

Annexe :

- *Annexe 1 : Théorie sur les concepts et modèles statistiques utilisés*

Dans le cadre d'études scientifiques ou plus globalement universitaires, J. Creswell, dans son ouvrage « Research Design » (2003), décrit la démarche à suivre pour établir et définir une recherche. Or une fois le squelette défini, il insiste encore sur 4 sujets à préciser clairement lorsque le modèle statistique a été identifié :

1. Les définitions des variables (qualitatives et quantitatives)
2. Les délimitations
3. Les limites
4. La significativité

Ce schéma est à suivre tant dans le cadre de la structure d'une recherche scientifique comme ce mémoire, mais aussi dans le cadre d'une analyse statistique inhérente à ce genre de recherche. Dans la partie pratique de ce mémoire, on retrouve chacun des 4 points ci-dessus. Cependant une définition des outils statistiques utilisés est nécessaire à la bonne compréhension de la démarche suivie.

Qualitatif – Quantitatif :

En pratique, on distingue les variables qualitatives et les variables quantitatives.

« Les données quantitatives sont des nombres qui représentent des comptages ou des mesures.

Les données qualitatives (ou catégorielles) peuvent être réparties en différentes catégories qui sont repérées par des caractéristiques non numériques. » (Triola & Triola, 2012)

Les données qualitatives sous-forme catégorielles peuvent aussi comprendre des catégories numériques (exemple des numéros d'identification).

Dans le premier chapitre de leur livre, Triola et Triola vont plus loin en disant que les données quantitatives peuvent être de deux types : discrètes ou continues. Dans le cadre de ce projet, les données quantitatives sont discrètes : c'est-à-dire des données quantitatives avec un nombre fini de valeurs (âge des patients, nombre de jours entre deux séjours, etc.) et non des valeurs continues.

Test d'hypothèse :

Chacun de ces tests utilise la méthode du test d'hypothèse qui suit les étapes globales suivantes :

- Enoncé de l'hypothèse nulle (H_0), de l'hypothèse alternative (H_1) et du niveau de risque acceptable (généralement 5 %)
- Détermination de la statistique de test, de sa distribution sous l'hypothèse nulle et de la région de rejet
- Calcul de la valeur observée de la statistique de test et calcul de la P-valeur
- Rejet ou non de l'hypothèse nulle

(Pigeon & Segers, 2012)

La P-valeur représente la probabilité en valeur absolue d'avoir une valeur supérieure à la statistique observée. Si la probabilité est inférieure au niveau de risque acceptable α (généralement $\alpha=5\%$), cela signifie que la valeur observée se trouve dans la zone de rejet de H_0 . On peut donc conclure au rejet de l'hypothèse nulle.

Régression linéaire :

La régression linéaire a pour but de modéliser la valeur de la moyenne d'une variable dépendante à l'aide d'une variable (modèle simple) ou plusieurs variables (modèle multiple) explicatives qu'elles soient qualitatives ou quantitatives. Cependant, le modèle simple n'est valable que si la variable explicative est quantitative.

Avant de réaliser ces tests statistiques, il est important de tester la corrélation entre la variable dépendante et la ou les variables explicatives. En effet le calcul de corrélation nous renseignera sur le degré exact de corrélation ainsi que sa direction. Cette dernière est définie par le signe de la corrélation (noté R) :

1. Modèle simple

En addition aux renseignements du test de corrélation, la régression linéaire simple « permet de décrire plus précisément comment les deux variable sont liées » (Rousson, 2013, p. 187). En effet la régression linéaire indique comment la variation de la variable explicative (X) explique la variation de la variable dépendante (Y).

Le modèle de régression linéaire simple suit une droite de régression suivant l'équation ci-dessous :

$$Y = \beta_0 + \beta_1 * X$$

où β_0 est appelé la constante (l'ordonnée à l'origine) et β_1 représente la pente de la droite de régression. C'est donc cette équation qui nous permet de décrire comment X explique Y. En effet on l'utilise pour estimer la variable Y en fonction de la grandeur de X.

2. Modèle multiple

Dans la même logique que le modèle simple, la régression linéaire multiple permet « d'exprimer linéairement la valeur moyenne d'une variable dépendante ($E[Y_i]$) par des variables explicatives ($X_1; \dots; X_p$). L'équation générale est donnée par :

$$E[Y_i] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p, \quad i = 1, \dots, n$$

Ou, de façon équivalente,

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon_i, \quad i = 1, \dots, n, \gg$$

(Pigeon & Segers, 2012, pp. 105-106). Le terme ϵ_i représente l'erreur ou le résidu. C'est la partie de la variabilité totale de Y qui n'a pas pu être expliquée par le reste de l'équation.

L'intérêt du modèle multiple est son aspect généralisation du modèle simple, à plusieurs variables explicatives. Cependant chacune de ces variables a une importance relative dans l'explication voire la prédiction de la variable dépendante.

Il y a donc une étape supplémentaire dans l'analyse statistique via la régression multiple que l'on n'a pas à effectuer dans la régression simple. Il va falloir établir le meilleur modèle explicatif de la variable dépendante en sélectionnant les variables explicatives.

De manière simplifiée, cette sélection se fait sur 2 critères principaux : la P-valeur de chaque coefficient (β_p) et le R^2 du modèle complet. La P-valeur indique le résultat d'un test d'hypothèse automatiquement généré pour chaque coefficient. Ce test d'hypothèse a pour but de vérifier que chaque β_p est significativement différent de 0, avec comme hypothèse nulle $\beta_p=0$. Or si la P-valeur est inférieure au niveau de risque acceptable ($\alpha=0,05$), cela signifie que le coefficient de la variable explicative étudiée est suffisamment significatif dans l'explication de la variabilité de Y (car rejet de H_0).

Pour ce qui est du R^2 , celui-ci représente la proportion de la variabilité de Y qui est expliquée par la variabilité des variables explicatives. Il représente donc l'ajustement du modèle. L'ajustement du modèle en fonction du nombre et du choix des variables explicatives doit donc toujours être un équilibre entre le R^2 du modèle et la pertinence des variables présentes. Plusieurs stratégies peuvent être suivies dans le choix du modèle. Les 2 principales sont « avant » ou « arrière » ; la première favorisant l'ajout d'une variable après l'autre, la deuxième démarrant avec toutes les variables disponibles pour en enlever petit à petit.

Régression logistique

La régression logistique ressemble à la régression linéaire mais en présentant moins de contraintes en ce qui concerne des variables. On peut plus facilement mélanger des variables quantitatives et qualitatives dans la construction d'une régression logistique. La première différence se marque par le fait que la régression logistique est généralement utilisée lorsque la variable dépendante Y est dichotomique. Une variable dichotomique est une variable catégorielle binaire, c'est-à-dire contenant 2 catégories possibles que l'on peut remplacer par 0 (exemple : non réadmis) ou 1 (exemple : réadmis). De plus « contrairement à la régression multiple et l'analyse discriminante, cette technique n'exige pas une distribution normale des prédicteurs ni l'homogénéité des variances » (Desjardins, 2005, p. 35). L'équation utilisée dans ce test est la suivante :

$$\ln \left(\frac{\Pr(Y = 1|X)}{1 - \Pr(Y = 1|X)} \right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p. \quad (\text{Pigeon \& Segers, 2012})$$

Cette équation représente dans sa partie gauche le logarithme népérien du rapport de la probabilité de faire partie d'une catégorie, par la probabilité de ne pas en faire partie. Et dans sa partie droite, appelée score, on retrouve toutes les variables explicatives avec leur coefficient ainsi que la constante β_0 .

Pour cette modélisation également, plusieurs stratégies peuvent être envisagées pour sélectionner le modèle final. La stratégie la plus courante, dans ce cas, est proche de celle dite « arrière » dans la régression multiple. Il s'agit de commencer par ajouter toutes les variables explicatives possibles en transformant les variables catégorielles en facteurs (et non plus en chiffres) pour ne pas être confondu avec des variables quantitatives. Ensuite, les variables qui ne sont pas liées suffisamment dans l'explication de la variable dépendante seront enlevées pas à pas (procédure automatisée dans certains programmes).

La dernière étape d'une régression logistique est le calcul du « score » d'un individu. Le score est la partie droite de l'équation de la régression. Il se calcule par la multiplication de chaque coefficient obtenu dans la régression par les données spécifiques de l'individu relatives à chaque variable. Pour obtenir la probabilité d'appartenir à la catégorie ($Y = 1$) en fonction du score de l'individu, il suffit ensuite de résoudre l'équation suivante : $P(Y = 1) = \frac{e^{\text{score}}}{1 + e^{\text{score}}}$

F-test ou test de Fisher

Le test de Fisher (ou F-test) est un test d'hypothèse sur l'égalité des variances de deux échantillons. Il est utilisé dans ce mémoire comme test préliminaire au test d'égalité des moyennes qui fait l'objet du prochain paragraphe. L'hypothèse nulle est l'égalité des variances, l'hypothèse alternative étant leur différence. On utilise la formule $F = \frac{\sigma(x)^2}{\sigma(y)^2}$ pour calculer la F_{obs} qui ne doit pas dépasser une valeur théorique trouvée dans la table de Fisher. Il est important que la variance la plus grande soit au numérateur et la plus petite au dénominateur.

En réalisant ce test, on veut prouver l'égalité des variances, à un seuil α deux fois plus élevé que le seuil fixé pour le test d'égalité des espérances. L'égalité des variances dans les deux distributions constitue une condition nécessaire pour pouvoir effectuer le test d'égalité des espérances, décrit ci-après. On cherche donc le non-rejet de l'hypothèse nulle.

Test d'égalité des espérances avec et sans variance égales

L'exactitude de l'analyse sur l'égalité des espérances (ou moyennes) nécessite deux conditions théoriques :

- La distribution normale des deux échantillons
- L'égalité des variances des deux échantillons

Cependant, le test est défini comme assez robuste que pour passer outre ces conditions lorsque les échantillons sont assez « grands » ; c'est-à-dire si les « n » des échantillons sont plus grands que 30 (Dress, 2007) en théorie.

Les tests sur les espérances avec et sans variances égales réalisés sont effectués via un test de Student. Celui-ci peut être défini comme « un test paramétrique qui compare, soit la moyenne observée d'un échantillon statistique à une valeur fixée, soit les moyennes observées de deux échantillons statistiques » (Dress, 2007, p. 167). L'hypothèse nulle de ce test est l'égalité des espérances, tandis que l'hypothèse alternative est leur différence. La valeur du test est calculée par les deux équations suivantes : $t = \frac{|m(x)-m(y)|}{s \sqrt{\frac{1}{n(x)} + \frac{1}{n(y)}}}$ et $s^2 = \frac{(n(x)-1)s^2(x) + (n(y)-1)s^2(y)}{n(x)+n(y)-2}$ (Dress, 2007) avec $m(x)$ = la moyenne observée de x ; $s^2(x)$ = la variance observée de x ; et $n(x)$ = la taille de l'échantillon de x. La notation est semblable pour y.

Le $T_{\text{observé}}$ est comparé à la valeur de référence dans la table de Student (le quantile de Student), défini par le nombre de degrés de liberté de l'échantillon ($ddl=n(x)+n(y)-2$) ainsi que le risque (généralement $\alpha=5\%$).

- Annexe 2 : Lien DRG's des séjours 1 avec DRG's des séjours 2

Code 261564 comme Séjour 1 avec DRG du S1 en ligne et DRG des séjours qui suivent en colonne :

Nombre de no_sejour	Étiquettes de colonnes															
Étiquettes de lignes	443	445	446	463	465	466	467	468	482	484	500	501	693	813	862	Total général
≤10j																
445							1							1		2
446			1													1
482		1		1		1	13	1	1			1		2		21
501									1							1
11 à 28j																
482							2			1						3
951					1			1								2
29 à 91j																
446									2						2	4
482							4	6				2	1		3	16
484									1							1
693			1													1
951											1					1
Total général		1	1	1	1	1	20	8	5	1	1	3	1	3	5	53

Code 260481 comme Séjour 1 avec DRG du S1 en ligne et DRG des séjours qui suivent en colonne :

Nombre de no_sejour	Étiquettes de colonnes				
Étiquettes de lignes	441	446	468	500	Total général
≤10j					
446			1		1
11 à 28j					
482				1	1
29 à 91j					
445		1	1		2
Total général	1	1	1	1	4

Codes 261391/261402 comme Séjour 1 avec DRG du S1 en ligne et DRG des S2 en colonne :

Nombre de no_sejour		Ét																			
Étiquettes de lignes	385	441	442	443	445	446	461	463	464	466	467	468	482	693	722	791	813	861	862	951	Total général
≤10j																					
443											1										1
446		1									7			1		1	3	1			14
484										1											1
11 à 28j																					
385	1																				1
445											1										1
446			1			1	1				3	3		4					2		15
466											1										1
468						1															1
501													1								1
951												1									1
29 à 91j																					
446		1	4	2	1	23	12	2	1		3	7	1	17	1			22	1		98
468											2										2
693														1							1
850																		2			2
861			1																		1
951																			1		1
Total général	1	2	6	2	1	25	13	2	1	1	18	11	2	23	1	1	3	1	27	1	142

- *Annexe 3 : Nomenclature des DRG's*

Liste des principaux DRG's (version 28) et leur libellé :

228	=	"228 Cures de hernie inguinale, crurale et ombilicale"
361	=	"361 Greffe cutanée pour affections de la peau et du tissu sous-cutané"
385	=	"385 Autres affections de la peau, du tissu sous-cutané et du sein"
441	=	"441 Interventions majeures sur la vessie"
442	=	"442 Interventions pour aff. malignes des reins et des voies urinaires"
443	=	"443 Interventions pour aff. non malignes des reins et des voies urinaires"
445	=	"445 Autre intervention de la vessie"
446	=	"446 Interventions urétrales et transuretrales"
447	=	"447 Autres interventions chirurg. des reins et des voies urinaires"
461	=	"461 Affections malignes des reins et voies urinaires"
463	=	"463 Infections des reins et voies urinaires"
465	=	"465 Lithiases urinaires et obstructions acquises des voies urinaires supérieures"
466	=	"466 Dysfct et complications d'une greffe ou appareil génito-urinaire"
468	=	"468 Autres diagnostics, signes et symptômes des reins et voies urinaires"
480	=	"480 Interventions majeure sur petit bassin, homme"
481	=	"481 Chirurgie du pénis "
482	=	"482 Prostatectomie transuretrale"
483	=	"483 Interventions sur les testicules"
484	=	"484 Autres interventions sur le système génital masculin"
500	=	"500 Affections malignes d'organes génitaux masculins"
501	=	"501 Autres affections des organes génitaux masculins"
514	=	"514 Interventions restructrices d'organes génitaux féminins"
532	=	"532 Troubles menstruels, autres troubles des organes génitaux féminins"
663	=	"663 Autres anémies et affections du sang et des organes hématopoïétiques"
681	=	"681 Autres interventions pour affections malignes des des organes lymphatiques et hématopoïétiques"
693	=	"693 Chimiothérapie"
720	=	"720 Septicémie et infection disséminée"
721	=	"721 Infections postopératoires, posttraumatiques ou sur matériel"
791	=	"791 Interventions pour autres complications de traitement"
813	=	"813 Autres complications de traitement"
850	=	"850 Interventions avec diagnostics de revalidation, de suites de soins ou d'autre contact avec services de santé"
860	=	"860 Revalidation"
861	=	"861 Signes, symptômes et autres facteurs influençant l'état de santé"
862	=	"862 Autres suites de soins et convalescence"
930	=	"930 Polytraumatisme significatif sans intervention"
950	=	"950 Intervention majeure sans relation avec le diagnostic principal"
951	=	"951 Intervention modérée non liée au diagnostic principal"
952	=	"952 Intervention mineure non liée au diagnostic principal"
999	=	"Autres";

- *Annexe 4 : Taux de réadmissions par intervention*

Codes INAMI chirurgie et taux de réadmission:

code INAMI	<=10j	% 10j	11 à 28j	% 28j	29 à 91j	% 91j	>3mois	non	Total séjours
220113	0	0%	0	0%	1	3%	3	25	29
220124	0	0%	0	0%	0	0%	1	0	1
220231	0	0%	0	0%	0	0%	0	7	7
220242	0	0%	1	100%	0	0%	0	0	1
220264	1	100%	0	0%	0	0%	0	0	1
220356	0	0%	0	0%	0	0%	3	6	9
220360	0	0%	0	0%	0	0%	0	2	2
221023	0	0%	0	0%	0	0%	0	1	1
227603	0	0%	0	0%	0	0%	0	1	1
235185	0	0%	0	0%	0	0%	0	1	1
240461	0	0%	0	0%	0	0%	3	7	10
240483	1	17%	1	17%	0	0%	1	3	6
241102	0	0%	0	0%	0	0%	0	1	1
241124	0	0%	0	0%	0	0%	0	1	1
241161	0	0%	0	0%	0	0%	1	11	12
241323	0	0%	0	0%	0	0%	0	2	2
241872	0	0%	0	0%	0	0%	0	2	2
241883	0	0%	0	0%	0	0%	1	9	10
241894	0	0%	0	0%	0	0%	0	4	4
241905	0	0%	0	0%	0	0%	0	2	2
241916	0	0%	0	0%	0	0%	1	0	1
241920	0	0%	0	0%	0	0%	0	2	2
241942	0	0%	0	0%	0	0%	0	2	2
242701	0	0%	0	0%	0	0%	0	1	1
243062	0	0%	1	100%	0	0%	0	0	1
243202	0	0%	0	0%	0	0%	0	1	1
243283	0	0%	0	0%	0	0%	1	0	1
243622	0	0%	0	0%	0	0%	0	1	1
243644	0	0%	0	0%	2	14%	1	11	14
243762	2	9%	0	0%	0	0%	5	15	22
243784	0	0%	0	0%	0	0%	0	1	1
244160	0	0%	0	0%	0	0%	4	38	42
244451	0	0%	0	0%	0	0%	0	1	1
245615	0	0%	0	0%	0	0%	0	1	1
245851	0	0%	0	0%	0	0%	0	2	2
246094	0	0%	1	50%	0	0%	0	1	2
246131	0	0%	1	50%	0	0%	0	1	2
246223	0	0%	0	0%	0	0%	0	2	2
246514	0	0%	0	0%	0	0%	0	1	1
246923	0	0%	1	20%	0	0%	0	4	5

248640	0	0%	0	0%	1	33%	0	2	3
248986	0	0%	0	0%	0	0%	0	1	1
249233	0	0%	0	0%	0	0%	0	2	2
249244	0	0%	1	11%	1	11%	0	7	9
250180	0	0%	0	0%	0	0%	0	2	2
251300	0	0%	0	0%	0	0%	0	1	1
251742	0	0%	0	0%	0	0%	0	1	1
251764	0	0%	0	0%	0	0%	0	1	1
252523	0	0%	0	0%	0	0%	0	1	1
254925	0	0%	0	0%	0	0%	0	1	1
255021	0	0%	0	0%	0	0%	1	0	1
257740	0	0%	0	0%	0	0%	1	0	1
258823	0	0%	0	0%	0	0%	1	0	1
260061	0	0%	0	0%	0	0%	0	1	1
260105	0	0%	4	9%	9	20%	1	30	44
260142	0	0%	0	0%	0	0%	0	1	1
260175	0	0%	0	0%	0	0%	0	2	2
260186	2	33%	1	17%	1	17%	1	1	6
260201	1	25%	0	0%	1	25%	0	2	4
260223	0	0%	2	50%	1	25%	0	1	4
260245	0	0%	0	0%	2	67%	1	0	3
260260	0	0%	1	20%	1	20%	0	3	5
260293	24	11%	24	11%	31	15%	72	61	212
260304	59	12%	93	19%	99	20%	102	140	493
260315	4	4%	9	8%	15	13%	37	48	113
260326	6	2%	15	6%	49	20%	88	93	251
260400	0	0%	1	13%	1	13%	0	6	8
260481	1	3%	1	3%	2	6%	9	18	31
260540	0	0%	0	0%	4	57%	0	3	7
260584	1	17%	1	17%	0	0%	1	3	6
260606	0	0%	0	0%	1	50%	0	1	2
260621	1	50%	1	50%	0	0%	0	0	2
260643	6	7%	1	1%	0	0%	7	71	85
260665	0	0%	0	0%	0	0%	15	98	113
260724	0	0%	0	0%	0	0%	1	7	8
260735	0	0%	0	0%	0	0%	0	9	9
260746	0	0%	0	0%	1	4%	3	21	25
260750	0	0%	0	0%	0	0%	0	1	1
260761	0	0%	0	0%	0	0%	0	2	2
260772	0	0%	0	0%	0	0%	0	2	2
260783	0	0%	0	0%	0	0%	0	16	16
260794	1	1%	1	1%	0	0%	2	93	97
260805	0	0%	0	0%	0	0%	0	5	5
260853	0	0%	0	0%	0	0%	0	8	8
260864	0	0%	0	0%	0	0%	3	34	37

260875	0	0%	0	0%	0	0%	0	3	3
260886	3	4%	1	1%	1	1%	5	75	85
260890	0	0%	0	0%	0	0%	0	1	1
260901	0	0%	0	0%	0	0%	0	5	5
260923	0	0%	0	0%	0	0%	0	4	4
260934	1	1%	0	0%	3	2%	6	170	180
260945	0	0%	1	4%	2	8%	1	22	26
260956	1	4%	2	8%	3	12%	9	10	25
260960	1	1%	8	5%	9	6%	56	83	157
260993	0	0%	0	0%	0	0%	0	1	1
261004	0	0%	1	50%	0	0%	0	1	2
261026	0	0%	1	10%	1	10%	0	8	10
261041	0	0%	0	0%	0	0%	0	3	3
261085	0	0%	0	0%	0	0%	1	0	1
261096	0	0%	0	0%	0	0%	0	4	4
261100	0	0%	0	0%	2	4%	4	50	56
261122	0	0%	0	0%	0	0%	0	3	3
261155	0	0%	0	0%	0	0%	1	12	13
261166	0	0%	0	0%	0	0%	0	1	1
261214	0	0%	0	0%	0	0%	0	1	1
261225	0	0%	0	0%	0	0%	0	2	2
261236	0	0%	0	0%	0	0%	0	1	1
261240	0	0%	0	0%	0	0%	0	2	2
261306	0	0%	0	0%	0	0%	0	1	1
261310	0	0%	0	0%	0	0%	0	1	1
261321	0	0%	0	0%	0	0%	0	1	1
261343	0	0%	0	0%	0	0%	1	1	2
261365	0	0%	0	0%	0	0%	0	3	3
261380	0	0%	1	3%	1	3%	16	12	30
261391	1	3%	1	3%	10	31%	11	9	32
261402	15	4%	20	5%	95	23%	133	142	405
261424	0	0%	0	0%	1	20%	0	4	5
261446	1	9%	3	27%	0	0%	2	5	11
261461	0	0%	0	0%	1	6%	6	9	16
261483	0	0%	0	0%	2	67%	0	1	3
261531	0	0%	0	0%	0	0%	0	3	3
261542	1	11%	0	0%	0	0%	0	8	9
261564	25	7%	5	1%	23	6%	60	266	379
261575	0	0%	0	0%	1	4%	2	22	25
261601	0	0%	0	0%	0	0%	0	1	1
261612	0	0%	0	0%	1	11%	1	7	9
261623	1	7%	0	0%	0	0%	1	13	15
261645	0	0%	0	0%	2	6%	4	30	36
261682	5	3%	5	3%	7	4%	24	149	190
261704	0	0%	0	0%	0	0%	1	2	3

261726	0	0%	0	0%	0	0%	0	5	5
261763	1	25%	2	50%	0	0%	0	1	4
261785	0	0%	0	0%	0	0%	0	1	1
261800	3	4%	3	4%	1	1%	5	68	80
261811	6	16%	3	8%	8	22%	13	7	37
261822	14	13%	28	25%	20	18%	15	34	111
261833	4	12%	3	9%	9	26%	11	7	34
261844	6	10%	11	18%	14	23%	19	12	62
261881	5	9%	2	4%	3	5%	6	39	55
261892	0	0%	0	0%	0	0%	1	0	1
261903	1	33%	0	0%	0	0%	0	2	3
262021	0	0%	0	0%	2	67%	0	1	3
262043	0	0%	0	0%	0	0%	0	2	2
262054	0	0%	0	0%	0	0%	0	1	1
262065	0	0%	0	0%	0	0%	1	2	3
262102	0	0%	0	0%	1	2%	4	36	41
262124	0	0%	0	0%	0	0%	4	8	12
262146	0	0%	1	6%	0	0%	3	12	16
262150	0	0%	0	0%	0	0%	2	2	4
262161	0	0%	0	0%	0	0%	12	4	16
262172	0	0%	0	0%	0	0%	0	3	3
262183	0	0%	0	0%	0	0%	0	6	6
262205	0	0%	2	6%	1	3%	4	27	34
262216	0	0%	0	0%	0	0%	0	3	3
262220	1	3%	1	3%	1	3%	5	22	30
262231	0	0%	0	0%	0	0%	0	1	1
262242	7	14%	4	8%	6	12%	8	26	51
262301	0	0%	0	0%	0	0%	0	1	1
262323	0	0%	0	0%	0	0%	0	1	1
262356	1	14%	0	0%	1	14%	2	3	7
262360	3	7%	8	19%	10	24%	5	16	42
262371	31	25%	27	22%	8	7%	8	48	122
262382	43	11%	69	18%	44	12%	34	192	382
262581	0	0%	0	0%	0	0%	0	8	8
262603	0	0%	0	0%	0	0%	0	2	2
269404	0	0%	0	0%	0	0%	0	2	2
297146	0	0%	0	0%	0	0%	1	0	1
318916	0	0%	0	0%	0	0%	1	5	6
318920	4	2%	4	2%	3	1%	15	185	211
431384	0	0%	0	0%	0	0%	0	2	2
431756	0	0%	0	0%	0	0%	0	1	1
431760	0	0%	0	0%	0	0%	1	2	3
431863	0	0%	0	0%	0	0%	0	2	2
431900	0	0%	0	0%	0	0%	0	3	3
432762	0	0%	0	0%	0	0%	3	82	85

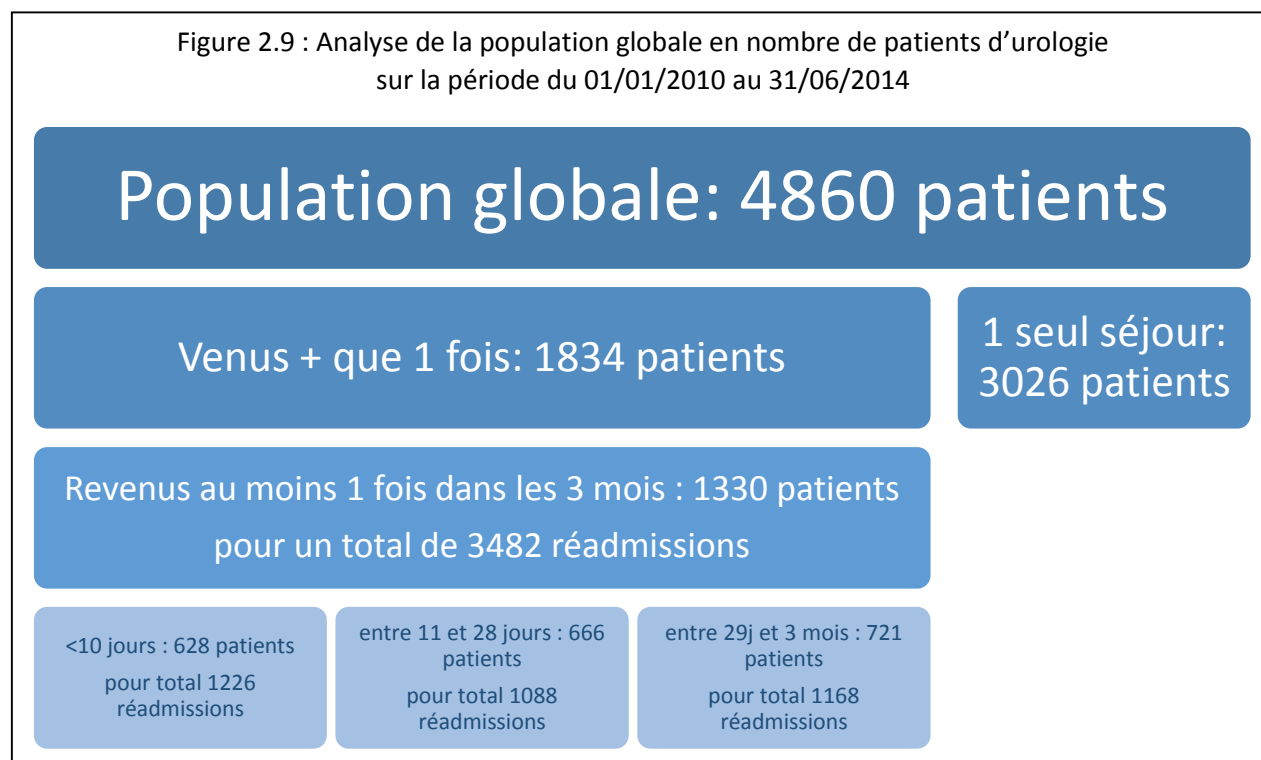
Codes INAMI médicotechniques et taux de réadmission :

Code INAMI	<=10j	% 10j	11 à 28j	% 28j	29 à 91j	% 91j	>3mois	non	Total séjours
260271	61	4%	161	10%	189	11%	492	764	1667
260282	5	8%	7	11%	12	19%	16	24	64
260330	2	4%	1	2%	6	13%	8	28	45
260341	0	0%	0	0%	2	11%	3	14	19
260352	1	3%	2	6%	3	9%	14	12	32
260680	0	0%	0	0%	0	0%	0	2	2
261914	7	4%	20	12%	23	14%	40	80	170
261925	0	0%	1	7%	0	0%	1	12	14
261973	0	0%	0	0%	0	0%	0	1	1
261984	0	0%	0	0%	0	0%	0	1	1
261995	3	1%	10	3%	28	9%	34	225	300
262006	0	0%	1	7%	2	14%	1	10	14
262393	2	2%	6	5%	10	8%	13	97	128
262426	0	0%	0	0%	1	25%	0	3	4
262430	0	0%	2	8%	2	8%	4	16	24
262441	0	0%	1	5%	5	24%	2	13	21
262452	0	0%	3	14%	2	9%	5	12	22
262463	1	25%	0	0%	0	0%	0	3	4
355865	0	0%	0	0%	0	0%	0	1	1
355073	474	29%	410	25%	229	14%	148	371	1632
355084	11	31%	4	11%	6	17%	8	6	35
355320	0	0%	0	0%	0	0%	0	1	1
355375	0	0%	0	0%	0	0%	2	4	6
355386	0	0%	1	20%	1	20%	1	2	5
355806	0	0%	1	4%	4	17%	2	17	24
355832	6	2%	1	0%	18	5%	89	256	370
355843	0	0%	0	0%	0	0%	2	11	13
355865	0	0%	0	0%	0	0%	0	1	1
Sans code	393	30%	143	11%	182	14%	131	448	1297

- Annexe 5 : Analyse de la population globale par patient

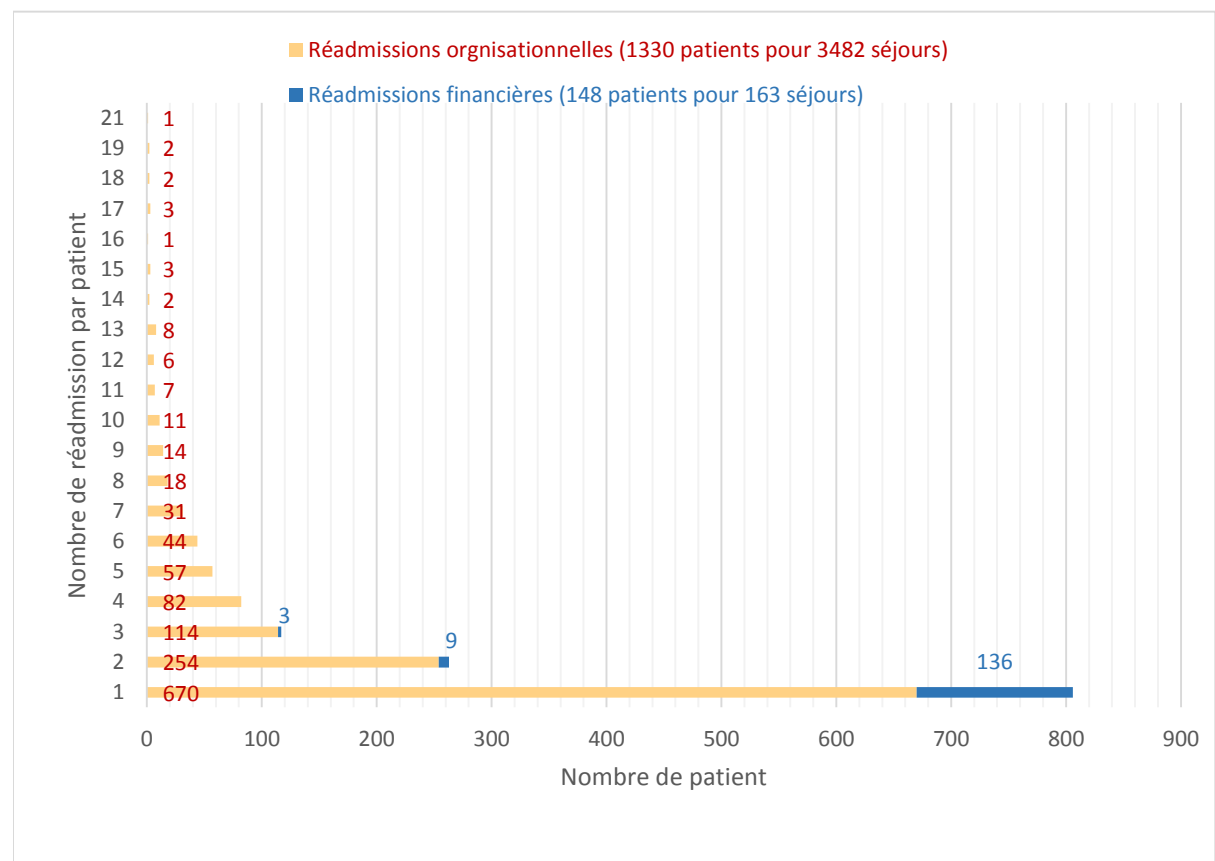
Dans l'hypothèse d'analyse suivie au départ, chaque séjour a la même importance, peu importe que ce soit le séjour d'un patient revenant régulièrement ou pas. En effet en analysant les réadmissions, seul le fait qu'un séjour soit lié au même patient, et endéans une certaine période de temps depuis le précédent est pris en compte. Cependant pour améliorer la compréhension du contexte du projet et mieux caractériser la population d'urologie, une analyse par patient a été réalisée.

Le tableau ci-dessous résume la répartition des patients en fonction de leur fréquentation du service d'urologie.



Le résultat le plus intéressant dans cette analyse montre que la majorité des patients en urologie (62%) ne viennent qu'une seule fois. En termes de patients c'est donc une minorité (38%) qui est réadmise au moins une fois, et seulement 27% (1330 patients) qui génèrent toutes les réadmissions organisationnelles (endéans les 3 mois). La figure ci-dessous permet de visualiser le nombre de réadmissions organisationnelles ou financières que chaque patient génère.

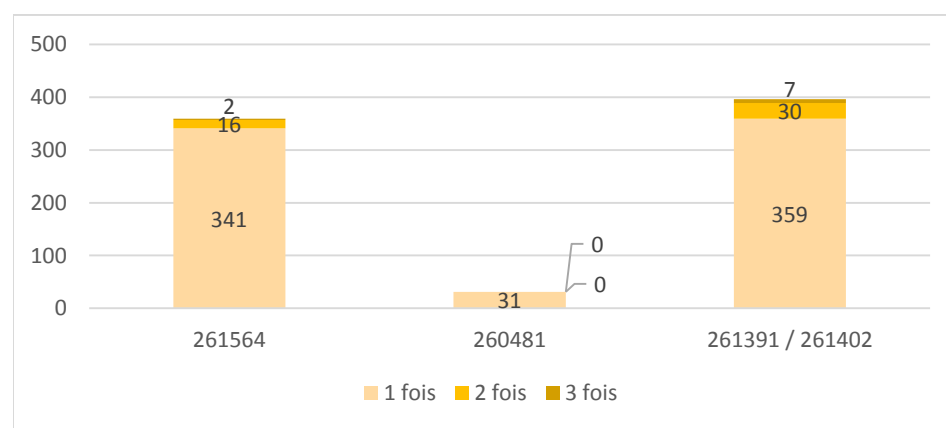
Figure 2.10 : Analyse du nombre de patients générant des réadmissions et à quelle fréquence



On remarque par exemple qu'un seul patient peut générer jusqu'à 21 réadmissions organisationnelles. Sans doute que ces valeurs extrêmes sont relatives à des patients revenus régulièrement pour des traitements medicotechniques comme par exemple le traitement de la lithiase. Mais ces analyses donnent une idée plus globale de la configuration de la population d'urologie.

Ci-dessous se trouve une analyse relative à la fréquentation des patients concernés par les trois interventions ciblées au départ : la résection endoscopique de la prostate par voie transurétrale (261564), la résection endoscopique du col de la vessie (260481), et la cure totale d'une tumeur vésicale par résection endoscopique (261391 / 261402).

Figure 2.11 : Fréquentations des patients des 3 interventions cibles



- *Annexe 6 : Modèles et résultats statistiques*

- Régression linéaire multiple :

Modèle 1 complet

```
lm(formula = Réad ~ Age + dureesej + chir + grave + Hospit +
    Numsej + prive + urg, data = Dataset2)

Residuals:
    Min       1Q   Median       3Q      Max
-1.0586 -0.3639 -0.1919  0.5490  0.9038

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.3552802   0.0176203   20.163 < 2e-16 ***
Age          -0.0003114   0.0002739   -1.137  0.25563
dureesej     -0.0048221   0.0029498   -1.635  0.10214
chir         -0.0976987   0.0125476   -7.786 7.58e-15 ***
grave         0.0255602   0.0224313    1.139  0.25453
Hospit       -0.0754825   0.0151548   -4.981 6.44e-07 ***
Numsej        0.0279743   0.0015144   18.472 < 2e-16 ***
prive        -0.0557442   0.0174005   -3.204  0.00136 **
urg           0.1526375   0.0183895    8.300 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4554 on 10113 degrees of freedom
Multiple R-squared:  0.08199, Adjusted R-squared:  0.08126
F-statistic: 112.9 on 8 and 10113 DF,  p-value: < 2.2e-16
```

Modèle 2 sans les paramètres âge, durée de séjour et gravité (sévérité)

```
Call:
lm(formula = Réad ~ chir + Hospit + Numsej + prive + urg, data = Dataset2)

Residuals:
    Min       1Q   Median       3Q      Max
-1.0612 -0.3659 -0.1821  0.5507  0.8763

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.338062   0.008269  40.881 < 2e-16 ***
chir         -0.096757   0.012448   -7.773 8.40e-15 ***
Hospit       -0.087044   0.013579   -6.410 1.52e-10 ***
Numsej        0.027813   0.001501   18.535 < 2e-16 ***
prive        -0.058369   0.017330   -3.368 0.000759 ***
urg           0.154348   0.018287    8.440 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4554 on 10116 degrees of freedom
Multiple R-squared:  0.08155, Adjusted R-squared:  0.0811
F-statistic: 179.7 on 5 and 10116 DF,  p-value: < 2.2e-16
```

- Régression logistique :

Modèle de base

Call:

```
glm(formula = Réad ~ Age + AgeCateg + DurSej + NumSej + Chambre +
    Urg + Traitement + Severite + TypeHosp, family = binomial(logit),
    data = Dataset)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.2142	-0.9293	-0.6613	1.2280	2.3161

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.810301	0.146729	-12.338	< 2e-16	***
Age	-0.008830	0.003985	-2.216	0.026688	*
AgeCateg[T.31-50]	0.816404	0.126713	6.443	1.17e-10	***
AgeCateg[T.51-70]	0.714342	0.178573	4.000	6.33e-05	***
AgeCateg[T.71->]	0.842578	0.239755	3.514	0.000441	***
DurSej	-0.031884	0.016822	-1.895	0.058039	.
NumSej	0.121217	0.007461	16.247	< 2e-16	***
Chambre[T.Privé]	-0.363432	0.099693	-3.646	0.000267	***
Urg[T.via urgence]	0.728109	0.087727	8.300	< 2e-16	***
Traitement[T.MedTec]	0.462925	0.060226	7.686	1.51e-14	***
Severite[T.Sévérité 2]	0.192048	0.049663	3.867	0.000110	***
Severite[T.Sévérité 3]	0.236118	0.113417	2.082	0.037356	*
Severite[T.Sévérité 4]	0.288554	0.556140	0.519	0.603864	
TypeHosp[T.JF]	0.394424	0.077083	5.117	3.11e-07	***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Modèle 2 sans Durée de séjour et les 4 degrés de sévérité

Call:

```
glm(formula = Réad ~ Age + AgeCateg + NumSej + Chambre + Urg +
    Traitement + TypeHosp, family = binomial(logit), data = Dataset)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.2701	-0.9369	-0.6555	1.2389	2.1843

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.835525	0.142172	-12.911	< 2e-16	***
Age	-0.007735	0.003968	-1.949	0.051270	.
AgeCateg[T.31-50]	0.800505	0.126444	6.331	2.44e-10	***
AgeCateg[T.51-70]	0.703173	0.178286	3.944	8.01e-05	***
AgeCateg[T.71->]	0.840435	0.239471	3.510	0.000449	***
NumSej	0.123512	0.007462	16.553	< 2e-16	***
Chambre[T.Privé]	-0.383666	0.099224	-3.867	0.000110	***
Urg[T.via urgence]	0.744382	0.087218	8.535	< 2e-16	***
Traitement[T.MedTec]	0.458668	0.060245	7.613	2.67e-14	***
TypeHosp[T.JF]	0.418327	0.068062	6.146	7.93e-10	***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Modèle final

Call:

```
glm(formula = Réad ~ AgeCateg + NumSej + Chambre + Urg + Traitement +  
    TypeHosp, family = binomial(logit), data = Dataset)
```

Deviance Residuals:

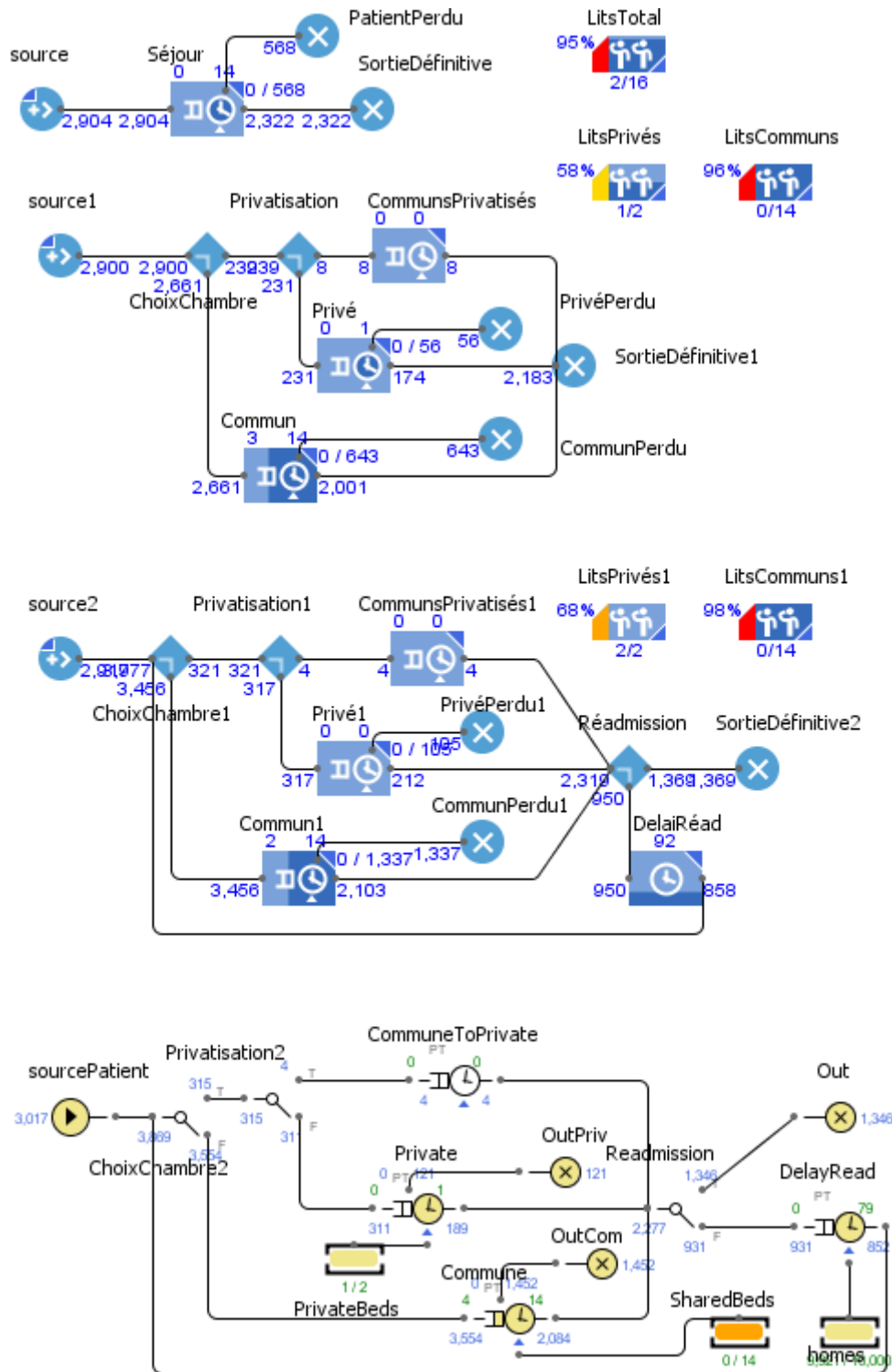
Min	1Q	Median	3Q	Max
-2.2641	-0.9341	-0.6413	1.2245	2.1808

Coefficients:

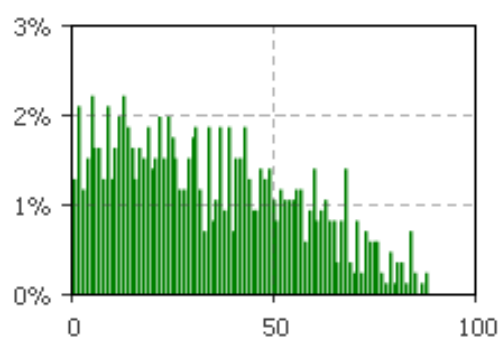
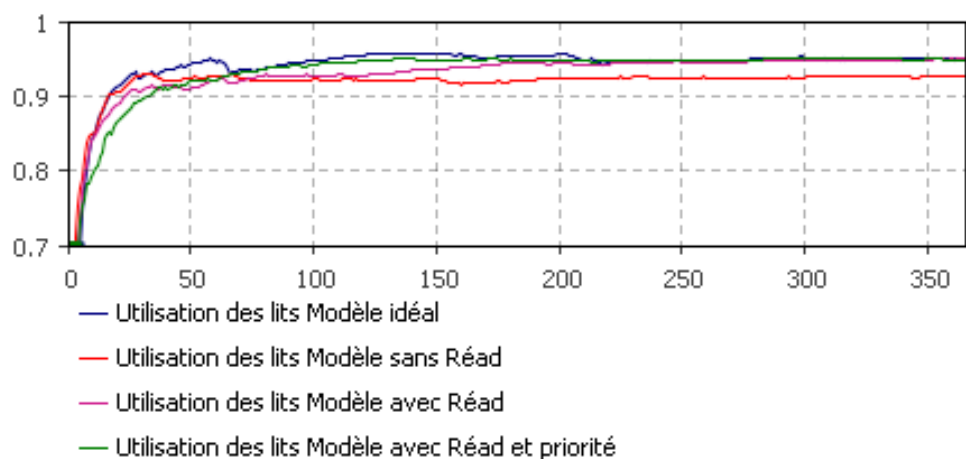
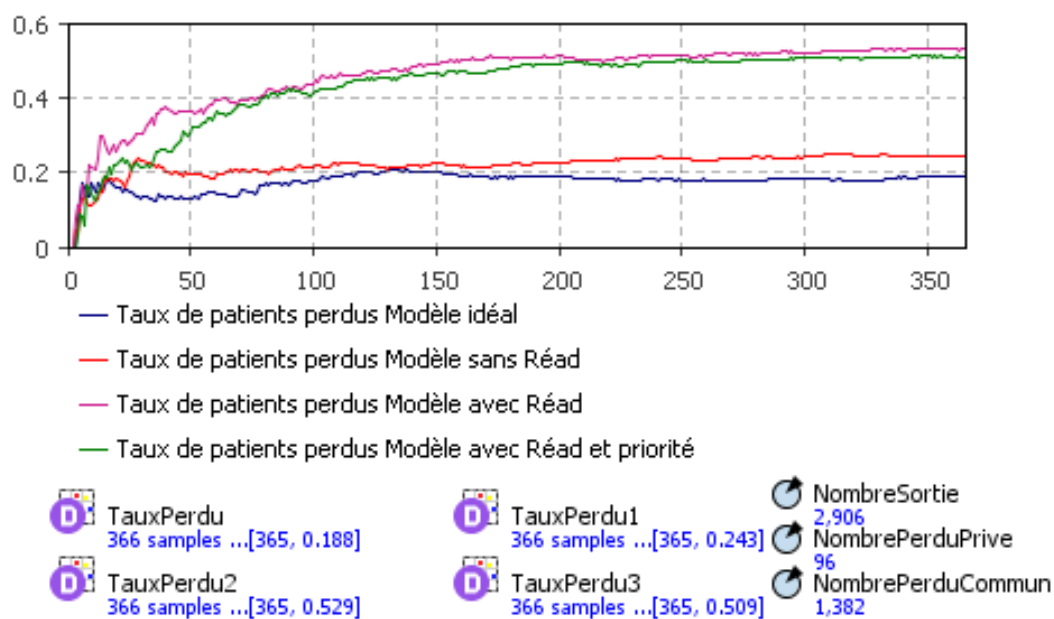
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.023188	0.104885	-19.290	< 2e-16 ***
AgeCateg[T.31-50]	0.666064	0.105977	6.285	3.28e-10 ***
AgeCateg[T.51-70]	0.418899	0.102629	4.082	4.47e-05 ***
AgeCateg[T.71->]	0.421481	0.105699	3.988	6.68e-05 ***
NumSej	0.123354	0.007461	16.533	< 2e-16 ***
Chambre[T.Privé]	-0.380809	0.099199	-3.839	0.000124 ***
Urg[T.via urgence]	0.743620	0.087193	8.528	< 2e-16 ***
Traitement[T.MedTec]	0.454722	0.060212	7.552	4.29e-14 ***
TypeHosp[T.JF]	0.422796	0.068029	6.215	5.14e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

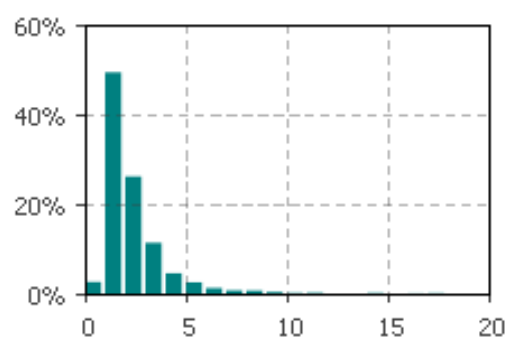
- Annexe 7 : Les modèles des simulations



Graphes des résultats d'une simulation :



Time before readmission distribution



Time des séjours distribution