

Devoir Statistiques Avancées

Thomas Husson, Groupe 52

Table des matières

1 Énoncé et présentation des échelles	3
1.A Énoncé du devoir	3
1.B Échelles	3
2 Gestion des données	4
2.A Présentation des fichiers de données	4
2.B Import des données et data management	4
3 Question 1 : Validation de l'échelle Hamilton	6
3.A Validation à J0	6
3.B Validation à J56	16
4 Question 2 : Comparaison de la réponse au traitement entre deux groupes de patients	23
4.A Gestion des données	23
4.B Scores bruts à J0 dans chaque groupe	23
4.C Évolution des scores bruts entre J0 et J56 dans chaque groupe	24
4.D Comparaison de la réponse au traitement entre deux groupes de patients	26
4.E Conclusion sur la comparaison de la réponse au traitement entre les deux groupes de patients	31
5 Question 3 : Analyse de survie	32
5.A Critère binaire de réponse (50% à la dernière visite observée)	32
5.B Intégration du critère binaire de réponse dans une analyse de survie	33
5.C Analyse de survie avec le critère binaire de réponse	34
6 Annexe – Code R	37

! Important

Utilisation de l'IA

Des LLMs ont été utilisés à plusieurs reprises dans ce devoir, pour deux tâches principales :

- En cas de problème d'exécution du code R (pour suggérer correction et amélioration)
- Pour amélioration du rendu depuis un fichier Quarto Markdown vers PDF. Notamment certaines fonctions dont l'output R n'est pas compatible avec le rendu pdf (par exemple, `factanal()` pour l'analyse factorielle ou le rendu des tableaux automatisé des coefficients alpha de Cronbach et leur IC)

Utilisation de Gemini pour le rendu Quarto et Github Copilot pour le code R.

1 Énoncé et présentation des échelles

1.A Énoncé du devoir

Consigne :

- Étude d'épidémiologie clinique avec mesures répétées
- Données :
 - 146 patients déprimés
 - Évaluations à J0, J4, J7, J14, J21, J28, J42, J56
 - Autoévaluation (SCL90) et hétéroévaluation (échelle de dépression de Hamilton)
- Questions :
 1. Validation de l'échelle de dépression de Hamilton aux temps J0 et J56
 2. Comparaison de la réponse au traitement entre deux groupes de patients (groupe=0 et groupe=1) en utilisant le score brut de Hamilton avec une approche LOCF puis un modèle mixte
 3. Réponse à la question 2 en utilisant un critère binaire censuré « réponse au traitement » défini par une chute de 50% à l'échelle de Hamilton par rapport à J0
- Fichiers :
 - Fichier groupe (`outil groupe.xlsx`) (2 sous-groupes de patients)
 - Fichier autoévaluation (`outil autoeval.xlsx`) (SCL 90)
 - Fichier hdrs (`outil hdrs.xlsx`) (échelle de Hamilton)

1.B Échelles

Table 1: Présentation des échelles utilisées dans le devoir

	Échelle de Hamilton (HDRS)	Échelle SCL90
Objectif	Mesure l'intensité de la symptomatologie dépressive	“Inconfort psychopathologique” selon plusieurs dimensions.
Type	Hétéro-évaluation	Autoévaluation
Méthode	17 items codés de 2 à 4- Score 7 : pas de dépression clinique- Score 8–15 : dépression mineure- Score > 15 : dépression majeure	10 dimensions : somatisation, symptômes obsessionnels, sensibilité interpersonnelle, dépression, anxiété, hostilité, phobies, traits paranoïaques, traits psychotiques et symptômes divers.

2 Gestion des données

2.A Présentation des fichiers de données

Les 3 fichiers sont en format “large” : chaque ligne correspond à une visite d'un patient et une colonne par item de l'échelle (sauf l'item 16 = PERTE DE POIDS qui est codé en deux variables HAMD16A et HAMD16B dans l'échelle de Hamilton, selon que la perte de poids est déclarée par le patient ou appréciée par le médecin)

On créera donc une colonne `hdrs$HAMD16` qui prendra la valeur de `hdrs$HAMD16A` si elle est remplie, sinon la valeur de `hdrs$HAMD16B`.

2.A.1 Fichier Hamilton

- 1053 observations, 20 variables pour 146 patients
- On ajoute une colonne `score` qui contient le score total de l'échelle de Hamilton (somme des items)
- Les données d'une ligne (J7 du 128ème patient) sont manquantes → on supprime cette ligne.

2.A.2 Fichier scl90

- 1034 observations, 92 variables, 146 patients.
- On crée 10 nouvelles variables représentant les scores moyen des 10 dimensions de l'échelle SCL90.
- Données aberrantes parfois, qui sont recodées en données manquantes et représentent ainsi 0.6% des données totales.

2.A.3 Fichier groupe

- Répartit les 146 patients en 2 groupes (1 et 0)
- Pas de NA

2.B Import des données et data management

Les données sont importées à partir de fichiers Excel.

2.B.1 SCL90

Le jeu de données `scl90` est traité de la manière suivante :

- Visites ordonnées chronologiquement
- Identification des doublons
- Visualisation et gestion des données aberrantes
- Imputation des données manquantes par le mode pour chaque question
- Création des scores moyens par dimension (10 dimensions)

- Nouveau dataframe `scl90_dim` avec uniquement les 10 dimensions

2.B.2 HDRS

Le jeu de données `hdrs` est traité de la manière suivante :

- Visites ordonnées chronologiquement
- Identification des doublons
- Fusion des variables HAMD16A et HAMD16B en une seule variable HAMD16
- Création du score total HDRS (ajouté dans la colonne `hdrs$score`)

2.B.3 Groupes

2.B.4 Fusion des 3 fichiers

Convertir `hdrs_groupe`, `scl90_groupe` et `df_total_wide` de format “large” à format “long”

3 Question 1 : Validation de l'échelle Hamilton

i Note

Consigne de la question 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.

- Vérification d'une échelle de mesure subjective = 1/ Que mesure l'instrument ? 2/ Que vaut la mesure ?
- Premier temps : Évaluation préliminaire des réponses aux items, puis chercher une corrélation entre eux par une matrice de corrélation 2 à 2
- Second temps : Analyse de la structure dimensionnelle = **que mesure l'instrument ?**
 - Exploration de la structure par analyse en composante principale : visualiser les relations entre les items
 - Détermination du nombre de dimensions : diagramme des valeurs propres (*scree plot*) permet de déterminer le nombre de dimensions (composantes principales)
 - Si structure dimensionnelle identifiée : **analyse factorielle** permet de déterminer quels items se regroupent dans chaque dimension
- Troisième temps : Évaluation de la fiabilité interne = **que vaut la mesure ?**
 - La consistance interne des items (évalue si les items sont cohérents entre eux) sera évaluée par le calcul de l'alpha de Cronbach, calculé sur l'échelle totale et sur chaque dimension identifiée précédemment.
- Quatrième temps : Évaluation de la validité = **l'instrument mesure-t-il ce qu'il est censé mesurer ?** (similaire à la question “que mesure l'instrument ?”)
 - Validité interne : déjà évaluée au cours du second temps (structure dimensionnelle)
 - Validité externe : corrélation avec d'autres instruments de mesure (ici les dimensions de l'échelle SCL90)

3.A Validation à J0

3.A.1 Description

Les réponses sont représentées :

- par des histogrammes pour chaque item de l'échelle de Hamilton à J0
- par une matrice de corrélation 2 à 2 entre les items

NB : le code R utilise une fonction pour faciliter la création des histogrammes pour chaque item.

La fonction crée un histogramme pour chaque item listés dans un vecteur créé précédemment (`hdrs_items`).

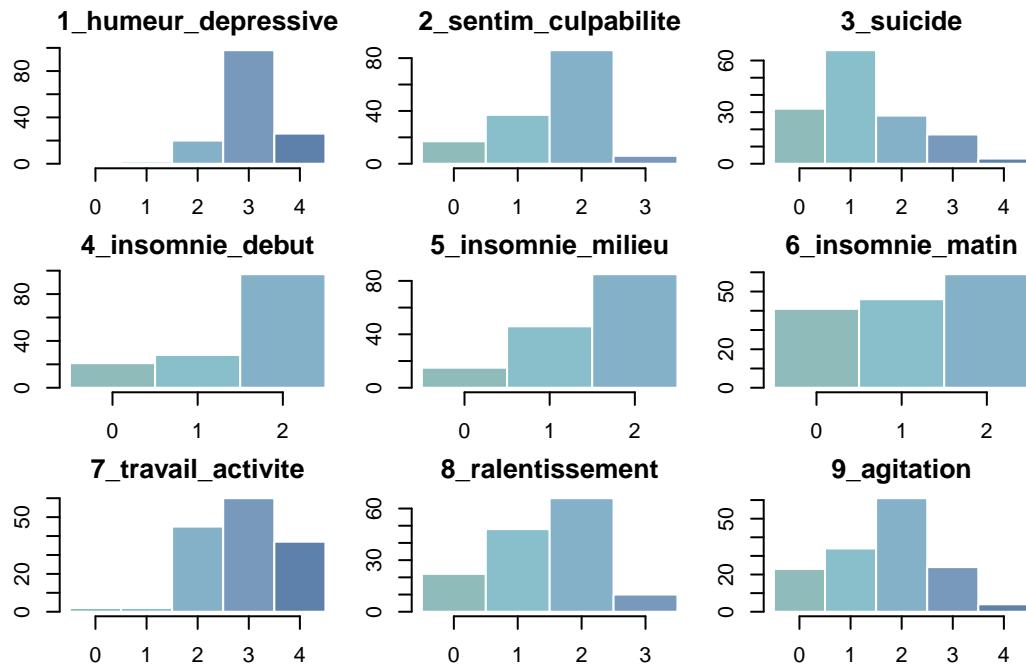


Figure 1: Histogrammes des scores des items de l'échelle de Hamilton à J0

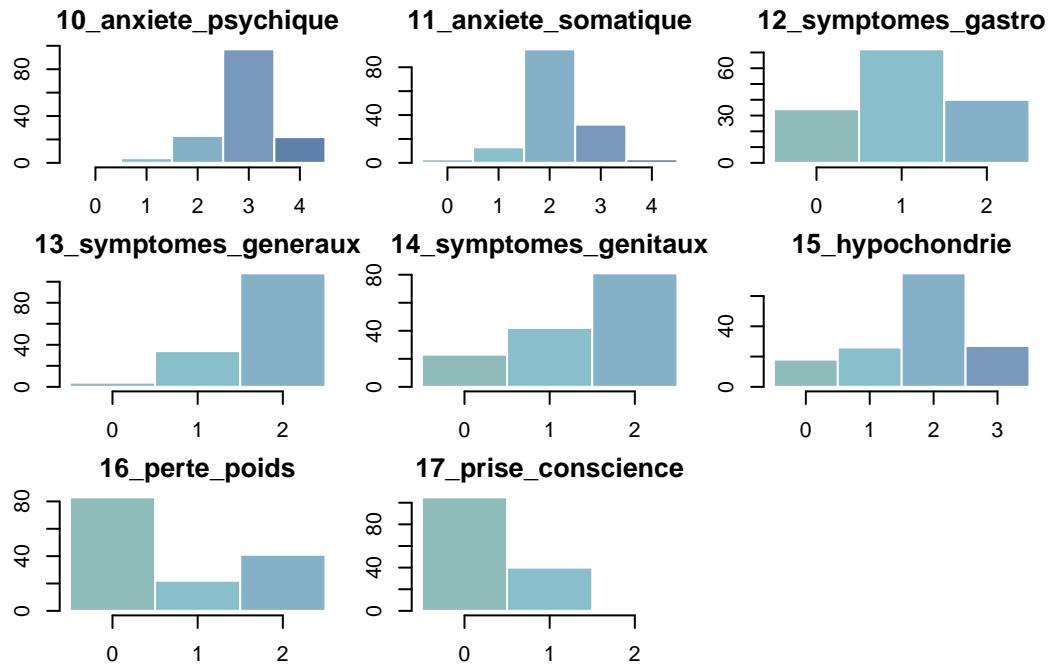


Figure 2: Histogrammes des scores des items de l'échelle de Hamilton à J0

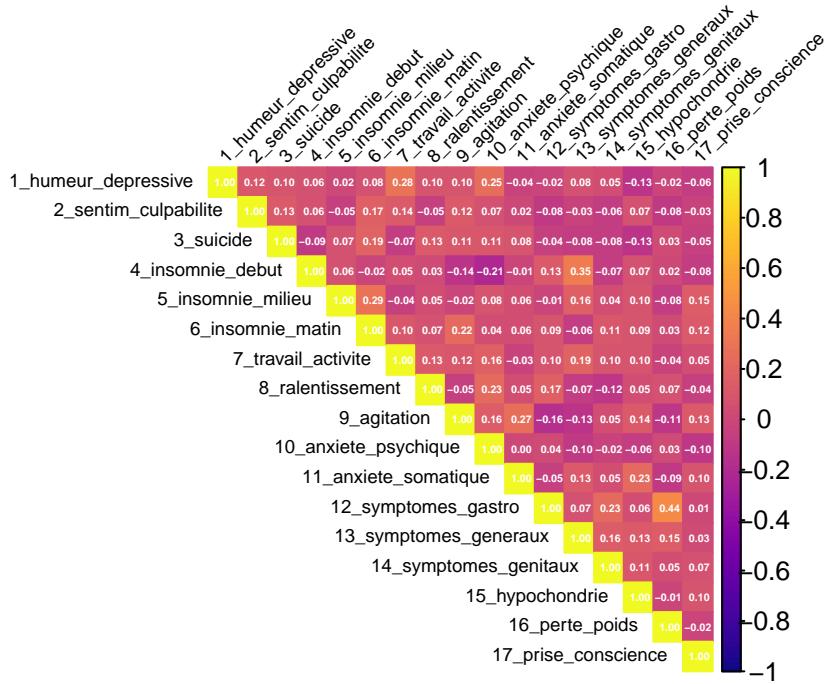


Figure 3: Matrice de corrélation entre les items de l'échelle de Hamilton à J0

- Il n'y a pas de données manquantes.
- Les histogrammes montrent que certains items ont une distribution asymétrique (ex : insomnie quelque soit le moment de la nuit, symptômes généraux, perte de poids...)
- La matrice de corrélation des items 2 à 2 ne retrouve pas de coefficient de corrélation supérieure à 0,50 en valeur absolue, il n'existe pas de redondance entre les items de l'échelle Hamilton.

3.A.2 Validité interne : structure dimensionnelle, analyse factorielle

3.A.2.1 Exploration de la structure dimensionnelle : analyse en composantes principales

- On peut réaliser une analyse en composantes principales (ACP) pour visualiser les relations entre les items.

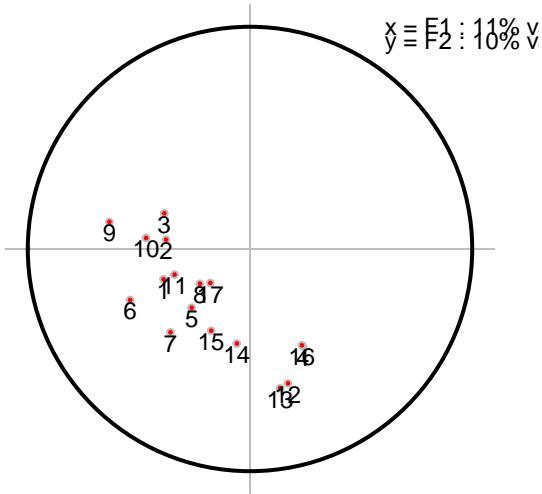


Figure 4: Analyse en composantes principales des items de l'échelle de Hamilton à J0

- Chaque point représente un item de l'échelle de Hamilton.
- Deux axes principaux :
 - l'axe horizontale x représente la première composante principale (PC1) qui explique 11% de la variance totale, l'axe vertical y représente la deuxième composante principale (PC2) qui explique 10% de la variance totale.
 - Ensemble, les deux premières composantes principales expliquent 21% de la variance totale, ce qui est relativement faible.
- La majorité des variables sont proches du centre, ce qui indique qu'elles ne contribuent pas fortement aux premières composantes principales.
- Au total, cette ACP ne révèle pas de structure dimensionnelle claire parmi les items de l'échelle de Hamilton à J0.

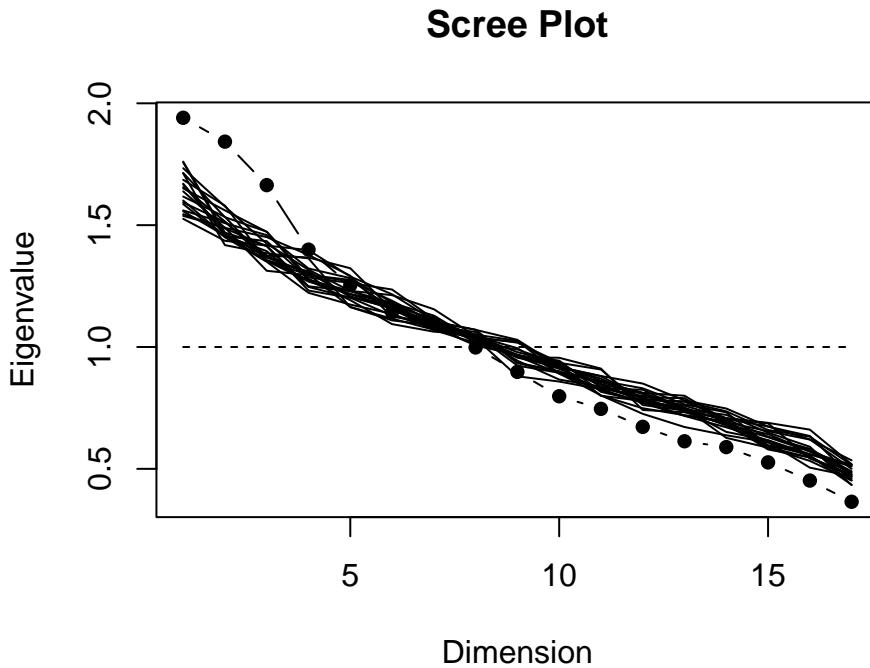


Figure 5: Diagramme des valeurs propres (scree plot) des items de l'échelle de Hamilton à J0 avec représentation de données simulées (analyse parallèle)

3.A.2.2 Exploration de la structure dimensionnelle : analyse factorielle

- À J0, le *scree plot* ne permet pas d'identifier un nombre clair de facteurs : les valeurs propres décroissent progressivement sans “coude” net.
- En analyse parallèle, on observe au moins 3 dimensions ayant une valeur propre supérieure à celle obtenue sur des données simulées.
- On pourrait réaliser des tests statistiques qui permettraient de déterminer le nombre optimal de dimensions, mais ces tests sont sujets à plusieurs biais :
 - on calculerait une p-value pour l'hypothèse “n facteurs sont suffisants”
 - mais ces tests sont difficiles à interpréter et sensibles à la taille de l'échantillon
 - On retient donc 3 facteurs principaux pour l'analyse factorielle.

Table 2: Contribution des facteurs à la variance de la réponse à chaque item du score de Hamilton évalué à J0 (analyse factorielle avec rotation varimax à 3 facteurs)

Variable	Factor1	Factor2	Factor3
1_humeur_depressive	0.006	0.028	0.263
2_sentim_culpabilite	-0.057	-0.076	0.240
3_suicide	-0.021	-0.120	0.207
4_insomnie_debut	0.110	0.359	-0.050
5_insomnie_milieu	0.004	0.118	0.240

Variable	Factor1	Factor2	Factor3
6_insomnie_matin	0.132	-0.155	0.427
7_travail_activite	0.124	0.121	0.320
8_ralentissement	0.183	-0.105	0.101
9_agitation	-0.107	-0.228	0.501
10_anxiete_psychique	0.066	-0.152	0.255
11_anxiete_somatique	-0.029	0.070	0.316
12_symptomes_gastro	0.992	0.044	-0.091
13_symptomes_generaux	0.046	0.977	0.194
14_symptomes_genitaux	0.238	0.120	0.147
15_hypochondrie	0.076	0.089	0.226
16_perte_poids	0.424	0.155	-0.100
17_prise_conscience	0.022	-0.004	0.176

- À J0, l'analyse factorielle exploratoire avec rotation varimax met en évidence 3 facteurs latents expliquant cumulativement 21,9 % de la variance des réponses aux items du score de Hamilton.
- Concernant l'interprétation du tableau des “loadings” (charges factorielles) :
 - Les “loadings” (charges factorielles) indiquent la contribution de chaque item à chaque facteur. Ils mesurent la force de la régression linéaire entre chaque item et chaque facteur latent. Plus l'item est proche de 1, plus il est expliqué par le facteur latent commun.
 - Il n'existe pas de “seuil” universel pour déterminer si un loading est “élevé” ou “faible”, mais généralement, un loading supérieur à 0,4 ou 0,5 est considéré comme significatif.
- On peut interpréter les 3 facteurs identifiés comme suit :
 - Facteur 1 : 8, 12, 14, 16 (principalement des symptômes somatiques).
 - Facteur 2 : 4, 13 (relatifs à l'asthénie).
 - Facteur 3 : les items restants (symptômes dépressifs psychiatriques proprement dits).

On peut rajouter 3 “sous-scores” au score total de Hamilton à J0, correspondant aux scores moyens des items chargés sur chaque facteur.

- A titre exploratoire, on peut refaire une ACP sur ces 3 sous-scores pour visualiser leur relation.

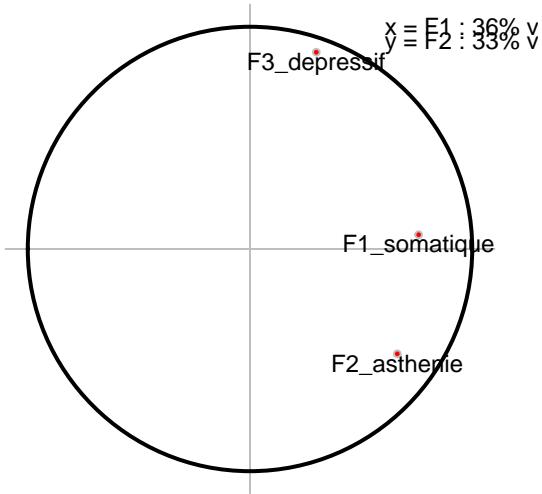


Figure 6: Analyse en composantes principales des sous-scores de l'échelle de Hamilton à J0

- Les 3 sous-scores sont bien représentés (proches du cercle). Le facteur “symptômes dépressifs” semble orthogonal aux deux autres facteurs.
- La variance totale expliquée par ces 3 sous-score est de 69%.

Une ACP focalisée sur ces 3 sous-scores et le score total de Hamilton permet de visualiser la relation entre le score total et les sous-scores.

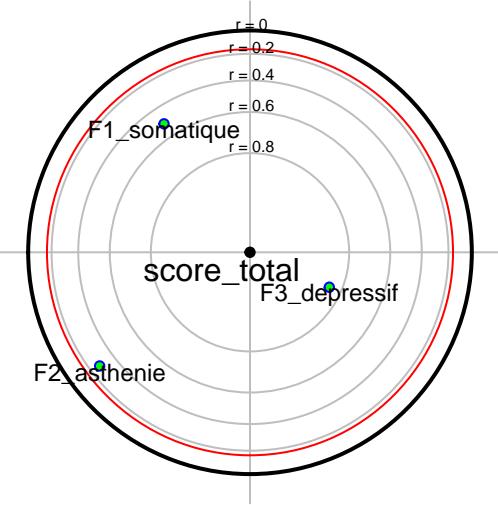


Figure 7: Analyse en composantes principales focalisée des sous-scores de l'échelle de Hamilton à J0

Le score total de Hamilton semble plus corrélé aux symptômes dépressifs (F3) qu'aux deux autres sous-scores.

3.A.3 Fiabilité interne = que vaut la mesure ?

3.A.3.1 Consistance interne : alpha de Cronbach La consistance interne des items de l'échelle de Hamilton à J0 est évaluée par le calcul de l'alpha de Cronbach, qui correspond globalement au pourcentage de « variance partagée » entre le score vrai (hypothétique) et la mesure obtenue.

Il permet ainsi de mesurer la cohérence entre les items d'une échelle de mesure, et est élevé lorsque les items sont fortement corrélés entre eux.

On peut donc calculer dans un premier temps l'alpha de Cronbach sur l'ensemble des items de l'échelle de Hamilton à J0, puis sur chacun des 3 facteurs identifiés précédemment.

Les intervalles de confiance (IC) à 95% des alpha de Cronbach sont estimés par la méthode du bootstrap avec 1000 rééchantillonnages. Le bootstrap est possible ici car il y a > 100 observations.

Table 3: Alpha de Cronbach et intervalles de confiance à 95% pour l'échelle de Hamilton à J0 et ses sous-échelles

Scale	Alpha	CI_lower	CI_upper
Global	0.456	0.262	0.624
F1_somatique	0.381	0.188	0.530
F2_asthenie	0.490	0.266	0.653
F3_depressif	0.486	0.349	0.622

Au total, quelque soit le niveau d'analyse (global ou par facteur), les alpha de Cronbach sont < 0.5 , indiquant une faible consistance interne des items de l'échelle de Hamilton à J0.

3.A.4 Validité externe = l'instrument mesure-t-il ce qu'il est censé mesurer ?

- Validité externe d'un instrument cherche à démontrer que l'instrument se comporte logiquement par rapport au réseau théorique qui lui est associé.
- Selon la théorie nomologique (c'est à dire selon les relations postulées entre les différents concepts d'une même discipline), la dépression mesurée par l'échelle de Hamilton doit être fortement liée à d'autres manifestations de la détresse psychologique générale (mesurée par le SCL-90), mais distincte de certains autres concepts.
 - Ici, la validité du construit peut être évaluée en évaluant la validité convergente (corrélation forte entre des concepts proches).
 - Il est plus difficile d'évaluer la validité divergente (corrélation faible entre des concepts différents) car le SCL-90 mesure principalement des dimensions de la détresse psychologique ; de même pour la validité concurrente (corrélation forte avec un *gold-standard*, car nous ne disposons pas d'un instrument de mesure de la dépression reconnu comme un *gold-standard* ici).

3.A.4.1 Corrélation entre le score de Hamilton et les dimensions du SCL-90 à J0

- On peut représenter une matrice de corrélation entre le score total de Hamilton et les 10 dimensions du SCL-90 à J0.

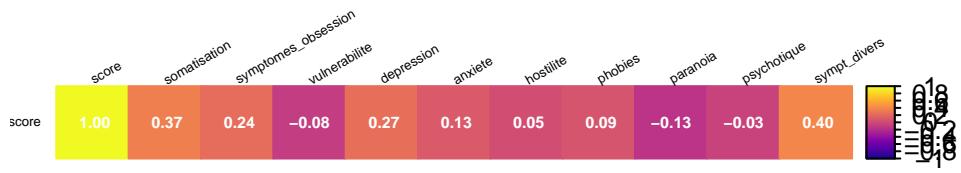


Figure 8: Corrélation entre le score total de l'échelle de Hamilton et les dimensions du SCL-90 à J0

La corrélation est au maximum à 0.37 avec la composante “somatisation” du SCL-90 et de 0.40 avec la composante “symptômes divers” du SCL-90, indiquant une validité convergente modérée entre le score total de Hamilton et cette dimension du SCL-90 à J0.

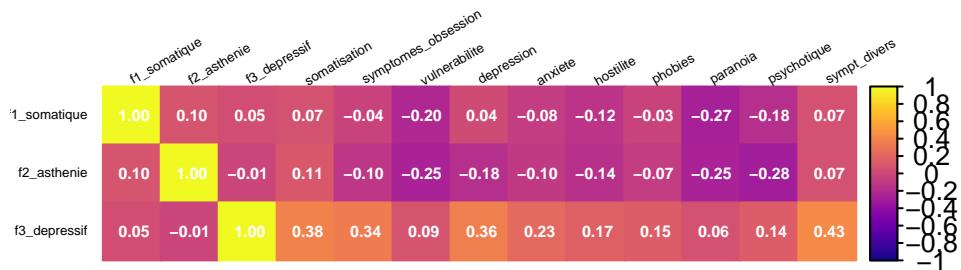


Figure 9: Corrélation entre les sous-scores de l'échelle de Hamilton et les dimensions SCL-90 à J0

- Il n'y a pas non plus de corrélation forte entre les sous-scores de Hamilton et les dimensions du SCL-90 à J0.
- Par exemple, le sous-score “symptômes somatiques” de Hamilton est faiblement corrélé avec la dimension “somatisation” du SCL-90 ($r = 0.07 !!$).

3.A.4.2 Conclusion validité externe à J0 La validité convergente entre le score total de Hamilton et les dimensions du SCL-90 à J0 est faible à modérée, suggérant que l'échelle de Hamilton mesure partiellement des aspects de la détresse psychologique générale, mais pas de manière très forte.

3.A.5 Conclusion globale à J0

À J0, l'échelle de dépression de Hamilton présente une structure dimensionnelle peu claire, avec une faible consistance interne des items (alpha de Cronbach < 0.5) et une validité convergente modérée

avec les dimensions du SCL-90.

Ces résultats suggèrent que l'échelle de Hamilton pourrait ne pas être un instrument optimal pour mesurer la dépression dans cette population à ce moment précis.

3.B Validation à J56

3.B.1 Description

[1] 120 [1] 146 [1] 82.19178

À J56, 120 des 146 patients présents à J0 ont répondu à tous les items : 17.8% des patients ont été perdus de vue.

Comme à J0, les réponses sont représentées :

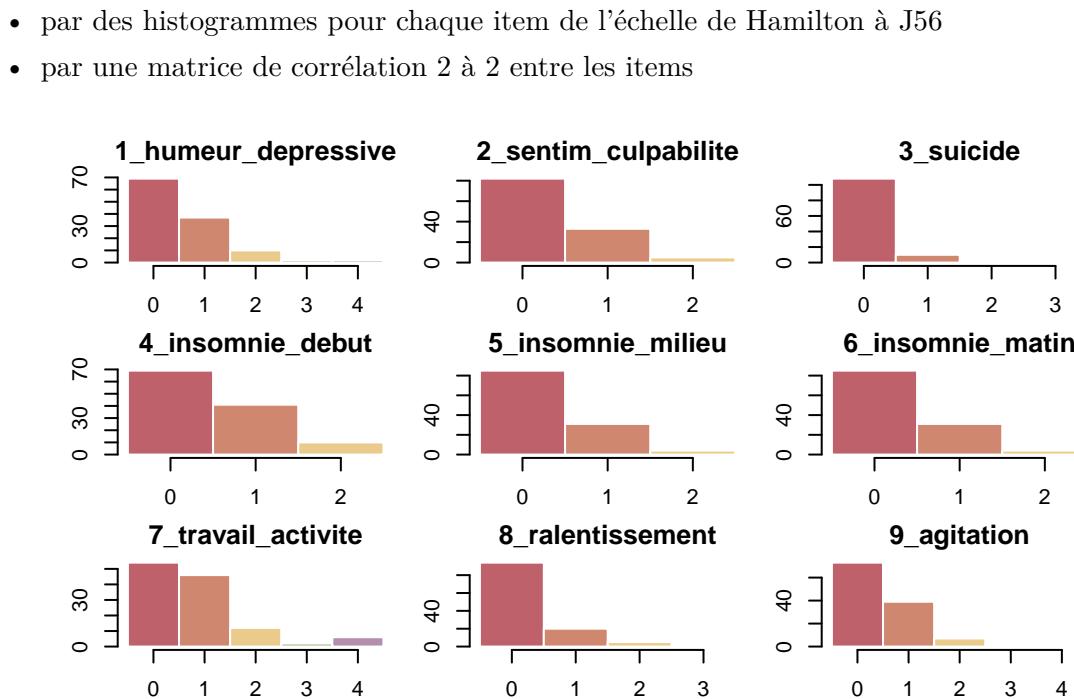


Figure 10: Histogrammes des scores des items de l'échelle de Hamilton à J56

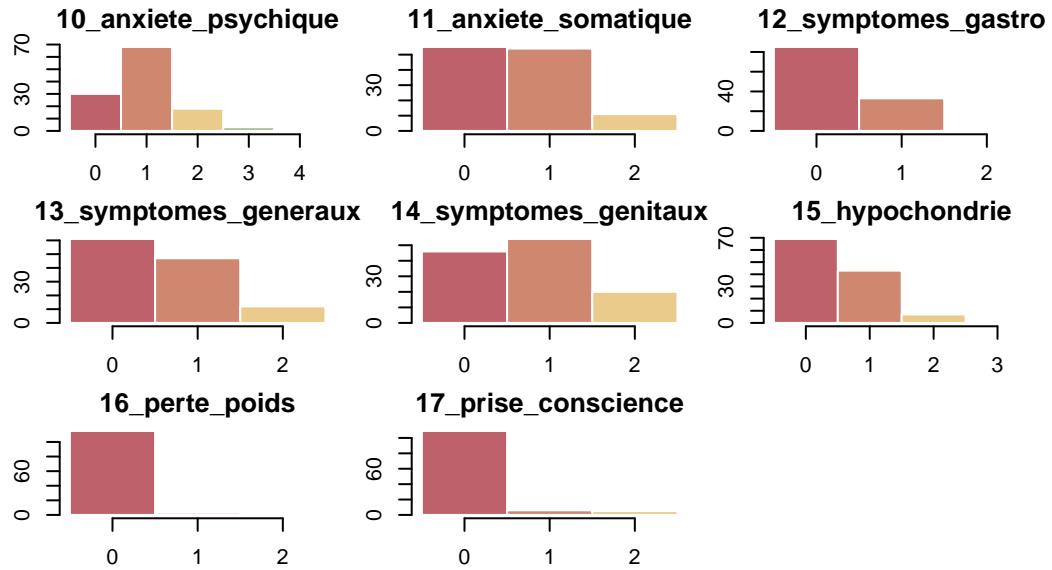


Figure 11: Histogrammes des scores des items de l'échelle de Hamilton à J56

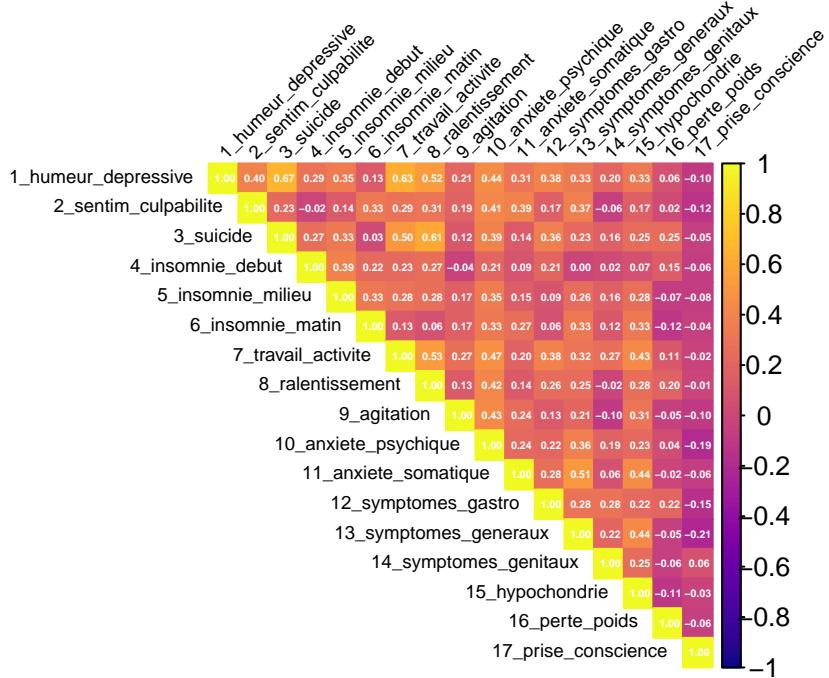


Figure 12: Matrice de corrélation entre les items de l'échelle de Hamilton à J56

- Sur les histogrammes, on observe un effet plancher, c'est à dire que de nombreux patients ont des scores faibles sur plusieurs items. Il est probable que ce soit lié à l'amélioration de la symptomatologie dépressive entre J0 et J56.
- La matrice de corrélation des items met en évidence des coefficients de corrélation plus élevés

que ceux observés à J0, avec plusieurs coefficients supérieurs à 0.5 en valeur absolue, indiquant une certaine redondance entre les items de l'échelle Hamilton à J56.

- Il est possible que cette redondance soit liée à l'effet plancher observé sur les histogrammes, où de nombreux patients ont des scores faibles sur plusieurs items, ce qui peut augmenter artificiellement les corrélations entre ces items.

3.B.2 Validité interne : structure dimensionnelle, analyse factorielle

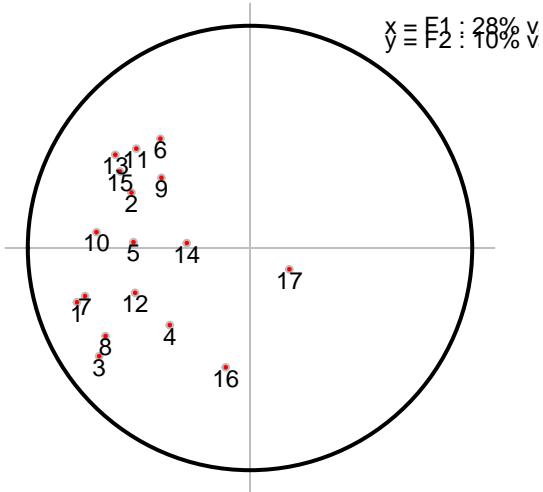


Figure 13: Analyse en composantes principales des items de l'échelle de Hamilton à J56

3.B.2.1 Exploration de la structure dimensionnelle : analyse en composantes principales

- La première composante principale explique 28% de la variance totale, la deuxième composante principale explique 10% de la variance totale.
 - Les deux premières composantes principales expliquent 38% de la variance totale, ce qui est une amélioration par rapport à J0.
 - Certaines variables sont un peu plus éloignées du centre qu'à J0, il est possible qu'elles contribuent davantage aux premières composantes principales.

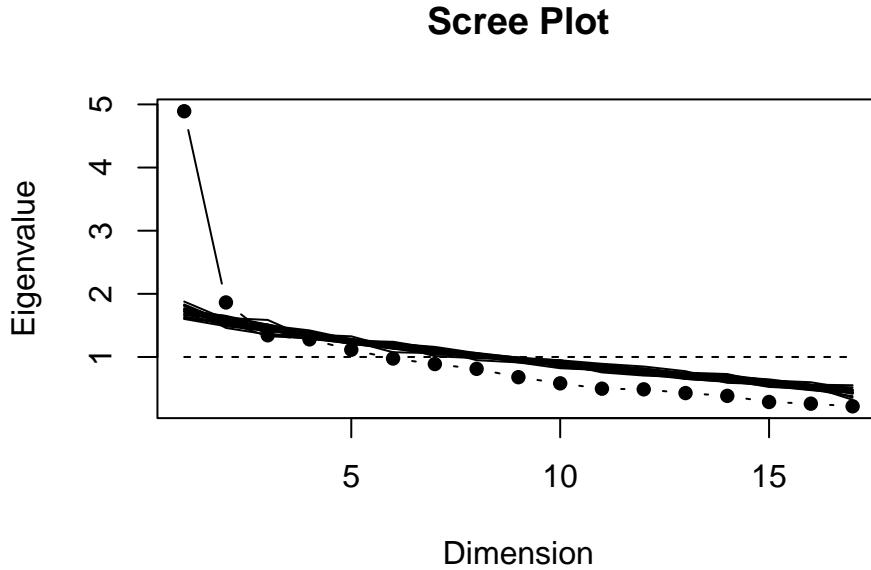


Figure 14: Diagramme des valeurs propres (scree plot) des items de l'échelle de Hamilton à J56 avec représentation de données simulées (analyse parallèle)

3.B.2.2 Exploration de la structure dimensionnelle : analyse factorielle

- À J56, le *scree plot* n'est pas du tout le même qu'à J0 : on observe un "coude" net après la première valeur propre.
- **L'échelle Hamilton à J56 semble donc essentiellement unidimensionnelle.**

Table 4: Contribution du facteur unique à la variance de la réponse à chaque item du score de Hamilton évalué à J56 (analyse factorielle avec rotation varimax à 1 facteur)

Variable	Factor1
1_humeur_depressive	0.802
2_sentim_culpabilite	0.468
3_suicide	0.706
4_insomnie_debut	0.333
5_insomnie_milieu	0.456
6_insomnie_matin	0.282
7_travail_activite	0.745
8_ralentissement	0.657
9_agitation	0.328
10_anxiete_psychique	0.624
11_anxiete_somatique	0.401
12_symptomes_gastro	0.471
13_symptomes_generaux	0.495
14_symptomes_genitaux	0.252
15_hypochondrie	0.494

Variable	Factor1
16_perte_poids	0.125
17_prise_conscience	-0.132

- Les valeurs numériques mesurent la force de la relation linéaire entre chaque item observé et le facteur latent unique.
- Les symptômes centraux de la dépression (humeur dépressive, ralentissement, baisse d'activité) présentent les charges factorielles les plus élevées, indiquant qu'ils sont de bons indicateurs du facteur latent unique (la dépression).
- Certains items périphériques (perte de poids, symptômes génitaux) ont une très faible charge, ils peuvent être peu informatifs et faiblement contribuer à la mesure du concept unique de dépression à J56. Leur pertinence dans le score total de l'instrument peut être remise en question.

3.B.3 Fiabilité interne = que vaut la mesure ?

3.B.3.1 Consistance interne : alpha de Cronbach La consistance interne des items de l'échelle de Hamilton à J56 est évaluée par le calcul de l'alpha de Cronbach, et son IC à 95% par la méthode du bootstrap avec 1000 rééchantillonnages.

Table 5: Alpha de Cronbach et intervalle de confiance à 95% pour l'échelle de Hamilton à J56

Scale	Alpha	CI_lower	CI_upper
Global	0.82	0.76	0.86

- A J56, l'alpha de Cronbach est de 0.82, avec un intervalle de confiance à 95% de [0.75 ; 0.86], indiquant une bonne consistance interne des items de l'échelle de Hamilton à ce moment. Elle n'est pas exagérément lourde ni redondante.

3.B.4 Validité externe = l'instrument mesure-t-il ce qu'il est censé mesurer ?

De la même manière qu'à J0, la validité externe de l'échelle de Hamilton à J56 peut être évaluée en examinant la corrélation entre le score total de Hamilton et les dimensions du SCL-90.



Figure 15: Corrélation entre score total de Hamilton et les dimensions du SCL-90 à J56

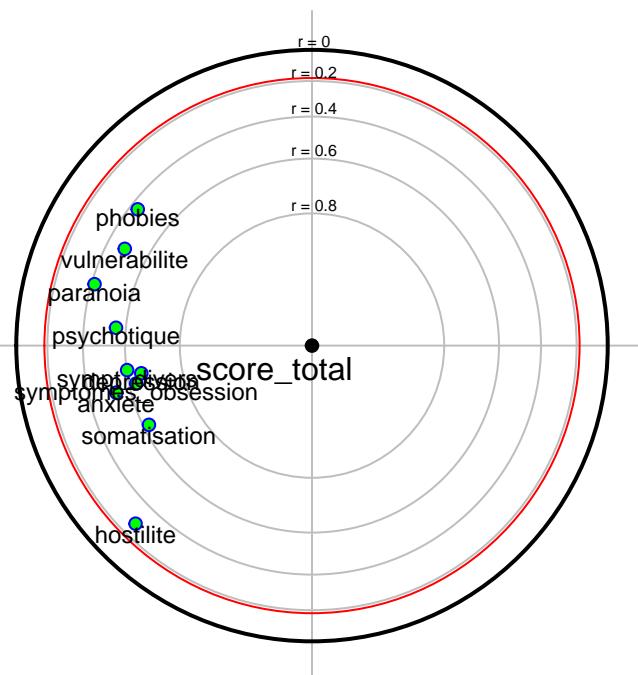


Figure 16: ACP focalisée du score total de l'échelle de Hamilton et des dimensions du SCL-90 à J56

3.B.4.1 Corrélation entre le score de Hamilton et les dimensions du SCL-90 à J56

- Les corrélations sont plus importantes entre les sous-scores de SCL90 et le score d'Hamilton à J56, : ces deux échelles mesurent probablement la même chose.
- La sous échelle dépression du SCL90 est particulièrement corrélée avec le score d'Hamilton.

3.B.4.2 Conclusion validité externe à J56 La validité convergente entre le score total de Hamilton et les dimensions du SCL-90 à J56 est modérée à forte, suggérant que l'échelle de Hamilton mesure des aspects pertinents de la détresse psychologique générale à ce moment.

3.B.5 Conclusion globale à J56

À J56, l'échelle de dépression de Hamilton présente une structure dimensionnelle essentiellement unidimensionnelle, avec une bonne consistance interne des items (alpha de Cronbach = 0.82) et une validité convergente modérée à forte avec les dimensions du SCL-90.

Au total, ces résultats suggèrent que l'échelle de Hamilton est un instrument valide et fiable pour mesurer la dépression dans cette population à J56.

3.B.6 Synthèse des résultats de validation de l'échelle de Hamilton à J0 et J56

Tableau récapitulatif des principaux résultats de validation de l'échelle de Hamilton à J0 et J56 :

Aspect	J0	J56
Structure dimensionnelle	Multidimensionnelle (3 facteurs)	Essentiellement unidimensionnelle
Consistance interne (<i>Alpha</i> de Cronbach)	Faible ($\alpha = 0.45$ [0.39 ; 0.51])	Bonne ($\alpha = 0.82$ [0.75 ; 0.86])
Validité convergente (corrélation avec SCL-90)	Modérée (max = 0.40)	Plus élevée (max = 0.85)
Conclusion globale	Structure dimensionnelle peu claire, faible consistance interne, validité convergente modérée	Essentiellement unidimensionnelle, bonne consistance interne, validité convergente modérée à forte

i Note

Validité divergente

La validité divergente (ou discriminante) fait référence à la capacité d'un instrument de mesure à ne pas être fortement corrélé avec des concepts ou des mesures qui sont théoriquement distincts. En d'autres termes, un instrument doit montrer qu'il mesure un concept spécifique sans être influencé de manière significative par d'autres concepts non liés.

En l'occurrence, on ne dispose pas d'instruments mesurant des concepts clairement distincts de la dépression dans ce jeu de données, ce qui limite notre capacité à évaluer la validité divergente de l'échelle de Hamilton dans ce contexte.

4 Question 2 : Comparaison de la réponse au traitement entre deux groupes de patients

Note

Consigne

“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.”

4.A Gestion des données

Il faut utiliser

4.B Scores bruts à J0 dans chaque groupe

- Présentation en tableau des scores bruts de l'échelle de Hamilton à J0 dans chaque groupe de traitement avec effectifs, moyennes, écarts-types, médianes et intervalles interquartiles.

Table 7: Scores bruts de l'échelle de Hamilton à J0 par groupe de traitement

GROUPE	N	Moyenne	Écart type	Médiane	Q1	Q3
0	75	28.1	4.43	27	26	30
1	71	27.8	3.97	27	25	30

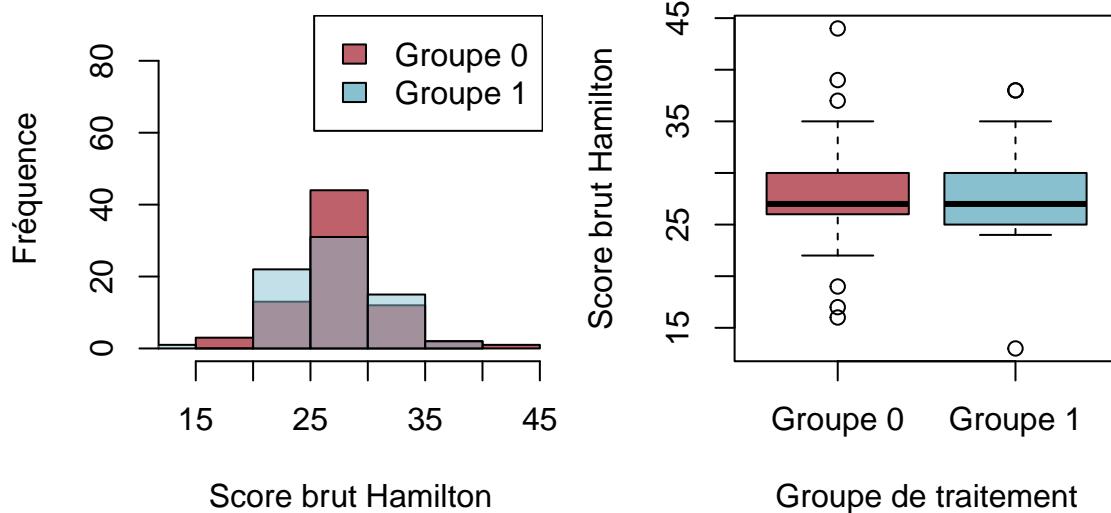


Figure 17: Histogramme et boxplot présentant les scores bruts de l'échelle de Hamilton à J0

4.B.1 Comparaison des scores bruts à J0 entre les deux groupes

- Un histogramme des scores bruts à J0 à déjà été présenté ci-dessus.
- La distribution semble approximativement normale dans chaque groupe.
- On peut donc comparer les scores bruts moyens entre les deux groupes par un test t de Student pour échantillons indépendants.

Characteristic	0 N = 75 ¹	1 N = 71 ¹	p-value ²
score	28.05 (4.43)	27.79 (3.97)	0.7

¹ Mean (SD)

² Two Sample t-test

NB : la fonction utilisée ici est `tbl_summary` du package `gtsummary` permet de réaliser des tableaux de synthèse avec des tests statistiques intégrés.

4.C Évolution des scores bruts entre J0 et J56 dans chaque groupe

4.C.1 Évolution du nombre de patients dans chaque groupe à chaque visite

- Nombre de données entre chaque visite entre J0 et J56 dans chaque groupe :

A la visite J56, les visites manquent pour 11 patients sur 75 dans le groupe 0 (14,7%), contre 15 patients sur 56 dans le groupe 1 (26,8%).

Table 8: Nombre de patients par groupe de traitement et visite

	J0	J4	J7	J14	J21	J28	J42	J56
0	75	72	71	70	69	68	65	64
1	71	67	65	63	60	59	57	56

A la visite J56, les visites manquent pour 11 patients sur 75 dans le groupe 0 (14,7%), contre 15 patients sur 56 dans le groupe 1 (26,8%).

4.C.2 Évolution des scores bruts moyens entre J0 et J56 dans chaque groupe

On peut représenter :

- L'évolution des scores bruts moyens dans chacun des groupes avec des boxplot à chaque visite

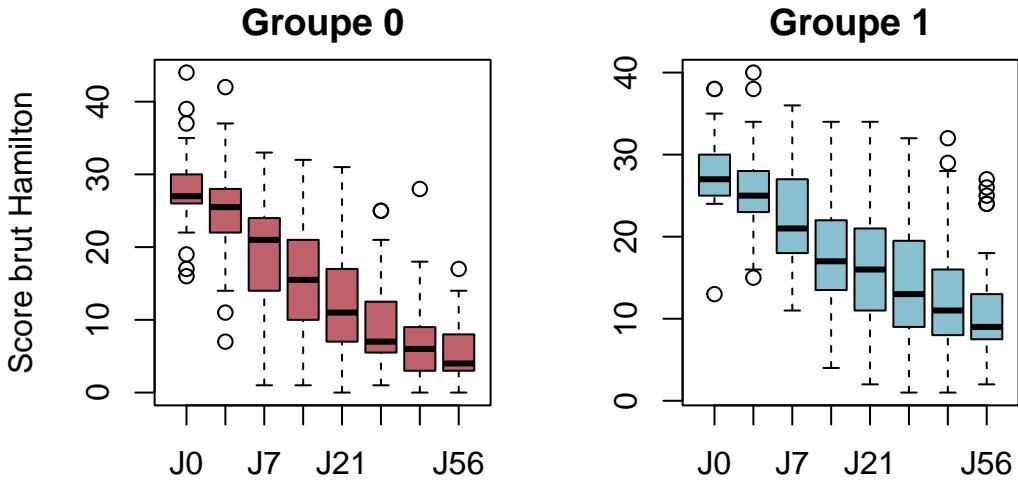


Figure 18: Évolution des scores moyens entre J0 et J56 dans chaque groupe

- L'évolution par patients (dans chacun des groupes), avec un diagramme en spaghetti, d'interprétation plus difficile mais qui montre les trajectoires individuelles.



Figure 19: Évolution individuelle des scores entre J0 et J56 dans chaque groupe

- De nouveaux boxplots qui montre le score à J56 côté à côté pour chaque groupe.

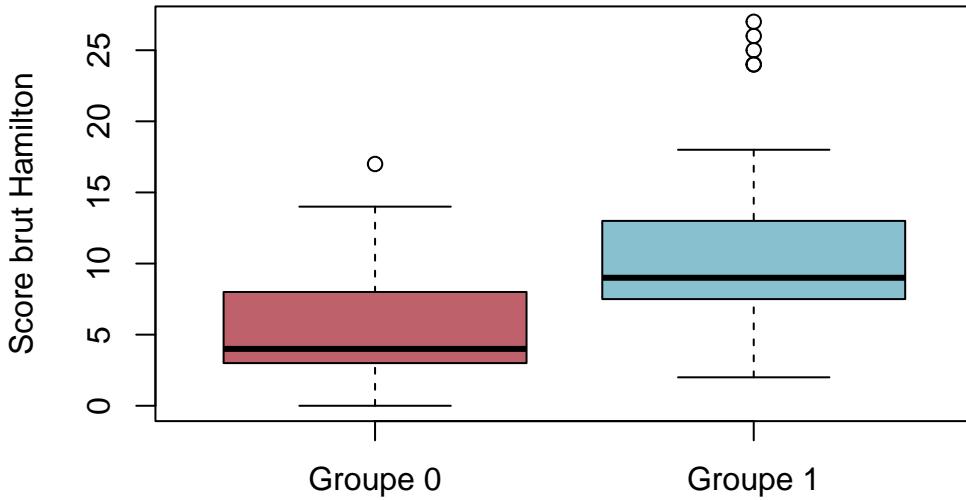


Figure 20: Scores bruts de l'échelle de Hamilton à J56 dans chaque groupe de traitement

Quelque soit la méthode de visualiation, on observe une tendance diminution des scores bruts moyens de l'échelle de Hamilton entre J0 et J56 dans les deux groupes de traitement.

La diminution semble plus marquée dans le groupe 0 que dans le groupe 1.

4.D Comparaison de la réponse au traitement entre deux groupes de patients

L'objectif est donc de déterminer si un des groupes répond mieux au traitement que l'autre, en comparant l'évolution des scores bruts entre J0 et J56 dans chaque group, en utilisant d'abord une approche LOCF (last observation carried forward), puis un modèle mixte.

4.D.1 Comparaison de la réponse en utilisant la méthode LOCF

4.D.1.1 Imputation des données manquantes par la méthode LOCF (last observation carried forward) Méthode LOCF = on remplace les valeurs manquantes par la dernière valeur observée pour chaque patient.

On applique la fonction `na.locf` du package `zoo` pour réaliser cette imputation.

Table 9: Nombre de patients par groupe de traitement et visite après imputation LOCF

	J0	J14	J21	J28	J4	J42	J56	J7
0	75	75	75	75	75	75	75	75
1	71	71	71	71	71	71	71	71

4.D.1.2 Comparaison de la réponse au traitement entre les deux groupes après imputation LOCF Il faut comparer l'évolution des scores bruts moyens entre J0 et J56, donc la différence de score entre J56 et J0, entre les deux groupes.

Pour ça, on crée une nouvelle variable “difference” qui correspond à la différence entre le score à J56 et le score à J0 (après imputation LOCF).

On fait maintenant un test t de Student pour comparer la différence moyenne de score entre J0 et J56 entre les deux groupes.

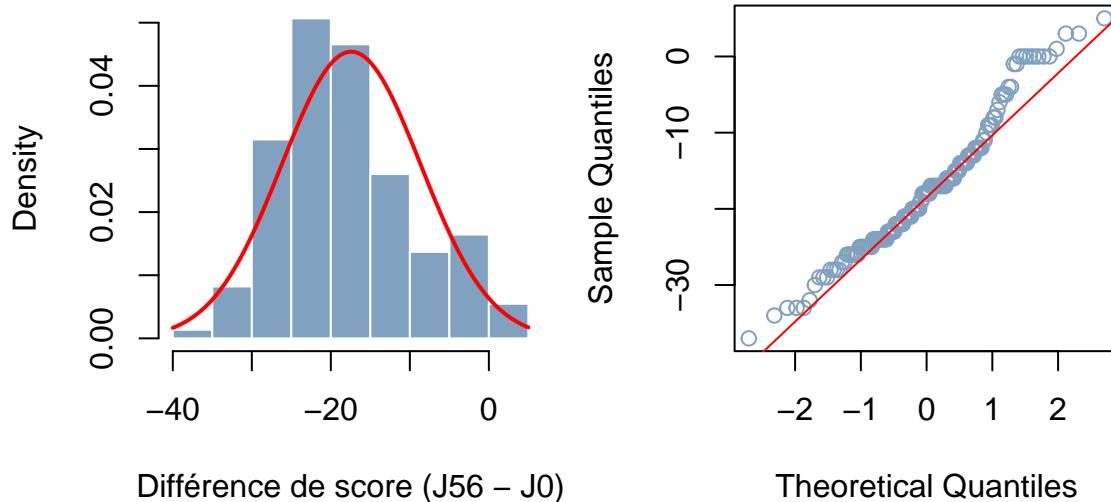


Figure 21: Histogramme et QQ-plot de la différence de score entre J0 et J56 (après imputation LOCF) sur l'échantillon complet

Les conditions de validité semblent respectées pour le test t de Student sur la différence de score entre J0 et J56 (après imputation LOCF) sur l'échantillon complet.

Characteristic	0 N = 75 ¹	1 N = 71 ¹	p-value ²
difference	-19.92 (8.68)	-14.79 (8.15)	<0.001

¹ Mean (SD)

² Two Sample t-test

Le test t montre une différence moyenne de score entre J0 et J56 significativement plus élevée dans le groupe 0 (-19,92) que dans le groupe 1 (-14,79), avec une p-value < 0.001.

Le problème majeur de cette méthode est qu'elle ne tient pas en compte que les mesures sont répétées chez les mêmes patients !

En utilisant un modèle mixte, on va pouvoir prendre en compte cette dépendance entre les mesures répétées en modélisant un effet aléatoire pour chaque patient.

4.D.2 Comparaison de la réponse en utilisant un modèle mixte

Ici on veut comparer l'évolution du score HDRS entre les groupes **en tenant compte des mesures répétées** (plusieurs visites par patient).

Un modèle linéaire mixte (LMM) est adapté car il modélise la corrélation intra-patient.

- Variable à expliquer : **score** (après imputation LOCF), quantitative, discrète.
- Effets fixes : **VISIT** et **GROUPE**.
- Effet aléatoire : un intercept par patient ($1|NUMERO$) pour tenir compte du niveau moyen différent selon les patients, donc une ordonnée à l'origine différente pour chaque patient.

On peut envisager deux modèles :

- **Sans interaction** : suppose des trajectoires « parallèles » à chaque visite (différence entre groupes constante au cours du temps).
- **Avec interaction VISIT*GROUPE** : teste si l'évolution au cours du temps diffère entre groupes (c'est le test principal pour « répond mieux »).

La question résiduelle est le codage de la variable **VISIT** : on peut la traiter comme une variable catégorielle (facteur) ou continue (temps).

Dans le code, on fera les deux pour voir si ça change les résultats.

4.D.2.1 Modèle mixte sans interaction

4.D.2.1.1 Modèle mixte sans interaction avec VISIT comme facteur et J0 comme référence

Table 10: Estimation des coefficients du modèle mixte sans interaction entre VISIT (facteur) et GROUPE

Coefficient	Estimate	Std_Error	t_value	CI_lower	CI_upper	p_value
(Intercept)	26.523	0.638	41.555	25.275	27.771	0e+00
VISITJ4	-2.889	0.498	-5.796	-3.863	-1.914	0e+00
VISITJ7	-7.142	0.502	-14.223	-8.123	-6.160	0e+00
VISITJ14	-11.250	0.506	-22.237	-12.238	-10.261	0e+00
VISITJ21	-13.607	0.511	-26.643	-14.605	-12.609	0e+00
VISITJ28	-16.298	0.513	-31.759	-17.301	-15.295	0e+00
VISITJ42	-18.546	0.519	-35.704	-19.561	-17.530	0e+00
VISITJ56	-20.010	0.522	-38.333	-21.030	-18.990	0e+00
GROUPE	2.883	0.795	3.627	1.324	4.440	4e-04

- La t-value et la p-value de l'intercept n'ont pas d'intérêt ici, car le test de Wald pour l'intercept n'a pas de signification pratique, il teste l'hypothèse $H_0 : \text{intercept} = 0$, ce qui n'a pas de sens dans ce contexte.
- Intercept (26.523) : score moyen estimé à J0 dans le groupe de référence (groupe 0).
- Effet du temps (VISIT) : tous les coefficients VISITJ_x sont négatifs et significatifs → le score diminue significativement par rapport à J0. À J56, la baisse moyenne est de -20.010 points (IC95% [-21.030 ; -18.990]).

- NB : ici l'effet du temps est fait quelque soit le groupe !
- Effet du groupe ($GROUPE = 2.883$) : à visite donnée, le groupe 1 a en moyenne +2.883 points de score par rapport au groupe 0 (IC95% [1.324 ; 4.440], $p = 0.0004$), donc des scores plus élevés en moyenne sur l'ensemble du suivi.

Ce modèle décrit une amélioration globale au cours du temps et un décalage moyen entre groupes, mais il ne permet pas de conclure sur une différence de réponse “en pente” entre groupes (il faut le modèle avec interaction $VISIT \times GROUPE$ pour ça)

4.D.2.1.2 Modèle mixte sans interaction avec VISIT comme variable continue

Table 11: Estimation des coefficients du modèle mixte sans interaction entre VISIT (continue) et GROUPE

Coefficient	Estimate	Std_Error	t_value	CI_lower	CI_upper	p_value
(Intercept)	22.968	0.581	39.543	21.831	24.106	0e+00
VISIT_num	-0.352	0.009	-40.792	-0.369	-0.335	0e+00
GROUPE	2.997	0.799	3.752	1.431	4.562	3e-04

- Intercept = 22.968 : score HDRS prédit pour le groupe de référence (groupe 0) quand VISIT_num = 0 (jour 0), selon le modèle linéaire en jours. Ce n'est pas forcément égal à la moyenne empirique à J0.
 - NB : l'intercept correspond donc ici à la valeur prédite par la droite ajustée sur toutes les visites !
- VISIT_num = -0.352 : le score diminue en moyenne de 0.352 point par jour (IC95% [-0.369 ; -0.335], $p < 0.001$). Sur 56 jours, ça correspond à une baisse attendue d'environ $0.352 \times 56 = 19.7$ points.
- GROUPE = +2.997 : à temps donné (même jour), le groupe 1 a un score moyen plus élevé d'environ 3 points que le groupe 0 (IC95% [1.431 ; 4.562], $p = 0.0003$). Sans interaction, cette différence est supposée constante dans le temps.

Le score baisse avec le temps (modélisé linéairement en jours) et le groupe 1 reste en moyenne plus haut que le groupe 0 ; ce modèle ne teste pas une différence d'évolution entre groupes (il faudrait ajouter l'interaction $VISIT_num * GROUPE$).

4.D.2.2 Modèle mixte avec interaction

4.D.2.2.1 Modèle mixte avec interaction avec VISIT comme facteur et J0 comme référence

Table 12: Estimation des coefficients du modèle mixte avec interaction entre VISIT (facteur) et GROUPE

Coefficient	Estimate	Std_Error	t_value	p_value
(Intercept)	28.053	0.699	40.115	0.0000

Coefficient	Estimate	Std_Error	t_value	p_value
VISITJ4	-3.364	0.674	-4.990	0.0000
VISITJ7	-8.663	0.677	-12.790	0.0000
VISITJ14	-12.587	0.680	-18.497	0.0000
VISITJ21	-15.640	0.683	-22.882	0.0000
VISITJ28	-18.875	0.686	-27.495	0.0000
VISITJ42	-21.328	0.696	-30.654	0.0000
VISITJ56	-22.307	0.699	-31.914	0.0000
GROUPE	-0.265	1.003	-0.264	0.7920
VISITJ4:GROUPE	0.949	0.969	0.979	0.3278
VISITJ7:GROUPE	3.141	0.977	3.216	0.0013
VISITJ14:GROUPE	2.778	0.984	2.822	0.0049
VISITJ21:GROUPE	4.289	0.994	4.313	0.0000
VISITJ28:GROUPE	5.464	0.999	5.469	0.0000
VISITJ42:GROUPE	5.889	1.011	5.824	0.0000
VISITJ56:GROUPE	4.853	1.016	4.775	0.0000

4.D.2.2.2 Modèle mixte avec interaction avec VISIT comme variable continue

Table 13: Estimation des coefficients du modèle mixte avec interaction entre VISIT (continue) et GROUPE

Coefficient	Estimate	Std_Error	t_value	p_value
(Intercept)	23.807	0.600	39.669	0.0000
VISIT_num	-0.394	0.012	-33.792	0.0000
GROUPE	1.251	0.863	1.450	0.1486
VISIT_num:GROUPE	0.090	0.017	5.283	0.0000

4.D.2.2.3 Test de l'effet d'interaction par rapport au modèle sans interaction Pour tester l'effet d'interaction, on peut faire un test F pour comparer l'effet de l'ajout d'un terme d'interaction entre VISIT et GROUPE par rapport au modèle sans interaction

Utilisation de la fonction `drop1` pour ça :

Single term deletions using Satterthwaite's method:

```
Model: score ~ VISIT_num * GROUPE + (1 | NUMERO)
Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
VISIT_num:GROUPE 672.72 672.72 1 924.77 27.908 1.587e-07 *** — Signif. codes: 0 ‘‘ 0.001 ’’
0.01 ’ 0.05 ? 0.1 ’ ’ 1
```

La p-value associée au test de l'effet d'interaction est < 0.001, ce qui indique que l'ajout du terme d'interaction entre VISIT et GROUPE améliore significativement le modèle par rapport au modèle sans interaction.

En pratique, cela signifie que l'évolution des scores HDRS au cours du temps diffère significativement entre les deux groupes de traitement.

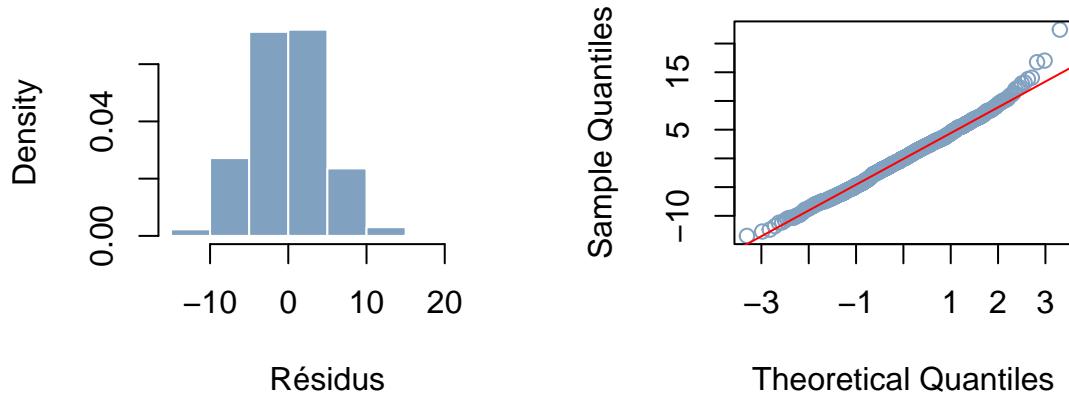


Figure 22: Normalité des résidus du modèle mixte avec interaction entre VISIT (continue) et GROUPE

Les résidus semblent suivre une distribution normale.

4.E Conclusion sur la comparaison de la réponse au traitement entre les deux groupes de patients

- La méthode LOCF suivie d'un test t de Student montre une différence moyenne de score entre J0 et J56 significativement plus élevée dans le groupe 0 que dans le groupe 1 ($p < 0.001$).
- Le modèle mixte avec interaction entre VISIT et GROUPE confirme que l'évolution des scores HDRS au cours du temps diffère significativement entre les deux groupes de traitement ($p < 0.001$ pour l'effet d'interaction).

5 Question 3 : Analyse de survie

i Note

Consigne

“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.”

5.A Critère binaire de réponse (50% à la dernière visite observée)

L'objectif est de définir un critère binaire de réponse au traitement basé sur une chute d'au moins 50% du score HDRS par rapport à J0, en utilisant la dernière visite observée pour chaque patient.

Les étapes sont :

- Obtention du score HDRS de la dernière visite observée (non manquante) pour chaque patient.
- Calcul de la chute par rapport à J0.
- Si la chute est d'au moins 50%, le patient est considéré comme ayant répondu au traitement (réponse = 1), sinon non (réponse = 0).
- Fusion de cette information avec les données au format long `hdrs_groupe`, dans une variable `reponse` (et un nouveau data.frame `hdrs_groupe_reponse`).

	NUMERO	GROUPE	VISIT	1_humeur_depressive	2_sentim_culpabilite	3_suicide
1	0	1	J0	3	2	2
2	10	1	J0	3	2	3
3	10	1	J4	3	2	2
4	10	1	J7	3	2	1
5	10	1	J14	2	2	1
6	10	1	J21	2	1	0
	4_insomnie_debut	5_insomnie_milieu	6_insomnie_matin	7_travail_activite		
1	1		1	0		3
2	1		2	2		2
3	0		2	2		2
4	1		1	2		2
5	1		1	1		1
6	1		1	1		1
	8_ralentissement	9_agitation	10_anxiete_psychique	11_anxiete_somatique		
1	1	2	3	4		
2	1	3	3	2		
3	0	1	3	1		
4	0	1	3	1		
5	0	1	2	1		
6	0	0	2	1		
	12_symptomes_gastro	13_symptomes_generaux	14_symptomes_genitaux			
1	0		2	2		
2	0		1	2		

```

3          0          1          2
4          0          0          2
5          0          0          2
6          0          0          2
  15_hypochondrie 17_perte_poids score VISIT_num score_J0
1          2          1          0    29      0    29
2          2          0          0    29      0    29
3          2          0          0    23      4    29
4          2          0          0    21      7    29
5          2          0          0    17     14    29
6          1          0          0    13     21    29
derniere_visite score_dernier reponse
1          J0         29        0
2          J56        9         1
3          J56        9         1
4          J56        9         1
5          J56        9         1
6          J56        9         1

```

[1] TRUE

FALSE	TRUE
0.8184411	0.1815589
0	1
0.1815589	0.8184411

5.B Intégration du critère binaire de réponse dans une analyse de survie

L'objectif est d'utiliser le critère binaire de réponse (chute d'au moins 50% du score HDRS par rapport à J0) dans une analyse de survie.

Il faut donc définir :

- Un temps de suivi pour chaque patient, qui correspond au délai jusqu'à la **première** réponse observée (chute 50% par rapport à J0). Si une visite est manquante, on considère une censure à la visite précédente (logique du PDF DEGACHI), sinon censure administrative à J56.
- Un indicateur d'événement, qui vaut 1 si la réponse est observée et 0 si le patient est censuré.

Il y a une censure administrative à 56 jours, et une censure au premier manquant quand une visite est absente.

	NUMERO	temps_suivi	evenement	reponse
1	0	0	0	0
2	3	7	0	0
3	4	28	1	1
4	7	14	1	1
5	8	21	1	1
6	9	42	1	1

0	4	7	14	21	28	42	56	
0	7	4	3	4	2	2	1	9
1	0	2	22	25	21	19	18	7

Table 14: Temps de suivi (en jours) par groupe de traitement

	0	4	7	14	21	28	42	56
0	3	3	16	17	13	12	9	2
1	4	3	9	12	10	9	10	14

5.C Analyse de survie avec le critère binaire de réponse

Censure administrative à 56 jours et censure au premier manquant.

Test de log-rank pour comparer les courbes de survie de la réponse entre les deux groupes :

Call:

```
survdiff(formula = surv_obj_reponse ~ GROUPE, data = hdrs_wide_reponse)
```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
GROUPE=0	75	65	48.6	5.52	13.9
GROUPE=1	71	49	65.4	4.10	13.9

```
Chisq= 13.9 on 1 degrees of freedom, p= 2e-04
```

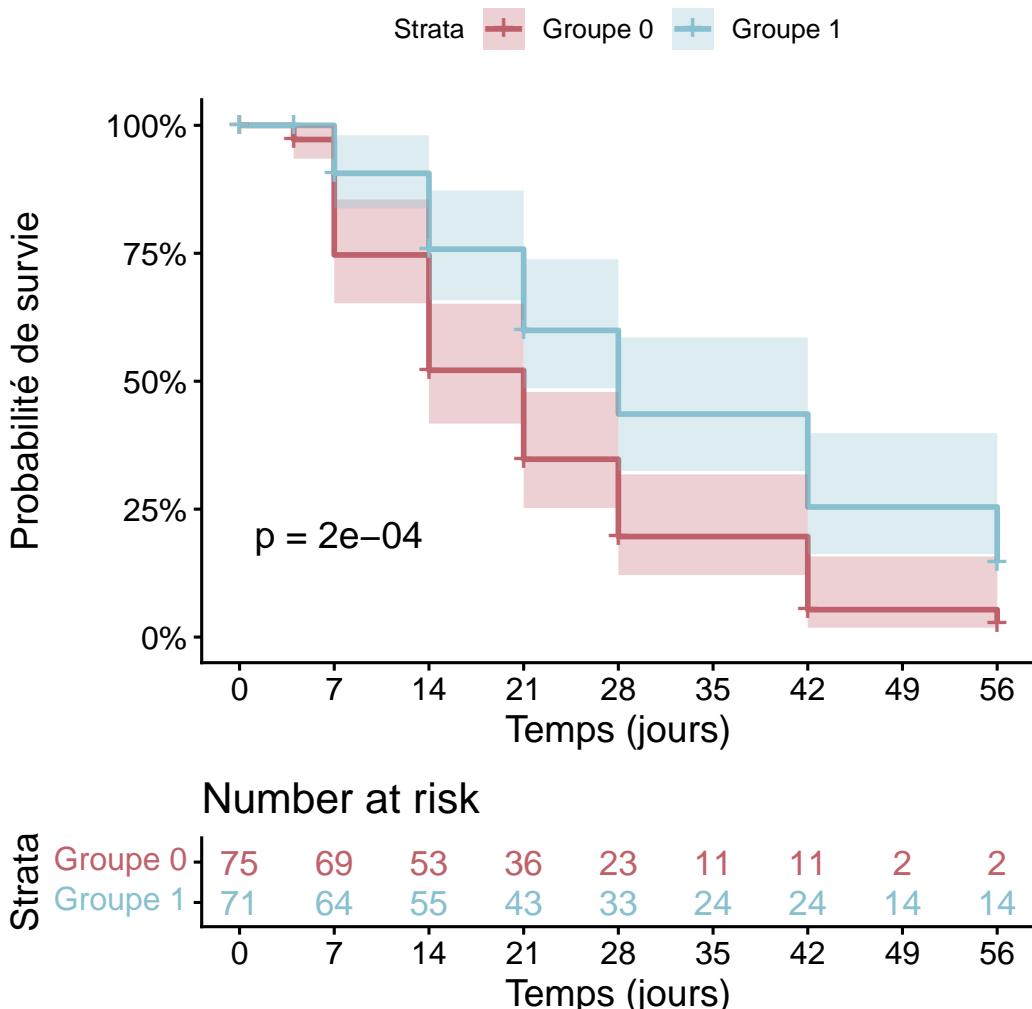


Figure 23: Courbes de survie de la réponse au traitement par groupe de traitement

Le test de log-rank montre une différence significative entre les deux groupes de traitement ($p < 0.01$), indiquant que la probabilité de réponse au traitement diffère entre les groupes.

Mais le test de log-rank permet seulement une comparaison globale des courbes de survie entre les groupes.

Pour évaluer l'effet du groupe de traitement sur le temps jusqu'à la réponse, on peut ajuster un modèle de Cox.

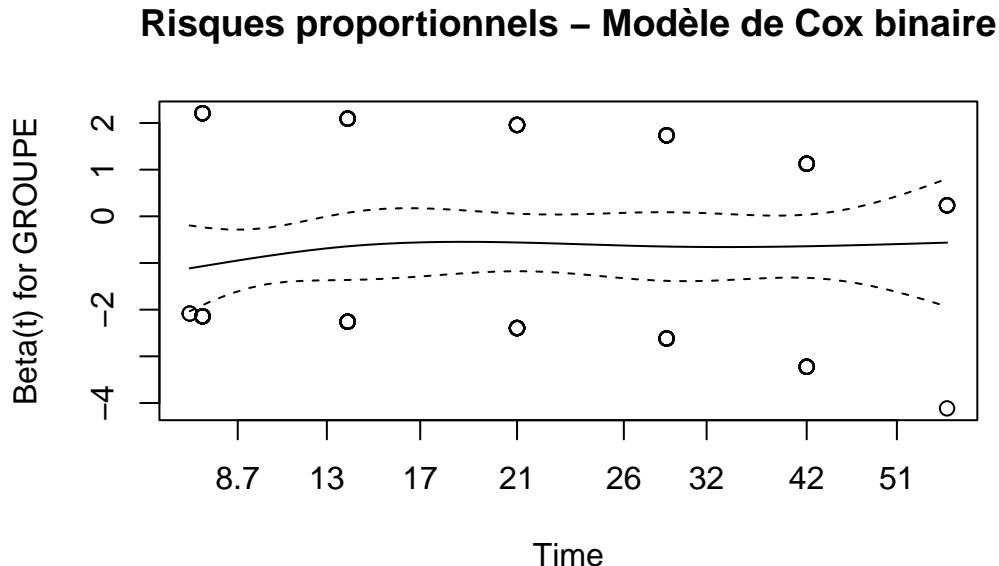
affichage dans un tableau des résultats du modèle de Cox

Table 15: Estimation des coefficients du modèle de Cox pour le critère binaire de réponse au traitement

Coefficient	Estimate	HR	CI_lower	CI_upper	p_value
GROUPE1	-0.714	0.49	0.334	0.718	3e-04

Pour vérifier l'hypothèse des risques proportionnels, on peut utiliser la fonction `cox.zph` :

```
plot(cox.zph(cox_model_reponse), main = "Risques proportionnels - Modèle de Cox  
↪ binaire")
```



6 Annexe – Code R