

The background of the entire page is a light beige color. It is decorated with several large, stylized tropical leaves in shades of green, yellow, and blue. Additionally, there are clusters of small, dark blue polka dots scattered across the page, particularly around the central text box.

---

UNIVERSITÉ PARIS SACLAY

# TP HAMILTON ET SCALE

MASTER METHODOLOGIE ET  
STATISTIQUE EN RECHERCHE  
BIOMÉDICALE

**NISRINE BENNOR**

GRUPE 2 - JANVIER 2021

## 1. Énoncé du devoir

1.Énoncé du devoir .....	3
2.Présentation du contexte .....	4
2.1. Échelle de Hamilton .....	4
2.2. SCL90 .....	5
3.Jeux de données .....	5
4.Objectif du travail.....	6
5.Techniques et Méthodes .....	6
6.Data Management .....	6
6.1. Le fichier groupe : .....	6
6.2. Le fichier d'hétéro_évaluation (questionnaire HRDS) :Échelle de Hamilton .....	7
6.2.1. <i>Analyse descriptive de la somme des items de l'échelle de Hamilton à J0 et J56</i>	
6.3. Le fichier d'auto évaluation (questionnaire SCL 90) .....	9
6.4. Description du fichier unifié .....	10
7.Validation de l'échelle de Hamilton .....	11
7.1. Analyse de la distribution d'items et de score à J0 et J56 .....	13
7.2 – Analyse de Corrélations entre les items (redondance) .....	16
7.3 -- Structure dimensionnelle : Unidimensionnalité .....	17
7.4 – Analyse factorielle .....	17
7.5 – Validation de consistance interne .....	18
7.6 -- Validité concourante .....	18
8.Comparaison des deux groupes : Analyse de réponse au traitement .....	20
8.1 – Approche LOCF .....	20
8.2 – Modèle mixte .....	22
9.Analyse de la Survie : Analyse de la réponse binaire au traitement .....	24

## 1. Énoncé du devoir

Dans une étude d'épidémiologie clinique, 146 patients déprimés sont évalués à J0 (début de traitement), J4, J7, J14, J21, J28, J42, J56 à l'aide d'une autoévaluation (SCL90) et d'une hétéroévaluation (échelle de dépression de Hamilton).

Vous disposez de trois fichiers : le fichier Tt (traitement des patients), le fichier autoévaluation (SCL 90) et le fichier hdrs (échelle de Hamilton).

**1/** lorsque l'on utilise un instrument de mesure subjective dans une étude clinique, il est toujours bon de le (re)valider rapidement. Procédez ici à cette vérification sur l'échelle de dépression de Hamilton, aux temps J0 et J56.

**2/** A partir du score brut de Hamilton, déterminez si les patients du groupe Tt=1 répondent mieux au traitement que les patients du groupe Tt=0. (utiliser d'abord une approche LOCF, puis un modèle mixte).

**3/** On répond aussi parfois à cette question en considérant le critère binaire censuré « réponse au traitement » défini par une chute de 50% à l'échelle de Hamilton par rapport à J0. En utilisant ce nouveau critère, répondez donc de nouveau à la question 2.

## 2. Présentation du contexte

### 2.1. Échelle de Hamilton :

L'échelle de Hamilton date de 1967. Elle repose sur l'évaluation par un clinicien de 17 critères représentant la symptomatologie dépressive. Les items de l'échelle sont cotés de 0 à 2 ou de 0 à 4, et la somme de tous les items représente le score de dépression total. Plus le score est élevé, plus la dépression est importante.

L'échelle de Hamilton est une échelle d'hétéroévaluation composée de 17 items cotés de 0 à 2 ou de 0 à 4 et dont la somme totale varie de 0 à 52. Les notes seuils pour la dépression (Bech, 1989) sont : absence de dépression ( $< 8$ ) ; dépression mineure ( $7 < \text{note} < 16$ ) et dépression majeure ( $> 15$ ).

Cependant, l'analyse de la note globale n'est pas sans difficulté et il semble plus prudent d'utiliser une sous-échelle constituée des 6 variables répondant aux critères d'unidimensionnalité : humeur dépressive (0 à 4), sentiments de culpabilité (0 à 4), travail et activités (0 à 4), ralentissement (0 à 4), anxiété psychique (0 à 4) et symptômes somatiques généraux (0 à 2). La sensibilité au changement de l'échelle de Hamilton semble bonne.

<i>Items</i>	<i>Symptômes</i>
<i>HAMD 1</i>	<i>Humeur dépressive</i>
<i>HAMD 2</i>	<i>Sentiments de culpabilité</i>
<i>HAMD 3</i>	<i>Suicide</i>
<i>HAMD 4</i>	<i>Insomnie du début de la nuit</i>
<i>HAMD 5</i>	<i>Insomnie du milieu de la nuit</i>
<i>HAMD 6</i>	<i>Insomnie du matin</i>
<i>HAMD 7</i>	<i>Travail et activités</i>
<i>HAMD 8</i>	<i>Ralentissement</i>
<i>HAMD 9</i>	<i>Agitation</i>
<i>HAMD 10</i>	<i>Anxiété psychique</i>
<i>HAMD 11</i>	<i>Anxiété somatique</i>
<i>HAMD 12</i>	<i>Symptômes somatiques gastro-intestinaux</i>
<i>HAMD 13</i>	<i>Symptômes somatiques généraux</i>
<i>HAMD 14</i>	<i>Symptômes génitaux</i>
<i>HAMD 15</i>	<i>Hypocondrie</i>
<i>HAMD 16</i>	<i>Perte de poids</i>
<i>HAMD 17</i>	<i>Prise de conscience</i>

Tableau 1 : Échelle de Hamilton (Items et Symptômes)

## 2.2. SCL90:

L'échelle SCL90 est un instrument d'autoévaluation, c'est-à-dire que c'est le patient lui-même qui le remplit. Cette échelle est composée de 90 phrases courtes décrivant des symptômes, les réponses sont cotées de 0 (« non ») à 4 (« oui, extrêmement »). Le patient choisit de cocher la case qui lui semble concorder le mieux avec son ressenti. Ici encore la somme de tous les items aboutit à un score global de dépression. Cependant on peut diviser les items de l'échelle en dix sous-scores représentant un aspect particulier de la dépression, comme la somatisation, l'anxiété, l'hostilité, etc.

<i>Items</i>	<i>Symptômes</i>
1-4-1-27-40-42-48-49-52-53-56-58	Somatisation
3-9-10-28-38-45-46-51-55-65	Symptômes obsessionnels
6-21-34-36-3-41-61-69-73	Sensibilité interpersonnelle ou vulnérabilité
5-14-15-20-22-26-29-30-31-32-54-71-79	Dépression
2-17-23-33-39-57-72-78-80-86	Anxiété
11-24-63-67-74-81	Hostilité
13-25-47-70-75-82-50	Phobies
8-18-43-68-76-83	Traits paranoïaques
7-16-35-62-77-84-85-87-90-88	Traits psychotiques
19-44-59-60-64-66-89	Symptômes divers

Tableau 2 : SCL\_90 (Items et Symptômes)

## 3. Jeux de données

Dans une étude d'épidémiologie clinique, 146 patients déprimés ont été évalués à J0 (début du traitement), J4, J7, J14, J21, J28, J42 et J56, à l'aide de deux outils : un outil d'hétéro-évaluation (échelle de dépression de Hamilton ou HDRS) ; et un outil d'auto-évaluation (SCL 90 pour Symptom distress Check-List de 90 items).

Le jeu de données étudié est composé de 3 fichiers csv :

- Un fichier séparant les patients en deux groupes hommes et femmes (groupe 1 et groupe0), c'est l'outil groupe.
- Un fichier de données sur les résultats de l'hétéro-évaluation (échelle de Hamilton HDRS) pour chaque visite : c'est l'outil hrds de 1967 modifié à 17 questions (HDRS-17), cette échelle est issue d'un questionnaire à choix multiples, mesurant la sévérité des symptômes associées à la dépression, le clinicien opte pour l'une des réponses des 3 à 5 réponses proposés en examinant son malade, ces réponses sont classées par ordre croissant de sévérité.

- Et un fichier de données sur les résultats de l'auto-évaluation (SCL 90 R) pour chaque visite : « outils autoeval ». Comme son nom l'indique, cette échelle comporte 90 descriptions courtes des plaintes et symptômes du patient.

## 4. Objectif du travail

L'objectif de ce travail est de déterminer si les patients du groupe 1 répondent mieux au traitement que les patients du groupe 0, d'après l'échelle de dépression de Hamilton qui est évaluée dans le temps sur plusieurs consultations, ces consultations ont été programmées à l'avance à J0.

## 5. Techniques et Méthodes

Après une revalidation rapide de l'échelle de dépression de Hamilton, des analyses sont effectuées afin de déterminer l'impact du groupe dans la réponse au traitement.

La revalidation de l'échelle de dépression de Hamilton (HDRS) a été réalisée sur les scores des temps J0 et J56.

Après une analyse descriptive du jeu de données, la structure dimensionnelle de l'échelle a été étudiée par des diagrammes de valeurs propres et des analyses factorielles. La consistance interne a été évaluée par le calcul des coefficients de Cronbach.

Enfin, la validité concourante de l'échelle a été appréciée à l'aide de corrélations avec un autre score (SCL 90 R : score d'autoévaluation) et d'une analyse en composante principale focalisée.

L'impact du groupe dans la réponse au traitement à partir du score de l'échelle de Hamilton a été évalué selon trois approches successives :

- Par une approche LOCF (Last Observation Carried Forward),
- Puis par un modèle mixte,
- Et enfin, en considérant le critère binaire censuré « réponse au traitement » défini par une chute de 50% à l'échelle de Hamilton par rapport à J0 (début de traitement).

Toutes les analyses statistiques ont été réalisées avec le logiciel R. studio (version1.0.16).

## 6. Data Management

Nous disposons pour l'analyse de 3 fichiers : le fichier groupe (2 sous-groupes de patients), le fichier autoévaluation (SCL 90) et le fichier hdrs (échelle de Hamilton).

### 6.1. Le fichier groupe :

Le fichier « outils groupe » correspond à la répartition des patients en deux groupes en fonction du traitement anti déprimeur administré. Il comporte 146 observations.

Groupe de Traitement	Groupe 0	Groupe 1
Nb.Sujets	75	71

Tableau 3 : Nb de sujets dans chaque groupe

On constate visuellement que les numéros semblent avoir été distribués de manière aléatoire et les différences entre les numéros des individus appartenant aux groupes 0 et 1 sont explicables par le hasard.

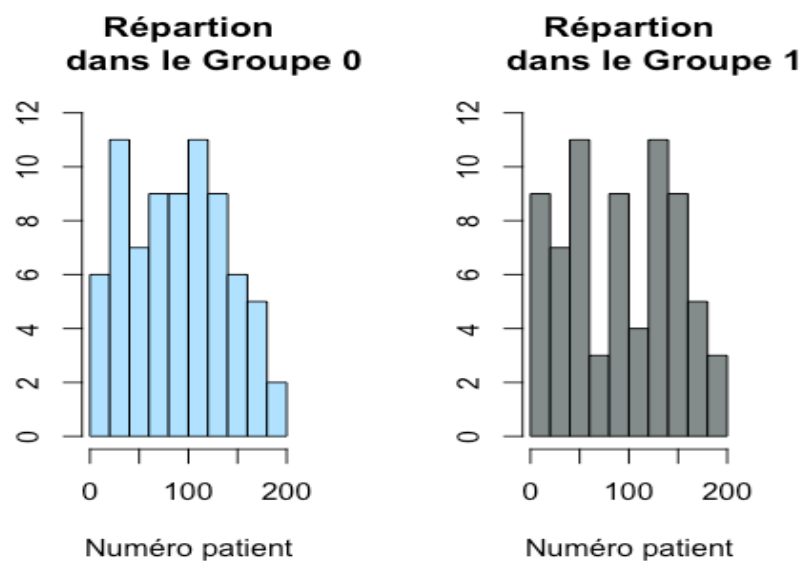


Figure 1 : Répartition des numéros en fonction des groupes

## 6.2. Le fichier d'hétéro\_évaluation (questionnaire HRDS) :Échelle de Hamilton

Le fichier comprend 20 variables pour 1053 observations :

- 18 items du score HAMILTON (du fait du 16A et 16B)
- 1 variable « VISIT » qui est une variable catégorielle à 8 classes
- 1 variable « NUMERO » qui correspond à l'identification des patients inclus.

On note en observant la répartition des patients entre J0 et J56 que 26 patients ont « disparus » soit 17.8% qui ont de valeurs manquantes à la fin de l'étude.

Visit	J0	J4	J7	J14	J21	J28	J42	J56	<NA>
Nb.Patients	146	139	137	133	129	127	122	120	0

Tableau 4 : Nb de patients pour chaque visite

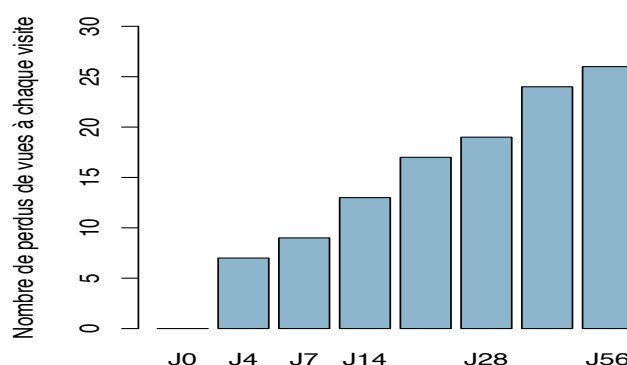


Figure 2 : Nb de patients perdus à chaque visite

**Il n'y a pas de données aberrantes** dans ce jeu de données (aucune valeur > 4)

On s'aperçoit qu'aucune donnée n'a été remplie pour le sujet numéro « 128 » à la visite de J7. Les items 16A(*perte de poids rapporté*) et 16B(*perte de poids mesurée*) sont mutuellement exclusifs dans le score HAMILTON ce qui explique que soit l'item 16A est rempli( auto evaluation) soit l'item 16B est rempli. On observe donc 964 données manquantes pour l'item 16B et 90 pour l'item 16A.

Pour pouvoir obtenir un seul item informatif et exploitable, les NA ont été substitués par des 0 et les deux items ont été fusionnés par une somme. L'item unique HAMD16, prenant des valeurs de 0 à 2 a ainsi été créé.

### 6.2.1. Analyse descriptive de la somme des items de l'échelle de Hamilton à J0 et J56 :

	Min	1 <sup>er</sup> Qu	Median	Moyenne	3 <sup>ème</sup> Qu	Max
J0	13.00	25.00	27.00	27.92	30.00	44.00



J56                      0.00                      4.00                      7.50                      7.08                      10.00                      27.00

Pour commencer, le tableau suivant présente les différents indicateurs statistiques de la somme du score de Hamilton à J0 et J56 pour l'ensemble des patients.

On peut noter sur ce tableau, une baisse globale du score moyen de Hamilton au fur et à mesure des visites(entre J0 et J56).

En ce qui concerne la variabilité des scores, celle-ci a tendance à être plus importante entre J0 et J56, avant de diminuer vers la fin du traitement J56.

On évalue graphiquement la somme des items de l'échelle de Hamilton à J0 et J56 :

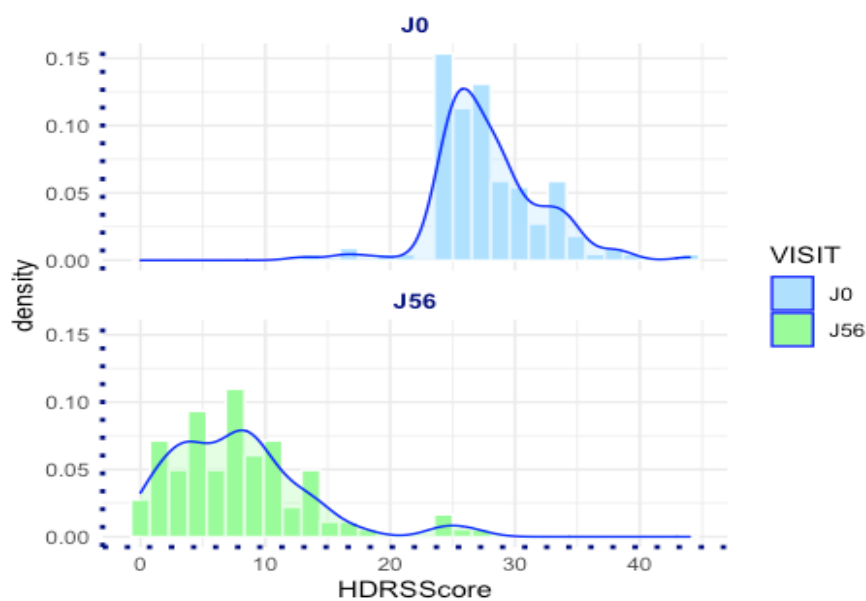


Figure 3 : Score globale de L'échelle Hamilton à la visite J0 et J56

Les courbes de densités ne semblent en aucun cas suivre la loi normale. On remarque aussi que le score global de Hamilton s'est réduit entre J0 et J56 .

On constate qu'à J56, il reste encore des scores élevés, Cela peut être expliquer par l'existence de certain individu qui sont toujours en dépression.

### 6.3. Le fichier d'auto évaluation (questionnaire SCL 90):

Le fichier « outils autoeval » contient 1034 observations, et 92 variables (NUMERO, VISIT et les réponses aux 90 questions du questionnaire SCL 90).

Les variables « question 1 » à « question 64 » sont codées comme des variables numériques.

Les variables « question 65 » à « question 90 » sont codées comme des variables catégorielles à plusieurs classes.

L'analyse des données montre qu'un certain nombre de variables prennent des valeurs  $> 4$  et l'existence de 19 valeurs manquantes récurrentes. Ceci n'est pas possible dans l'échelle SCL 90 et il s'agit donc de **données « aberrantes »**. Je fais le choix de considérer et **recoder ces valeurs en tant que données manquantes (NA)**.

#### 6.4. Description du fichier unifié :

Avant de commencer l'analyse à proprement parlée des données, je procède à la fusion des 3 fichiers pour obtenir un fichier unifié. Cela sera utile pour les analyses ultérieures. Le fichier unifié ne contient pas de doublons

Je crée deux versions du fichier fusionné:

- Un fichier normalisé (1 ligne par visite et par patient, soit 110 variables de 1053 observations) qui sera utilisé pour l'analyse du modèle mixte.
- Un fichier dénormalisé (1 ligne par sujet, 873 variables et 146 observations) qui sera utilisé pour l'analyse LOCF (question 2).

## 7. Validation de l'échelle de Hamilton

### 7.1. Analyse de la distribution d'items et de score à J0 et J56

On veut évaluer si l'échelle de Hamilton est interprétable dans notre population d'étude. Nous ne vérifierons ici que les temps J0 et J56. L'existence d'une seule valeur manquante n'influencera pas les résultats

L'évaluation de la validité de l'échelle s'effectue en plusieurs étapes complémentaires : - Vérifier que l'instrument mesure bien ce que l'on veut mesurer

- Analyse individuelle des items inclus dans l'échelle, recherche d'effets plafonds / plancher
- Analyse de la structure multidimensionnelle de l'échelle - Etude de la reproductibilité de la mesure

La lecture des items inclut dans l'échelle confirme la validité de construction de cette échelle et montre qu'elle est effectivement adaptée à l'évaluation de la dépression. L'échelle comprend essentiellement des items portant sur la composante somatique de la dépression.

#### **A J0 :**

Il y a 146 patients inclus à J0. Ce nombre est « relativement » faible pour évaluer un instrument de mesure (on recommande généralement un chiffre > 300). A J0, les données sont complètes. La majorité des items suivent globalement une distribution normale.

Suite à une description visuellement de la figure 4, on remarque que :

- Un effet plafond est remarqué pour l'item 1, 4, 10, 13,16,15.
- Un effet plancher est remarqué pour l'item 3, 7 ,8,11, 17.

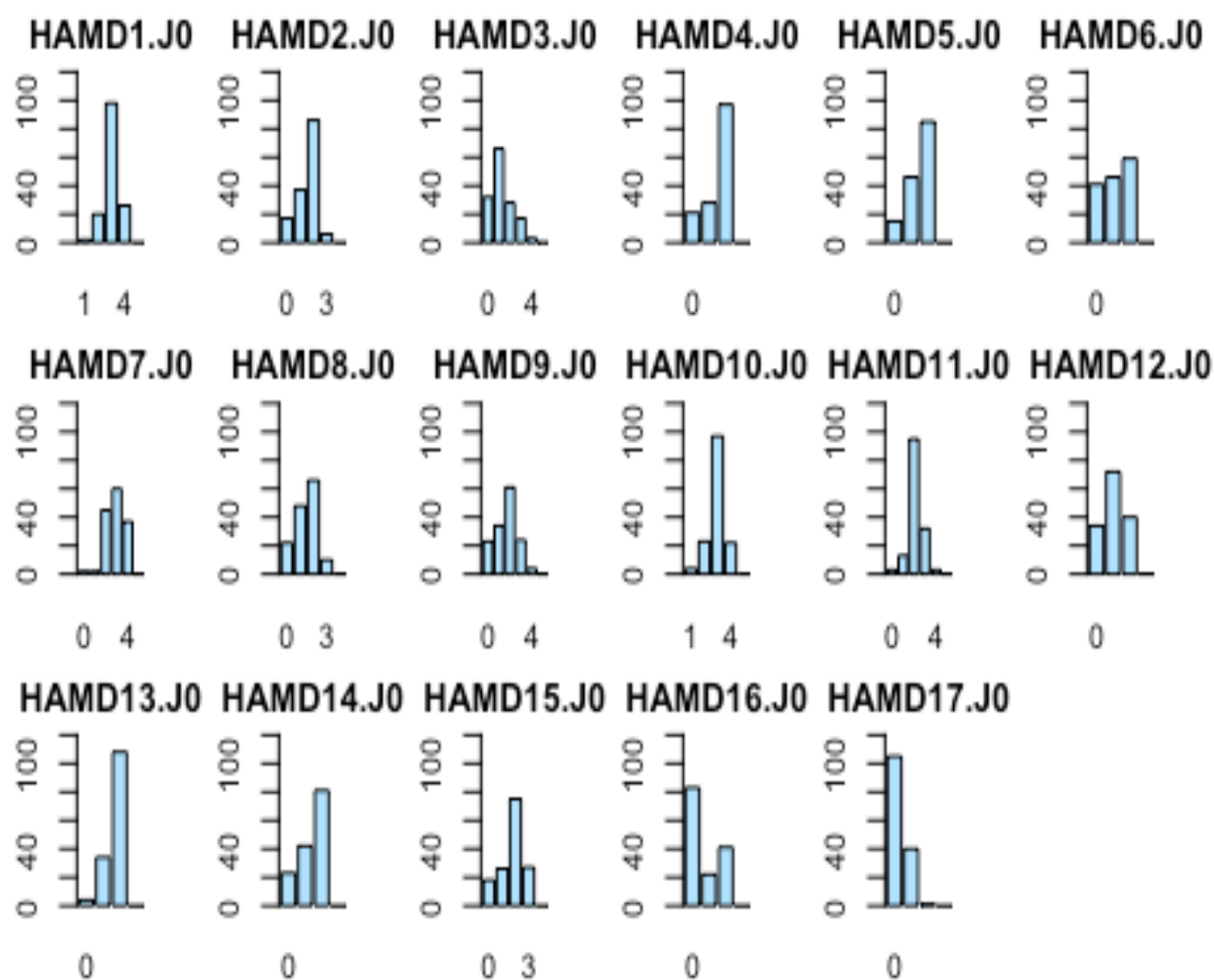


Figure 4 : Barplots des items de l'échelle de Hamilton à J0(pas de données manquantes)

#### A J56 :

Les mêmes analyses sont réalisées à J56.

A J56, il reste 120 sujets inclus analysables. Les données des patients restants sont complètes.

- On remarque un effet plancher des items 3, 16 et 17.

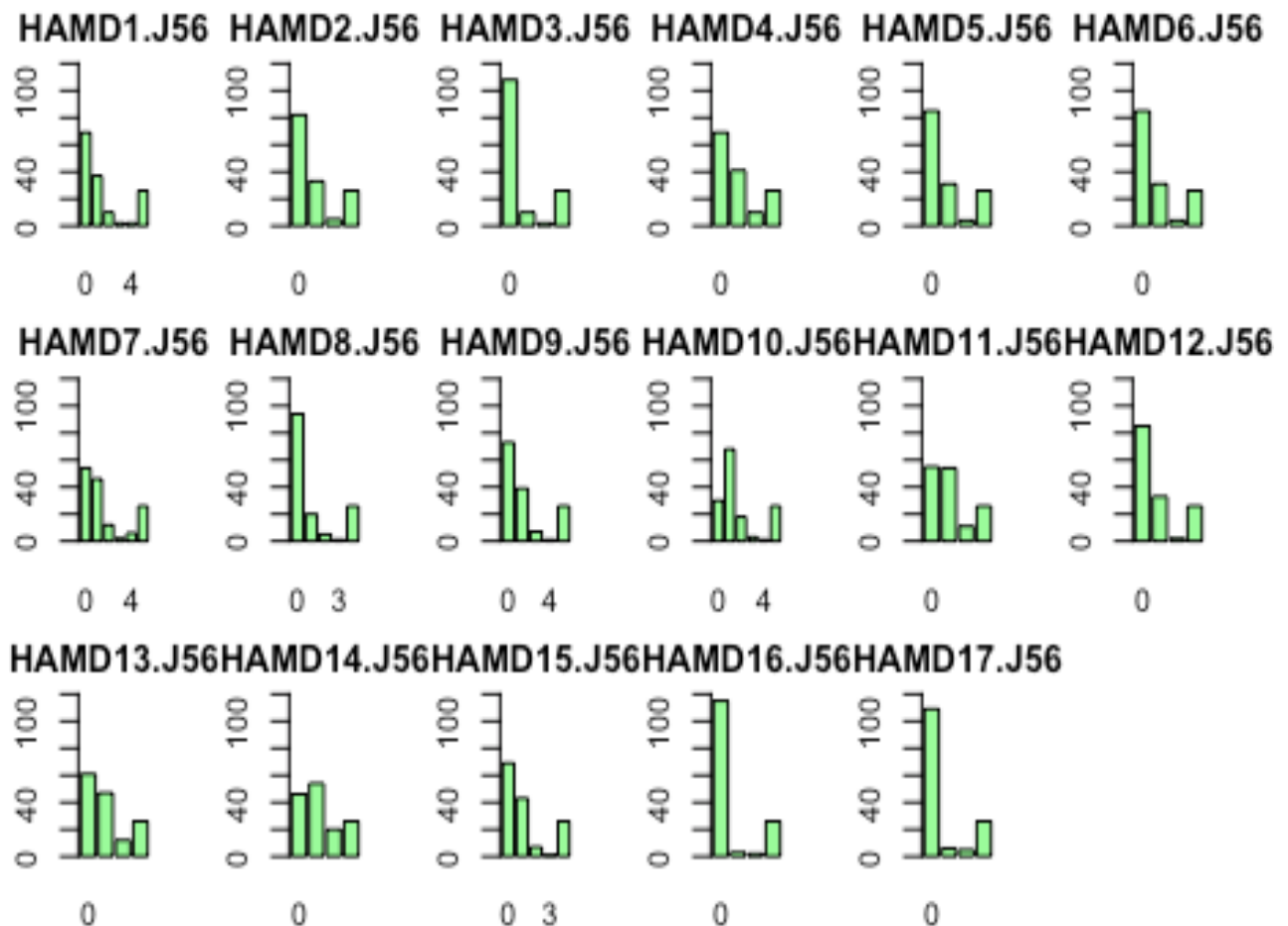


Figure 5: Barplots des items de l'échelle de Hamilton à J56(26 données manquantes)

## 7.2 – Analyse de Corrélations entre les items (redondance)

Afin de rechercher la présence d'items redondants, une matrice de corrélation a été réalisée grâce à la fonction « corplot » entre les différents items (deux à deux) à J0 et J56.

### A J0 :

- On constate qu'entre l'item HAMD12 et HAMD16 , la corrélation est égale à 0.44 (la plus élevée).
- qu'entre l'item HAMD13 et HAMD4 , la corrélation est égale à 0.35.
- Aucune corrélation n'est supérieur ou égale à 0,5 nous pouvons dire qu'il n'y a pas de redondance d'items à J0.

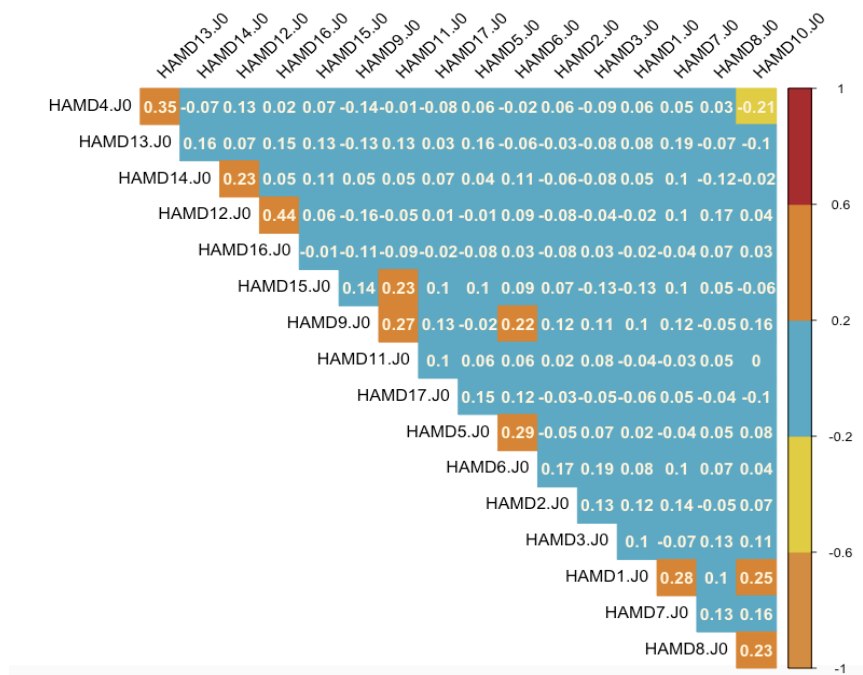


Figure 6 : Matrice de corrélation à J0

#### A J56 :

- On constate qu'entre l'item HAMD1 et HAMD3 , la corrélation est égale à 0.67 (la plus élevée).
- qu'entre l'item HAMD1 et HAMD7 , la corrélation est égale à 0.63.
- qu'entre l'item HAMD3 et HAMD8 , la corrélation est égale à 0.61.

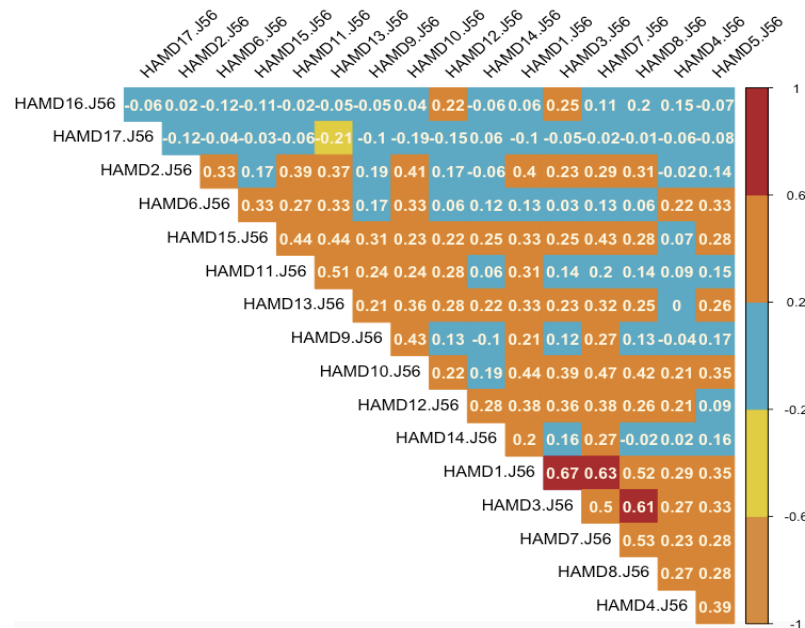


Figure 6 : Matrice de corrélation à J56

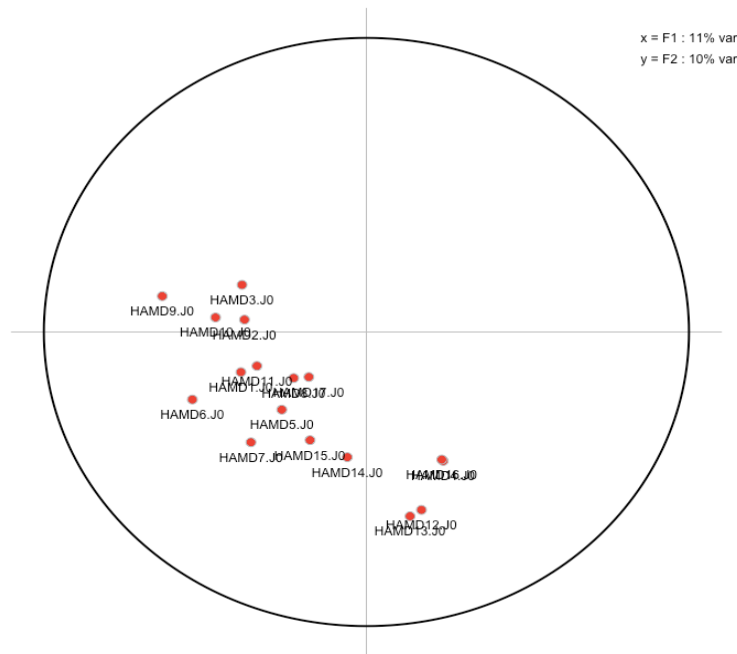


Figure 7 : Analyse de composante principale à J0

### A J0 :

- Cette outil semble moins adéquat pour montrer les relation entre les items puisque le nombre de ces derniers est un peu grand. Les items sont également un peu loin de bord du cercle.

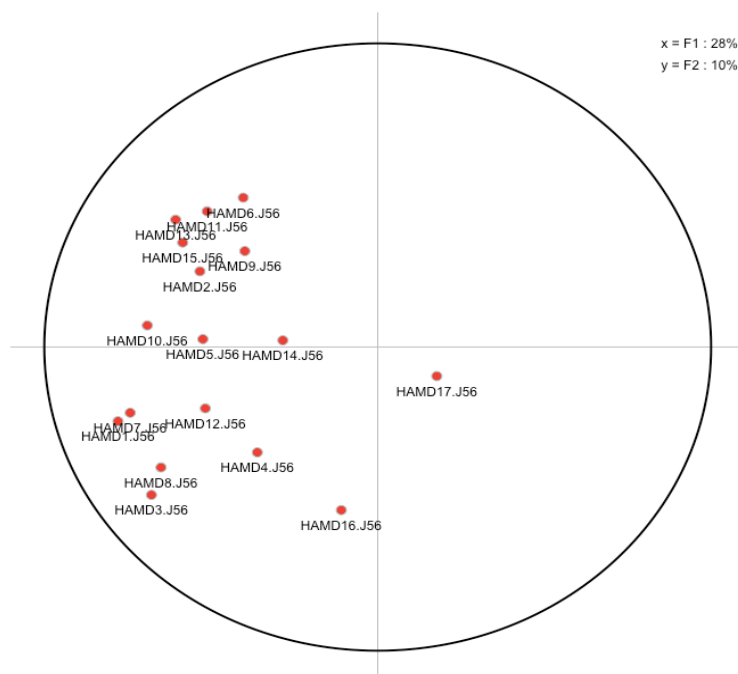


Figure 8 : Analyse de composante principale à J56

### **A J56 :**

- Les items sont un peu plus proches du bord du cercle.
- On confirme la corrélation par la distance trop proche entre HAMD1 et HAMD7.
- On confirme également la corrélation par la distance trop proche entre HAMD3 et HAMD8.

**Un plus :** La classification hiérarchique peut aussi être un très bon moyen de visualisation pour confirmer la corrélation entre les items .

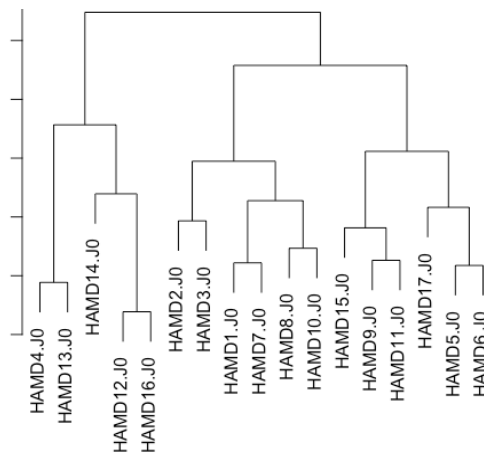


Figure 9 : Classification hiérarchique à J0.

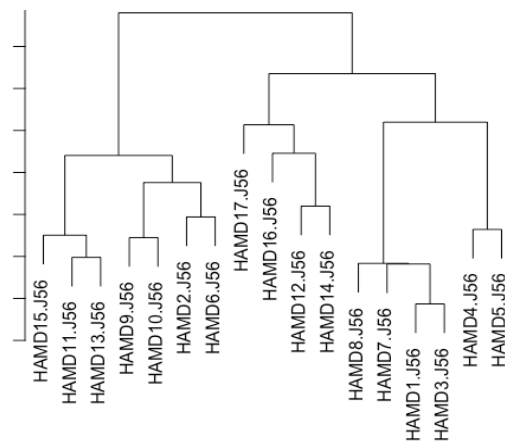


Figure 10: Classification hiérarchique à J56

### **A J0 :**

- On confirme les résultats de corrélations trouvées au-dessus (HAMD12 et HAMD16 , HAMD 4 et HAMD13)

### **A J56 :**

- On confirme les résultats de corrélations trouvées au-dessus (HAMD1 et HAMD7 , HAMD 3 et HAMD8, HAMD 1 et HAMD3).

## **7.3 -- Structure dimensionnelle : Unidimensionnalité**

Dans cette partie, afin de développer les résultats obtenus précédemment, une analyse factorielle a été réalisée, pour laquelle il a fallu, au préalable, sous R, construire les diagrammes des valeurs propres à J0 et J56 afin de déterminer le nombre de facteur à retenir par la suite.

Pour savoir combien de dimensions on évalue, on réalisera les diagrammes des valeurs propres :



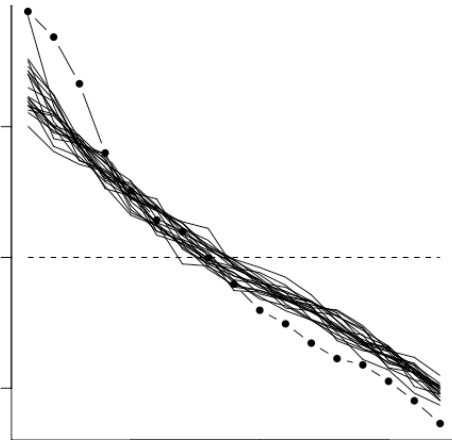


Figure 11 : Diagramme de valeurs propres HDRS à J0

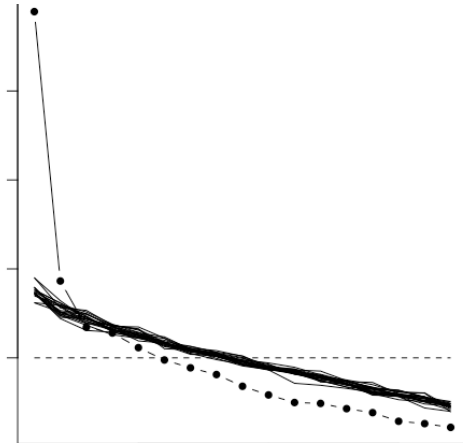


Figure 12 : Diagramme de valeurs propres HDRS à J56

**A J0**, on peut qualifier la structure de cette échelle de tridimensionnelle : en effet, 3 dimensions semblent se « détacher ».

**A J56**, On distingue une dimension principale à J56 et une deuxième beaucoup moins importante. A chacun son point de vue, on peut considérer la structure du score HDRS à J56 comme unidimensionnelle.

## 7.4 – Analyse factorielle

	Factor1	Factor2	Factor3
HAMD1.J0	0.006	0.028	0.263
HAMD2.J0	-0.057	-0.076	0.240
HAMD3.J0	-0.021	-0.120	0.207
HAMD4.J0	0.110	0.359	-0.050
HAMD5.J0	0.004	0.118	0.240
HAMD6.J0	0.132	-0.155	0.427
HAMD7.J0	0.124	0.121	0.320
HAMD8.J0	0.183	-0.105	0.101
HAMD9.J0	-0.107	-0.228	0.501
HAMD10.J0	0.066	-0.152	0.255
HAMD11.J0	-0.029	0.070	0.316
HAMD12.J0	0.992	0.044	-0.091
HAMD13.J0	0.046	0.977	0.194
HAMD14.J0	0.238	0.120	0.147
HAMD15.J0	0.076	0.089	0.226
HAMD16.J0	0.424	0.155	-0.100
HAMD17.J0	0.022	-0.004	0.176

Tableau 6 : Analyse factorielle à J0

	Factor1
HAMD1.J56	0.802
HAMD2.J56	0.468
HAMD3.J56	0.706
HAMD4.J56	0.333
HAMD5.J56	0.456
HAMD6.J56	0.282
HAMD7.J56	0.745
HAMD8.J56	0.657
HAMD9.J56	0.328
HAMD10.J56	0.624
HAMD11.J56	0.401
HAMD12.J56	0.471
HAMD13.J56	0.495
HAMD14.J56	0.252
HAMD15.J56	0.494
HAMD16.J56	0.125
HAMD17.J56	-0.132

Tableau 7: Analyse factorielle à J0

### A J0 :

Une analyse factorielle à 3 facteurs avec rotation varimax permettra de voir à quoi correspondent ces 3 dimensions.

Les 3 dimensions relevées pourraient correspondre à ces 3 facteurs :

- Facteur 1 : items 8 ; 12 ; 14 ; 16 (ralentissement, symptômes somatiques gastro-intestinaux, symptômes génitaux, perte de poids) : symptômes somatiques spécifiques.
- Facteur 2 : items 4 et 13 (insomnies, symptômes somatiques généraux) : symptômes somatiques généraux.
- Facteur 3 : items 1 ; 2 ; 3 ; 5 ; 6 ; 7 ; 9 ; 10 ; 11 ; 15 ; 17 (humeur dépressive, sentiments de culpabilité, suicide, insomnie du milieu de la nuit, insomnie du matin, travail et activités, agitation, anxiété psychique, anxiété somatique, hypocondrie, prise de conscience) : symptômes psychologiques.

#### **A J56 :**

L'échelle est unidimensionnelle et différente de J0.

### **7.5 – Validation de consistance interne**

**A J0**, afin d'évaluer la reproductibilité de la mesure, on utilise le coefficient de Cronbach. Sa valeur alpha est évaluée à 0.45594. Ce résultat ne paraît pas très satisfaisant, mais peut être lié au caractère multidimensionnel de l'instrument et donc **moins de reproductibilité**.

Par contre à **J56**, le coefficient est à 0.819954. Étant supérieur à 0.7, on peut dire que la consistance interne est bonne, ce qui correspond au caractère unidimensionnel observé.

### **7.6 -- Validité concourante**

Afin d'étudier la validité de la concourante de l'échelle de dépression de Hamilton, la corrélation du score total de l'échelle avec une autre échelle d'évaluation connue a été étudié.

L'échelle à laquelle les données ont été comparée correspond à une auto-évaluation : la Symptom check List SCL-90.

Dans cette partie, le but est d'analyser les corrélations entre score obtenu à l'échelle de Hamilton (aux visites J0 et J56), et les différents sous scores obtenus lors de l'autoévaluation SCL90, afin de démontrer si les réponses aux deux questionnaires (HAMILTON et SCL-90) sont corrélées, en particulier au niveau du score de dépression.

Voici la matrice de corrélations pour estimer les corrélations entre les échelles et sous-échelles(deux à deux).

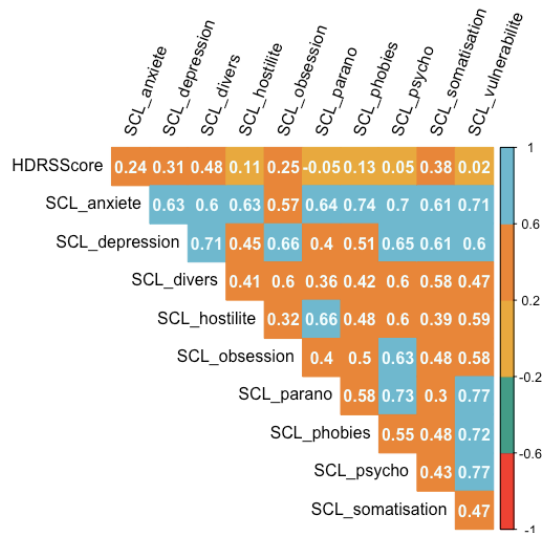


Figure 13 : Matrice de corrélation à J0

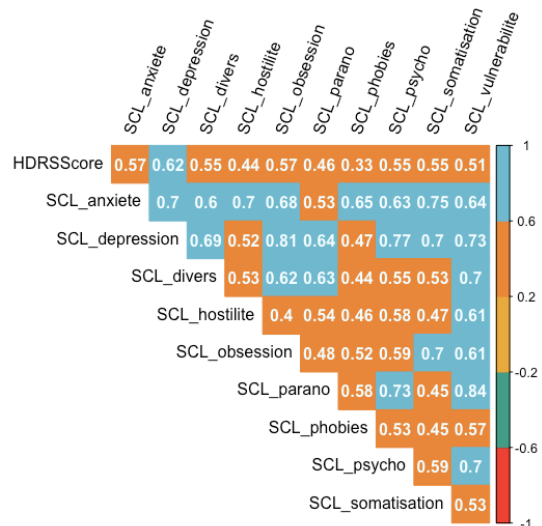


Figure 14 : Matrice de corrélation à J56

### A J0 :

On remarque que le score de Hamilton (HDRSScore) est **faiblement corrélé** à J0 aux sous-échelles anxiété, hostilité, obsession, parano, phobies, traits psychotiques, et vulnérabilité. Et Moyennement corrélé avec symptômes divers(0.48). Ce résultat est un peu banal si on prend en compte les données en avant et aussi l'absence de l'unidimensionnalité de l'échelle de Hamilton à J0.

### A J56 :

Il est **fortement corrélé** aux sous-échelles de dépression(0.62), anxiété, obsession, somatisation, traits psychotiques et symptômes divers.

*N.B : Une ACP à J0 et J56 a été également réalisée. (voir le script)*

## 8. Comparaison des deux groupes : Analyse de réponse au traitement

### 8.1 – Approche LOCF

La suite du devoir est consacrée à la comparaison des groupes de traitement 0 et 1. On recherche un lien entre le groupe de traitement et la diminution de la symptomatologie dépressive évaluée par l'échelle Hamilton.

Compte-tenu des 26 données manquantes à J56, une approche peut consister à considérer la dernière valeur disponible d'un sujet comme valeur du score HDRS à J56: c'est l'approche LOCF(**L**ast **O**bservation **C**arry **F**orward) .

On crée une variable dont la valeur est la différence entre le score à J56 et le score à J0.

Après avoir traité les données manquantes par la méthode LOCF, on réalise un test T de Student entre les moyennes de cette variable « différence » dans les deux groupes de traitement. On vérifie également des conditions de réalisation de ce test.

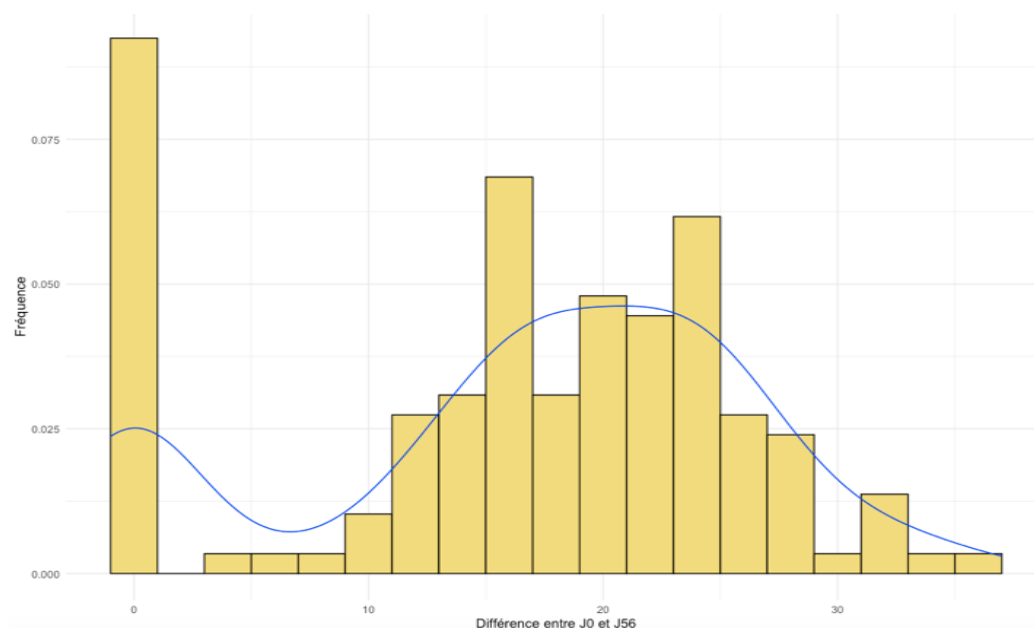


Figure 15 : Histogramme et courbe de densité de la différence entre J0 et J56

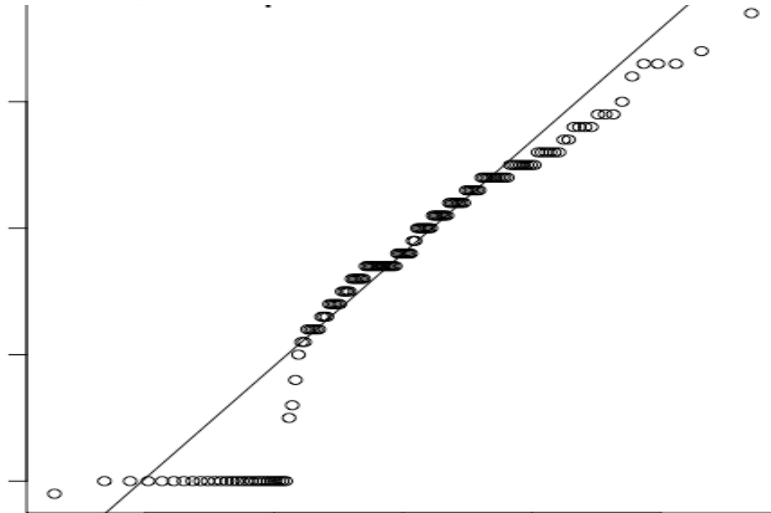


Figure 17 : *QQplot des résidus de la différence entre J0 et J56*

D'après l'histogramme, la distribution de la variable différence ne semble pas suivre une loi normale. Mais on a des effectifs dans chaque groupe (Groupe0 = 75 et Groupe1 = 71) qui sont quand même grand et aussi les variances entre les 2 groupes sont proches (Groupe0 = 9.638268 et Groupe1 = 8.966201). Donc, on utilisera le test de Student.

On obtient **0.00061**. Les moyennes des différences J0-J56 sont donc **statistiquement significatives** (au seuil de 5%) entre les groupes. Le test Student montre également que la **moyenne de différence** est plus élevée chez le Groupe0 (**19.32000**) avec une moyenne de différence chez le Groupe1 qui est égale à **13.91549**.

## 8.2 – Modèle mixte

L'utilisation d'un modèle mixte permet d'évaluer la réponse au traitement selon le groupe, en tenant compte de l'existence d'un effet sujet aléatoire. On effectuera un modèle mixte où les individus représentent la variable aléatoire, tandis que les visites et les groupes sont les variables fixes.

L'évolution du score total de l'échelle de Hamilton en fonction du temps selon le groupe est représentée graphiquement sur la figure 18.

Vérification de type de l'allure de scores en fonction du temps (à chaque visite) : Allure à peu près linéaire.

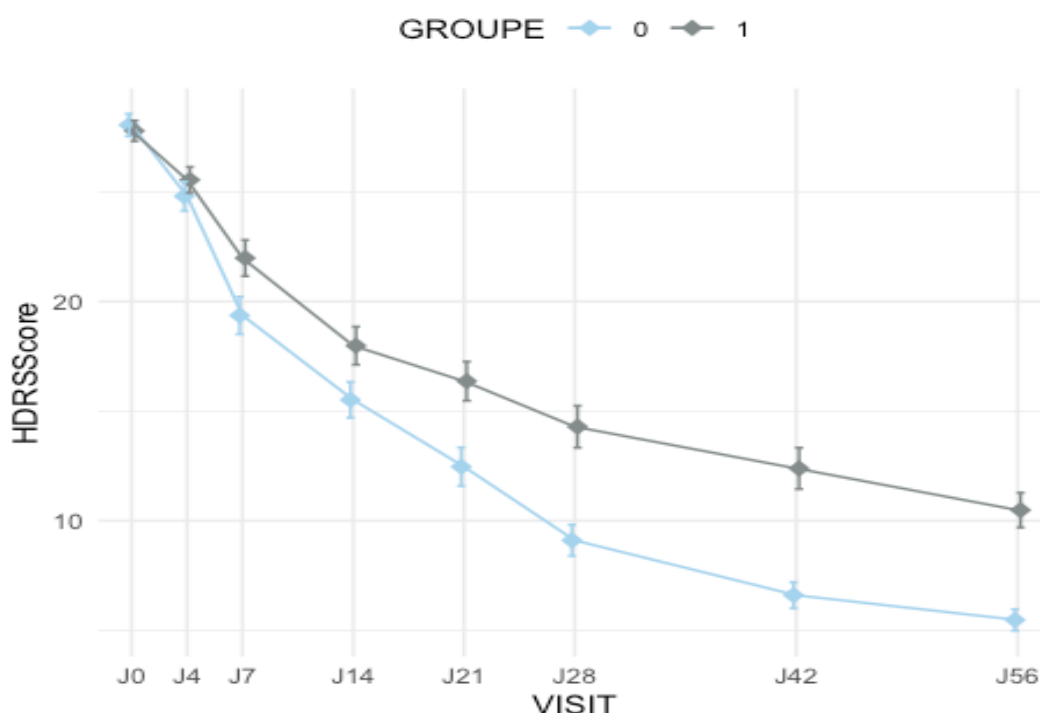


Figure 18 : Evolution temporelle du score de l'échelle de Hamilton dans les deux groupes

On réalise également un modèle mixte où on suppose qu'il y a une **interaction** entre la variable groupe et la variable visite :

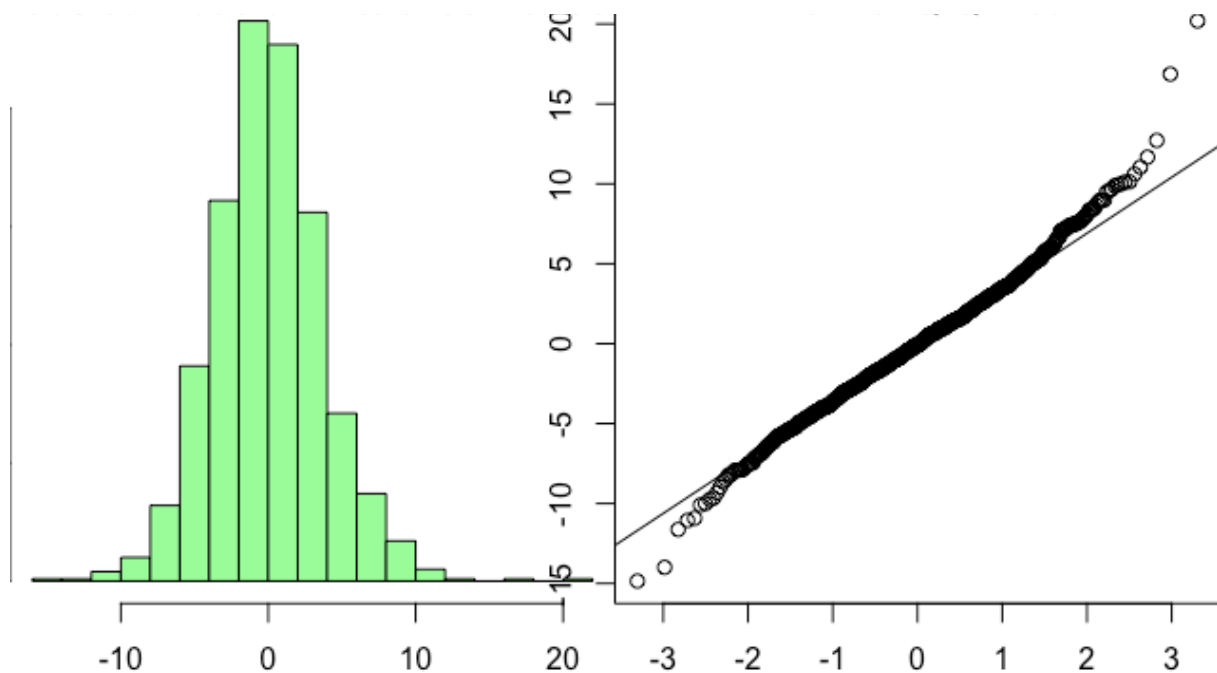
$$\text{HDRS} = a + b \cdot \text{GROUPE} + c \cdot \text{temps} + d \cdot (\text{GROUPE} \cdot \text{temps}) + \text{effet-sujet} + \varepsilon$$

- Temps : numéro de la visite (1 à 8)
- GROUPE : numéro du groupe (0 ou 1)

- Effet-sujet : Effet aléatoire correspondant à la variable NUMERO  
Avec ce modèle, avec un intercept « effet sujet » aléatoire, tester un effet du groupe revient à rechercher s'il y a une interaction significative entre le temps et le GROUPE, c'est-à-dire de tester si « d » est significativement différent de 0. La réponse au traitement est significativement différente entre le groupe 0 et le groupe 1 en fonction du temps ( $p = 0.003$ ) à un seuil de 5%.

#### ***Vérification des conditions de validité :***

On vérifie la normalité des résidus graphiquement . La normalité semble très bien respectée ici, le modèle peut donc être considéré comme valide.



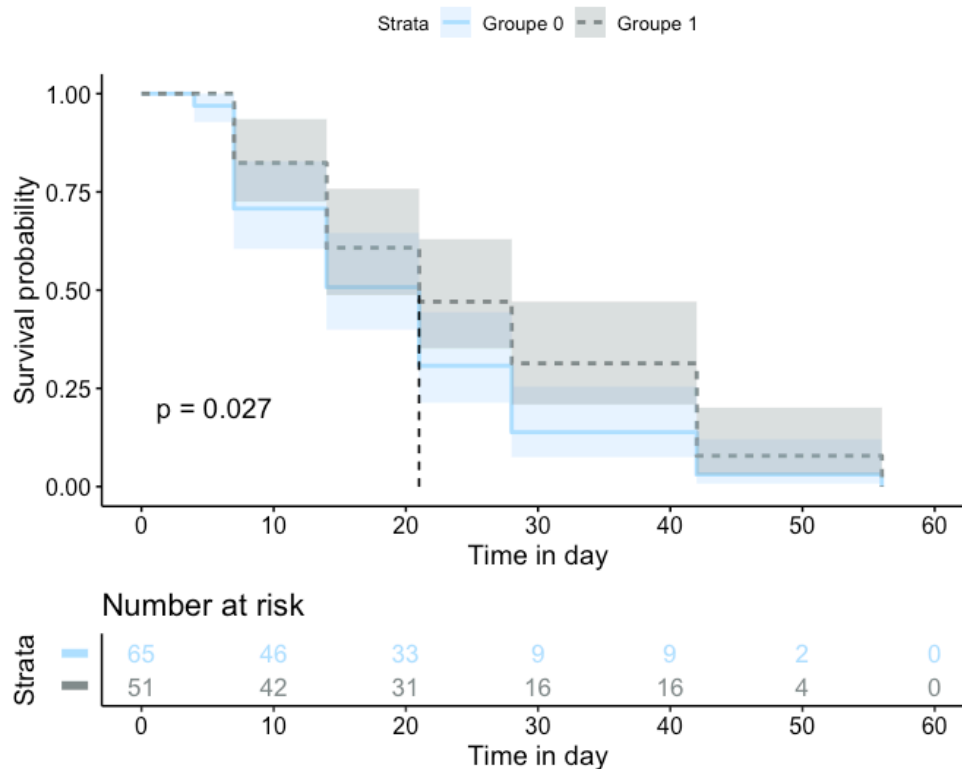
*Figure19 : Représentations graphiques des résidus du modèle mixte.*

## 9. Analyse de la Survie : Analyse de la réponse binaire au traitement

La réponse au traitement peut être considérée comme une diminution d'au moins 50% du score HDRS entre J0 et la dernière visite. Une nouvelle variable binaire est donc créée. Chaque patient perdu de vue est censuré à la date de perte de vue.

Afin d'analyser ces données, nous réaliserons une **courbe de Kaplan-Meier** puis un **test du Log-Rank** pour comparer les proportions de « survie » entre les deux groupes en fonction du temps.

La courbe ci-dessous représente la réponse au traitement dans les deux groupes, selon la méthode de Kaplan-Meier :



On réalise un log-rank entre les deux groupes et on observe une meilleure réduction du score dans le groupe 0. La médiane de temps avant le succès est de **14 jours** dans le groupe 0, tandis qu'elle est de **28 jours** dans le groupe 1

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
GRUPE=0	65	65	55.8	1.51	4.89
GRUPE=1	51	51	60.2	1.40	4.89

Chisq= 4.9 on 1 degrees of freedom,  $p = 0.03$



La réponse dans le groupe 0 semble plus importante et plus rapide sur ce graphique. La médiane de survie est de 21 jours dans le groupe 0 et de 28 jours dans le groupe 1. Le test du LogRank confirme cette impression en comparant les 2 groupes.

## 10.Script

```
install.packages("prettyR")
install.packages("corrgram")
install.packages("corrplot")
install.packages("dplyr")
install.packages("psy")
install.packages("ggplot2")
install.packages("Hmisc")
install.packages("lme4")
install.packages("survival")
install.packages("ltm")
install.packages("psych")
install.packages("tidyverse")
install.packages("summarytools")
install.packages("reshape2")
install.packages("wesanderson")
install.packages("survminer")
```

### Choix et téléchargement des librairie indispensables par ordre de besoin :

```
library(prettyR)
library(corrgram)
library(corrplot)
library(dplyr)
library(psy)
library(ggplot2)
library(Hmisc)
library(lme4)
library(survival)
library(ltm)
library(psych)
```

```
library(tidyverse)

library(survminer)

library(wesanderson)
```

```
##=====Data management=====
```

```
###=====Importation des fichiers=====
```

```
df_grp<-read.csv2("outils_groupe.csv",sep = ",")
df_hamilton<-read.csv2("outils_hdrs.csv",sep = ",")

## le fichier df_scl90 contient des "ND":
df_scl90<-read.csv2("outils_autoeval.csv",sep="," ,na.strings=c("NA","ND"))
```

```
##=====Fusion des fichiers=====
```

```
### Verification des données, colonnes, lignes
```

```
str(df_grp)
str(df_hamilton)
str(df_scl90)
```

```
### Recherche de données manquantes
```

```
sum(is.na(df_grp))
sum(is.na(df_hamilton))
sum(is.na(df_scl90))
```

```
###===== Données manquantes et aberrantes=====
```

```
#### Fichier df_Hamilton
```

```
### Fusion de HAMD16A et HAMD16B en HAMD16 :
```

```

## Données manquantes

## Je constate que l'item 16 de l'échelle de Hamilton est scindé en 2(HAMD16A & HAMD16B)
donc on le

## fusionne en une seule colonne HAMD16:

df_hamilton$HAMD16 <-
ifelse(is.na(df_hamilton$HAMD16A)==F,df_hamilton$HAMD16A,df_hamilton$HAMD16B)

head(df_hamilton)

### Suppression des 2 colonnes HAMD16A et HAMD16B

df_hamilton <- df_hamilton[,names(df_hamilton)!="HAMD16A"&
names(df_hamilton)!="HAMD16B"]

## Réorganiser l'ordre de les colonnes: HAMD16 et HAMD17

df_hamilton<-df_hamilton%>%select("NUMERO": "HAMD15", "HAMD16", "HAMD17")


### Valeurs aberrantes

### je recherche les valeurs aberrantes: pas de valeurs aberrantes pour le fichier hamilton( scores des
items entre 0 et 4)

apply(df_hamilton[,3:16],2,table,useNA="always")


## L'existence d'une seule valeur manquantes pour tout les items du Score de Hamilton laisse lace
au doute s'il s'agit pas d'une valeur manquante récurrente

## On cherche l'individu qui pour une visite le Score Hamilton n'a pas été mesuré (valeur NA
Conjointe):

## patient numéro 128

df_hamilton[is.na(df_hamilton$HAMD1), 1:19]

#####Fichier df_sc190=====

## On constate l'existance de plusieurs valeurs abberantes qui normalement ne doit pas dépasser un
score de 4

## Et on trouve plusieurs valeurs manquantes également

apply(df_sc190[,3:92],2,table,useNA="always")


## En analysant les résultats l'item Q4 présente plus de nomre de données aberrantes = 13 valeurs
ce qui représente un proportion de 1.26%

```

#### Pour gérer l'ensemble des valeurs aberrantes, on va les considérer comme valeurs manquantes( on les transforme)

```
is.na(df_scl90[,3:92]) <- df_scl90[,3:92] > 4
```

```
## verification des valeurs aberrantes == 0 valeurs aberrantes
```

```
df_scl90 %>% select("Q1":"Q90") %>%
```

```
  filter_all(all_vars(> 4))
```

```
#### Fusionner des fichiers en format long
```

```
df_intermediaire <- merge(df_hamilton,df_scl90, by=c("NUMERO","VISIT"),all=TRUE)
```

```
df_tot <- merge(df_intermediaire,df_grp, by="NUMERO",all=TRUE)
```

```
## Réorganiser l'ordre de les colonnes:
```

```
df_tot <- df_tot %>% select("NUMERO", "VISIT", "GROUPE", "HAMD1":"Q90")
```

```
#### Verification du nouveau fichier "df_tot":
```

```
length(df_tot$NUMERO) ## 1053
```

```
length(unique(df_tot$NUMERO)) ## 146
```

```
length(df_tot$NUMERO) / length(unique(df_tot$NUMERO)) #### Moyenne de visite par patient
```

```
sum(duplicated(df_tot)) ## aucune valeur dupliquée
```

```
str(df_tot)
```

```
#### on regarde une 2ème fois les valeurs manquantes après fusion pour le score de SCL90 pour
```

```
## voir les valeurs manquantes existe pour le même sujet de Hamilton ( 19 valeurs manquantes récurrentes)
```

```
df_tot %>% select("Q1":"Q90") %>% filter_all(all_vars(is.na(.)))
```

```
NA_SCL90 <- df_tot[is.na(df_tot$Q11), c(1:2, 21:110)]
```

```
NA_SCL90
```

```
####=====Nombre d'individus dans Chaque groupe=====
```

```
####=====Nombre d'individus dans Chaque groupe=====
```

```
table(df_grp$GROUPE)
```

### présentation graphique de la répartition des individus dans chaque groupe en fonction de leur Numéro

```
par(mfrow=c(1,2))
```

```
hist(unique(df_tot$NUMERO[df_tot$GROUPE == 0]), main = "Répartition
dans le Groupe 0",xlab = "Numéro patient", xlim =
c(0,max(df_tot$NUMERO)+10),ylab = "", ylim = c(0,12), col =
"lightskyblue1")
```

```
hist(unique(df_tot$NUMERO[df_tot$GROUPE == 1]), main = "Répartition
dans le Groupe 1",xlab = "Numéro patient", xlim =
c(0,max(df_tot$NUMERO)+10),ylab = "", ylim = c(0,12), col =
"azure4")
```

```
###=====Nombre de Chaque visite pour chaque groupe(groupe1 et groupe0)=
table(df_tot$GROUPE, df_tot$VISIT, deparse.level = 2)
```

```
## En faisant le calcul , on découvre qu'on passant de J0 à J56
```

```
## on passe de 146 à 120 patients et ce qui veut dire qu'on perd 17.8% de la populations
```

```
###=====Nombre de patients à chaque visite=====
```

```
df_hamilton$VISIT<-
ordered(df_hamilton$VISIT,levels=c("J0","J4","J7","J14","J21","J28","J42","J56"))
```

```
tab<-table(df_hamilton$VISIT)
```

```
barplot(146-tab,main = "Nombre de perdus de vues à chaque visite",ylim = c(0,30),
ylab = "Nombre de perdus de vues à chaque visite",col = "lightskyblue3")
```

```
###=====Echelle de Hamilton=====
```

```
## Création d'une variable de score de Hamilton des 17 Questions d'une visite
```

```
df_tot$HDRSScore <-apply(df_tot[,4:20],1,sum, na.rm=TRUE)
```

```
## Réorganisation d'ordre de la variable dans le dataframe
```

```
df_tot <- df_tot %>% select("NUMERO":"HAMD17","HDRSScore","Q1":"Q90")
```

```
### on crée des 2 sous-groupes entre les 2 visites J0 et J56
```

```

HDRSScore_J0 <- df_tot$HDRSScore[df_tot$VISIT=="J0"]
HDRSScore_J56 <- df_tot$HDRSScore[df_tot$VISIT=="J56"]

summary(HDRSScore_J0)
summary(HDRSScore_J56)

####==== Repartition de la somme du Score Hamilton à J0 et J56 en fonction des 2 groupes 0 et 1
par(mfrow=c(2,2))

hist(HDRSScore_J0[df_tot$GROUPE=="0"],main = "Répartition du score Hamilton à
J0/Groupe0",
     xlab = "Score de Hamilton",ylim=c(0,40),
     col = "lightskyblue1",ylab = "Effectif")

hist(HDRSScore_J0[df_tot$GROUPE=="1"],main = "Répartition du score Hamilton à
J0/Groupe1",
     xlab = "Score de Hamilton",ylim=c(0,40),
     col = "lightskyblue1",ylab = "Effectif")

hist(HDRSScore_J56[df_tot$GROUPE=="0"],main = "Répartition du score Hamilton à
J56/Groupe0",
     xlab = "Score de Hamilton",ylim=c(0,40),
     col = "palegreen1",ylab = "Effectif")

hist(HDRSScore_J56[df_tot$GROUPE=="1"],main = "Répartition du score Hamilton à
J56/Groupe1",
     xlab = "Score de Hamilton",ylim=c(0,40),
     col = "palegreen1",ylab = "Effectif")

####=====Distribution de la somme du Score Hamilton à J0 et J56=====
ggplot(df_tot[df_tot$VISIT == "J0"|df_tot$VISIT == "J56", ], aes(x = HDRSScore, fill=VISIT)) +
  geom_histogram(aes(y =..density..), alpha = 1,color="white",bins=30) +
  geom_density(alpha =0.3,color="blue")+
  theme_minimal()+ theme(axis.line = element_line(size = 1,linetype = "dotted",colour =
"navyblue"))+

```

```

    facet_wrap(~VISIT)+facet_wrap(~VISIT, ncol = 1)+theme(strip.text =
element_text(colour="navyblue", size=10,face = "bold"))+

    scale_fill_manual(values = c("lightskyblue1", "palegreen1"))+

    ggtitle("Distribution de la somme du Score Hamilton à J0 et J56")

#### =====Fichier Dénormalisé(format large)=====

HDRS_wide<-reshape(df_hamilton,direction="wide",idvar="NUMERO",timevar="VISIT")

ncol(HDRS_wide)

nrow(HDRS_wide)

str(HDRS_wide)

#OU en utilisant somme du Score Hamilton=====

library(reshape2) # Chargement de la librairie

df_tot_wide<- dcast(df_tot, NUMERO + GROUPE ~ VISIT, value.var = "HDRSScore")

ncol(df_tot_wide)

nrow(df_tot_wide)

str(df_tot_wide)

##### =====VALIDATON DE L'ÉCHELLE=====

#####====Analyse d'items: effets planchées et plafond, redondance =====

####====recherche d'effets plancher, plafond et données manquantes=====


### échelle hamilton 17 items à J0

par(mar=c(2,2,2,2))

par(mfrow=c(4,6))

hdrs_j0<-
c("HAMD1.J0","HAMD2.J0","HAMD3.J0","HAMD4.J0","HAMD5.J0","HAMD6.J0","HAMD7.J0","HA
MD8.J0"
,"HAMD9.J0","HAMD10.J0","HAMD11.J0","HAMD12.J0","HAMD13.J0","HAMD14.J0","HAMD15.J0
","HAMD16.J0","HAMD17.J0")

for (i in hdrs_j0){barplot(table(HDRS_wide[,i], useNA = "always"), col =
"lightskyblue1",names.arg = c(names(table(HDRS_wide[,i])), "NA"), ylim = c(0,120), main =i, sub =
"graphique")}}

### échelle hamilton 17 items à J56

```



```

par(mar=c(2,2,2,2))

par(mfrow=c(4,6))

hdrs_j56<-
c("HAMD1.J56","HAMD2.J56","HAMD3.J56","HAMD4.J56","HAMD5.J56","HAMD6.J56","HAMD7.J
56","HAMD8.J56","HAMD9.J56","HAMD10.J56","HAMD11.J56","HAMD12.J56","HAMD13.J56","HA
MD14.J56","HAMD15.J56","HAMD16.J56","HAMD17.J56")

for (i in hdrs_j56){barplot(table(HDRS_wide[,i], useNA = "always"), col = c("palegreen1")
, names.arg = c(names(table(HDRS_wide[,i])), "NA"), ylim = c(0,120), main = i)}

#####=====ANALYSE DES CORRELATIONS ENTRE LES ITEMS(redondance)=
### recherche de la redondance des items: on constate que aucune correlation ne dépasse 0.44 à J0

hdrs_j0 <- c("HAMD1.J0", "HAMD2.J0", "HAMD3.J0", "HAMD4.J0",
"HAMD5.J0", "HAMD6.J0",
"HAMD7.J0", "HAMD8.J0", "HAMD9.J0", "HAMD10.J0", "HAMD11.J0", "HAMD12.J0",
"HAMD13.J0", "HAMD14.J0", "HAMD15.J0", "HAMD16.J0", "HAMD17.J0")

COR_hdrs_J0 <- cor(subset(HDRS_wide[,hdrs_j0]), use = "complete.obs")

corrplot(COR_hdrs_J0, method = "color", addCoef.col = "cornsilk1", type="upper",
order="hclust",

col = wes_palette(name = "FantasticFox1", type = "discrete"), tl.col = "black", tl.srt = 45,
diag = FALSE)

### ===== On répète la même opération pour J56=====

hdrs_j56<-c("HAMD1.J56", "HAMD2.J56", "HAMD3.J56", "HAMD4.J56",
"HAMD5.J56", "HAMD6.J56",
"HAMD7.J56", "HAMD8.J56", "HAMD9.J56", "HAMD10.J56", "HAMD11.J56",
"HAMD12.J56",
"HAMD13.J56", "HAMD14.J56", "HAMD15.J56", "HAMD16.J56", "HAMD17.J56")

COR_hdrs_J56<-cor(na.omit(subset(HDRS_wide[,hdrs_j56])))

corrplot(COR_hdrs_J56, method = "color", addCoef.col = "cornsilk1", type="upper",
order="hclust", col = wes_palette(name = "FantasticFox1", type = "discrete"),
tl.col = "black", tl.srt = 45, diag = FALSE

```

```

#### ===== ACP=====

## On réalise une ACP pour estimer le nombre de dimensions du score total de Hamilton

### Pour J0

mdspca(HDRS_wide[,hdrs_j0])

### Pour J56

mdspca(HDRS_wide[,hdrs_j56])

#### =====CLASSIFICATION HIERARCHIQUE=====

## On utilise cette technique pour vocation de déterminer de façon homogène
## l'existence de lien entre nos différents items. l'idée est la présentation par arbre

## POUR J0

cah<-hclust(dist(t(scale(HDRS_wide[,hdrs_j0]))),method = "ward.D")

plot(cah,xlab="",ylab = "",main="Classification hiérarchique de SCore de hamilton à J0")

## POUR J56

cah2<-hclust(dist(t(scale(HDRS_wide[,hdrs_j56]))),method = "ward.D")

plot(cah2,xlab="",ylab = "",main="Classification hiérarchique de SCore de hamilton à J56")

##### ===== Évaluation du caractère du score de Hamilton à J0 et J56 (Unidimensionnel???)=

### Structure dimensionnelle

## POUR J0

scree.plot(HDRS_wide[,hdrs_j0], simu =20, title = "Scree Plot des items de Hamilton à J0")

## POUR J56

scree.plot(HDRS_wide[,hdrs_j56], simu =20, title = "Scree Plot des items de Hamilton à J56")

fa.parallel(HDRS_wide[,hdrs_j0], main = "Diag. valeurs propres avec données simulées à
J0",fa="pc", n.iter = 100)

fa.parallel(HDRS_wide[,hdrs_j56], main = "Diag. valeurs propres avec données simulées à
J56",fa="pc", n.iter = 100)

#### ===== Analyse factorielle =====

```

```

## POUR J0 = à 3 facteurs

print(factanal(na.omit(HDRS_wide[,hdrs_j0]),factors=3,rotation="varimax"),cutoff=0)


## POUR J56 = à 1 facteur

print(factanal(na.omit(HDRS_wide[,hdrs_j56]),factors=1,rotation="varimax"),cutoff=0)

####=====Cronbach:=====

## Evaluation de la consistance interne (Coefficient alpha de cronbach)

## POUR J0

cronbach(na.omit(HDRS_wide[,hdrs_j0])) ## $alpha= 0.45594

## POUR J0

cronbach(na.omit(HDRS_wide[,hdrs_j56])) ### $alpha = 0.819954


####===== Validité Concourante et divergente : Corrélation score total de Hamilton avec sous
dimensions de SCL_90 =====


## On créer des sous-échelles , sommes des items composant différents symptômes

## Cependant, à cause des valeurs manquantes,

## on s'affranchit des problèmes causés par la somme et calculant des moyennes


df_pool<- df_tot

df_pool <- mutate(df_tot, SCL_somatisation = Q1 + Q4 + Q12 + Q27 + Q40 + Q42 + Q48 + Q49
+ Q52 + Q53 + Q56 + Q58,

SCL_obsession = Q3 + Q9 + Q10 + Q28 + Q38 + Q45 + Q46 + Q51 + Q55 + Q65,

SCL_vulnerabilite = Q6 + Q21 + Q34 + Q36 + Q37 + Q41 + Q61 + Q69 + Q73,

SCL_depression = Q5 + Q14 + Q15 + Q20 + Q22 + Q26 + Q29 + Q30 + Q31 + Q32 +
Q54 + Q71 + Q79, SCL_anxiete = Q2 + Q17 + Q23 + Q33 + Q39 + Q57 + Q72 + Q78 + Q80 + Q86,

SCL_hostilite = Q11 + Q24 + Q63 + Q67 + Q74 + Q81,

SCL_phobies = Q13 + Q25 + Q47 + Q70 + Q75 + Q82 + Q50,

SCL_parano = Q8 + Q18 + Q43 + Q68 + Q76 + Q83,

SCL_psycho = Q7 + Q16 + Q35 + Q62 + Q77 + Q84 + Q85 + Q87 + Q90 + Q88,

```

$SCL\_divers = Q19 + Q44 + Q59 + Q60 + Q64 + Q66 + Q89$

```
## On décompte le nombres de NA

sous_echelles_SCL <- c("SCL_somatisation", "SCL_obsession", "SCL_vulnerabilite",
"SCL_depression", "SCL_anxiete",
      "SCL_hostilite", "SCL_phobies", "SCL_parano", "SCL_psycho", "SCL_divers")

apply(df_pool[df_pool$VISIT == "J0",], sous_echelles_SCL, 2, function(y) sum(is.na(y)))
apply(df_pool[df_pool$VISIT == "J56",], sous_echelles_SCL, 2, function(y) sum(is.na(y)))

## On regarde les correspondances entre les sous-échelles du SCL-90 et le score total de Hamilton

COR_SCL_Hamilton_J0 <- cor(subset(df_pool[df_pool$VISIT == "J0",], select =
c("HDRSScore", sous_echelles_SCL)), use = "complete.obs", method = "spearman")

corrplot(COR_SCL_Hamilton_J0, method = "color", addCoef.col = "azure1", type="upper",
order="alphabet",
      col = wes_palette(name = "Darjeeling1", type = "discrete"), tl.col = "black", tl.srt = 70, diag
= FALSE)

COR_SCL_Hamilton_J56 <- cor(subset(df_pool[df_pool$VISIT == "J56",], select =
c("HDRSScore", sous_echelles_SCL)), use = "complete.obs", method = "spearman")

corrplot(COR_SCL_Hamilton_J56, method = "color", addCoef.col = "azure1", type="upper",
order="alphabet",
      col = wes_palette(name = "Darjeeling1", type = "discrete"), tl.col = "black", tl.srt = 70, diag
= FALSE)

##=====

### Validité concourante à J0 = ACP

fpca(HDRSScore_J0~SCL_somatisation+SCL_obsession+SCL_vulnerabilite+SCL_depression+S
CL_anxiete+
      SCL_hostilite+SCL_phobies+SCL_parano+SCL_psycho+SCL_divers,data=df_pool[df_pool
$VISIT == "J0",])
```

```

### Validité concourante à J56 = ACP

fPCA(HDRSScore_J56~SCL_somatisation+SCL_obsession+SCL_vulnerabilite+SCL_depression+
SCL_anxiete+

SCL_hostilite+SCL_phobies+SCL_parano+SCL_psycho+SCL_divers,data=df_pool [df_pool
$VISIT == "J56",])

###=====QUESTION2=====

###===== Analyse de la réponse au traitement : LOCF =====

## On remplace les valeurs manquantes par les dernières valeurs connues

# On remplace les données manquantes avec l'option locf

hdrsgroupe<-merge(HDRS_wide, df_grp, by= "NUMERO")

sum(duplicated(hdrsgroupe))

names(hdrsgroupe)

hdrsgroupe$HDRSScore.J0 <- rowSums(hdrsgroupe[2:18])

hdrsgroupe$HDRSScore.J56<- rowSums(hdrsgroupe[121:137])

HDRS_wide_locf <- subset(hdrsgroupe, select=c("NUMERO","HDRSScore.J0",
"HDRSScore.J56" , "GROUPE"))

HDRS_wide_locf2<-HDRS_wide_locf[order(HDRS_wide_locf$GROUPE,
HDRS_wide_locf$NUMERO),]

library(zoo)

HDRS_wide_locf2 <- t(na.locf(zoo(t(HDRS_wide_locf))))

HDRS_wide_locf2<- data.frame(HDRS_wide_locf2)

## création d'une variable différence entre les deux scores

## On créer une variable = différences entre J0 et J56

HDRS_wide_locf2$difference <- HDRS_wide_locf2$HDRSScore.J0 -
HDRS_wide_locf2$HDRSScore.J56

## On regarde les conditions de validité du test t de Student

## Distribution des différences entre J0 et J56

ggplot(HDRS_wide_locf2, aes(x = difference)) +

geom_histogram(aes(y=..density..), bins = 20, fill = wes_palette(n = 1, name = "Moonrise1"),

colour = "black")+

geom_density(alpha = 0.3, color="blue") +

```

```

theme_minimal()+

labs(title = "Distribution des différences entre J0 et J56",
      x = "Différence entre J0 et J56", y = "Fréquence") ## ça suit pas une loi normale

## Estimation par rapport à la loi normale par visualisation du qqnorm et de la qqline
qqnorm(HDRS_wide_locf2$difference)
qqline(HDRS_wide_locf2$difference)

## On teste l'égalité des variances
by(HDRS_wide_locf2$difference,HDRS_wide_locf2$GROUPE,sd,na.rm=TRUE)

## On compare la moyenne des différences
by(HDRS_wide_locf2$difference, HDRS_wide_locf2$GROUPE, mean)
table( HDRS_wide_locf2$GROUPE)
t.test(HDRS_wide_locf2$difference~HDRS_wide_locf2$GROUPE, var.equal=TRUE)

boxplot(HDRS_wide_locf2$difference~HDRS_wide_locf2$GROUPE, ylab = "Difference du
score de Hamilton entre J0 et J56", xlab="Groupe",main="Diminution du score HDRS entre J0 et J56 selon
le

groupe",col=c("lightskyblue1","azure4"))

#####=====MODELE MIXTE=====

df_tot$GROUPE <- factor(df_tot$GROUPE)

ggplot(df_tot, aes(x = VISIT, y = HDRSScore, group = GROUPE, colour = GROUPE)) +
  stat_summary(fun.y = mean, geom = "point", shape = 18, size = 3, position =
position_dodge(.20)) +
  stat_summary(fun.y = mean, geom = "line") +
  theme_minimal()+

  stat_summary(fun.data = mean_cl_normal, fun.args = list(mult = 1), geom = "errorbar", width =
.3, position = position_dodge(.20)) +

  scale_x_discrete(limits = c("J0", "J4", "J7", " ", "J14", " ", "J21", " ", "J28", " ", " ", " ", "J42", " ",
" ", " ", "J56"),

breaks = c("J0", "J4", "J7", "J14", "J21", "J28", "J42", "J56")) +

```

```

scale_colour_manual(values = c("lightskyblue2", "azure4")) + theme(legend.position = "top")

## On réalise un premier modèle en mettant fixe Groupe et Visite, et aléatoire les sujets
modele_mixte_simple <- lmer(HDRSScore ~ GROUPE + VISIT + (1|NUMERO), data = df_tot)
summary(modele_mixte_simple)

par(mfrow = c(1, 2))

hist(resid(modele_mixte_simple), breaks = 20, col = "palegreen1", main = "Résidus du modèle
mixte simple", xlab = "Résidus")

qqnorm(resid(modele_mixte_simple))
qqline(resid(modele_mixte_simple))

## On réalise un deuxième modèle en ajoutant une interaction entre groupe et visite
modele_mixte_interaction <- lmer(HDRSScore ~ GROUPE * VISIT + (1|NUMERO), data =
df_tot)

summary(modele_mixte_interaction)

par(mfrow = c(1, 2))

hist(resid(modele_mixte_interaction), breaks = 20, col = "palegreen1", main = "Résidus du modèle
mixte avec interaction", xlab = "Résidus")

qqnorm(resid(modele_mixte_interaction))
qqline(resid(modele_mixte_interaction))

## On teste l'ajout de l'interaction sur le modèle
anova(modele_mixte_simple, modele_mixte_interaction, test = "Chisq")

```

=====**QUESTION3**=====

```

##### Creation de la variable binaire guérison : diminution > 50% du HDRS à J0:

## On créer la variable reponse binaire que l'on met en NA
df_tot_denormalise_binaire <- df_tot_wide %>% mutate(reponse_binaire = NA)

jours_visites <- c("J0", "J4", "J7", "J14", "J21", "J28", "J42", "J56")

## On créer la variable réponse seuil qui est la moitié de la valeur de J0
df_tot_denormalise_binaire <- df_tot_denormalise_binaire %>% mutate(seuil_reponse = J0 / 2)

## mettre la variable à 1 si on a une diminution de 50 % du score à J0
for (k in 1:length(df_tot_denormalise_binaire$NUMERO)) {
  for (l in jours_visites) {

```

```

    if (!is.na(df_tot_denormalise_binaire[k, l]) & df_tot_denormalise_binaire[k, l] <=
df_tot_denormalise_binaire[k, "seuil_reponse"] & is.na(df_tot_denormalise_binaire[k, "reponse_binaire"]))
{
    df_tot_denormalise_binaire[k, "reponse_binaire"] <- 1 }
} }

## changer la var reponse_binaire en int puisqu'elle en Chr

for (k in 1:length(df_tot_denormalise_binaire$NUMERO)) {

    df_tot_denormalise_binaire[k, "reponse_binaire"] <-
str_extract_all(df_tot_denormalise_binaire[k, "reponse_binaire"], "\\d+") }

    df_tot_denormalise_binaire$reponse_binaire <-
as.integer(df_tot_denormalise_binaire$reponse_binaire)

    summary(df_tot_denormalise_binaire)

    ## On créer un objet survfit

    survie_bin<-survfit(Surv(time=reponse_binaire)~GROUPE,df_tot_denormalise_binaire)

    summary(survie_bin)

    summary(survie_bin)$table

    ggsvplot(survie_bin, pval = TRUE, conf.int = TRUE, risk.table = TRUE, risk.table.col =
"strata", linetype = "strata",

    surv.median.line = "v", palette = c("lightskyblue1", "azure4"),main="Courbe de guérison
dans le temps",

    xlab = "Time in day", risk.table.y.text = FALSE, legend.labs = c("Groupe 0", "Groupe 1"))

    survie_bin_difference <- survdiff(Surv(time = reponse_binaire) ~ GROUPE, data =
df_tot_denormalise_binaire)

    survie_bin_difference

```