

Master 2 : Méthodologie et statistiques en recherche
biomédicale

Devoir n°2 : Statistiques avancées

Décembre 2022

Baptiste Grenier – Groupe 43

Table des matières

1. Introduction

- a. Énoncé
- b. Présentation des échelles
- c. Jeu de données

2. Question 1 : Validation de l'échelle Hamilton

- a. Validation à J0
 - i. Description des items
 - ii. Structure dimensionnelle, analyse factorielle, validité interne
 - iii. Validité externe, validité concourante
- b. Validation à J56
 - i. Description des items
 - ii. Structure dimensionnelle, validité interne
 - iii. Validité externe, validité concourante

3. Question 2 : Comparaison entre les groupes à J0 et J56

- a. Présentation des groupes et des perdus de vue
- b. Comparaison entre les groupes
 - i. Test de Student, imputation par LOCF
 - ii. Modèles mixtes
 - 1. Modèle sans interaction
 - 2. Modèle avec interaction
 - 3. Conditions de validité

4. Question 3 : Comparaison entre les groupes sur un critère binaire

- a. Présentation de la population
- b. Comparaison de la survie dans les deux groupes

5. Conclusion

6. Script R

INTRODUCTION

1.1 Énoncé

« *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 groupe (2 sous-groupes de 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 groupe=1 répondent mieux au traitement que les patients du groupe groupe=0. (utilisez d'abord une approche LOCF (last observation carried forward), 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. »

1.2 Présentation des échelles

L'**échelle de dépression d'Hamilton (HDRS)** est une échelle clinique conçue pour mesurer les changements d'intensité de la symptomatologie dépressive. La version originale possédait 21 items (1960), réduite à 17 items dans sa version définitive (1967). Il s'agit d'une hétéroévaluation, réalisée par un clinicien aguerri, au cours d'un entretien d'environ 30 minutes.

De nombreuses études ont permis de démontrer une validité interne suffisante à l'échelle (test-retest et fidélité inter-juges élevés). Il existe néanmoins une grande variabilité dans les études quant à la structure factorielle de l'échelle.

Les items sont côtés de 0 à 2 ou de 0 à 4, et la somme totale des 17 items permet d'établir un score, reflet de l'intensité de la symptomatologie dépressive, avec :

- Score ≤ 7 : absence de dépression clinique
- Score entre 8 et 15 : dépression mineure
- Score > 15 : dépression majeure.

La **Symptom Distress Check-List 90** (SCL 90) est une échelle éditée en 1965, visant à retrouver un certain nombre de dimensions d'inconfort psychopathologiques, évaluées par 90 questions courtes cotées de 0 à 4. Il s'agit d'une auto-évaluation, remplie par le patient lui-même, en une quinzaine de minutes. Sa structure factorielle représente 10 dimensions : somatisation, symptômes obsessionnels, sensitivité interpersonnelle ou vulnérabilité, dépression, anxiété, hostilité, phobies, traits paranoïaques, traits psychotiques et symptômes divers.

Sa validité concourante, sa stabilité factorielle et sa traductibilité en français ont été jugées suffisantes.

1.3 Jeux de données

Les fichiers sont des fichiers **normalisés**, comprenant une ligne par visite par patient, et une colonne par item des questionnaires, à l'exception de l'item 16 (« *Perte de poids* ») de l'échelle Hamilton, qui est représentée par 2 colonnes (« *HAMD16A* » et « *HAMD16B* ») selon que la perte ait été rapportée par le patient (A) ou mesurée par le clinicien (B). La variable « *HAMD16* » est créée, fusion des deux précédentes, qu'elle remplace.

Le fichier **Hamilton** contient 1053 observations pour 20 variables chez 146 patients. Une 21^e variable, « *Score* », est ajoutée, correspondant au score total de chaque patient. Seules les données d'une ligne (correspondant à la visite de J7 du 128^e patient) sont manquantes, il est décidé de supprimer cette ligne ; le fichier est ensuite complet et ne contient aucune donnée aberrante.

Le fichier **SCL90** contient 1034 observations pour 92 variables chez 146 patients. Il comprend un certain nombre de données aberrantes et manquantes, qui représentent, après recodage des données aberrantes en données manquantes, 0,6% de données totales. Dix nouvelles variables sont créées, représentant les scores moyen propres aux dix dimensions retrouvées dans la littérature.

Un troisième jeu de données, **Groupe**, réparti les 146 patients en 2 groupes, 1 et 0. Il ne contient pas de données manquantes.

Pour cette étude, les fichiers ont été **dénormalisés** et **fusionnés**, permettant la comparaison de HDRS à SCL90 et à ses dimensions propres, à différents temps du suivi (J0, J56), ou en fonction de l'attribution du patient au groupe 0 ou au groupe 1.

QUESTION 1 : Validation de l'échelle de Hamilton

Intitulé de la question : « 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. »

- Procéder à la vérification d'une échelle de mesure subjective consiste à se poser deux questions : Que vaut-elle ? Que mesure-t-elle ?
- Il convient de procéder dans un premier temps à la **description** de l'échelle en représentant la répartition des réponses item par item, puis de rechercher une redondance entre les items à l'aide d'une matrice de corrélation de ces derniers deux à deux.
- Dans un second temps, la **structure dimensionnelle** de l'échelle est étudiée, à l'aide d'un diagramme des valeurs propres, et, si besoin, d'une analyse factorielle, avant validation intrinsèque de la reproductibilité de l'échelle par mesure des coefficients de Cronbach.
- Enfin, la **validité externe** de l'échelle, sa validité concourante / divergente, est évaluée en la comparant à d'autres échelles connues, ici en la comparant à SCL90 et ses différents sous-scores.

2.1. Vérification au temps J0

2.1.1 Description des items

- La répartition des réponses des patients à chaque item est représentée par les diagrammes suivants. Il n'y a aucune donnée manquante.
- La matrice de corrélation des items 2 à 2 ne retrouve pas de coefficient de corrélation supérieure à 0,44 en valeur absolue, il n'existe **pas de redondance** entre les items de l'échelle Hamilton.

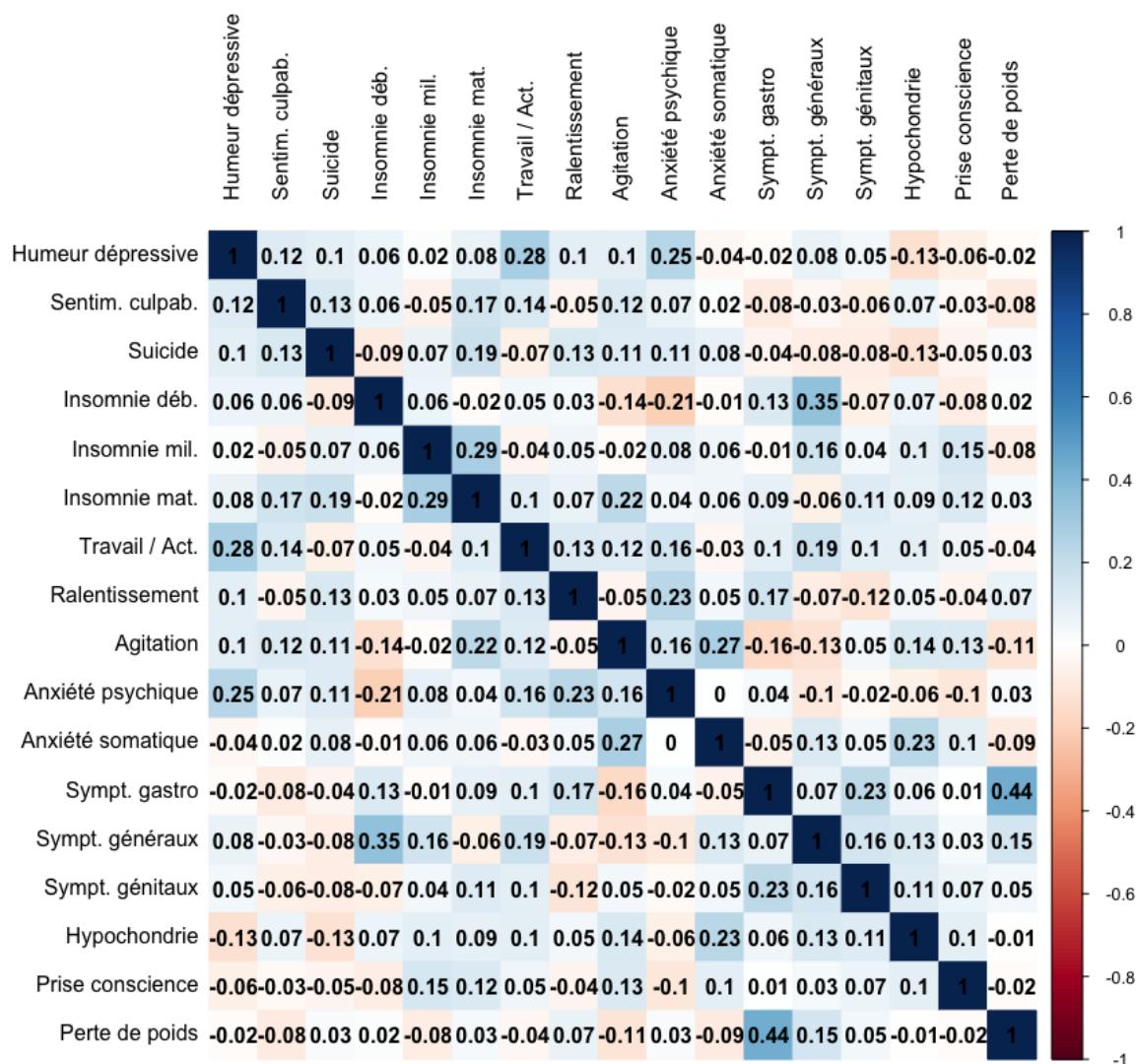


Figure 1 : Matrice de corrélation 2 à 2 des items d'Hamilton à J0

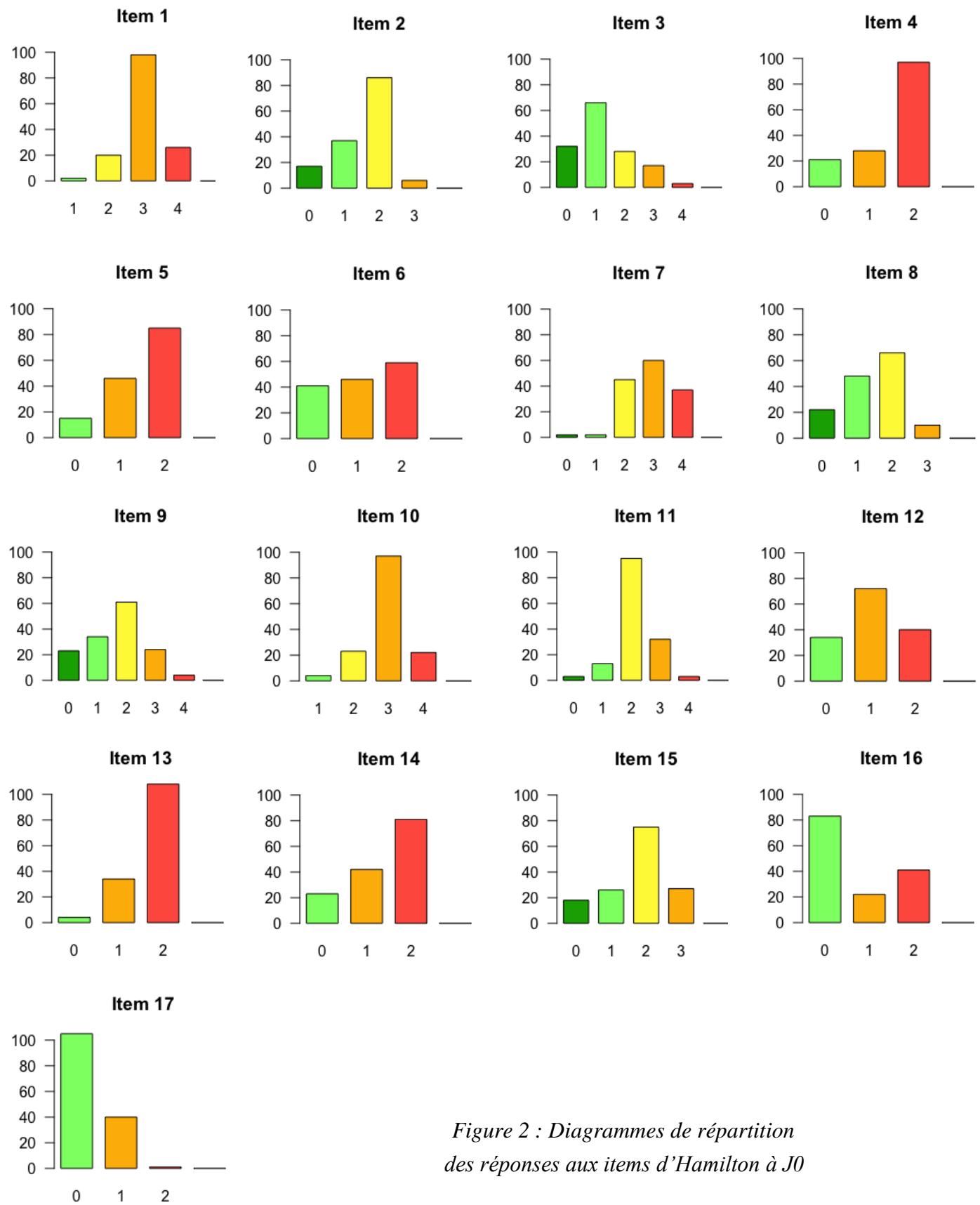
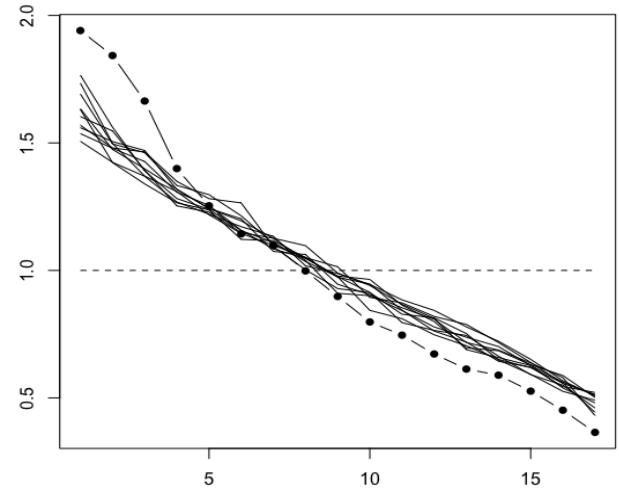


Figure 2 : Diagrammes de répartition des réponses aux items d'Hamilton à J0

2.1.2 Structure dimensionnelle, analyse factorielle, validité interne

Figure 3 : Diagramme des valeurs propres d'Hamilton à J0



- Le diagramme des valeurs propres met en évidence l'**absence de score global** à l'échelle Hamilton à J0. Il est possible de constater aux moins 3, sinon 4 points au-dessus de l'analyse parallèle des 10 droites tracées par simulation.

	Dimension 1	Dimension 2	Dimension 1	Dimension 2	Dimension 3	Dimension 1	Dimension 2	Dimension 3	Dimension 4
Humeur dépressive	-0,016	0,251	0,006	0,028	0,263	-0,022	0,117	0,562	-0,122
Sentim. culpab.	-0,098	0,246	-0,057	-0,076	0,24	-0,086	-0,038	0,232	0,091
Suicide	-0,056	0,238	-0,021	-0,12	0,207	-0,047	-0,08	0,231	0,038
Insomnie déb.	0,144	-0,115	0,11	0,359	-0,05	0,127	0,35	-0,072	-0,028
Insomnie mil.	-0,01	0,203	0,004	0,118	0,24	-0,02	0,128	0,066	0,216
Insomnie mat.	0,077	0,473	0,132	-0,155	0,427	0,077	-0,124	0,245	0,331
Travail / Act.	0,1	0,288	0,124	0,121	0,32	0,09	0,178	0,401	0,09
Ralentissement	0,173	0,155	0,183	-0,105	0,101	0,167	-0,07	0,246	-0,052
Agitation	-0,198	0,513	-0,107	-0,228	0,501	-0,174	-0,201	0,29	0,455
Anxiété psychique	0,026	0,297	0,066	-0,152	0,255	0,033	-0,081	0,484	-0,047
Anxiété somatique	-0,073	0,27	-0,029	0,07	0,316	-0,066	0,052	0,017	0,448
Sympt. gastro	0,921	0,033	0,992	0,044	-0,091	0,997	0,013	0,011	0,025
Sympt. généraux	0,107	-0,036	0,046	0,977	0,194	0,054	0,979	-0,021	0,181
Sympt. génitaux	0,234	0,136	0,238	0,12	0,147	0,221	0,108	0,02	0,216
Hypochondrie	0,059	0,191	0,076	0,089	0,226	0,049	0,054	-0,082	0,425
Prise conscience	-0,001	0,165	0,022	-0,004	0,176	-0,001	-0,026	-0,06	0,306
Perte de poids	0,474	-0,068	0,424	0,155	-0,1	0,438	0,143	-0,025	-0,067

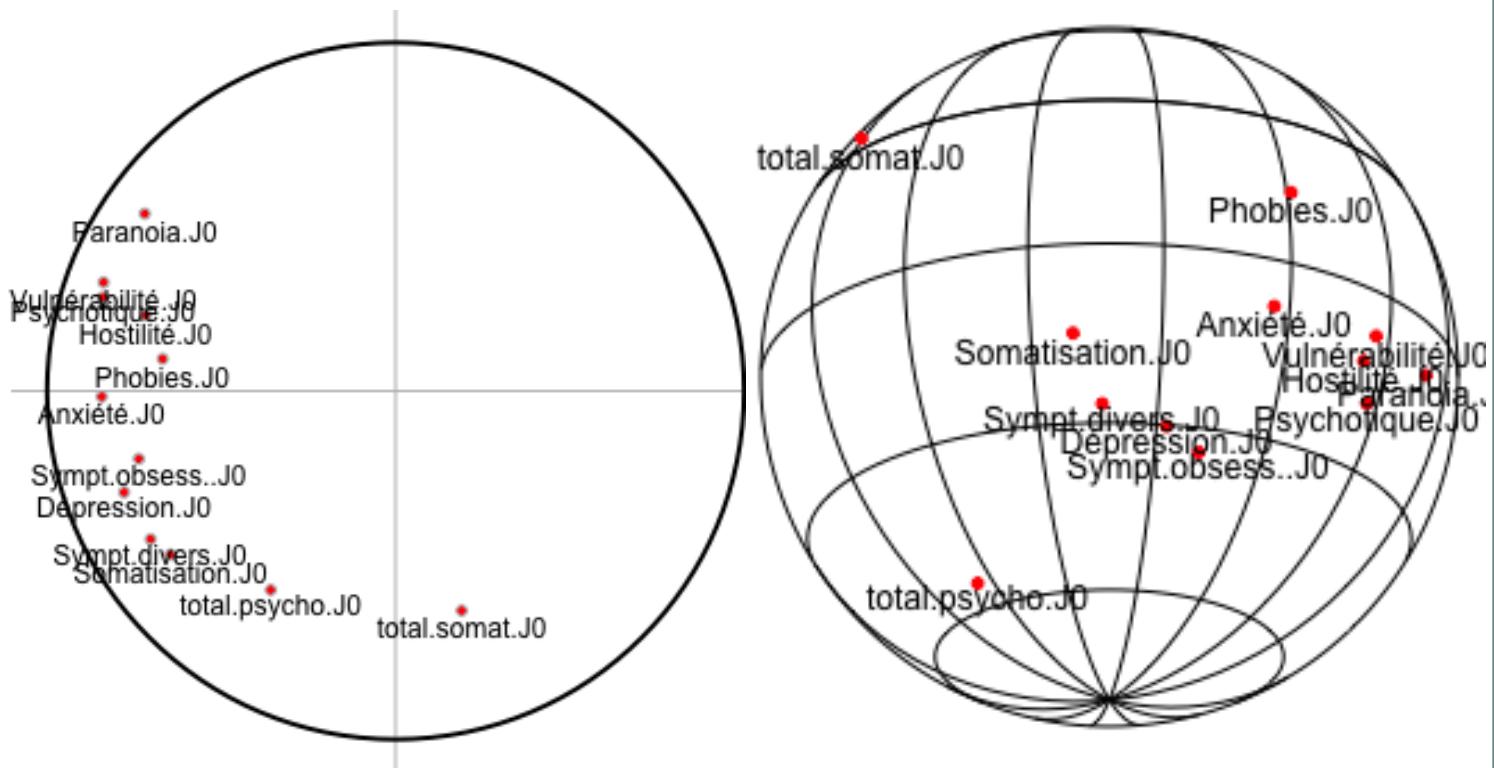
Figure 4 : Analyse factorielle à 2, 3 et 4 dimensions de l'échelle Hamilton à J0

- Les *p-value* respective du test de l'hypothèse « *n facteurs sont suffisants* » valent :
 - à 2 facteurs : 0,0002
 - à 3 facteurs : 0,02
 - à 4 facteurs : 0,24
- L'analyse factorielle détermine si l'outil parvient à identifier des dimensions cliniquement pertinentes. **L'analyse à 2 facteurs semble être la plus discriminante.**

- Deux sous-scores à l'outil Hamilton ont donc été mis en évidence :
 - Échelle 1 : « *Sous-score somatique* », regroupant les items 4, 8, 12, 13, 14 et 16.
 - Échelle 2 : « *Sous-score psychologique* », regroupant les items 1, 2, 3, 5, 6, 7, 9, 10, 11, 15 et 17.
- La cohérence interne de chaque échelle est évaluée à l'aide de coefficients de corrélation de Cronbach dont les intervalles de confiances sont estimés à l'aide de Bootstrap :
 - Échelle 1 : Coefficient $\alpha = 0.49$, intervalle = [0.32 ; 0.60]
 - Échelle 2 : Coefficient $\alpha = 0.40$, intervalle = [0.20 ; 0.54]
- Les coefficients sont faibles, la validité interne de ces deux échelles est discutable, elles semblent peu reproductibles.

2.1.3 Validité externe, validité concourante

- L'analyse de la validité externe consiste à se poser la question : que mesure l'outil Hamilton ?
- En comparant les sous-échelles d'Hamilton aux différents sous-scores connus de SCL90, il est possible de vérifier si Hamilton est effectivement un outil pertinent de mesure de la symptomatologie d'un tableau dépressif.



p. 9 Figure 5 : Analyse en composante principale des sous-échelles d'Hamilton et SCL90 à J0

- Sur ces APC, le sous-score « *Somatique* » semble parfaitement indépendant des dimensions de SCL, et le sous-score « *Psychologique* » leur semble faiblement corrélé.

	total.somat.J0	total.psycho.J0	Somatisation.J0	Sympt.obsess..J0	Vulnérabilité.J0	Dépression.J0	Anxiété.J0	Hostilité.J0	Phobies.J0	Paranoïa.J0	Psychotique.J0	Sympt.divers.J0
total.somat.J0	1	0.05	0.1	-0.08	-0.29	-0.05	-0.11	-0.17	-0.06	-0.34	-0.28	0.09
total.psycho.J0	0.05	1	0.38	0.34	0.09	0.36	0.23	0.17	0.15	0.06	0.14	0.43

Figure 6 : Matrice de corrélation des sous-échelles d'Hamilton aux sous-échelles de SCL90 à J0

- La matrice ne montre **pas de corrélation** supérieure à 0,4 en valeur absolue, l'outil Hamilton semble peu adapté à évaluer la dépression à J0.

	score.J0	Somatisation.J0	Sympt.obsess..J0	Vulnérabilité.J0	Dépression.J0	Anxiété.J0	Hostilité.J0	Phobies.J0	Paranoïa.J0	Psychotique.J0	Sympt.divers.J0
score.J0	1	0.37	0.24	-0.08	0.27	0.13	0.05	0.09	-0.13	-0.03	0.4

Figure 7 : Matrice de corrélation du score total d'Hamilton aux sous-échelles de SCL90 à J0

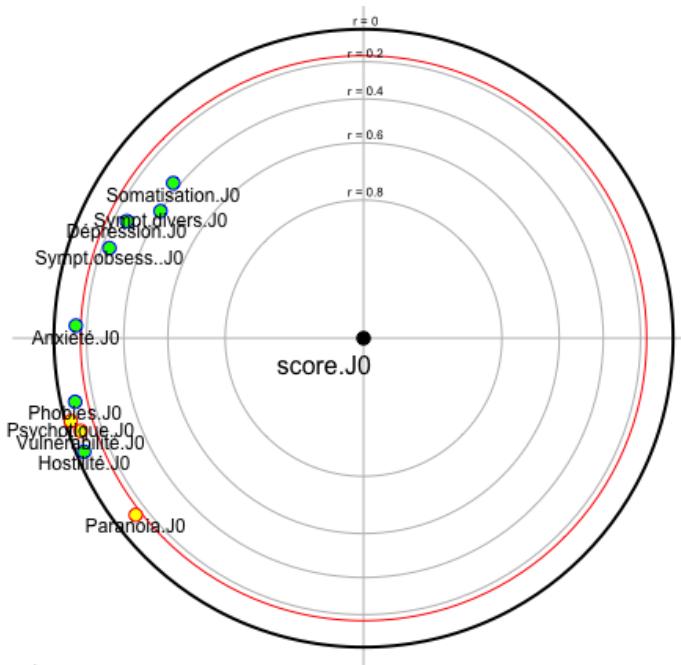


Figure 7 : APC focalisée de corrélation du score total d'Hamilton aux sous-échelles de SCL90 à J0

- De même, le score total sur l'échelle d'Hamilton n'est que très peu corrélé aux sous-scores de SCL90 à J0, sur la matrice comme sur l'APC focalisée.

Ainsi, l'outil Hamilton possède une mauvaise validité interne et externe à J0, il semble peu adapté à l'évaluation de la dépression à J0.

- La validité divergente resterait néanmoins à évaluer.

2.2. Vérification au temps J56

2.2.1 Description des items

- À J56, seuls 120 des 146 patients présents à J0 ont répondu à tous les items, signifiant que **17.8% des patients ont été perdus de vue.**
- La répartition des réponses des patients à chaque item est représentée par les diagrammes suivants.
- La matrice de corrélation des items 2 à 2 met en évidence des coefficients de corrélation supérieurs à 0,6 entre les items « *Humeur dépressive* » et « *Suicide* », « *Humeur dépressive* et « *Travail / Act.* », « *Suicide* » et « *Ralentissement* », mais jamais supérieurs à 0,8 ; il n'y a **pas de redondance** significative entre les items d'Hamilton à J56.

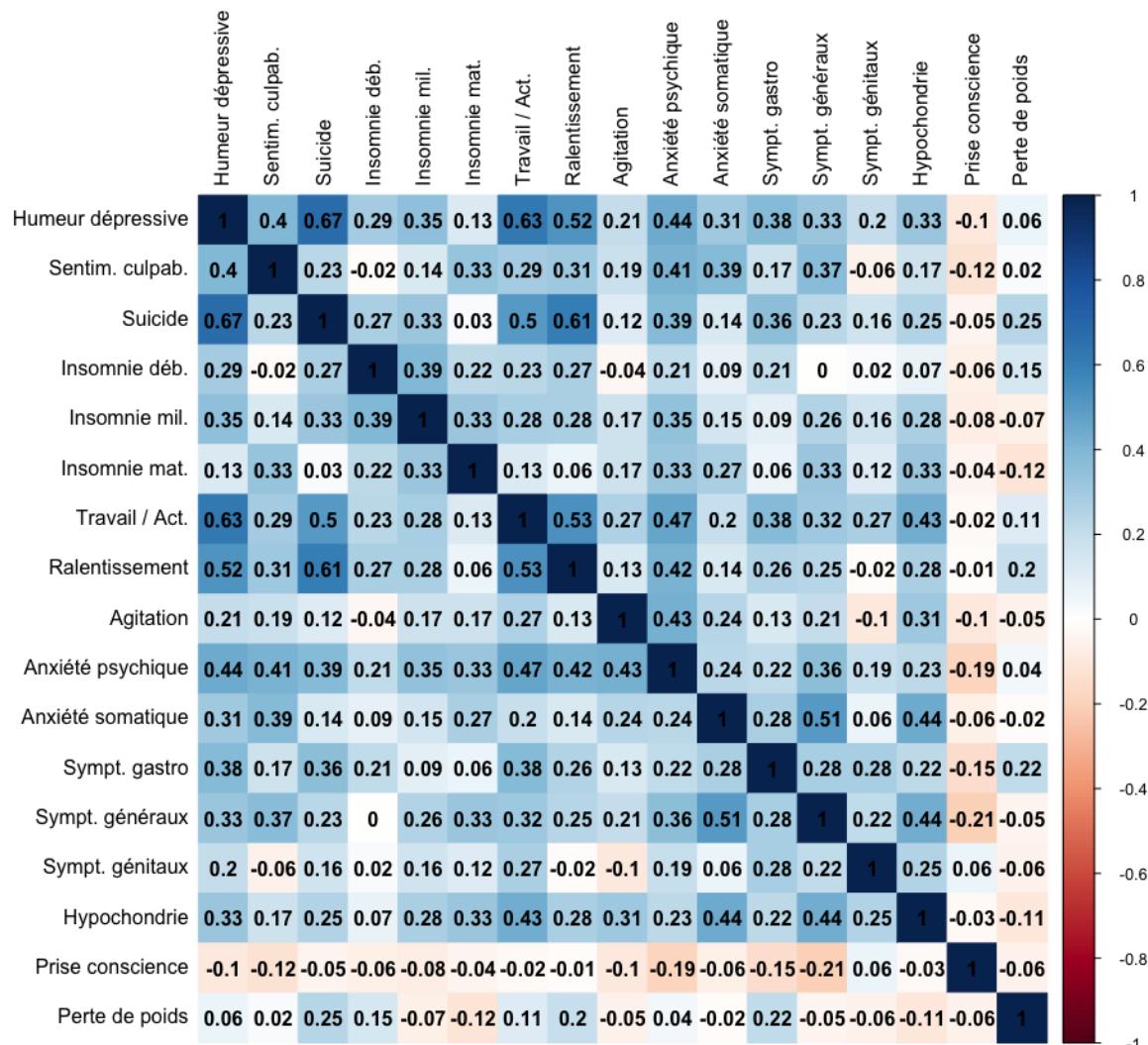


Figure 8 : Matrice de corrélation 2 à 2 des items d'Hamilton à J56

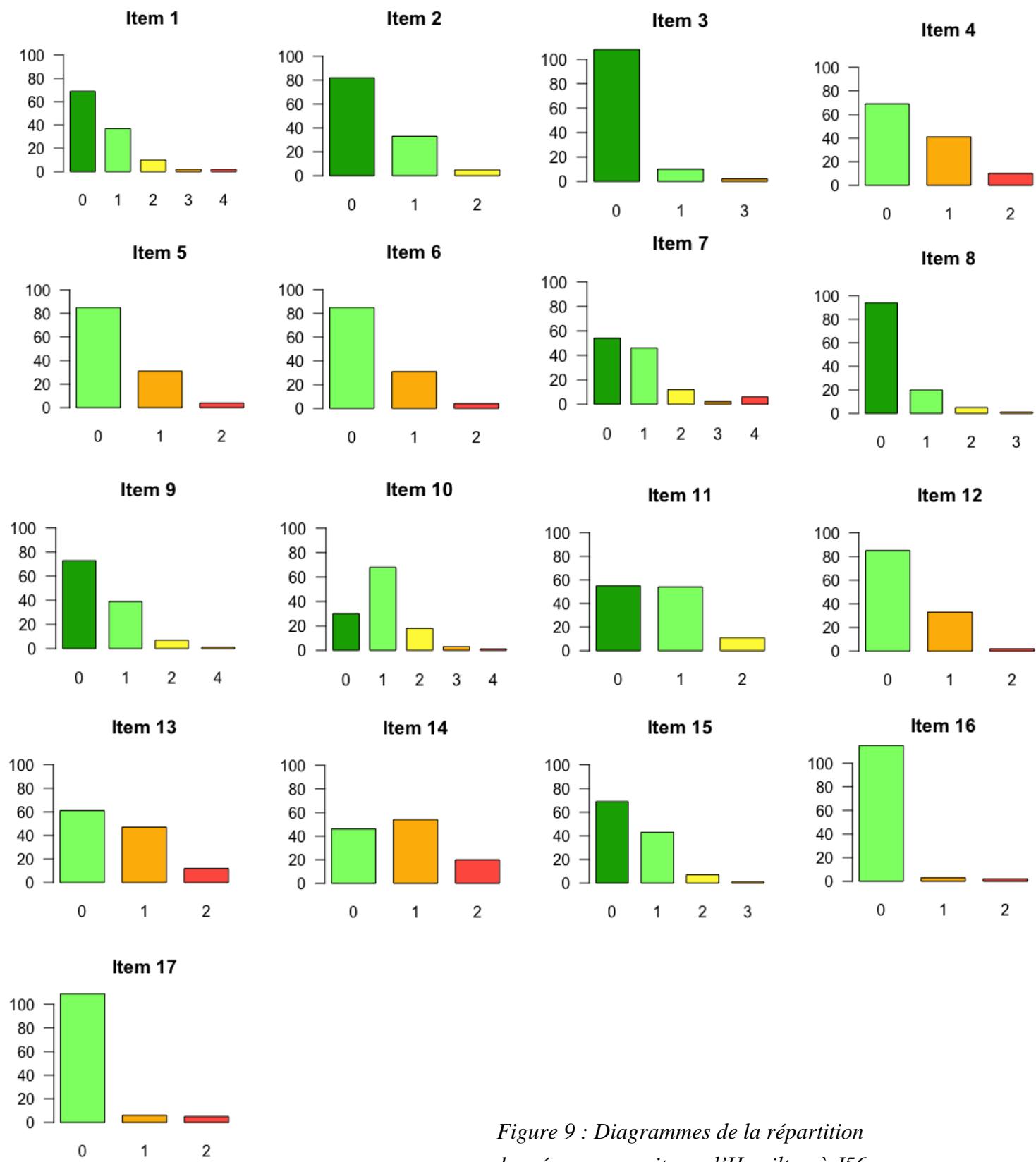
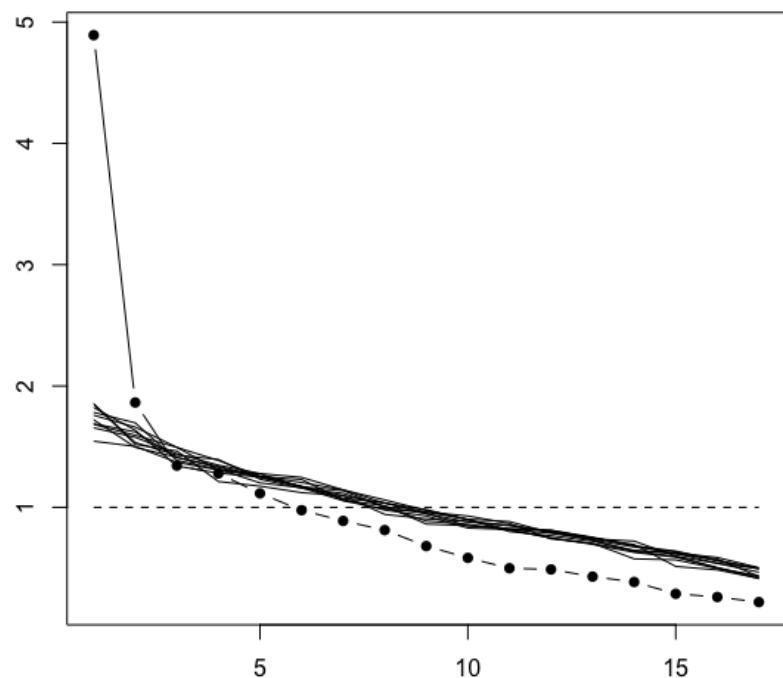


Figure 9 : Diagrammes de la répartition des réponses aux items d'Hamilton à J56

2.2.2 Structure dimensionnelle, validité interne

Figure 10 : Diagramme des valeurs propres d'Hamilton à J56



- Le diagramme des valeurs propres met en évidence le caractère nettement **unidimensionnel** de l'échelle Hamilton à J56. Il existe donc un **score global** d'intensité de la symptomatologie dépressive.
- De fait, il n'est pas nécessaire de procéder à une analyse factorielle pour identifier de potentiels sous-scores.
- La cohérence interne de l'échelle est évaluée à l'aide du coefficient de corrélation de Cronbach dont l'intervalles de confiance est estimé à l'aide d'un Bootstrap :
 - Coefficient $\alpha = 0.82$, intervalle = [0.75 ; 0.86]
- Le coefficient est important, supérieur à 0.8, l'échelle est **reproductible**. Le coefficient n'est pas supérieur à 0.9, l'échelle n'est pas exagérément lourde ou redondante.

2.2.3. Validité externe, validité concourante

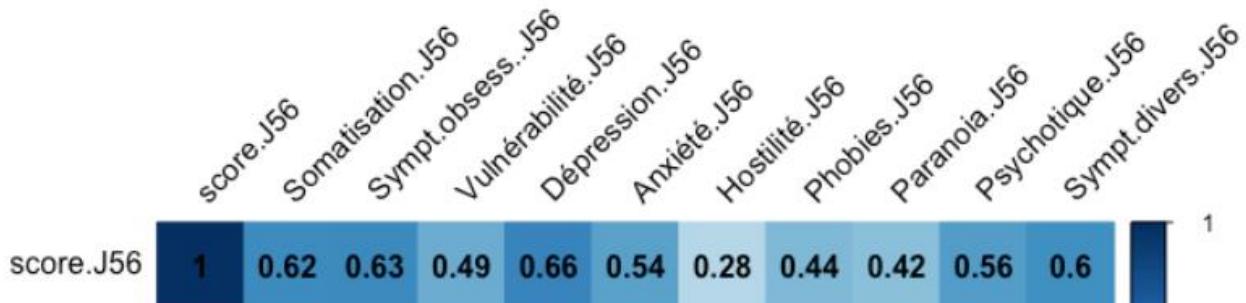
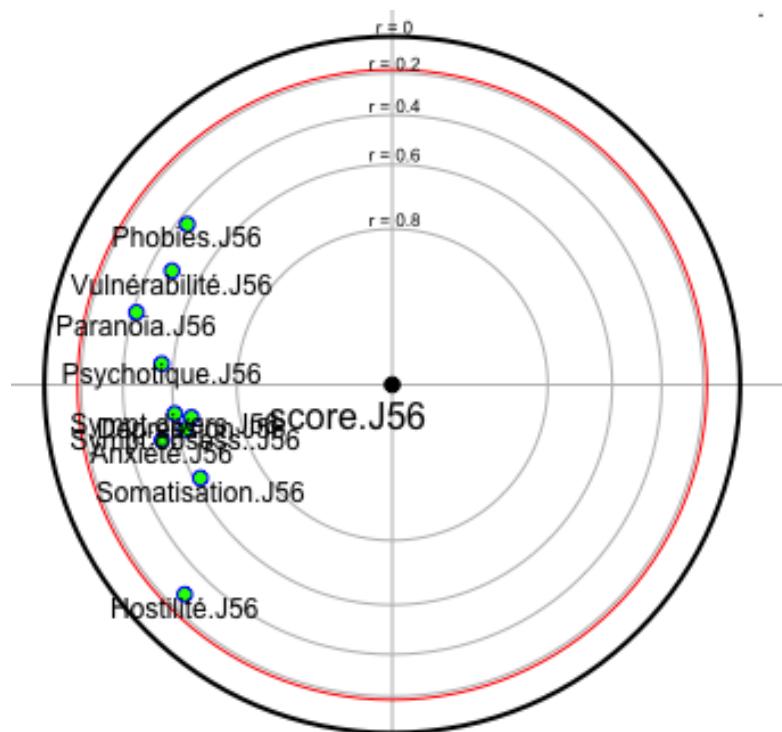


Figure 11 : Matrice de corrélation du score total d'Hamilton aux sous-échelles de SCL90 à J56

Figure 12 : Analyse en composante principale focalisée du score total d'Hamilton en fonction des sous-échelles de SCL90 à J56



- Les corrélations sont globalement importantes entre les sous-scores de SCL90 et le score total d'Hamilton à J56, suggérant que **ces deux échelles mesurent la même chose**.
- De plus, la sous-échelle « Dépression » de SCL90 est la plus fortement corrélée au score global d'Hamilton à J56.

Ainsi, l'outil Hamilton possède une bonne validité interne et externe à J56, il semble bien adapté à l'évaluation de la dépression à J56.

- La validité divergente resterait néanmoins à évaluer.

QUESTION 2 : Comparaison du score de Hamilton entre les groupes

Initié de la question 2 : « A partir du score brut de Hamilton, déterminez si les patients du groupe groupe=1 répondent mieux au traitement que les patients du groupe groupe=0. (utilisez d'abord une approche LOCF (last observation carried forward), puis un modèle mixte).

- L'ensemble des sujets est partagé en deux groupes, nommés groupes 0 et 1, recevant chacun une prise en charge spécifique.
- Chez ces sujets, les mesures du score total d'Hamilton sont à considérer comme des **mesures répétées** et non comme des mesures parfaitement indépendantes les unes des autres. En effet, les bruits \mathcal{E} d'une mesure à l'autre chez un même patient sont corrélés, et cette corrélation est à prendre en compte dans la rédaction du modèle de comparaison des groupes.
- Comme vu précédemment, à J56, seuls 120 des 146 patients présents à J0 ont répondu à tous les items, signifiant que **17.8% des patients ont été perdus de vue**. Le profil de ces perdus de vue doit être analysé, afin d'évaluer si le caractère « perdu de vue » est dépendant du groupe auquel ces patients étaient attribués, ou de leur score initial.
- Enfin, la **gestion de ces perdus de vue** est une question primordiale, il est décidé d'utiliser une approche LOCF, puis un modèle mixte.

3.1 Présentation des groupes et des perdus de vue

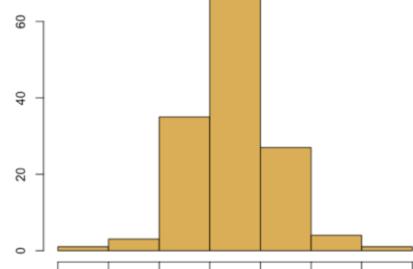
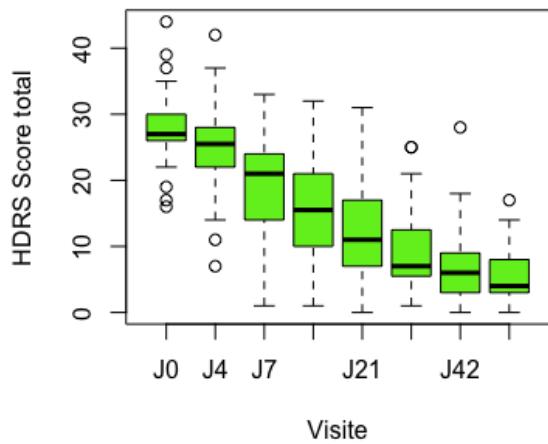
GROUPE	NON-PERDUS	PERDUS DE VUE	TEST STATISTIQUE	P-VALUE	CONDITIONS DE VALIDITE
GROUPE 0	64	11	Test du Chi2	$p=0.31$	Plus de 5 patients par groupe : vérifiée.
GROUPE 1	56	15			
SCORE MOYEN A J0	27.0	28.1	Test de Student	$p=0.24$	

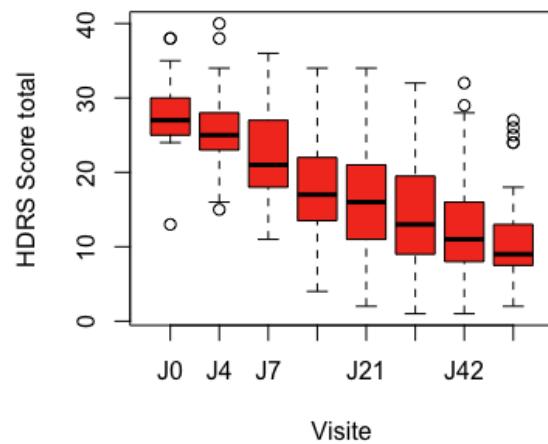
Figure 13 : Répartition des scores à J0

- Il n'existe donc **pas de différence significative entre les deux groupes** concernant le nombre de perdus de vie.
- Il n'existe **pas de différence** entre les perdus de vus et les patients complétant tout le suivi **concernant le score moyen à J0**.
- La censure liée aux perdus de vue n'est **pas informative**, on peut comparer les deux groupes.

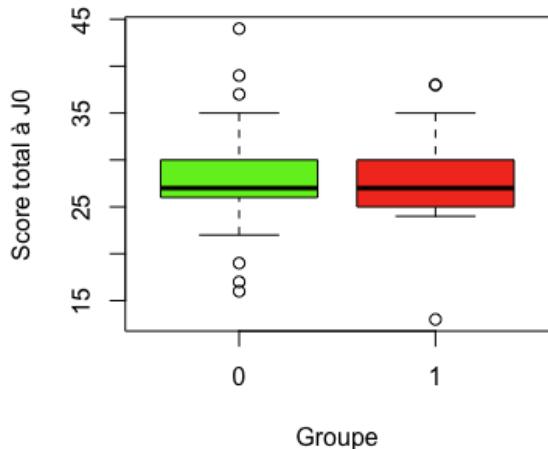
Evolution du score dans le groupe 0



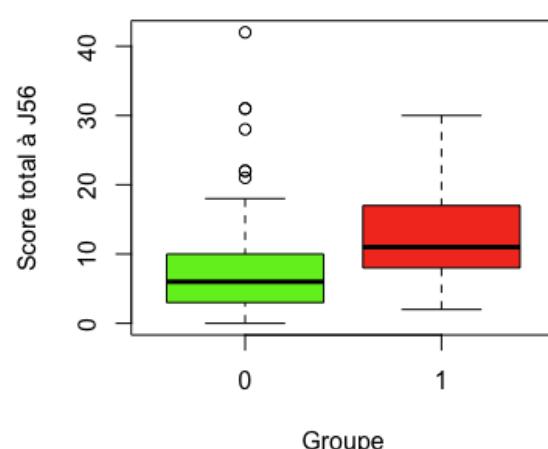
Evolution du score dans le groupe 1



Score Hamilton à J0



Score Hamilton à J56



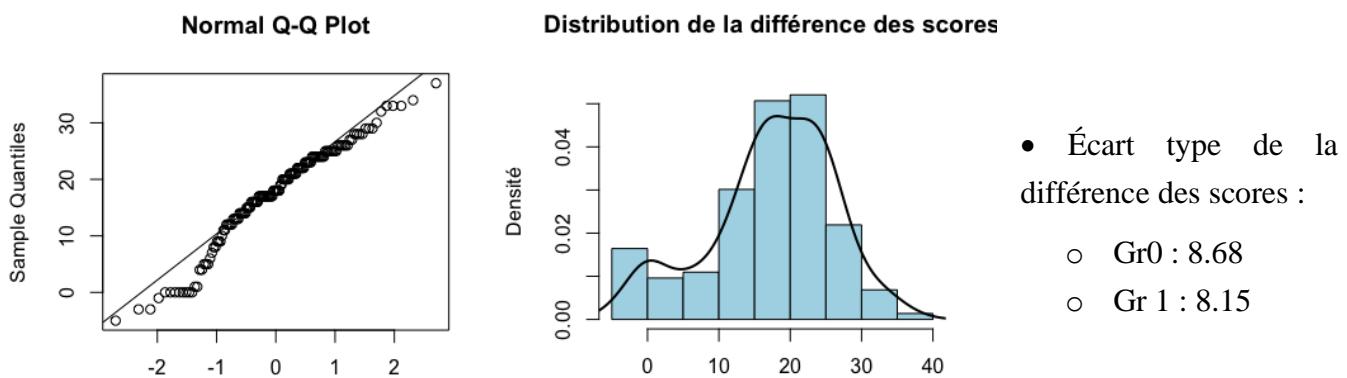
Figures 14 à 17 : Évolution des scores moyen de chaque groupe au cours des visites

- De prime abord, les scores moyens à l'échelle d'Hamilton semblent **comparables** dans les deux groupes à J0.
- Au cours du suivi, la prise en charge des patients du groupe 0 semble leur permettre de mieux s'améliorer cliniquement que la prise en charge des patients du groupe 1.

3.2 Comparaison entre les groupes

3.2.1 Test de Student, imputation par approche LOCF

- Une nouvelle variable « *Difference* » est créée, représentant la différence entre les scores à J56 et à J0 d'un patient, reflétant son amélioration au cours du suivi. Pour les patients perdus de vue, il s'agit de la différence entre le score à J0 et le score lors de la dernière consultation (considéré stable jusqu'à J56 par la suite, selon l'approche **LOCF**).
- Une première comparaison « brute » entre les scores moyens des deux groupes à J0 est réalisée à l'aide d'un test de Student.



Figures 18 et 19 : Normalité de la variable « *Difference* »

- Les conditions de validité de réalisation du test sont remplies.

Différence entre les scores à J0 et J56	Groupe 0	Groupe 1	T-test	p-value
Moyenne	19.9	14.8	3.68	3×10^{-4}
Écart-type	8.7	8.2		

- La comparaison brute des différences de scores moyennes deux groupes trouve une **différence statistiquement significative** entre les groupes.
- Néanmoins, ce modèle ne tient pas compte du fait que les mesures à J0 et J56 soient faites chez un même patient, il est nécessaire de construire un modèle prenant en compte l'**« effet sujet »**. Il est décidé d'avoir recours à un **modèle mixte**, avec une **variable aléatoire [sujet]** suivant une loi normale de moyenne 0 et d'écart-type σ .

3.2.2 Modèle Mixte

- La variable étudiée étant une variable quantitative continue, les modèles utilisés correspondront à des **régessions linéaires généralisées**.
- Il convient dans un premier temps d'étudier l'influence de l'appartenance à un des deux groupes indépendamment de l'effet sujet, puis, dans un second temps, de chercher une éventuelle interaction entre ces deux termes.

3.2.2.1 Modèle mixte sans interaction

	Coefficient	t-value
Groupe	2.91	3.66
Visite	-2.97	-50.37

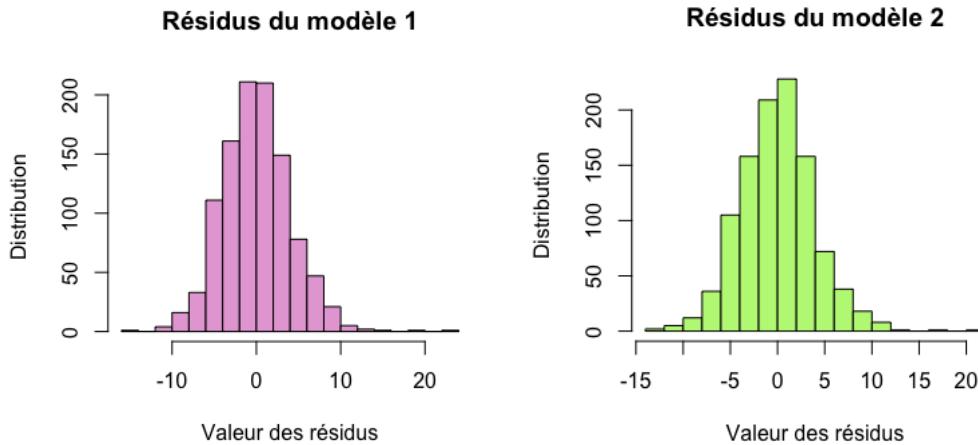
- Le coefficient de la variable « *Groupe* » vaut 2,91. Le groupe de référence étant le groupe numéro 0, et la variable étant binaire, cela signifie que **les patients du groupe 1 ont un score d'Hamilton supérieur à ceux du groupe 0 de 2,91 points**, après ajustement sur l' « *effet sujet* ». La *t-value* à 3,66 correspond à une *p-value* < 5%, le résultat est statistiquement significatif.
- Le coefficient de la variable « *Visite* » vaut -2,97, signifiant qu'indépendamment du groupe auquel ils appartiennent, **le score moyen des patients diminue de 2,97 points à chaque visite**.

3.2.2.2 Modèle mixte avec interaction

	Coefficient	t-value
Groupe	-0.47	-0,51
Visite	-3.35	0,08
Interaction groupe – visite	0,81	7.00

- Le coefficient de l'interaction des deux variables vaut 0,81. La *t-value* à 7 correspond à une *p-value* < 5%, le résultat est statistiquement significatif, permettant de conclure qu'il existe une **interaction significative** entre les deux termes.

3.2.2.3 Conditions de validité



*Figures 20 et 21 :
Répartition des résidus des
modèles mixtes*

- La normalité des résidus des modèles est vérifiée.

Ainsi, les patients du groupe 0 répondent significativement mieux à la prise en charge que les patients du groupe 1, au risque 5%.

QUESTION 3 : Régression linéaire multiple, interactions

*Intitulé de la question 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. »*

- Pour étudier non plus uniquement une différence dans l'intensité de l'amélioration de la dépression entre les deux groupes à l'issu de la prise en charge mais également une **différence dans le délai de survenue** de cette amélioration, il est nécessaire de se prêter à une **analyse de survie** entre les deux groupes.
- Pour ce faire, le **critère binaire** « Réponse » a été créé, défini par une diminution d'au moins **50% du score total** à l'échelle d'Hamilton au cours du suivi des patients, en comparaison au score à J0.
- Ce critère présente une **pertinence clinique**, il est le reflet d'une amélioration significative de l'état psychologique des sujets.
- Dans l'analyse de survie, les patients ont trois devenirs possibles : **réaliser l'évènement** « Réponse » au cours du suivi (considérés comme « Morts » dans l'analyse), arriver au bout du suivi (J56) sans avoir réalisé l'évènement (considérés alors « **Exclus vivants** »), ou être perdus de vue au cours du suivi avant d'avoir réalisé l'évènement (« **Perdus de vue** »). Les patients ayant interrompu le suivi clinique après avoir préalablement diminué leur score de plus de 50% sont considérés « Morts » à la date de l'évènement et non « Perdus de vue ».

4.1 Présentation de la population

- Les groupes 0 et 1 contiennent respectivement 75 et 71 patients.
- Après création de la variable binaire, le nombre de patients **réalisant l'évènement** « diminution d'au moins 50% du score total à l'échelle d'Hamilton » dans chaque groupe est respectivement de 65 (87%) et 51 (72%).
- Le nombre de patients **exclus vivants** dans chaque groupe est respectivement de 1 et 8. Le nombre de patients **perdus de vue** dans chaque groupe est respectivement de 9 (12%) et 12 (17%).

4.2 Comparaison de la survie dans les deux groupes

- De nombreux patients ont réalisé l'évènement « *diminution d'au moins 50% du score total à l'échelle d'Hamilton* » à chaque consultation, hormis lors de la première et de la dernière consultation pour le groupe 0 (respectivement 2 et 1 évènement, contre 6 à 16 pour les autres consultations et l'autre groupe). Ceci rend raisonnable la réalisation d'un **test du Log-Rank** pour comparer les courbes.
- La figure 22 rend compte de la vérification de l'hypothèse des risques instantanés proportionnels, permettant la réalisation d'un **modèle de Cox bivarié**, préférable dans certaines situations au test du Log Rank pour son aptitude à donner des **Hazard Ratio**.

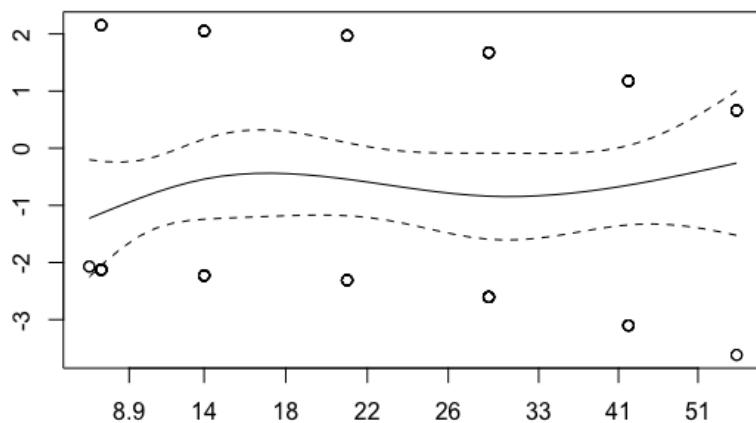


Figure 22 : Diagramme des risques instantanés proportionnels

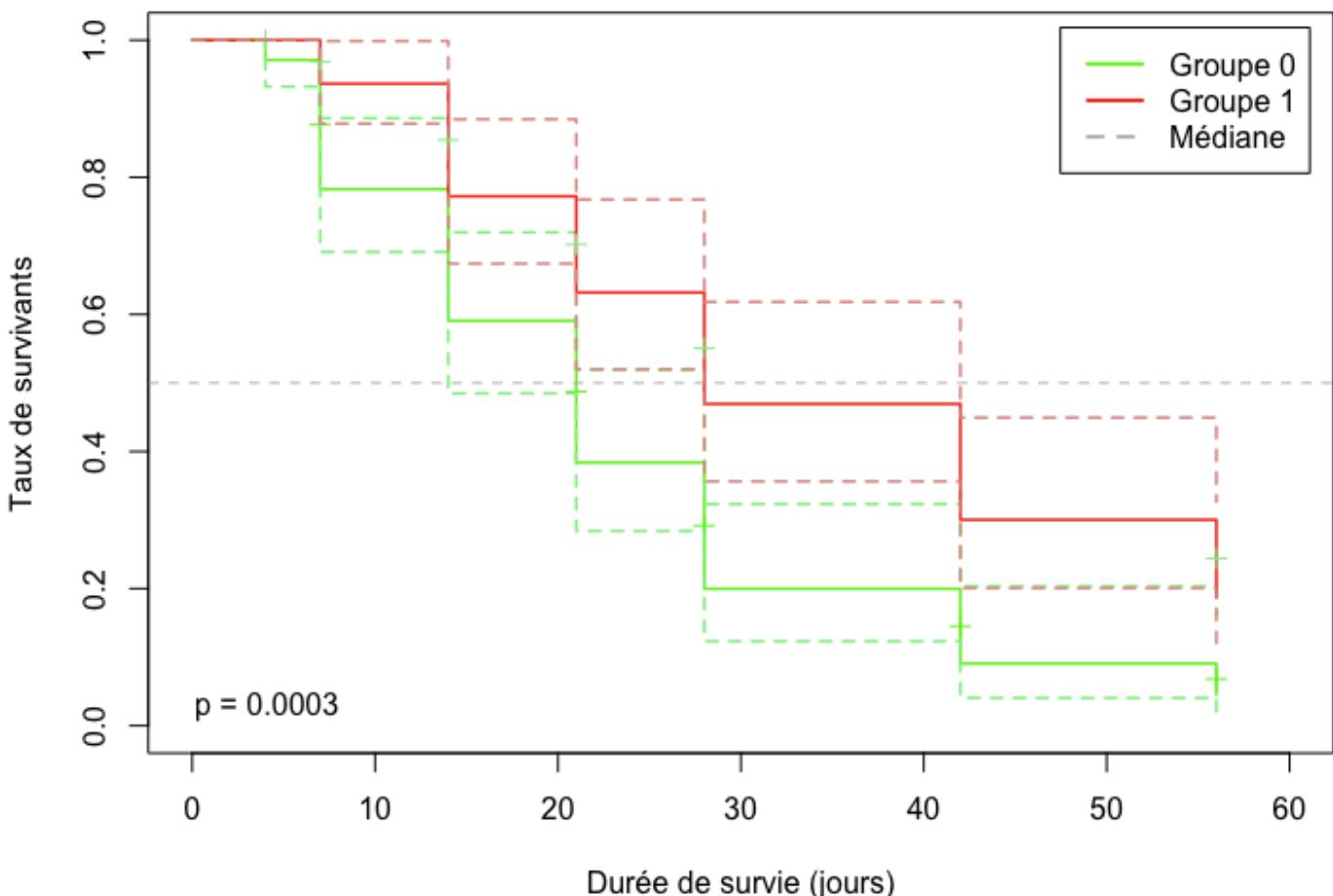


Figure 23 : Courbes de Kaplan-Meyer de survie dans les groupes avec comparaison par test du Log-Rank

- A vue d'œil, les courbes ne semblent pas se croiser, et leurs intervalles de confiance à 95% ne se chevauchent pas sur la majeure partie du graphique. Il semble possible de dire que **la baisse du score d'Hamilton dans le groupe 0 est statistiquement plus rapide** que dans le groupe 1.
- Le test du Log-Rank réalisé est **significatif au risque de 3e-04**.
- Le modèle de Cox bivarié retrouve une **exponentielle du coefficient valant 0.49** avec un degré de significativité inférieur à 1e-03, ce qui correspond à un « risque » instantané diminué de moitié pour les patients du groupe 1 de présenter l'événement, on peut conclure que :

Les patients du groupe 0 s'améliorent cliniquement deux fois plus rapidement que les patients du groupe 1 de manière significative.

CONCLUSION

- L'outil Hamilton est une échelle d'hétéroévaluation de l'intensité de la symptomatologie dépressive d'un patient. Bien que rencontrant quelques difficultés de cohérence interne initialement, il devient **fiable et de structure stable** au cours du suivi. Il est comparable à d'autres instruments communément admis tels que l'échelle SCL90.
- La comparaison entre les deux groupes de sujets étudiés par l'échelle d'Hamilton, à l'aide de régression linéaire avec une approche LOCF, de modèles mixtes ou d'une analyse de survie, retrouve dans tous les cas **une différence significative** entre les deux groupes, avec **une plus franche et rapide amélioration de la dépression chez les patients du groupe 0.**

SCRIPT R

Data-Management

```
groupes<-read.csv2("Ham Groupes.csv")
ham<-read.csv2("Ham HDRS.csv")
scl<-read.csv2("Ham SCL90.csv", na.strings=c("NA","ND",""))

library(Hmisc)
library(dplyr)
library(corrplot)
library(psy)
library(boot)
library(survminer)
library(survival)
library(zoo)
library(xts)
library(lme4)
library(gplots)
```

1.A/ Pour HDRS :

```
ham$VISIT <- factor(ham$VISIT, levels=c("J0","J4","J7","J14","J21","J28","J42","J56"))
ham <- ham[order(ham$NUMERO, decreasing=F),]
ham$NUMERO[duplicated(ham)]
dim(ham)
lapply (ham[1], function(x) length(unique(x)))
subset(ham,ham$NUMERO==0)

### Fusion des colonnes 16A et 16B
ham$HAMD16=ham$HAMD16A
ham$HAMD16[!is.na(ham$HAMD16B)]=ham$HAMD16B[!is.na(ham$HAMD16B)]
ham<- subset(ham, select = -c(18,19))
```

```
## Gestion des données manquantes (ligne 741)
describe(ham)
which(is.na(ham$HAMD1))
ham<-ham[-741,]
describe(ham)
```

```
# Création de la variable correspondant aux scores totaux :
ham$score <- apply(ham[,3:19],1,sum,na.rm=T)
```

```
# Changement de nom des variables :
# colnames(hamJ0)[4:20] <- c("Humeur dépressive", "Sentim. culpab.", "Suicide", "Insomnie déb.", "Insomnie mil.", "Insomnie mat.", "Travail / Act.", "Ralentissement", "Agitation", "Anxiété psychique", "Anxiété somatique", "Sympt. gastro", "Sympt. généraux", "Sympt. génitaux", "Hypochondrie", "Prise conscience", "Perte de poids")
```

```
# Vérification du jeu de données
str(ham)
summary(ham)
describe(ham)
```

1.B/ Pour SCL90

```
summary(scl)
scl$VISIT <- factor(scl$VISIT, levels=c("J0","J4","J7","J14","J21","J28","J42","J56"))
scl <- scl[order(scl$NUMERO, decreasing=F),]
scl$NUMERO[duplicated(scl)]
```

```

### Gestion des données aberrantes
apply(scl,2,table,useNA="always")
scl[,3:92] <- replace(scl[,3:92],scl[,3:92]>4, NA)
apply(scl,2,table,useNA="always")

### Gestion des données manquantes
sum(is.na(scl))
# Proportion de données manquantes :
100-sum(!is.na(scl))*100/(1034*92)
# Répartition des données manquantes :
sapply(scl,function(x) sum(is.na(x))/100/length(x))
# Imputation des données manquantes par le mode :
scl$questions <- c(names(select (scl,starts_with("Q"))))
for (i in scl$questions){scl[,i]<-as.factor(scl[,i])}
for (i in scl$questions){scl[,i]<-impute(scl[,i],mode)}
for (i in scl$questions){scl[,i]<-as.integer(scl[,i])}
describe(scl)

### Création des sous-scores de SCL90 :
scl$Somatisation <- apply(scl[,c(3,6,14,29,44,50,51,54,55,58,60,42)],1,mean,na.rm=TRUE)
scl$Sympt.obsess. <- apply(scl[,c(11,12,30,40,5,47,48,53,57,67)],1,mean,na.rm=TRUE)
scl$Vulnérabilité <- apply(scl[,c(8,23,36,38,39,43,63,71,75)],1,mean,na.rm=TRUE)
scl$Dépression <- apply(scl[,c(7,16,17,22,24,28,31,32,33,34,56,73,81)],1,mean,na.rm=TRUE)
scl$Anxiété <- apply(scl[,c(4,19,25,35,41,59,74,80,82,88)],1,mean,na.rm=TRUE)
scl$Hostilité <- apply(scl[,c(13,26,65,69,76,83)],1,mean,na.rm=TRUE)
scl$Phobies <- apply(scl[,c(15,27,49,72,77,84,52)],1,mean,na.rm=TRUE)
scl$Paranoïa <- apply(scl[,c(10,20,45,70,78,85)],1,mean,na.rm=TRUE)
scl$Psychotique <- apply(scl[,c(9,18,37,64,79,86,87,89,92,90)],1,mean,na.rm=TRUE)
scl$Sympt.divers <- apply(scl[,c(21,46,61,62,66,68,91)],1,mean,na.rm=TRUE)
scl.dim<-scl[,c(1:2, 93:102)]

```

1.C/ Pour Groupes

```

summary(groupes)
describe(groupes)
str(groupes)
groupes$NUMERO[duplicated(groupes)]
groupes <- groupes[order(groupes$NUMERO, decreasing=F),]

```

2. Fusion des fichiers

```

ham.gr <- merge(groupes, ham, by="NUMERO", all=T)
ham.tot <- merge(ham.gr, scl, by=c("NUMERO","VISIT"), all=T)

## Conversion des fichiers complets larges en fichiers longs
# Sans SCL90 :
ham.gr.wide <- reshape(ham.gr, direction="wide", timevar="VISIT",idvar="NUMERO",
v.names=c("HAMD1","HAMD2","HAMD3","HAMD4","HAMD5","HAMD6","HAMD7","HAMD8","HAMD9","HAMD10","HAMD11","HAMD12","HAMD13","HAMD14","HAMD15","HAMD16","HAMD17","score"))

# Avec SCL90 :
ham.tot.wide <- reshape(ham.tot, direction="wide", timevar="VISIT",idvar="NUMERO",
v.names=c("HAMD1","HAMD2","HAMD3","HAMD4","HAMD5","HAMD6","HAMD7","HAMD8","HAMD9","HAMD10","HAMD11","HAMD12","HAMD13","HAMD14","HAMD15","HAMD16","HAMD17","score","Q1", "Q2", "Q3", "Q4","Q5","Q6","Q7","Q8","Q9","Q10","Q11","Q12","Q13","Q14","Q15","Q16","Q17","Q18","Q19","Q20","Q21","Q22","Q23","Q24","Q25","Q26","Q27","Q28","Q29","Q30","Q31","Q32","Q33","Q34","Q35","Q36","Q37","Q38","Q39","Q40","Q41","Q42","Q43","Q44","Q45","Q46","Q47","Q48","Q49","Q50","Q51","Q52","Q53","Q54","Q55","Q56",

```

"Q57", "Q58", "Q59", "Q60", "Q61", "Q62", "Q63", "Q64", "Q65", "Q66", "Q67", "Q68", "Q69", "Q70", "Q71", "Q72", "Q73", "Q74", "Q75", "Q76", "Q77", "Q78", "Q79", "Q80", "Q81", "Q82", "Q83", "Q84", "Q85", "Q86", "Q87", "Q88", "Q89", "Q90", "Somatisation", "Sympt.obsess.", "Vulnérabilité", "Dépression", "Anxiété", "Hostilité", "Phobies", "Paranoïa", "Psychotique", "Sympt.divers"))

QUESTION 1

```
## Création d'une base de données à J0
hamJ0 <- subset(ham.gr, VISIT=="J0")
table(hamJ0$VISIT)
nrow(hamJ0)
sum(is.na(hamJ0))

## Création d'une base de données à J56
hamJ56 <- subset(ham.gr, VISIT=="J56")
table(hamJ56$VISIT)
nrow(hamJ56)
sum(is.na(hamJ56))

# Nombre de patients n'ayant pas achevé le suivi sur 56 jours :
nrow(hamJ0)-nrow(hamJ56)
```

1. Evaluation de l'échelle d'Hamilton à J0

1.1 Validité interne : Que vaut-il ?

Distribution des réponses par item :

```
barplot(table(hamJ0$HAMD1, useNA = "ifany"), col=c("#8BFE73", "#FEF744", "#FEB901", "#FF5E4D", "grey"),
main="Item 1", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD2, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#FEF744", "#FEB901",
"#FF5E4D", "grey"), main="Item 2", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD3, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#FEF744", "#FEB901",
"#FF5E4D", "grey"), main="Item 3", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD4, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item 4",
las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD5, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item 5",
las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD6, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item 6",
las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD7, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#FEF744", "#FEB901",
"#FF5E4D", "grey"), main="Item 7", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD8, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#FEF744", "#FEB901",
"#FF5E4D", "grey"), main="Item 8", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD9, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#FEF744", "#FEB901",
"#FF5E4D", "grey"), main="Item 9", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD10, useNA = "ifany"), col=c("#8BFE73", "#FEF744", "#FEB901", "#FF5E4D", "grey"),
main="Item 10", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD11, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#FEF744", "#FEB901",
"#FF5E4D", "grey"), main="Item 11", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD12, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item
12", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD13, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item
13", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD14, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item
14", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD15, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#FEF744", "#FEB901",
"#FF5E4D", "grey"), main="Item 15", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD16, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item
16", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ0$HAMD17, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item
17", las=1, ylim=c(0,100), space=0.4)
```

```

## Nom des items :
colnames(hamJ0)[4:20] <- c("Humeur dépressive", "Sentim. culpab.", "Suicide", "Insomnie déb.", "Insomnie mil.", "Insomnie mat.", "Travail / Act.", "Ralentissement", "Agitation", "Anxiété psychique", "Anxiété somatique", "Sympt. gastro", "Sympt. généraux", "Sympt. génitaux", "Hypochondrie", "Prise conscience", "Perte de poids")

## Matrice de correlation des items entre eux
matrix<-hamJ0[,c(4:20)]
round(cor(matrix), 3)
corrplot(cor(matrix), method="color",addCoef.col = "black", tl.col="black")
# Pas de fortes corrélation, pas de redondance des items.

## Diagramme des valeurs propres
head(hamJ0[,c(4:20)])
mdspca(hamJ0[,c(4:20)])
scree.plot(hamJ0[,c(4:20)], simu=10, use = "P")
# On estime à la louche 2 à 4 dimensions proéminentes.

## Analyse factorielle avec rotation Varimax
# A 4 facteurs :
AFJ0.4F<-print(factanal(na.omit(hamJ0[,c(4:20)]), factor=4), cutoff=0)
write.csv2(x=round(AFJ0.4F$loadings, 3), "AFJ0.4F.nom.csv")
# A 3 facteurs :
AFJ0.3F<-print(factanal(na.omit(hamJ0[,c(4:20)]), factor=3), cutoff=0)
write.csv2(x=round(AFJ0.3F$loadings, 3), "AFJ0.3F.nom.csv")
# A 2 facteurs :
AFJ0.2F<-print(factanal(na.omit(hamJ0[,c(4:20)]), factor=2), cutoff=0)
write.csv2(x=round(AFJ0.2F$loadings, 3), "AFJ0.2F.nom.csv")
# L'analyse factorielle à deux facteurs semble discriminer le mieux les dimensions de HDRS (p<0,001).

## Evaluation de la cohérence interne des échelles de chaque dimension : le coefficient alpha de Cronbach
# Création des échelles :
Echelle1<-hamJ0[,c(4:6, 8:10, 12:14, 18:19)]
Echelle2<-hamJ0[,c(7, 11, 15:17, 20)]

# Calcul des coefficients alpha de chaque échelle :
cronbach(Echelle1)
cronbach(Echelle2)

# Création des sous-scores totaux de chaque dimension :
hamJ0$total.somat <- apply(hamJ0[,c(7, 11, 15, 16, 17, 20)], 1, sum, na.rm=FALSE)
hamJ0$total.somat <- as.numeric(hamJ0$total.somat)
describe(hamJ0$total.somat)
hamJ0$total.psychosocial <- apply(hamJ0[,c(4, 5, 6, 8, 9, 10, 12, 13, 14, 18, 19)], 1, sum, na.rm=FALSE)
hamJ0$total.psychosocial <- as.numeric(hamJ0$total.psychosocial)
describe(hamJ0$total.psychosocial)

# Bootstrap pour estimer les intervalles de confiance des coefficients :
# Echelle 1 :
a1.boot<-function(data,x) {cronbach(data[x,])[3]}
res.a1<-boot(Echelle1, a1.boot, 1000)
quantile(res.a1$t, c(0.025, 0.975))
hist(res.a1$t)
# Echelle 2 :
a2.boot<-function(data,x) {cronbach(data[x,])[3]}
res.a2<-boot(Echelle2, a2.boot, 1000)
quantile(res.a2$t, c(0.025, 0.975))
hist(res.a2$t)

# Coefficients <0,6, mauvaise validité interne de cette échelle bidimensionnelle.

```

```

### 1.2. Validité externe : Que mesure-t-il ?
## Validité concourante et divergeante :

## Comparaison aux différentes dimensions connues de SCL90 :
# Création du fichier large des sous-scores et des échelles :
Ham.SCL.norm <- merge(hamJ0, scl.dim, by=c("NUMERO", "VISIT"), all=TRUE)
Ham.SCL.wide <- reshape(Ham.SCL.norm, direction="wide", timevar="VISIT", idvar="NUMERO",
v.names=c("GROUPE", "Humeur dépressive", "Sentim. culpab.", "Suicide", "Insomnie déb.", "Insomnie mil.", "Insomnie mat.", "Travail / Act.", "Ralentissement", "Agitation", "Anxiété psychique", "Anxiété somatique", "Sympt. gastro", "Sympt. généraux", "Sympt. génitaux", "Hypochondrie", "Prise conscience", "Perte de poids", "score", "total.somat", "total.psychos", "Somatisation", "Sympt.obsess.", "Vulnérabilité", "Dépression", "Anxiété", "Hostilité", "Phobies", "Paranoïa", "Psychotique", "Sympt.divers"))
Ham.SCL.wide.J0 <- subset(Ham.SCL.wide, select = -c(2, 33:249))

# Matrice de corrélation des sous-échelles de Hamilton et SCL :
var<-c("total.somat.J0", "total.psychos.J0",
"Somatisation.J0", "Sympt.obsess..J0", "Vulnérabilité.J0", "Dépression.J0", "Anxiété.J0", "Hostilité.J0", "Phobies.J0", "Paranoïa.J0", "Psychotique.J0", "Sympt.divers.J0")
corrplot(cor(Ham.SCL.wide.J0[,var],use="complete.obs"), method="color",addCoef.col = "black", tl.col="black",
tl.srt=45)

# Analyse en composante principale de comparaison des sous-échelles :
mdspca(Ham.SCL.wide.J0[,var])
sphpca(Ham.SCL.wide.J0[,var], v=30)
# Pas de lien évident entre les deux sous-échelles de Hamilton (total somatique et total psychologique) et les sous-scores de SCL.

# Comparaison du score global d'Hamilton aux sous-dimensions de SCL :
corrplot(cor(Ham.SCL.wide.J0[,c(19, 22:31)],use="complete.obs"), method="color",addCoef.col = "black",
tl.col="black")
fPCA(data=Ham.SCL.wide.J0, y="score.J0",
x=c("Somatisation.J0", "Sympt.obsess..J0", "Vulnérabilité.J0", "Dépression.J0", "Anxiété.J0", "Hostilité.J0", "Phobies.J0", "Paranoïa.J0", "Psychotique.J0", "Sympt.divers.J0"), partial="No")

# Pas de franche corrélation entre le score global de l'échelle Hamilton et les sous-dimension de SCL.
# Ainsi, le score Hamilton à J0 ne semble pas très performant pour évaluer la dépression, en terme de validité interne comme externe.

### 2. Evaluation de l'échelle à J56

### 2.1 Validité interne : Que vaut-il ?
## Distribution des réponses par item :

barplot(table(hamJ56$HAMD1, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#fef744", "#FEB901",
"#FF5E4D", "grey"), main="Item 1", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD2, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#fef744", "#FEB901",
"#FF5E4D", "grey"), main="Item 2", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD3, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#FEB901", "#FF5E4D", "grey"),
main="Item 3", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD4, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item 4",
las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD5, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item 5",
las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD6, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item 6",
las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD7, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#fef744", "#FEB901",
"#FF5E4D", "grey"), main="Item 7", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD8, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#fef744", "#FEB901",
"#FF5E4D", "grey"), main="Item 8", las=1, ylim=c(0,100), space=0.4)

```

```

barplot(table(hamJ56$HAMD9, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#fef744", "#FEB901",
 "#FF5E4D", "grey"), main="Item 9", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD10, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#fef744", "#FEB901",
 "#FF5E4D", "grey"), main="Item 10", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD11, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#fef744", "#FEB901",
 "#FF5E4D", "grey"), main="Item 11", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD12, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item
 12", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD13, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item
 13", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD14, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item
 14", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD15, useNA = "ifany"), col=c("#0DAA00", "#8BFE73", "#fef744", "#FEB901",
 "#FF5E4D", "grey"), main="Item 15", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD16, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item
 16", las=1, ylim=c(0,100), space=0.4)
barplot(table(hamJ56$HAMD17, useNA = "ifany"), col=c("#8BFE73", "#FEB901", "#FF5E4D", "grey"), main="Item
 17", las=1, ylim=c(0,100), space=0.4)

## Nom des items :
colnames(hamJ56)[4:20] <- c("Humeur dépressive", "Sentim. culpab.", "Suicide", "Insomnie déb.", "Insomnie mil.",
 "Insomnie mat.", "Travail / Act.", "Ralentissement", "Agitation", "Anxiété psychique", "Anxiété somatique", "Sympt.
 gastro", "Sympt. généraux", "Sympt. génitaux", "Hypochondrie", "Prise conscience", "Perte de poids")

## Matrice de correlation des items entre eux
matrix2<-hamJ56[,c(4:20)]
round(cor(matrix2), 2)
corrplot(cor(matrix2), method="color",addCoef.col = "black", tl.col="black")
# Peu de fortes corrélation, peu de redondance des items.

## Diagramme des valeurs propres
scree.plot(hamJ56[,c(4:20)], simu=10, use = "P")
# L'outil semble parfaitement unidimensionnel à J56, pas de nécessité de faire d'analyse factorielle.

# Coefficient alpha de Cronbach, Bootstrap pour estimer son intervalle de confiance :
cronbach(hamJ56[,c(4:20)])
a3.boot<-function(data,x) {cronbach(data[x,])[3]}
res.a3<-boot(hamJ56[,c(4:20)], a3.boot, 1000)
quantile(res.a3$t, c(0.025, 0.975))
hist(res.a3$t)
# Alpha = 0.82, très bonne validité interne.

### 2.2. Validité externe : Que mesure-t-il ?
## Validité concourante et divergente :

## Comparaison aux différentes dimensions connues de SCL90 :
# Création du fichier large du score de Hamilton total et des échelles de SCL :
Ham.SCL.norm56 <- merge(hamJ56, scl.dim, by=c("NUMERO","VISIT"), all=TRUE)
Ham.SCL.wide56 <- reshape(Ham.SCL.norm56, direction="wide", timevar="VISIT", idvar="NUMERO",
 v.names=c("GROUPE", "Humeur dépressive", "Sentim. culpab.", "Suicide", "Insomnie déb.", "Insomnie mil.", "Insomnie
 mat.", "Travail / Act.", "Ralentissement", "Agitation", "Anxiété psychique", "Anxiété somatique", "Sympt. gastro", "Sympt.
 généraux", "Sympt. génitaux", "Hypochondrie", "Prise conscience", "Perte de poids", "score",
 "Somatisation", "Sympt.obsess.", "Vulnérabilité", "Dépression", "Anxiété", "Hostilité", "Phobies", "Paranoïa", "Psychotique"
 , "Sympt.divers"))
Ham.SCL.wide.J56 <- subset(Ham.SCL.wide56, select = -c(1:222))

# Matrice de corrélation du score total de Hamilton et des sous-échelles de SCL :
var56<-c("score.J56",
 "Somatisation.J56", "Sympt.obsess..J56", "Vulnérabilité.J56", "Dépression.J56", "Anxiété.J56", "Hostilité.J56", "Phobies.J5
 6", "Paranoïa.J56", "Psychotique.J56", "Sympt.divers.J56")

```

```

corrplot(cor(Ham.SCL.wide.J56[,var56],use="complete.obs"), method="color",addCoef.col = "black", tl.col="black",
tl.srt=45)

# Explication du score global d'Hamilton par les sous-dimensions de SCL :
fPCA(data=Ham.SCL.wide.J56, y="score.J56",
x=c("Somatisation.J56","Sympt.obsess.J56","Vulnérabilité.J56","Dépression.J56","Anxiété.J56","Hostilité.J56","Phobies.J56","Paranoia.J56","Psychotique.J56","Sympt.divers.J56"))

# Bonnes corrélations entre les sous-scores de SCL et le score d'Hamilton à J56 :
# bonne fiabilité de l'outil pour évaluer la dépression à J56.

##### Question 2 #####

```

Répartition des Perdus de Vue

```

Ham.gr.wide.scores <- ham.gr.wide[,c(1, 2, 20, 38, 56, 74, 92, 110, 128, 146)]
table(is.na(Ham.gr.wide.scores[,3]))
table(is.na(Ham.gr.wide.scores[,10]))
table(is.na(subset(Ham.gr.wide.scores[,10], Ham.gr.wide.scores$GROUPE=="0")))
table(is.na(subset(Ham.gr.wide.scores[,10], Ham.gr.wide.scores$GROUPE=="1")))
Ham.gr.wide.scores$Perdus <- factor(ifelse(is.na(Ham.gr.wide.scores$score.J56),1,0), levels=c("1","0"),
labels=c("PDV", "non PDV"))
table(Ham.gr.wide.scores$Perdus)

```

Conditions de validité

```

table(Ham.gr.wide.scores$GROUPE, Ham.gr.wide.scores$Perdus, deparse.level = 2)
by(Ham.gr.wide.scores$score.J0, Ham.gr.wide.scores$Perdus, mean,na.rm=TRUE)
by(Ham.gr.wide.scores$score.J0, Ham.gr.wide.scores$Perdus, sd,na.rm=TRUE)
hist(Ham.gr.wide.scores$score.J0, col = "#e0b04b", main="")

```

Test statistique

Comparaison du nombre de PDV entre les groupes :

```

chisq.test(Ham.gr.wide.scores$GROUPE, Ham.gr.wide.scores$Perdus, correct=F)

```

Comparaison des scores à J0 entre les perdus de vue :

```

t.test(Ham.gr.wide.scores$score.J0~Ham.gr.wide.scores$Perdus, var.equal=TRUE)

```

Description de l'évolution du score en fonction du temps

```

HDRSGp0 <- ham.gr [ham.gr $ GROUPE==0,]
HDRSGp1 <- ham.gr [ham.gr $ GROUPE==1,]
head(factor(HDRSGp0 $ VISIT))
HDRSGp0$VISIT <- factor(HDRSGp0 $ VISIT, c ("J0", "J4", "J7", "J14", "J21", "J28", "J42", "J56"))
head(factor(HDRSGp1 $ VISIT))
HDRSGp1$VISIT <- factor(HDRSGp1$VISIT, c("J0", "J4", "J7", "J14", "J21", "J28", "J42", "J56"))
boxplot(HDRSGp0$score~HDRSGp0$VISIT, col="#70ED24", ylab = "HDRS Score total", xlab = "Visite",
main="Evolution du score dans le groupe 0")
boxplot(HDRSGp1$score~HDRSGp1$VISIT, col="#F33D23", ylab = "HDRS Score total", xlab = "Visite",
main="Evolution du score dans le groupe 1")

```

Approche LOCF

```

HDRSG.LOCF <- as.data.frame(na.locf(t(ham.gr.wide[1:146, c(1,2,20,38,56,74,92,110,128,146)])))
HDRSG.LOCF <- as.data.frame(t(HDRSG.LOCF))
HDRSG.LOCF$Diff <- apply(HDRSG.LOCF[,c(10,3)], 1, diff, na.rm=T)
HDRSG.LOCFGp0 <- HDRSG.LOCF[HDRSG.LOCF$GROUPE== 0,]
HDRSG.LOCFGp1 <- HDRSG.LOCF[HDRSG.LOCF$GROUPE== 1,]

```

Représentations graphiques

```

boxplot(HDRSG.LOCF$score.J0~HDRSG.LOCF$GROUPE, col=c("#70ED24", "#F33D23"), ylab="Score total à J0",
xlab = "Groupe", main = "Score Hamilton à J0")
boxplot(HDRSG.LOCF$score.J56~HDRSG.LOCF$GROUPE, col=c("#70ED24", "#F33D23"), ylab="Score total à J56",
xlab = "Groupe", main = "Score Hamilton à J56")

```

```

## Test t ##
t.test(HDRSG.LOCF$Diff~HDRSG.LOCF$GROUPE, var.equal = TRUE)

## Conditions de validité du test t ##
hist(HDRSG.LOCF$Diff, col="light blue", main = "Distribution de la différence des scores", xlab = "Différence", ylab = "Densité", ylim = c(0, 0.055), xlim = c(-5,40), prob = "TRUE" )
lines(density(HDRSG.LOCF$Diff, na.rm=TRUE), lwd=2)
aggregate(Diff~GROUPE, data=HDRSG.LOCF, sd)
qqnorm(HDRSG.LOCF$Diff)
qqline(HDRSG.LOCF$Diff)

### Modèle mixte avec effet sujet aléatoire ###
HDRSG = ham.gr
HDRSG $ NUMERO <- factor (HDRSG $ NUMERO)
HDRSG $ VISIT.f <- factor(HDRSG$VISIT, levels = c("J0","J4","J7","J14","J21","J28","J42","J56"), labels = c("1","2","3","4","5","6","7","8"))
HDRSG $ VISIT.car <- as.character(HDRSG $ VISIT.f)
HDRSG $ VISIT.num <- as.numeric(HDRSG$VISIT.car)
plotmeans(HDRSG$score~HDRSG$VISIT.f, xlab = "Temps", ylab = "Score Hamilton", ylim=c(0,30))

## Sans interaction ##
mod1 <- lmer(score~GROUPE+VISIT.num+(1|NUMERO), data = HDRSG)
summary(mod1)
hist(resid(mod1), col="#e6a8d7", main = "Résidus du modèle 1", xlab = "Valeur des résidus", ylab = "Distribution", nclass=20)

## Avec interaction ##
mod2 <- lmer(score~GROUPE+VISIT.num+GROUPE*VISIT.num+(1|NUMERO), data = HDRSG)
summary(mod2)
hist(resid(mod2), col="#bcf585", main = "Résidus du modèle 2", xlab = "Valeur des résidus", ylab = "Distribution", nclass=20)

```

Question 3

Crédation du fichier de survie :

```

ham.gr.surv <- ham.gr[,c("NUMERO", "VISIT", "score", "GROUPE")]
j0 <- subset(ham.gr.surv, VISIT == "J0")
j0$SCORE_J0 <- j0$score
ham.gr.surv <- merge(ham.gr.surv, j0[,c("NUMERO", "SCORE_J0")], by = c("NUMERO"), all=T)

```

Crédation de la variable "Réponse" :

```

ham.gr.surv$RESPONSE <- ifelse(ham.gr.surv$score <= ham.gr.surv$SCORE_J0 / 2, 1, 0)
ham.gr.surv$VISIT<-as.numeric(ham.gr.surv$VISIT)

```

Recodage de chaque variable quantitative "cs" au jour correspondant pour une échelle appropriée :

```

ham.gr.surv$VISIT <- ifelse(ham.gr.surv$VISIT == "1", 0,
  ifelse(ham.gr.surv$VISIT == "2", 4,
  ifelse(ham.gr.surv$VISIT == "3", 7,
  ifelse(ham.gr.surv$VISIT == "4", 14,
  ifelse(ham.gr.surv$VISIT == "5", 21,
  ifelse(ham.gr.surv$VISIT == "6", 28,
  ifelse(ham.gr.surv$VISIT == "7", 42,
  ifelse(ham.gr.surv$VISIT == "8", 56,
  ham.gr.surv$VISIT)))))))

```

```

ham.gr.wide$J4 <- ifelse(ham.gr.wide$score.J4 < ham.gr.wide$score.J0/2, 1, 0)
ham.gr.wide$J7 <- ifelse(ham.gr.wide$score.J7 < ham.gr.wide$score.J0/2, 1, 0)
ham.gr.wide$J14 <- ifelse(ham.gr.wide$score.J14 < ham.gr.wide$score.J0/2, 1, 0)
ham.gr.wide$J21 <- ifelse(ham.gr.wide$score.J21 < ham.gr.wide$score.J0/2, 1, 0)

```

```

ham.gr.wide$J28 <- ifelse(ham.gr.wide$score.J28 < ham.gr.wide$score.J0/2, 1, 0)
ham.gr.wide$J42 <- ifelse(ham.gr.wide$score.J42 < ham.gr.wide$score.J0/2, 1, 0)
ham.gr.wide$J56 <- ifelse(ham.gr.wide$score.J56 < ham.gr.wide$score.J0/2, 1, 0)

ham.gr.wide$delay <- apply(ham.gr.wide[,c("J4","J7","J14","J21","J28","J42","J56")], 1, function(x) which(x == 1)[1])
ham.gr.wide$delay <- ifelse(ham.gr.wide$delay == 1, 4,
  ifelse(ham.gr.wide$delay == 2, 7,
    ifelse(ham.gr.wide$delay == 3, 14,
      ifelse(ham.gr.wide$delay == 4, 21,
        ifelse(ham.gr.wide$delay == 5, 28,
          ifelse(ham.gr.wide$delay == 6, 42,
            ifelse(ham.gr.wide$delay == 7, 56,
              ham.gr.wide$delay)))))))

```

Création et importation d'un dataframe manuellement complété contenant les données de survie et son délai :

```

H6<-read.csv2("H6.csv")
HH<- merge(H6, ham.gr.wide, by = "NUMERO")

```

Courbes de survie en fonction du groupe :

```

survie2<-survfit(Surv(delay2, event2) ~ GROUPE.x, data = HH)
plot(surveie2,main = "Courbe de survie", xlab = "Durée de survie (jours)", ylab = "Taux de
survivants",xlim=c(0,60),lty=c(1,2,2,1, 2, 2), lwd=1.5, col=c("#70ED24", "light green","light green", "#F33D23",
"#DF998F", "#DF998F"), conf.int = T, mark.time=T)
abline(h=0.5,lty=2:3,col="grey")
legend("topright",lty=c(1,1, 2:3),col=c("#70ED24","#F33D23", "grey"), lwd=2, c("Groupe 0","Groupe 1",
"Médiane"),inset = .02)

```

Test de logrank :

```

Logrank<-survdiff(Surv(delay2, event2) ~ GROUPE.x, data = HH)
Logrank

```

```

text(0.15, 0.05, "p = 0.0003", adj = c(0,1), cex = 1)

```

#Modèle de Cox :

```

modele_c <- coxph(Surv(delay2, event2) ~ GROUPE.x, data = HH)
summary(modele_c)

```

```

plot(cox.zph(modele_c), main = "Hypothèse des risques instantanés proportionnels")

```