant pas une loi normale, des tests paramétriques peuvent être utilisés. Le test de Kruskal Wallis permet de tester l'égalité des moyennes entre une variable nominales qui a au moins 2 modalités et une variable numérique. Le calcul de la valeur de test s'appuie sur la valeur des rangs de chaque observation.

L'hypothèse nulle est donc : H0 : Egalité des moyennes de chaque modalité ou population



Kruskal-Wallis Test	
Chi-Square	15,57
DF	1
Pr > Chi-Square	<.0001

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent les populations n'ont pas des comportements identiques.

Dans la partie suivante, les établissements de la catégorie « HL MECS POUPONNIERES » pour le secteur DAF et les établissements MECS pour le secteur OQN sont retirés de l'analyse pour plusieurs raisons :

- La première partie a monté qu'il formait une catégorie à part.
- Les MECS faisant partie des établissements pédiatriques, cette dernière catégorie risque d'avoir un résultat biaisé.

2 La diversité d'activité des établissements

2.1 Les données

Les établissements qui ont plus de 50% de leurs journées dans la seule CMC numérotée 60 (CMC dit « autres ») se voient attribués comme valeur pour la variable GMD_50, la moyenne de cette dernière.

Les établissements sont classés en cinq catégories :

- Les établissements pédiatriques (plus de 50% de leur journée sont consacrés à la pédiatrie).
- Les établissements avec très peu de diversité (1 ou 2 GMD constituant plus de 50% des journées)
- Les établissements avec peu de diversité (3 GMD constituant plus de 50% des journées)
- Les établissements avec de la diversité (4 à 5 GMD constituant plus de 50% des journées)
- Les établissements avec beaucoup de diversité (Plus de 5 GMD constituant plus de 50% des journées)



2.2 Analyse descriptive

Secteur DAF

Les effectifs de chaque catégorie sont respectivement les suivants :

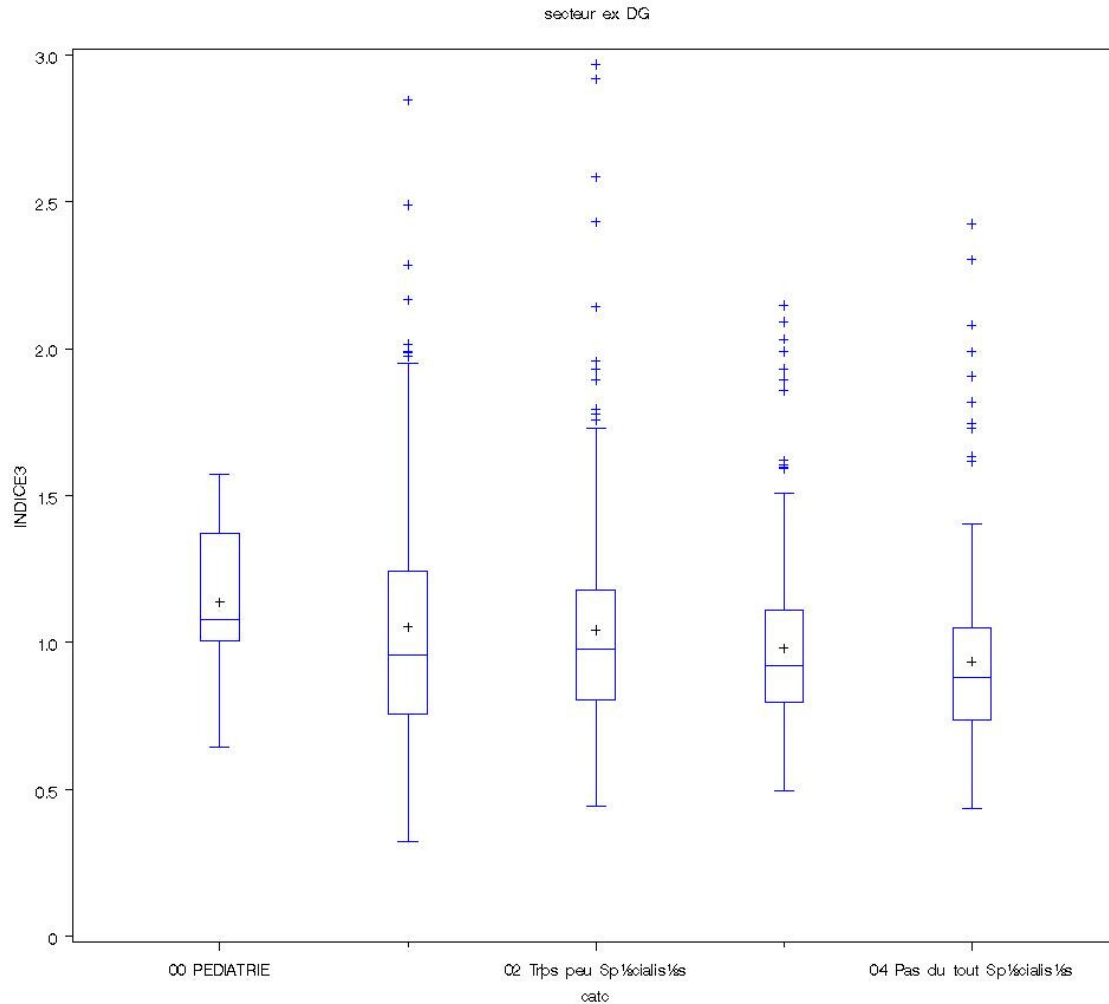
	Fréquence	Pourcentage
00 PEDIATRIE	22	2.69
01 Très peu diversifiés	173	21.12
02 Peu diversifiés	240	29.30
03 Diversifiés	200	24.42
04 Très Diversifiés	184	22.47

La catégorie Pédiatrie ne comprend que 22 établissements. Aussi les résultats sur cette catégorie devront être regardés avec précaution.

Le tableau ci-dessous indique que les établissements non diversifiés sont en moyenne relativement proche et sur dotés. A l'inverse les établissements diversifiés sont en moyenne sous dotés.

Analyses Variable : Coefficient		
	Effectif	Moyenne
00 PEDIATRIE	22	1.14
01 Très peu diversifiés	173	1.05
02 Peu diversifiés	240	1.04
03 Diversifiés	200	0.98
04 Très Diversifiés	184	0.93

Ces tendances sont confirmées par le tracé de la boîte à moustache



Les boîtes à moustaches sont de gauche à droite la catégorie « pédiatrie » et ensuite des catégories pas diversifiées à très diversifiées. Le graphique montre que les très peu diversifiés apparaissent les plus sur-dotés même si en moyenne l'indice moyen est proche de 1. Dans cette catégorie quelques établissements très sous dotés tirent la moyenne vers le bas.

Secteur OQN

Les établissements sont comme pour le secteur DAF classés en cinq catégories :



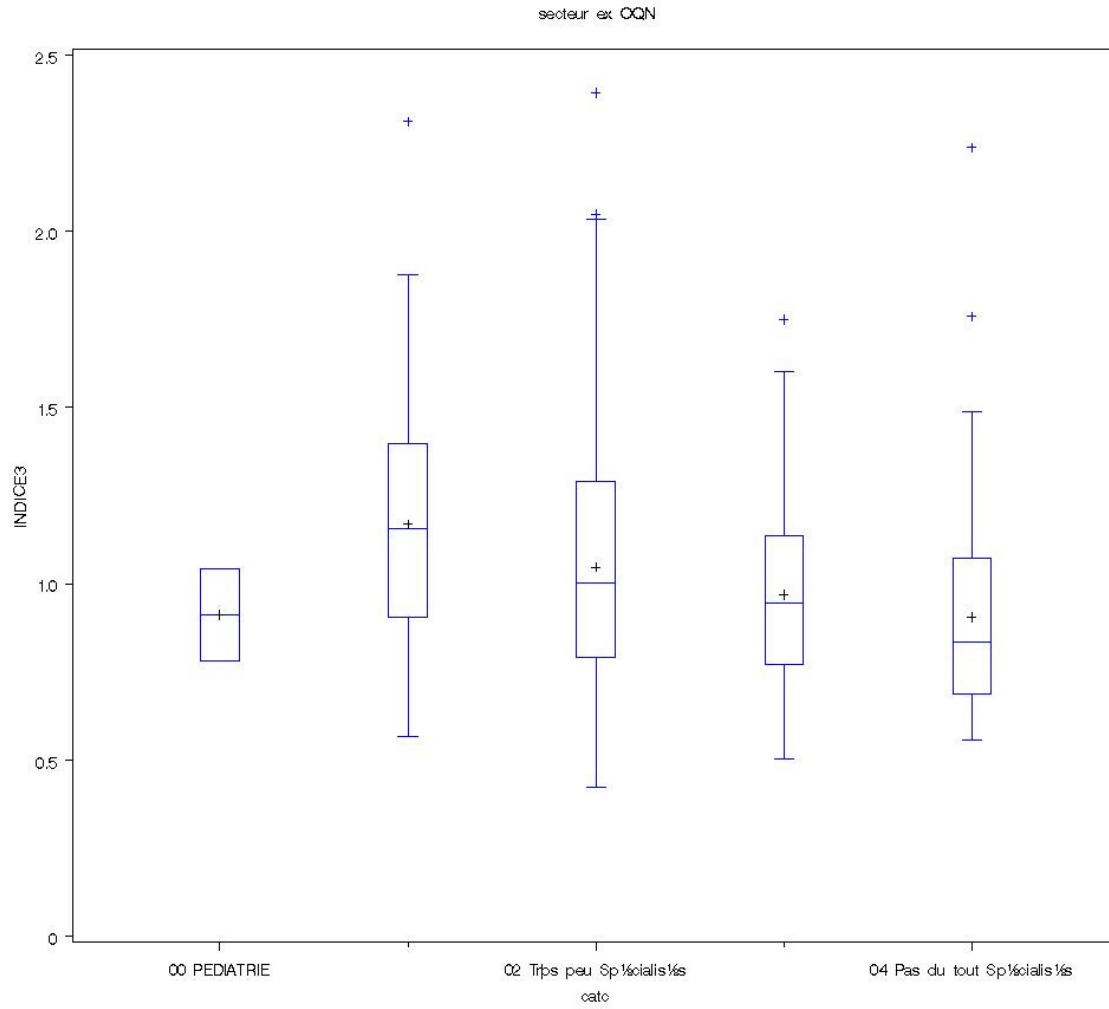
	Fréquence	Pourcentage
00 PEDIATRIE	2	0,5
01 Très peu diversifiés	74	18,41
02 Peu diversifiés	115	28,61
03 Diversifiés	135	33,58
04 Très Diversifiés	76	18,91

Etant donné que la catégorie pédiatrie hors MECS est composée de 2 établissements, cette catégorie est exclue de l'analyse, faute d'effectifs suffisants.

Comme pour le secteur DAF, il y a une opposition entre les établissements selon leur degré de diversité de l'activité

Analyses Variable : Coefficient		
	Effectif	Moyenne
01 Très peu diversifiés	74	1.17
02 Peu diversifiés	115	1.05
03 Diversifiés	135	0.97
04 Très Diversifiés	76	0.9

Ces tendances sont confirmées par le tracé de la boîte à moustache





2.3 Analyse de la Variance

Secteur DAF

Analyse Paramétrique

La procédure utilisée est toujours le test ANOVA.

Test ANOVA

Les résultats sont les suivants :

Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	4	2,12	0,53	4,33	0,0018
Error	814	99,97	0,12		
Corrected Total	818	102,10			

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent au moins une des moyennes est différente des autres.

L'intensité de la liaison entre la variable numérique et ordinale, mesurée à 2% reste très faible.

Afin de savoir quelles moyennes diffèrent significativement, on recourt au test de Scheffe.

Les moyennes avec la même lettre ne sont pas très différentes,			
Scheffe Groupement	Moyenne	Effectif	
A	1.14	22	00 PEDIATRIE
A B	1.05	173	01 Très peu diversifiés
A B	1.04	239	02 Peu diversifiés
A B	0.98	200	03 Diversifiés
B	0.93	184	04 Très Diversifiés

Seules les catégories « Pédiatrie » et « Très diversifiés » ont des moyennes significativement différentes. Toutefois ce résultat est à prendre avec précaution pour la catégorie « Pédiatrie » vu la faiblesse de ses effectifs. Les autres catégories ont des moyennes non significativement différentes.



Vérification des hypothèses de départ

Test d'homoscédasticité

Levene's Test for Homogeneity of Valeur du Point Variance					
ANOVA of Squared Deviations from Group Means					
Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
FILTRE_NAT_ACT	4	0,8242	0,2061	2,06	0,0842
Error	814	81,4222	0,1		

La P-Value étant supérieure à 0.05, l'hypothèse nulle est acceptée. Par conséquent le test d'homoscédasticité est concluant.

Test de normalité de la valeur du point

Tests de normalité				
Test	Statistique		p Value	
Shapiro-Wilk	W	0,88	Pr < W	<0,0001

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent la valeur du point ne suit pas une loi normale.

Comme l'hypothèse de normalité n'est pas vérifiée, il faut passer à une analyse non paramétrique.

Analyse non paramétrique

Kruskal-Wallis Test	
Chi-Square	24.34
DF	4
Pr > Chi-Square	<.0001

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent les populations n'ont pas des comportements identiques. Ce résultat est peut être dû à la présence de la catégorie « Pédiatrie » qui a une moyenne significativement différente et des effectifs faibles.



On relance le test de Kruskal-Wallis sans les établissements dits « Pédiatrique »

Kruskal-Wallis Test	
Chi-Square	16,25
DF	3
Pr > Chi-Square	0,0010

La valeur de la statistique de test a largement baissé. Par conséquent on peut largement mettre en doute de la présence d'une corrélation entre la diversité d'activité des établissements et l'indice de modulation sur le secteur DAF.

Secteur OQN

Analyse Paramétrique

Test ANOVA

Les résultats sont les suivants :

Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	3	3.03	1.01	10.86	<.0001
Error	396	36.85	0.093		
Corrected Total	399	39.88			

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent au moins une des moyennes est différente des autres.

L'intensité de la liaison entre la variable numérique et ordinaire est aux alentours de 16%.



Afin de savoir quelles moyennes diffèrent significativement, on recourt au test de Scheffe.

Les moyennes avec la même lettre ne sont pas très différentes.			
Scheffe Groupement	Moyenne	Nb	Catégorie
A	1.17	74	01 Très peu diversifiés
A B	1.05	115	02 Peu diversifiés
B C	0.97	135	03 Diversifiés
C	0.9	76	04 Très Diversifiés

Seules les catégories « Très peu diversifiés » et « Pas du tout diversifiés » ont des moyennes significativement différentes. Les autres catégories ont des moyennes non significativement différentes. Les établissements très peu diversifiés sont significativement sur dotés en moyenne. A l'inverse les établissements très diversifiés sont sous dotés en moyenne.

Vérification des hypothèses de départ

Test d'homoscédasticité

Levene's Test for Homogeneity of Valeur du Point Variance					
ANOVA of Squared Deviations from Group Means					
Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
FILTRE_NAT_ACT	3	0.1902	0.0634	2.03	0.1090
Error	396	12.3578	0.0312		

La P-Value étant supérieure à 0.05, l'hypothèse nulle est acceptée. Par conséquent le test d'homoscédasticité est concluant.

Test de normalité de la valeur du point

Tests de normalité				
Test	Statistique		p Value	
Shapiro-Wilk	W	0.95	Pr < W	<0.0001

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent la valeur du point ne suit pas une loi normale.



Comme l'hypothèse de normalité n'est pas vérifiée, il faut passer à une analyse non paramétrique pour vérification.

Analyse non paramétrique

Kruskal-Wallis Test	
Chi-Square	32,17
DF	3
Pr > Chi-Square	<.0001

La P-Value étant inférieure à 0.05, l'hypothèse nulle est rejetée. Par conséquent les populations n'ont pas des comportements identiques.

Sur le secteur DAF, les tests paramétriques et non paramétriques infirment l'hypothèse que les établissements ont des indices de modulations moyens identiques suivant leur degré de diversité de l'activité. Il convient de noter que l'intensité de liaison entre le degré de diversité de l'activité et l'indice de modulation est extrêmement faible.

Sur le secteur OQN, les tests paramétriques et non paramétriques confirment l'hypothèse que les établissements ont des indices de modulations moyens différents suivant leur degré de diversité de l'activité. Cependant l'intensité de liaison reste trop faible, pour en déduire que le degré diversité de l'activité influe fortement les variations de l'indice de modulation.

La catégorie pédiatrique hors MECS n'a pas pu être prise en compte dans l'analyse, en raison de la faiblesse de ses effectifs.