

# Devoir Statistiques Avancées

Thomas Husson, Groupe 52

## Table des matières

<b>1 Énoncé et présentation des échelles</b>	<b>2</b>
1.A Énoncé du devoir . . . . .	2
1.B Échelles . . . . .	2
<b>2 Gestion des données</b>	<b>3</b>
2.A Présentation des fichiers de données . . . . .	3
2.A.1 Fichier Hamilton . . . . .	3
2.A.2 Fichier scl90 . . . . .	3
2.A.3 Fichier groupe . . . . .	3
2.B Import des données et data management . . . . .	3
2.B.1 SCL90 . . . . .	3
2.B.2 HDRS . . . . .	4
2.B.3 Groupes . . . . .	4
2.B.4 Fusion des 3 fichiers . . . . .	4
<b>3 Question 1 : Validation de l'échelle Hamilton</b>	<b>5</b>
3.A Validation à J0 . . . . .	5
3.A.1 Description . . . . .	5
3.A.2 Validité interne : structure dimensionnelle, analyse factorielle . . . . .	8
3.A.3 Validité externe . . . . .	8
3.B Validation à J56 . . . . .	8
3.B.1 Description . . . . .	8
3.B.2 Validité interne : structure dimensionnelle, analyse factorielle . . . . .	8
3.B.3 Validité externe . . . . .	8
<b>4 Question 2 : Comparaison de la réponse au traitement entre deux groupes de patients</b>	<b>8</b>

# 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

	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 (évaluée si les items sont cohérents entre eux) sera évaluée par le calcul de l'alpha de Cronbach
  - Échelle de Hamilton = hétéro-évaluation donc on peut aussi évaluer la concordance inter-juges (ici entre les items) par le coefficient de corrélation intraclasse (CCI)
- 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`).

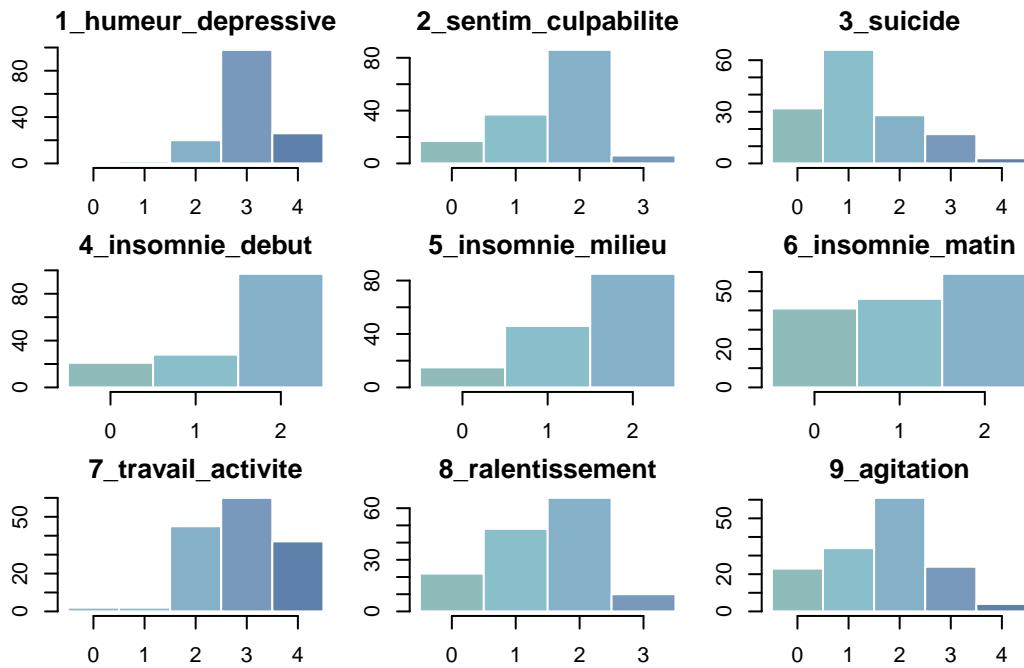


Figure 1: Histogrammes des scores des items de l'échelle de Hamilton à J0

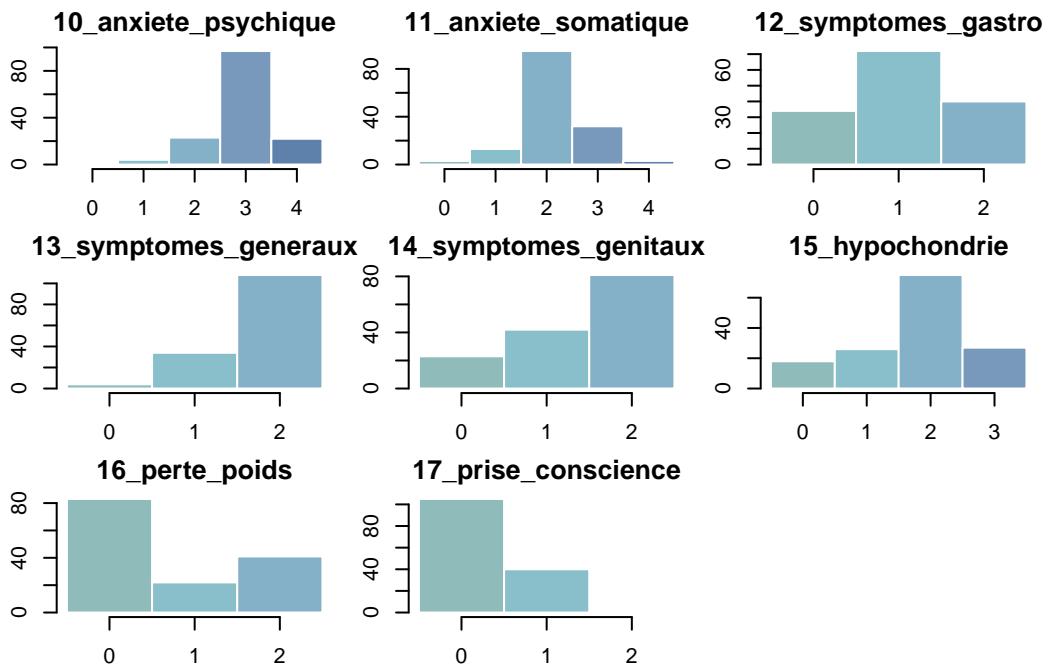


Figure 2: Histogrammes des scores des items de l'échelle de Hamilton à J0

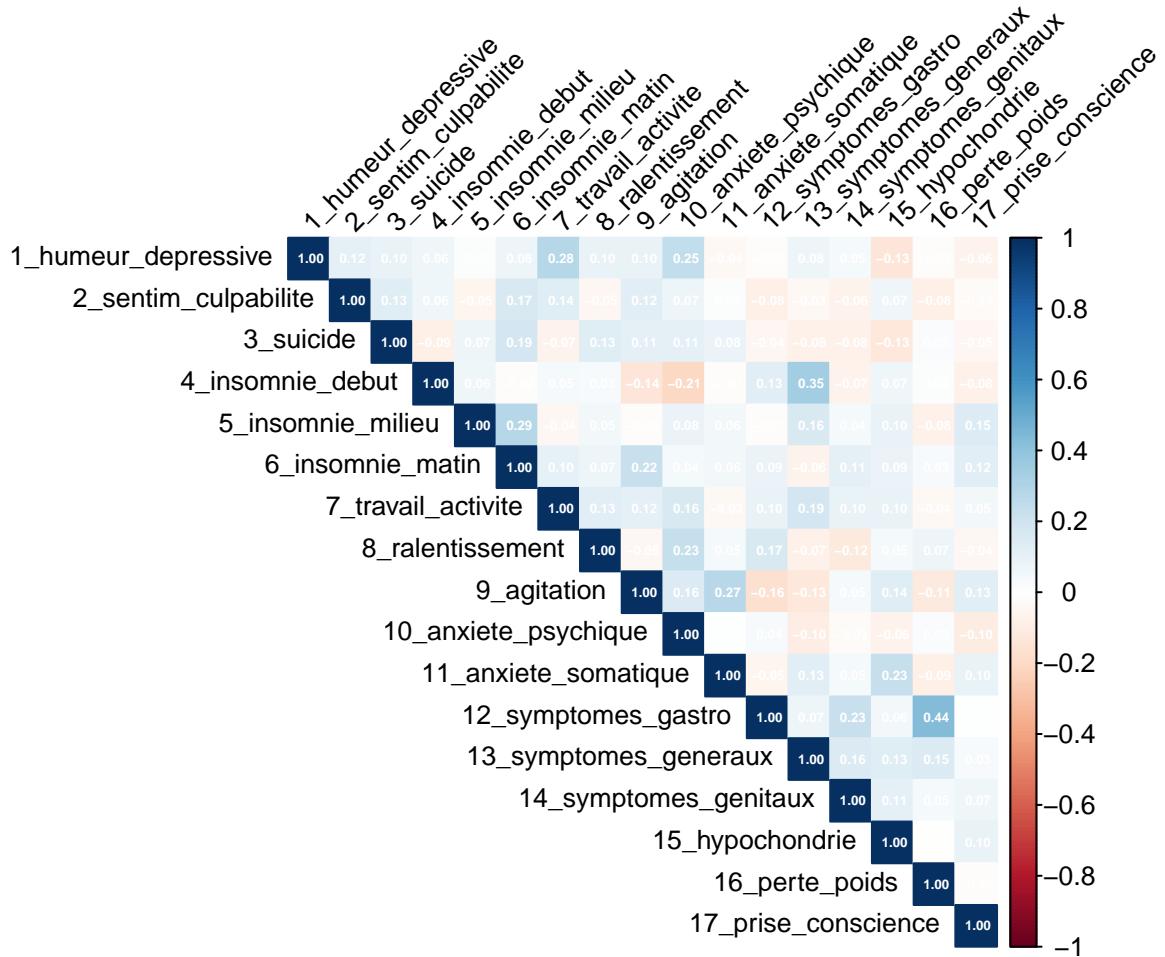


Figure 3: Matrice de corrélation entre les items de l'échelle de Hamilton à J0

- 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.
  - 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

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

**3.A.3 Validité externe**

**3.B Validation à J56**

**3.B.1 Description**

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

**3.B.3 Validité externe**

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