

14/06/2023



Etude observationnelle en épidémiologie Clinique

Est-ce que l'utilisation de la technique de réanimation, telle que la pose d'une sonde de Swan Ganz ou une cathétérisme cardiaque droit (Right Heart Catheterization - RHC), pourrait- elle améliorer la survie des patients face à leur état de santé ?

Samuel DARMALINGON
Maximilien SOUMAHORO
Cheikh GUEYE
Groupe 14
1ère année BUT SD

Nous avons analysé un jeu de données qui provient d'une étude observationnelle multicentrique (5 centres hospitaliers) réalisée aux Etats-Unis.

Cette étude porte sur 5 735 patients adultes (≥ 18 ans) admis en unité de soins intensifs (USI) entre juin 1989 et janvier 1994, et suivis pendant 6 mois.

A l'admission en USI, les données sociodémographiques des patients ont été recueillies par questionnaire. Le jour suivant l'admission en USI, différentes caractéristiques cliniques (quantitatives et qualitatives) ont été mesurées.

Le jeu de données comporte 67 variables dont le descriptif est présenté ci -dessous.

Variable	Type	Libellé	Variable	Type	Libellé
PTID	Texte	Identifiant Patient	Caractéristiques physiologiques (dans les 24h)		
Diagnostic à l'admission			WTKILO1	Num.	Poids (kg)
CAT1	Texte	Catégorie principale de maladie	TEMP1	Num.	Température corporelle (°C)
CAT2	Texte	Catégorie secondaire de maladie	MEANBP1	Num.	Tension Artérielle (mm Hg)
CA	Texte	Cancer	RESP1	Num.	Fréquence respiratoire (resp/mn)
RESP	Texte	Respiratoire	HRT1	Num.	Fréquence cardiaque (Bpm)
CARD	Texte	Cardiovasculaire	PAFI1	Num.	Ratio PaO2/FIO2 (mm Hg)
NEURO	Texte	Neurologique	PACO21	Num.	PaCo2 (mm Hg)
GASTR	Texte	Gastro-intestinal	PH1	Num.	PH
RENAL	Texte	Rénal	WBLC1	Num.	Leucocytes (10 ⁹ cell./L)
META	Texte	Métabolique	HEMA1	Num.	Hématocrite (%)
HEMA	Texte	Hématologique	SOD1	Num.	Sodium (mmol/L)
SEPS	Texte	Infectieux	POT1	Num.	Potassium (mmol/L)
TRAUMA	Texte	Traumatique	CREA1	Num.	Créatinine (mg/L)
ORTHO	Texte	Orthopédique	BILI1	Num.	Bilirubine (mg/L)
			ALB1	Num.	Albumine (g/dL)

			URIN1	Num.	Diurèse (mL)
Maladies associées					
CARDIOHX	Num.	Atteinte vasculaire, cardiovasc.			
CHFHX	Num.	Crise cardiaque	Evénements		
DEMENTHX	Num.	Démence, infarctus céréb., Parkinson	SWANG1	Texte	Right Heart Catheterization (RHC)
PSYCHHX	Num.	Psychose, dépression	DEATH	Texte	Décès durant le suivi
CHRPULHX	Num.	Atteinte pulmonaire	Dates		
RENALHX	Num.	Atteinte rénale	d_SADMDTE	Num.	Date inclusion étude (jour)
LIVERHX	Num.	Cirrhose, atteinte hépatique	m_SADMDTE	Num.	Date inclusion étude (mois)
GIBLEDHX	Num.	Hémorragie gastro-intestinale Haute	y_SADMDTE	Num.	Date inclusion étude (année)
MALIGHX	Num.	Tumeur solide, hémopathie maligne	d_DSCHDTE	Num.	Date de sortie de l'hôp. (jour)
IMMUNHX	Num.	Immunosuppression , VIH, diabète...	m_DSCHDTE	Num.	Date de sortie de l'hôp. (mois)
TRANSHX	Num.	Transfert d'un autre hôpital (>24h)	y_DSCHDTE	Num.	Date de sortie de l'hôp. (année)
AMIHX	Num.	Infarctus du myocarde	d_DTHDTE	Num.	Date de décès (jour)
Scores Cliniques			m_DTHDTE	Num.	Date de décès (mois)
SURV2MD1	Num.	Probabilité Estimée de survie à 2 mois	y_DTHDTE	Num.	Date de décès (année)
DAS2D3PC	Num.	Score DASI (range	d_LSTCTDTE	Num.	Date de dernière nouvelle (jour)
APS1	Num.	Score Apache 3	m_LSTCTDTE	Num.	Date de dernière nouvelle (mois)
SCOMA1	Num.	Score Glasgow	y_LSTCTDTE	Num.	Date de dernière nouvelle (année)
ADLD3P	Num.	Score ADL	Caractéristiques sociodémographiques		
			AGE	Num.	Age (année)
			SEX	Texte	Sexe
			RACE	Texte	Ethnie
			EDU	Num.	Année d'étude (année)

			INCOME	Texte	Revenu
			NINSCLAS	Texte	Type d'assurance médicale

Avant de procéder à l'analyse de ces données, nous avons effectué un processus de nettoyage à l'aide du logiciel RStudio. L'objectif était de préparer les données de manière à ce qu'elles puissent être exploitées pour des analyses statistiques.

Avant d'entamer le nettoyage des données dans le logiciel R, nous avons d'abord, installer toutes les fonctions qu'il existe sur le logiciel R en utilisant : `installed.packages()`

Puis, nous avons exporter le jeu de données sur le logiciel R que les enseignants nous ont donné à l'aide de la fonction "`read.table`". Ainsi nous avons pu exporter notre jeu de données sur le logiciel R en utilisant ce script ci-dessous :

```
test = read.table("Epi_Clin projet.txt", sep="\t", header=TRUE)
```

Puis, nous avons converti notre jeu de donnée en utilisant la fonction "`as.data.frame.matrix()`" qui est une fonction qui convertit une matrice en trame de données. Chaque colonne du bloc de données résultant correspondra à une colonne de la matrice d'origine, et les noms de colonne seront les numéros de colonne de la matrice d'origine ou les noms des colonnes de la matrice d'origine s'ils existent. Par conséquent, nous avons pu convertir notre jeu de donnée par ce script : `test_df = as.data.frame.matrix(test)`.

Comme vous pouvez le voir ci- dessus nous avons appelé notre jeu de données : `test_df`. Et donc notre jeu de donnée, `test_df` s'est affiché sous forme de tableau

#	PTID	CAT1	CAT2	CA	DEATH	CARDIOHX	CHPHX	DEMENTHX	PSYCHHX	CHSPULHX	RENALHX	LIVERHX	GIBLHX	MALIGHX	IMMUNHX	TRANSFX	AMBIHX	AGE
1	3	COPD		Yes	No	0	0	0	0	1	0	0	0	1	0	0	0	70.2
2	7	MDSF w/Septo		No	Yes	1	1	0	0	0	0	0	0	0	1	1	0	79.1
3	8	MDSF w/Malignancy	MDSF w/Septo	Yes	No	0	0	0	0	0	0	0	0	0	1	1	0	48.0
4	10	ARF		No	Yes	0	0	0	0	0	0	0	0	0	0	1	0	75.3
5	11	MDSF w/Septo		No	Yes	0	0	0	0	0	0	0	0	0	0	0	0	87.9
6	12	COPD		No	No	0	1	0	0	1	0	0	0	0	0	0	0	86.0
7	13	MDSF w/Malignancy		Metastatic	No	0	0	0	0	0	0	0	0	0	1	0	0	34.9
8	14	ARF	Conti	No	Yes	0	0	0	0	0	0	0	0	0	0	0	1	43.6
9	16	MDSF w/Malignancy		Yes	No	0	0	0	0	0	0	0	0	0	0	0	0	18.0
10	17	ARF		Yes	No	0	0	0	0	0	0	0	0	0	1	0	0	48.4
11	18	ARF		No	No	0	0	0	0	0	0	0	0	0	0	0	1	34.4
12	21	ARF		No	No	0	0	0	0	0	0	0	0	0	0	0	1	88.0
13	22	MDSF w/Septo		No	Yes	1	0	0	0	0	0	0	0	0	0	0	0	74.7
14	34	MDSF w/Septo		No	Yes	0	0	0	0	0	0	0	0	0	0	1	0	42.2
15	25	ARF		No	No	0	0	0	0	0	0	0	0	0	0	0	1	81.9
16	28	MDSF w/Malignancy		Metastatic	Yes	0	0	0	0	0	0	0	0	0	1	0	0	78.3
17	29	ARF		No	Yes	0	0	1	0	0	0	0	0	0	0	0	0	88.4
18	30	MDSF w/Septo		No	Yes	0	0	1	0	0	0	0	0	0	0	0	0	89.0
19	31	MDSF w/Septo		No	No	0	0	0	0	0	1	0	0	0	0	0	0	41.4
20	32	ARF	MDSF w/Septo	No	No	0	1	0	0	1	0	0	0	0	0	0	0	67.5
21	101	LYMPH w/Septo		No	No	1	0	0	0	0	1	0	0	0	0	1	0	85.5

Comme vous pouvez le voir, voici un extrait de notre jeu de donnée sous forme de tableau, par conséquent nous constatons que cette capture d'écran le nom des variables est affiché sur la première ligne suivit des variables.

Après avoir effectué, le nettoyage de ce jeu de données sur le logiciel R, le logiciel R nous a affiché le jeu de données sous forme de tableau comme vous pouvez voir ci-dessous un extrait de notre jeu de données après le nettoyage :

Identifiant_Patient	Catégorie_principale_de_maladie	Catégorie_secondaire_de_maladie	Cancer	Décès_durant_le_suivi	Atteinte_susceptible_cardiovasc	Onco_cardiaque	Délivrance_intra-hospitalier_Parkinson	Pochette_dépense
1	5	CCRO	Non	Non	0	0	0	0
2	7	MDSF n/Sepsis	Non	Non	1	1	0	0
3	8	MDSF n/Malignant	Non	Non	0	0	0	0
4	13	ABP	Non	Non	0	0	0	0
5	11	MDSF n/Sepsis	Non	Non	0	0	0	0
6	12	CCRO	Non	Non	0	1	0	0
7	13	MDSF n/Malignant	Metastatic	Non	0	0	0	0
8	14	ABP	Non	Non	0	0	0	0
9	19	MDSF n/Malignant	Non	Non	0	0	0	0
10	17	ABP	Non	Non	0	0	0	0
11	18	ABP	Non	Non	0	0	0	0
12	21	ABP	Non	Non	0	0	0	0
13	22	MDSF n/Sepsis	Non	Non	1	0	0	0
14	24	MDSF n/Sepsis	Non	Non	0	0	0	0
15	25	ABP	Non	Non	0	0	0	0
16	28	MDSF n/Malignant	Metastatic	Non	0	0	0	0
17	20	ABP	Non	Non	0	0	1	0
18	33	MDSF n/Sepsis	Non	Non	0	0	1	0
19	31	MDSF n/Sepsis	Non	Non	0	0	0	0
20	15	ABP	Non	Non	0	1	0	0
21	32	MDSF n/Sepsis	Non	Non	0	0	0	0
22	35	ABP	Metastatic	Non	0	0	0	0
23	39	CHF	Non	Non	0	1	0	0
24	40	Come	Non	Non	0	0	0	0
25	41	CHF	Non	Non	1	1	0	0
26	42	ABP	Non	Non	0	0	0	0
27	43	Come	Non	Non	0	0	0	0

Par conséquent, nous avons effectué des analyses statistiques afin d'étudier l'efficacité de la technique de réanimation appelée pose d'une sonde de Swan Ganz, également connue sous le nom de cathétérisme cardiaque droit (Right Heart Catheterization - RHC). Notre objectif principal était d'évaluer l'efficacité de cette technique sur les patients, ce qui a donné naissance à notre problématique suivante : **Est-ce que l'utilisation de la technique de réanimation, telle que la pose d'une sonde de Swan Ganz ou une cathétérisme cardiaque droit (Right Heart Catheterization - RHC), pourrait-elle améliorer la survie des patients face à leur état de santé ?**

De ce fait nous avons utilisé certaines pour faire ces analyses tel que :

- Le genre des patients (variable qualitative)
- La tranche d'âge des patients (variable quantitative discrète)
- Le décès des patients durant le suivi (variable qualitative)
- L'utilisation de la technique de réanimation (Right Heart Catheterization - RHC) par les patients (variable qualitative)
- Le score Apache 3 des patients (variable quantitative continue)

Sommaire

I. Analyses statistiques

- a) Analyses univariées
- b) Analyses bivariées

II. Conclusion

III. Annexes

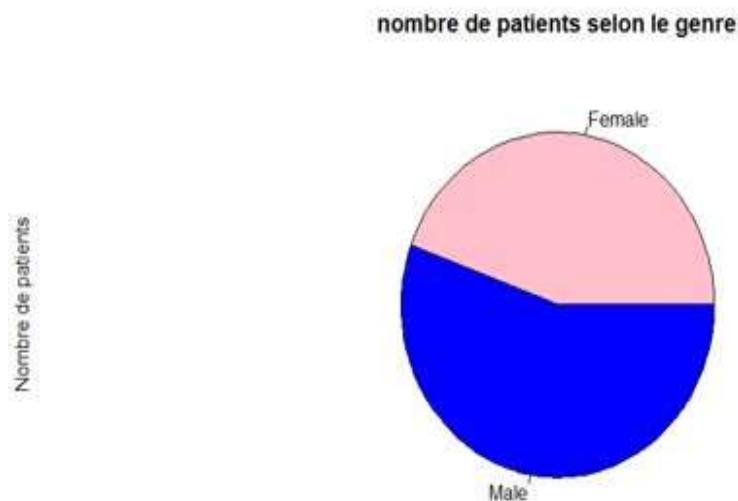
I- Analyses statistiques

a) Analyses univariées

- **Variable du genre**

Tout d'abord, nous avons entrepris une analyse exploratoire de nos données afin d'obtenir un aperçu de la répartition relative des données selon le genre

```
> prop.table(table(test_df$Sexe, useNA = "ifany"))  
      Female      Male  
0.4434176 0.5565824
```

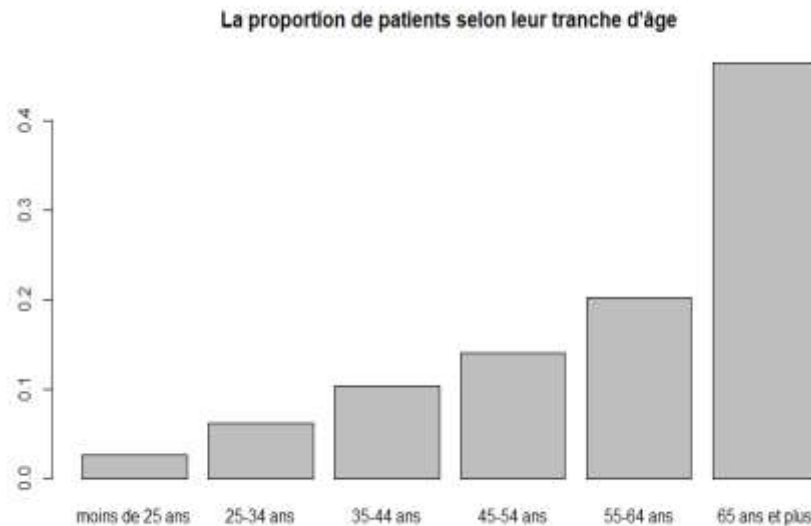


En examinant ce diagramme circulaire, il est observable que les hommes représentent environ 56% de la population étudiée, tandis que les femmes représentent 44%. Cette répartition suggère une prédominance du genre masculin au sein de l'étude.

- **Variable d'âge**

Étant conscients de l'impact du vieillissement sur la santé des hommes, nous avons décidé d'examiner de plus près les différentes tranches d'âge des patients inclus dans notre étude.

```
#Création de la variable tranche d'âge  
test_df$TRANCHE_AGE <- cut(test_df$Age_(année), breaks = c(0, 25, 35, 45, 55, 65, Inf),  
                           labels = c("moins de 25 ans", "25-34 ans", "35-44 ans", "45-54 ans", "55-64 ans", "65 ans et plus"))  
> prop.table(table(test_df$TRANCHE_AGE))  
  
moins de 25 ans    25-34 ans    35-44 ans    45-54 ans    55-64 ans    65 ans et plus  
0.02702703      0.06155187    0.10357454    0.14001744    0.20296425    0.46486486  
> |
```

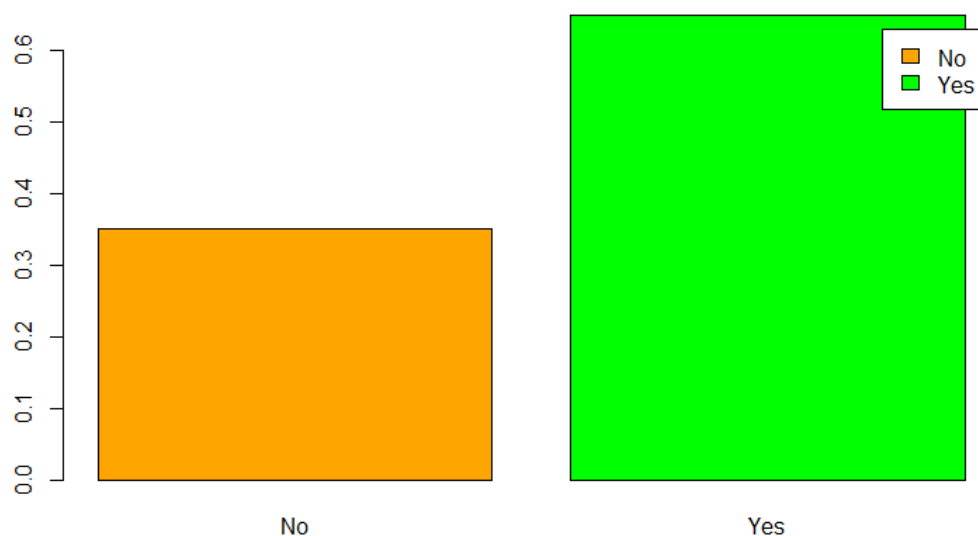


Dans ce diagramme à barres verticales, nous observons que les patients âgés de 65 ans et plus sont les plus représentés, avec une présence de 46,5 %. Ce chiffre dépasse plus du double de celui des patients âgés de 55 à 64 ans, qui constitue la deuxième catégorie la plus représentée, avec une présence de 20,3%. Les patients de moins de 25 ans sont les moins représentés, avec seulement 2,7% de présence.

- **Variable « Mort pendant le suivi »**

Un histogramme peut être utilisé ici pour représenter le nombre de patients décédés pendant le suivi, en distinguant les patients décédés avec la couleur verte et, ceux qui ont survécu avec la couleur orange. Voici le graphique suivi d'une explication détaillée de ce graphique :

Le nombre de patients qui sont mort durant leur suivi et ceux qui ne le sont pas



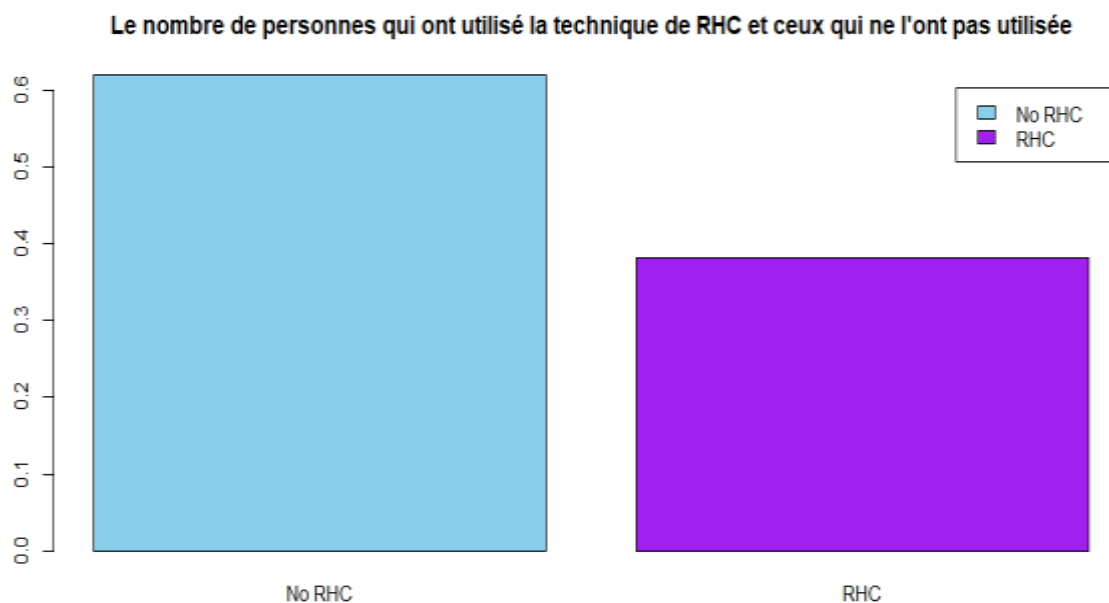
L'axe horizontal de l'histogramme représente deux catégories : "No" et "Yes", qui correspondent aux résultats du suivi des patients. L'axe vertical représente la fréquence de patients. La hauteur des barres sur l'axe vertical indique la fréquence de patients appartenant à chaque catégorie. On remarque que 65 % des patients sont morts pendant le suivi de l'étude contre 35 % qui ont survécu. Durant le suivi beaucoup de personnes sont décédées.

- **Variable « Utilisation de RHC »**

Effectivement, dans le cadre de notre étude centrée sur l'utilisation de la RHC, il est essentiel de connaître la fréquence à laquelle cette technique est utilisée.

```
> prop.table(table(test_df$`Right_Heart_Catheterization_(RHC)`, useNA = "ifany"))
```

No RHC	RHC
0.6191805	0.3808195



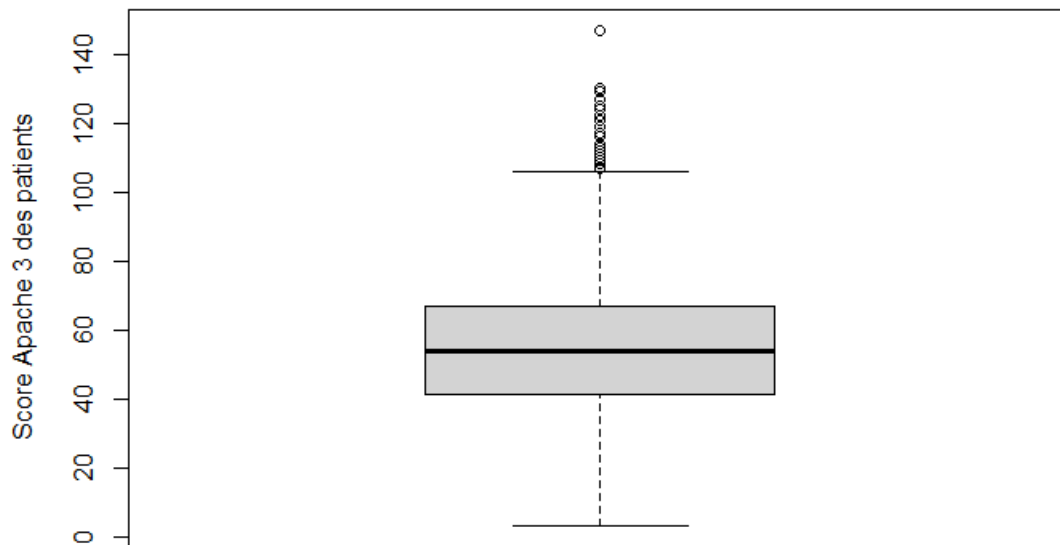
Ce diagramme à barres verticales met en évidence la proportion de patients ayant subi la technique de RHC, qui est inférieure à celle des patients qui ne l'ont pas utilisée, avec respectivement 38,1% et 61,9%.

- **Variable du « Score Apache 3 »**

Ensuite nous avons la variable du score Apache 3 qui est une variable quantitative continu. La score Apache 3 (Acute Physiology and Chronic Health Evaluation III) est un système de notation utilisé en médecine pour évaluer la gravité de la maladie et prédire la mortalité chez les patients admis en unité de soins intensifs (USI) ou en réanimation. Elle est basée sur plusieurs critères, notamment les paramètres physiologiques, les antécédents médicaux, les résultats de laboratoire et les traitements en cours. Les critères pris en compte dans le score

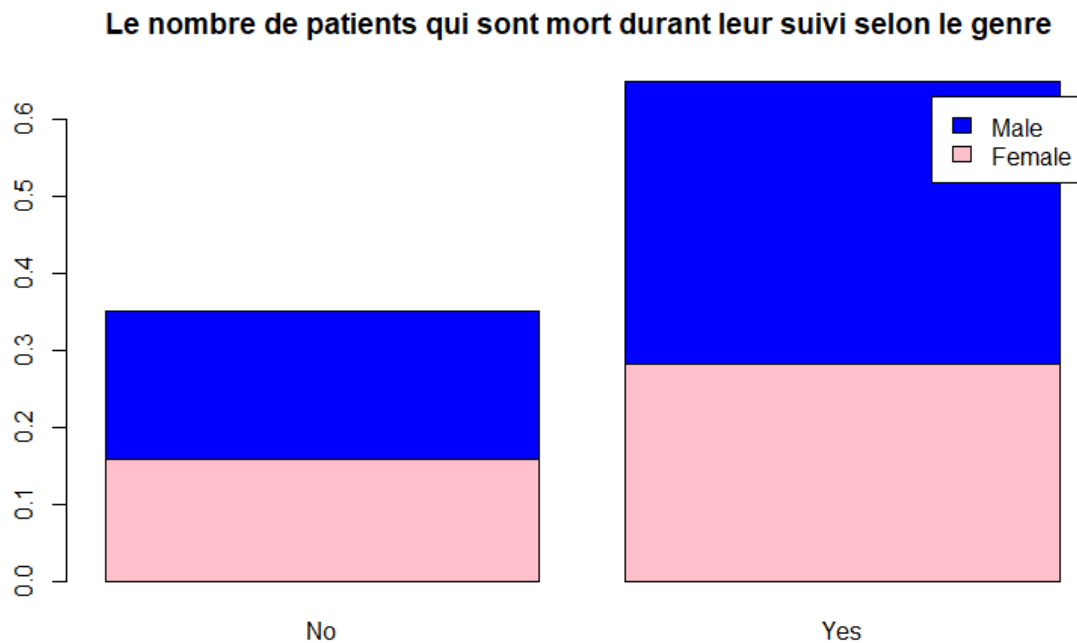
APACHE III comprennent des mesures telles que la fréquence cardiaque, la pression artérielle, la température corporelle, la fréquence respiratoire, les niveaux d'oxygène dans le sang, les taux sanguins de diverses substances, les résultats de tests de laboratoire, les comorbidités et les interventions thérapeutiques. Ainsi ce score est évalué sur des valeurs entre 0 et 299 et les valeurs élevées indiquant un pronostic péjoratif

La répartition des patients selon leur score Apache 3



Nous sommes ici en présence d'une boîte à moustache représentant le score Apache III des patients. La médiane qui est un indicateur de la tendance centrale du jeu de données se situe autour de 50, cela signifie que la moitié des patients ont un score Apache 3 inférieur à 50 et l'autre moitié a un score supérieur à 50. Pour le premier quartile le score est d'environ 38 et pour le troisième le score d'Apache III est d'environ 65. On peut aussi mentionner la présence de nombreuses valeurs aberrantes, cela est indiqué par les points de données qui s'écartent considérablement de la tendance générale. Les valeurs aberrantes peuvent indiquer des cas extrêmes, des erreurs de mesure ou des situations inhabituelles.

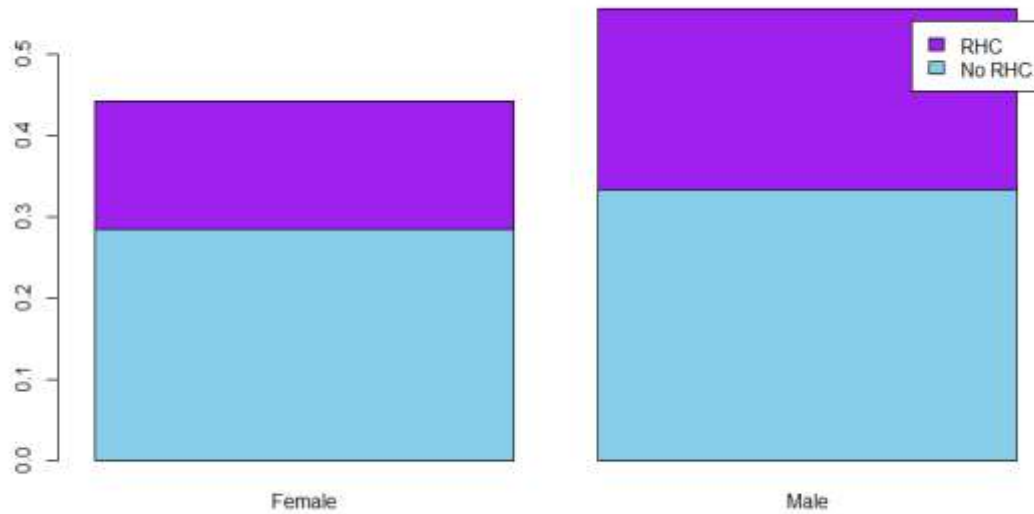
b) Analyses bivariées



Nous sommes en présence d'un graphique en barres empilé qui montre le nombre de patients qui sont mort durant leur suivi. Ainsi, nous pouvons voir comme précédemment qu'il n'y a plus de personne qui sont décédé durant le suivi. Néanmoins, on observe qu'il y a plus de personne de sexe masculin qui sont morte durant leur suivi que de femme, soit 28 % de femmes contre 37% d'hommes. Mais est-ce que le genre de l'individu a un impact sur son décès ? De ce fait, nous avons utilisé le test de khi-deux à l'aide de la fonction `chisq.test()` avec un risque de 5 % de se tromper. Après avoir effectué ce test, la p-valeur = 0.1492. Cette p-valeur est supérieure à 0.05, donc le genre des individus n'a pas d'individu sur leur décès. Donc les variables sont indépendantes.

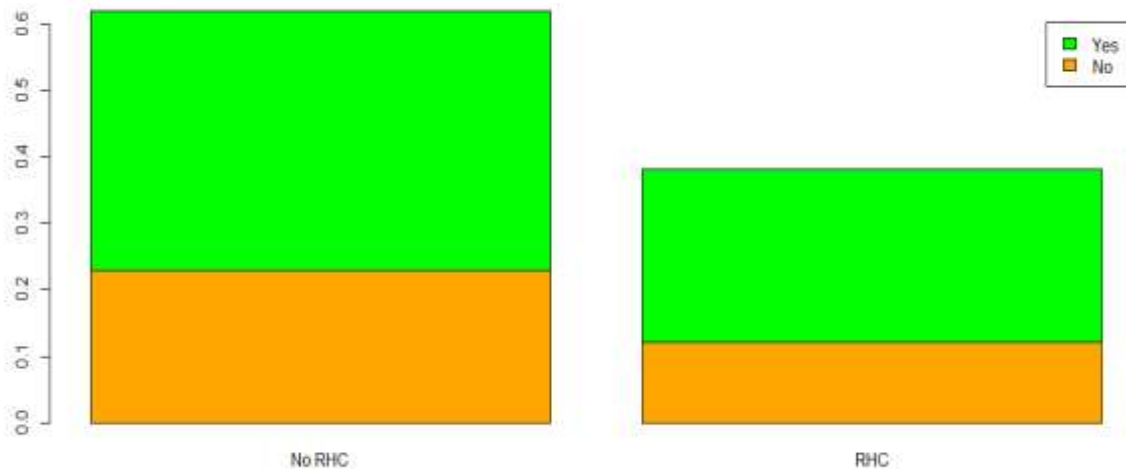
L'analyse du graphique diagramme en bars empilés représentant l'utilisation de la technique de réanimation RHC selon le genre des patients révèlent les informations suivantes :

Nombre de patients qui ont utilisé la technique de réanimation RHC selon leur genre



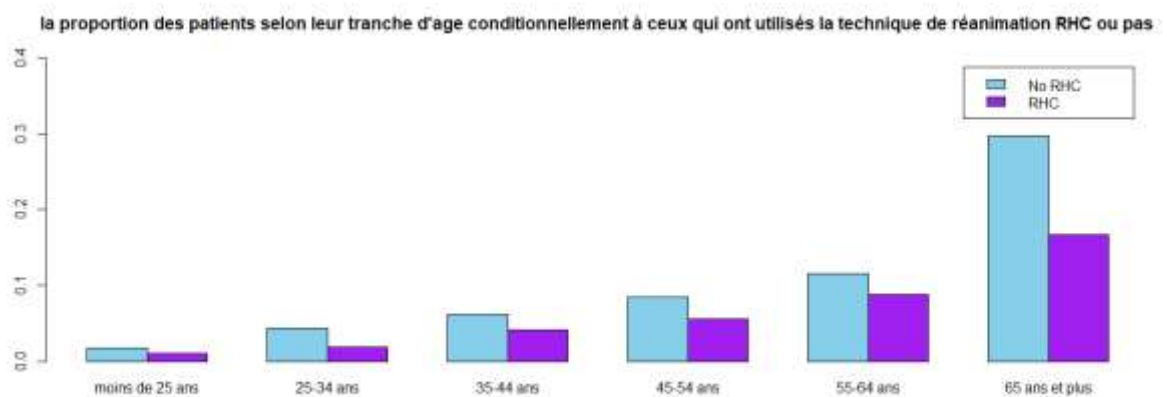
En examinant les deux diagrammes en bars empilés pour chaque genre, on constate que la fréquence des patients masculins utilisant la technique RHC (représentée en violet) est plus élevée que celle des patients féminins. Cependant, il convient de noter que la fréquence des patients masculins n'ayant pas bénéficié de la technique RHC (représentée en bleu) est également plus élevée que celle des patients féminins. Cela s'explique peut-être par une plus grande présence de patient masculins dans l'étude. En termes de pourcentages, environ 22% des patients masculins ont utilisé la technique RHC, tandis que pour les patients féminins, cette proportion est d'environ 17 %. Cela suggère qu'il y a une tendance plus marquée chez les hommes à recourir à la réanimation RHC par rapport aux femmes. Même si la plupart des patients n'en n'ont pas bénéficié. 33 % chez les hommes contre environ 28 % chez les femmes.

Le nombre de patients qui ont utilisés la technique de réanimation RHC ou pas conditionnellement au décès durant le suivi



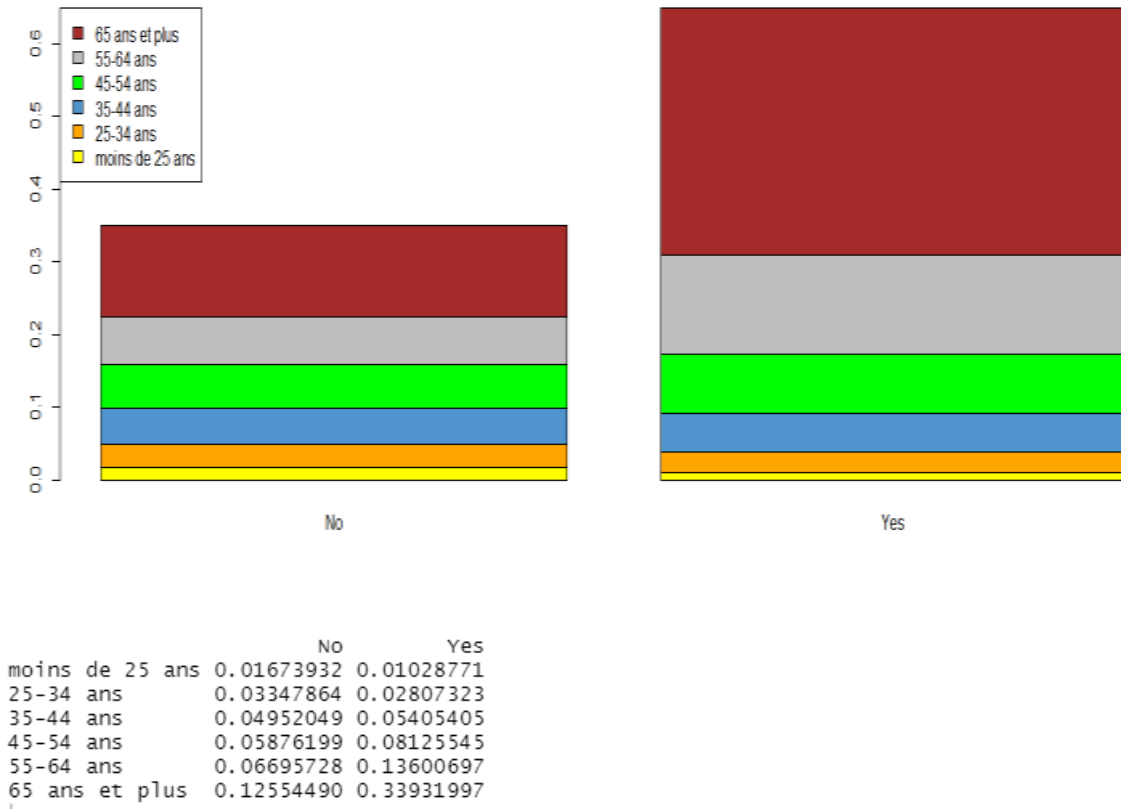
	No RHC	RHC
No	0.2292938	0.1217088
Yes	0.3898867	0.2591107

L'objectif de ce diagramme à barres empilées est de mettre en évidence la différence entre les patients ayant reçu une RHC en fonction de leur survie ou non. Après une observation attentive, nous constatons que les patients n'ayant pas obtenu de la RHC présentent un taux de survie de 22,9 %, tandis que ceux ayant reçu la RCP affichent un taux de survie de seulement 12,2 %. En revanche, le taux de décès des patients n'ayant pas utilisé la RHC est plus élevé que celui des patients ayant eu recours à la RHC, respectivement de 40% et 25,9%. Par la suite, nous avons réalisé un test du khi-deux afin de déterminer s'il existait une corrélation significative entre la RHC et le décès des patients. La p-valeur, égale à $1,04e-4$, est inférieure au seuil de risque de 5%, ce qui indique une liaison significative entre les deux variables.



Ici, nous observons la proportion des patients selon leur tranche d'âge conditionnellement à ceux qui ont utilisés la technique de réanimation RHC ou pas sous forme de diagramme en barres. Nous constatons que plus les patients sont âgés, plus la proportion de patients qui ont utilisés la technique de réanimation RHC et ceux qui ne l'ont pas utilisé augmentent au fil du temps. Mais, le taux de personnes qui ont utilisé la technique de réanimation sont inférieure à celle qui ne l'ont pas utilisé. Comme par exemple, dans la tranche d'âge 65 ans et plus il y a 30 % patients qui n'ont pas utilisé cette technique contre 17% qui ont utilisé la technique réanimation. En outre, avec le test de khi-deux, nous avons calculé la p-valeur qui est de $1.052e-05$ avec un risque de 5%. La p-valeur est inférieur au niveau de risque de 5 %, donc les 2 variables sont liées et pas indépendantes. Cela veut dire que le taux de patients qui ont utilisé la technique de réanimation RHC et ceux qui ne l'ont pas utilisé augmentent au fil de l'âge des patients.

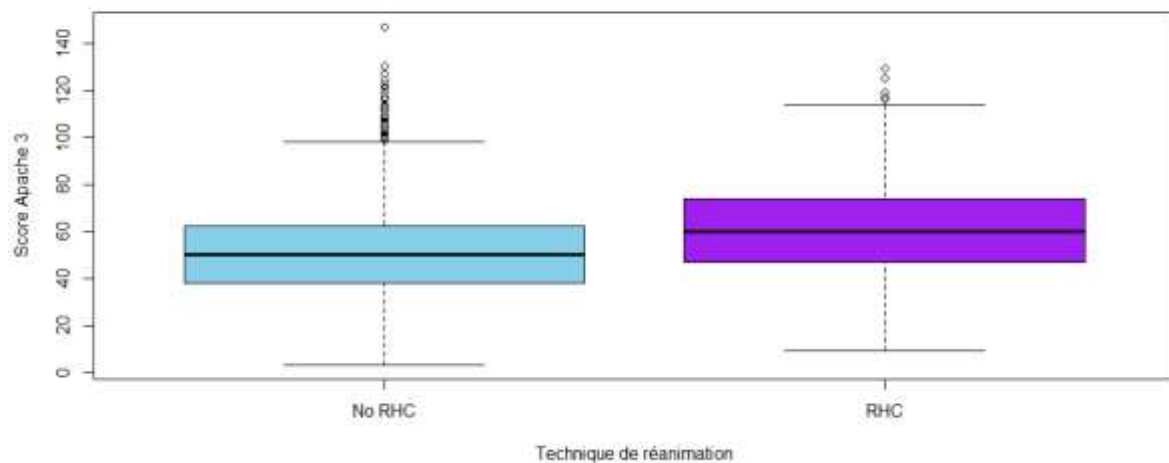
La proportion de patients qui sont morts durant la survie selon leur tranche d'âge



Dans ce diagramme à barres empilées représentant la proportion des patients décédés ou non en fonction des tranches d'âge, nous observons plusieurs tendances. Parmi les patients qui ne survivent pas, on constate que 33,9% sont âgés de 65 ans ou plus. En revanche, parmi les patients qui survivent, cette catégorie d'âge représente seulement 12,6%, soit deux fois moins. La deuxième tranche d'âge la plus représentée parmi les patients décédés est celle des patients âgés de 55 à 64 ans, avec un taux de décès de 13 %, comparé à seulement 6 % chez les patients qui survivent. Ce graphique met en évidence le fait que plus l'âge diminue, plus la proportion de patients survivants augmente et plus la proportion de patients décédés diminue.

L'analyse de la boîte à moustache du score Apache III selon l'utilisation de la réanimation RHC révèle les informations suivantes :

la boîte à moustaches du score Apache 3 des patients qui ont fait la technique de réanimation des patients ou pas



En comparant les médianes des boîtes à moustache, on constate que les patients ayant bénéficié de la réanimation RHC présentent un score Apache III plus élevé (score d'environ 60) que ceux qui n'ont pas eu recours à cette technique (score d'environ 50).

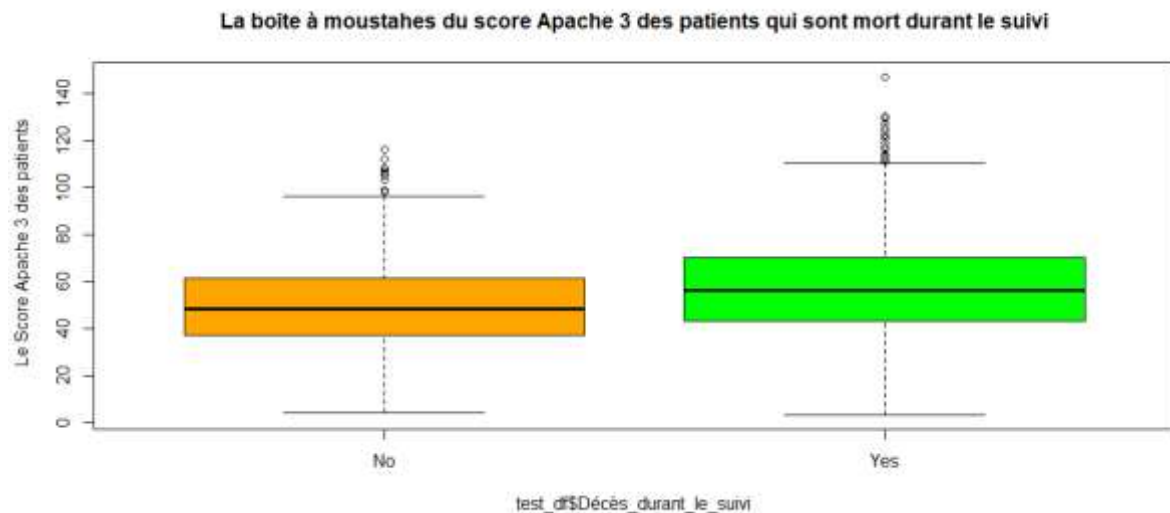
De plus, l'écart interquartile (la distance entre le premier quartile et le troisième quartile) est légèrement plus large pour les patients ayant utilisé la réanimation RHC, ce qui indique une plus grande dispersion des scores Apache III dans ce groupe. Cela suggère une plus grande variabilité dans la gravité des cas parmi les patients ayant bénéficié de la RHC.

Les bornes supérieures et inférieures des boîtes à moustache représentent les valeurs maximales et minimales respectives. Par exemple pour ceux n'ayant pas utilisé la RHC le score maximal est de 100 et environ 120 pour ceux qui en ont bénéficié.

On peut observer que pour les patients ayant utilisé la RHC, les valeurs aberrantes sont plus rares, tandis que pour les patients n'ayant pas bénéficié de la RHC, il y a un plus grand nombre de valeurs aberrantes.

En calculant la P-valeur des variables on a trouvée qu'elle était $< 2,2e-16$ c'est à dire qu'elle est inférieure à $2,2e-16$. Cela veut dire qu'il y a un lien entre les variables qui n'est pas significatif. Pour interpréter ce résultat on peut dire que les personnes ayant utilisé la technique de réanimation RHC ont un score d'apache plus élevé que ceux qui ne l'ont pas utilisé.

Ce graphique soulève d'intéressant faits sur la relation entre l'utilisation de la réanimation RHC et la gravité des cas évaluée par le score Apache III.



Nous observons une boîte à moustache du score Apache 3 des patients qui sont mort durant le suivi. Le score Apache 3 qui des patients qui sont mort durant le suivi est plus élevé que ceux qui ne sont pas mort durant le suivi. En effet, le score médian des patients qui ne sont pas mort est d'environ 50 alors que ceux qui sont mort durant le suivi ont un score médian d'environ 55. De plus, la taille de la boîte des personnes qui ne sont pas mort est légèrement moins épaisse que les personnes qui ne sont pas mort durant le suivi. Cela signifie qu'il y a une plus grande variabilité du score Apache 3 des personnes qui ne sont pas morts durant le suivi contrairement à ceux qui sont morts. Par ailleurs, il y a plus de valeurs aberrantes concernant les personnes qui sont décédées pendant le suivi à l'opposé des personnes qui ne sont pas décédées pendant le suivi.

En outre nous avons étudié la p-valeur du score des patients qui sont morts ou pas pendant le suivi où cette p-valeur est de $2,506 \times 10^{-16}$ avec un risque de se tromper de 5%. Ainsi on constate que cette p-valeur est inférieure au risque de 5 % donc les variables sont liées et nous pouvons interpréter que les personnes qui sont mortes pendant le suivi ont un score Apache plus élevées que les personnes qui ne sont pas mortes pendant le suivi.

II- Conclusion

En conclusion, nous avons réalisé une analyse statistique d'un jeu de données provenant d'une étude observationnelle multicentrique aux États-Unis portant sur des patients adultes admis en unité de soins intensifs. Nous avons effectué un nettoyage des données dans RStudio et procédé à des analyses univariées et bivariées.

Dans les analyses univariées, nous avons examiné différentes variables telles que le genre des patients, la tranche d'âge, le décès pendant le suivi, l'utilisation de la technique de réanimation (RHC) et le score Apache 3. Nous avons observé une prédominance du genre masculin dans l'étude, une concentration des patients dans la tranche d'âge de 65 ans et plus, une proportion élevée de décès pendant le suivi, une utilisation relativement faible de la RHC et une distribution variée des scores Apache 3 avec la présence de valeurs aberrantes.

Dans les analyses bivariées, nous avons constaté que le genre des patients n'avait pas d'impact significatif sur leur décès. Cependant, il y avait une tendance plus marquée chez les hommes à recourir à la réanimation RHC. De plus, nous avons observé une corrélation significative entre l'utilisation de la technique RHC et le décès des patients, avec un taux de survie plus faible pour les patients ayant reçu la RHC.

En ce qui concerne l'âge, nous avons constaté que plus les patients étaient âgés, plus la proportion de ceux utilisant la technique RHC augmentait, mais le taux de survie était plus faible pour ces patients.

L'analyse des données a révélé que les patients ayant bénéficié de la réanimation RHC présentaient un score Apache III plus élevé par rapport à ceux qui n'ont pas eu recours à cette technique. De plus, la dispersion des scores Apache III était plus importante chez les patients ayant utilisé la réanimation RHC. Les valeurs aberrantes étaient également moins fréquentes dans ce groupe. À l'aide du test de khi-deux, il y avait plus nous avons conclu que ces variables étaient dépendantes les unes des autres. Et donc, les personnes qui sont en mauvaise santé auraient forcément recours à cette technique.

Néanmoins, pendant ce projet, nous avons eu quelques difficultés. En effet, nous avons eu du mal au début à travailler avec le logiciel RStudio comme c'était notre premier gros projet sur RStudio. Mais au fil du temps nous avons compris que le but était que nous nous formions sur le logiciel en testant des choses. De plus le fait d'avoir plusieurs projets en cours a également entraîné une dispersion de notre attention et de notre énergie. Il était parfois difficile de se concentrer sur chaque projet et de maintenir un niveau optimal de productivité. Nous devons jongler entre les différentes tâches, les échéances et les demandes spécifiques de chaque projet, ce qui a généré un sentiment de perturbation et de pression accrue.

III- Annexes

```
test = read.table("Epi_Clin projet.txt", sep="\t", header=TRUE)
test_df = as.data.frame.matrix(test)
```

#Analyses statistiques

#Le nombre de personnes qui ont utilisé la technique de RHC et ceux qui ne l'ont pas utilisée

```
prop.table(table(test_df$`Right_Heart_Catheterization_(RHC)`, useNA = "ifany"))
barplot(prop.table(table(test_df$`Right_Heart_Catheterization_(RHC)`, useNA =
"ifany")),
        main = "Le nombre de personnes qui ont utilisé la technique de RHC et ceux qui ne
l'ont pas utilisée",
        col=c("skyblue", "purple"),legend.text = TRUE)
```

#nombre de patients selon le genre

```
prop.table(table(test_df$Sexe,useNA= "ifany"))
pie(table(test_df$Sexe,useNA= "ifany"), main= "nombre de patients selon le genre", ylab
= "Nombre de patients",
    col=c("pink", "blue"),legend.text = TRUE)
```

#Le nombre de patients qui sont mort durant le suivi et ceux qui ne le sont pas

```
barplot(prop.table(table(test_df$Décès_durant_le_suivi, useNA="ifany")),
        main = "Le nombre de patients qui sont mort durant leur suivi et ceux qui le sont
pas",
        legend.text = TRUE, col = c( "orange", "green"))
```

#Nombre de patients qui ont utilisé la technique de réanimation RHC selon leur genre

```
prop.table(table(test_df$`Right_Heart_Catheterization_(RHC)`,test_df$Sexe))
barplot(prop.table(table(test_df$`Right_Heart_Catheterization_(RHC)`,test_df$Sexe)),
        main= "Nombre de patients qui ont utilisé la technique de réanimation RHC selon
leur genre",
        col=c("skyblue", "purple"),legend.text = TRUE)
```

```
chisq.test(test_df$`Right_Heart_Catheterization_(RHC)`,test_df$Sexe)
```

#p-value = 0.0006999

#Les variables ne sont pas indépendantes entre la technique de réanimation (RHC) et le genre

#Le nombre de patients qui sont mort durant leur survie selon le genre

```
prop.table(table(test_df$Sexe, test_df$Décès_durant_le_suivi))
barplot(prop.table(table(test_df$Sexe, test_df$Décès_durant_le_suivi)),
  main = "Le nombre de patients qui sont mort durant leur suivi selon le genre ",
  col=c("pink", "blue"), legend.text = TRUE)
```

```
chisq.test(test_df$Sexe, test_df$Décès_durant_le_suivi)
#p-value = 0.1492
#Les variables sont indépendantes entre le genre et les patients qui sont mort durant le suivi
```

#bivariés entre le décès durant le suivi et le RHC

```
prop.table(table(test_df$Décès_durant_le_suivi,
  test_df$`Right_Heart_Catheterization_(RHC)`))
barplot(prop.table(table(test_df$Décès_durant_le_suivi,
  test_df$`Right_Heart_Catheterization_(RHC)`)),
  main = "Le nombre de patients qui ont utilisés la technique de réanimation RHC ou pas conditionnellement au décès durant le suivi",
  legend.text = TRUE, col = c("orange", "green"))
```

```
chisq.test(test_df$Décès_durant_le_suivi, test_df$`Right_Heart_Catheterization_(RHC)`)
#p-value = 0.0001047
# les variables ne sont pas indépendantes entre le décès durant le suivi et les patients qui ont utilisé la technique de réanimation RHC
```

#Création de la variable tranche d'âge

```
test_df$TRANCHE_AGE <- cut(test_df$`Age_(année)`, breaks = c(0, 25, 35, 45, 55, 65, Inf),
  labels = c("moins de 25 ans", "25-34 ans", "35-44 ans", "45-54 ans", "55-64 ans", "65 ans et plus"))
```

#La proportion de patients selon leur tranche d'âge

```
prop.table(table(test_df$TRANCHE_AGE))
barplot(prop.table(table(test_df$TRANCHE_AGE)), main = "La proportion de patients selon leur tranche d'âge")
```

#la proportion des patients selon leur tranche d'âge conditionnellement à ceux qui ont utilisés la technique de réanimation RHC ou pas

```
prop.table(table(test_df$`Right_Heart_Catheterization_(RHC)`, test_df$TRANCHE_AGE))
barplot(prop.table(table(test_df$`Right_Heart_Catheterization_(RHC)`, test_df$TRANCHE_AGE)),
  col = c("skyblue", "purple"),
  ylim = c(0.0, 0.4), beside = TRUE, legend.text = TRUE,
  main = "la proportion des patients selon leur tranche d'âge conditionnellement à ceux qui ont utilisés la technique de réanimation RHC ou pas")
```

```
chisq.test(test_df$`Right_Heart_Catheterization_(RHC)`, test_df$TRANCHE_AGE)
```

```
#p-value = 1.052e-05
```

```
#les variables ne sont pas indépendantes entre l'âge des patients et les patients qui ont utilisé la technique de réanimation RHC
```

```
#La proportion de patients qui sont morts durant la survie selon leur tranche d'âge
```

```
barplot(prop.table(table(test_df$TRANCHE_AGE,test_df$Décès_durant_le_suivi)),  
  main = "La proportion de patients qui sont morts durant la survie selon leur tranche  
d'âge",  
  col = c('yellow','orange','steelblue3', 'green', 'grey','brown'),  
  legend.text = c("moins de 25 ans", "25-34 ans", "35-44 ans", "45-54 ans", "55-64  
ans", "65 ans et plus"),  
  args.legend = list(x = "topleft"))
```

```
chisq.test(test_df$TRANCHE_AGE,test_df$Décès_durant_le_suivi)
```

```
#p-value < 2.2e-16
```

```
#les variables ne sont pas indépendantes entre les patients qui sont morts durant le suivi et leur tranche d'âge
```

```
##score Apache3 univarié
```

```
boxplot(test_df$Score_Apache_3, main = "La répartition des patients selon leur score  
Apache 3",  
  ylab = "Score Apache 3 des patients")
```

```
##bivarié
```

```
#la boîte à moustaches du score Apache 3 des patients qui ont fait la technique de  
réanimation des patients ou pas
```

```
boxplot(test_df$Score_Apache_3 ~ test_df$`Right_Heart_Catheterization_(RHC)` ,  
  main= "la boîte à moustaches du score Apache 3 des patients qui ont fait la technique  
de réanimation des patients ou pas ",  
  xlab="Technique de réanimation", ylab = "Score Apache 3",  
  col = c("skyblue", "purple"))
```

```
chisq.test(test_df$Score_Apache_3 ,test_df$`Right_Heart_Catheterization_(RHC)`)
```

```
#p-value < 2.2e-16
```

```
# les variables ne sont pas indépendantes entre le score Apache 3 des patients et les  
patients qui ont fait la technique de réanimation des patients ou pas
```

```
#le score Apache 3 des patients qui sont mort durant le suivi
```

```
boxplot(test_df$Score_Apache_3~ test_df$Décès_durant_le_suivi,  
  main= "La boîte à moustaches du score Apache 3 des patients qui sont mort durant le  
suivi",  
  ylab = " Le Score Apache 3 des patients" ,  
  col = c( "orange", "green"))
```

```
chisq.test(test_df$Score_Apache_3 , test_df$Décès_durant_le_suivi)
```

```
#p-value = 2.506e-16
```

#les variables ne sont pas indépendantes entre le Score Apache 3 des patients et les patients qui sont morts durant le suivi