

IUT	Robert Schuman	
	Institut universitaire de technologie	
	Département informatique	
	Université de Strasbourg	

M31: Probabilités et Statistiques

Mini-projet SAS

Contexte :

L'entreprise pharmaceutique UPBRAIN veut vérifier l'efficacité de son nouveau médicament MENTALOR dans la maladie d'Alzheimer. Le médicament sera testé avec deux doses différentes et un placebo. Cette étude comporte 102 patients randomisés et se déroule sur une durée de 88 semaines couvrant plusieurs visites médicales pour chaque patient.

Le but du mini-projet est de délivrer un rapport statistique pour étudier l'efficacité du traitement ainsi que le profil de certains patients. Pour cela diverses opérations d'import, de manipulation, d'analyse statistique et de mise en page en SAS seront nécessaires.

1 Import des données

Un jeu de données (*datasets*) vous est fourni pour ce mini-projet. Importez-les dans une bibliothèque ainsi que le catalogue de formats associé, en suivant la procédure décrite sur Moodle.

À l'aide d'une procédure SAS vous permettant de décrire les *datasets* et le nom des variables, identifiez les *datasets* et variables dont vous aurez besoin pour ce mini-projet.

2 Manipulation de données et statistiques

2.1 Démographie

1° À partir du *dataset* Demography :

- Créez une variable AGE_TODAY qui donne l'âge des patients par rapport à la date courante.
- Créez une variable ED_GROUP qui soit égale à 1 si le nombre d'année d'étude du sujet est entre 0-15 et égale à 2 si cette valeur est supérieure 15. Appliquez un format (au sens SAS du terme) sur cette variable correspondant à « Niveau d'étude supérieur » pour la valeur 2 et « Niveau d'étude secondaire » si la valeur est égale à 1.
- Nommez ce nouveau *dataset* DM.

2° Fusionnez le *dataset* DM et les informations de traitement du *dataset* Treatment_Assignment par patient et nommez le DM_TRT.

3° Décrivez statistiquement la variable AGE_TODAY (min, max, moyenne, écart-type, variance et médiane) pour chaque groupe de traitement et commenter les résultats obtenus.

2.2 Examens Physiques

Dans cette section, on utilise le *dataset* Physical_Exam.

1° Calculez l'IMC : poids (en kg)/(taille (en m))² pour les visites où le poids est renseigné.

Convertir le poids en kg (1 kilogramme = 35,2739619 onces) si besoin.

2° Fusionnez les *datasets* Physical_Exam, Date_of_Visit et DM_TRT de la section précédente. Nommez PE le *dataset* obtenu.

3° En utilisant le *dataset* PE, créez un nouveau dataset (PE_Last_Visit) sélectionnant l'enregistrement de la dernière visite ayant un IMC calculé par patient.

A. Quel est l'IMC moyen des trois groupes de traitements **pour la dernière visite (ayant un IMC calculé) de chaque patient** ? Faites une représentation graphique de l'IMC en différenciant les trois groupes de traitements.

B. Testez l'égalité de l'IMC des trois groupes de traitements au moyen d'un test ANOVA.

- Précisez les hypothèses du test et les vérifications à faire pour pouvoir l'utiliser.
- Réalisez le test et tirez-en la conclusion.

4° En utilisant le dataset PE (initialement créé à la question 2), faites une analyse descriptive des variables numériques pertinentes par rapport à l'ensemble des patients, puis par rapport aux modalités de la variable *sexe*. Commentez les résultats obtenus.

Faites une représentation graphique de ses variables, en différenciant les trois groupes de traitements.

2.3 Signes Vitaux

1° À partir du *dataset* Vital_Signs, créez une nouvelle variable : Tension (de label « Diagnostique tension »).

Indiquez dans cette variable le diagnostic (colonne définitions) retenu en fonction du classement ci-dessous. Si les résultats des paramètres systolique et diastolique ne correspondent pas à la même définition, sélectionnez le diagnostic le plus défavorable :

Définitions	Systolique	Diastolique
Tension optimale	< 120	< 80
Tension normale	120-129	80-84
Tension normale élevée	130-139	85-89
Hypertension légère	140-159	90-99
Hypertension modérée	160-179	100-109
Hypertension sévère	> 180	> 110

- 2° Fusionnez le *dataset* obtenu à la question 2.3.1° avec le *dataset* Date_of_Visit et nommez VS le résultat de cette fusion.
Si la visite existe mais qu'il n'y a pas d'enregistrement de signes vitaux, il faut indiquer dans *Tension* la valeur « Non Calculé » et laisser les autres items vides.
- 3° Fusionnez le *dataset* VS avec le *dataset* DM_TRT et nommez VS_DM le nouveau *dataset*.
- 4° Créez une table calculant les fréquences des diagnostics par rapport aux trois groupes de traitements (en utilisant une PROC FREQ).
- 5° Testez l'égalité des moyennes de la tension systolique puis la diastolique des trois groupes de traitements en utilisant un test ANOVA.
Et dans un deuxième temps, étudiez les diagnostics en fonction des différents groupes de traitement en utilisant un test de χ^2 (PROC FREQ, option χ^2).
Comparer et commenter les résultats obtenus par ces deux méthodes.

2.4 Tests Cognitifs - Mini-Mental test

Le *Mini-Mental Test de Folstein* ou MMSE (*Mini Mental State Examination*) est un test simple comportant 30 questions (score de 0 à 30) qui explorent l'orientation, la mémoire immédiate, l'attention, le calcul, le langage et les praxies visuo-constructives. Il peut être pratiqué en moins de 15 minutes.

De façon très schématique, la démence est considérée :

- débutante : > 26 et < 30;
- légère : entre 20 et 26;
- modérée : entre 10 et 20;
- sévère : < 10.

1° À partir du *dataset* MMSE_Result, créez un nouveau *dataset* MMSE_P_VQS, en transposant une ligne par patient / visite - question - score.

On rappelle que la *transposition* des données consiste à passer, par exemple, d'une présentation verticale à une présentation horizontale des données; ici :

Patient	Visite	Label de la Question	Score
0020029	Visit 1	WHAT IS THE (YEAR), (SEASON), (DATE), (DAY), (MONTH)	0
0020029	Visit 2	WHERE ARE WE (STATE), (COUNTY), (TOWN), (HOSPITAL), (FLOOR)	4

- 2° À partir du *dataset* MMSE_Result, créez un nouveau *dataset* MMSE_PQ_S, en transposant par patient / question et en créant une colonne par visite contenant le score (vide si non réalisé).
- 3° Créez un nouveau *dataset* que vous nommerez MMSE_Score, avec uniquement patient / visite - score total - groupe de traitement.
- 4° Faites une représentation temporelle (par visite) des moyennes des scores totaux de chaque groupe de traitement
- 5° Créez un *dataset* MMSE_freq contenant le numéro de patient, son groupe de traitement, son premier score total MMSE, et son dernier score MMSE.
- 6° A l'aide d'une ANOVA, comparez les scores totaux lors de la première visite et la dernière visite en fonction du traitement. Commentez les résultats obtenus après avoir exposé les hypothèses du test.
- 7° Comparer les scores totaux de la première et de la dernière visite par groupe de traitement en fonction du sexe, de l'âge ou de l'éducation à l'aide des outils statistiques de votre connaissance. Commenter les résultats. Y-a-t-il un effet du traitement en fonction d'un de ces trois paramètres ?

2.5 Événements Indésirables

Lors d'une étude clinique, un événement indésirable (AE) se définit comme une manifestation nocive survenant chez un patient inclus dans l'étude, qu'elle soit liée ou non au produit sur lequel porte cette recherche. Ces événements sont décrits dans le *dataset* `adverse_event`.

1° Des informations temporelles sont recueillies pour chaque événement indésirable, dont la date de début (Start date) et de fin (end date) de l'évènement. Pour faciliter notre analyse, nous allons compléter les dates incomplètes (jour ou mois non renseigné).

Créez une variable `ae_start_date` ayant un format date. Celui-ci contiendra la date de début de l'AE. En cas de date incomplète, substituez le jour manquant par la valeur 01 et le mois manquant par Janvier.

Créez une variable `ae_stop_date` ayant un format date. Celui-ci contiendra la date de fin de l'AE. En cas de date incomplète, substituez le mois manquant par décembre et le jour manquant par le dernier jour du mois.

Exclure les AE ayant une date de fin inférieure à la première visite du patient et remplacer les date de debut de l'AE par la date de première visite pour ceux dont la date de debut est antérieure (et date de fin postérieure).

2° Fusionnez à ce *dataset* `adverse_event`, le groupe de traitement, la date d'assignation de ce groupe de traitement ainsi que la dernière date de visite du patient et nommez ce dataset AE. Si la date de fin de l'AE est vide (AE en cours), complétez `ae_stop_date` par la dernière date de visite

3° Représentez graphiquement de manière temporelle, par groupe de traitement, les AE de chaque patient en fonction de leur sévérité (distinction des sévérités par un code couleur). Commentez ce graphique

4° Ajouter sur ce graphique une "icône" représentant la date de prise de traitement et la date de dernière visite.

Afin d'uniformiser la terminologie médicale utilisée, ces événements sont codés. Une hiérarchie fournit des degrés ou niveaux de « sur-ordination » et de subordination des termes médicaux. Les niveaux hiérarchiques représentent donc des liens verticaux dans la terminologie. L'ordre hiérarchique est du plus large au plus précis : SOC - HLGT - HLT - PT - LLT.

5° Sur l'ensemble des patients, faites un tableau et un graphique qui représente la proportion des « SOC termes » en fonction du groupe traitement. Commentez-les.

3 Reporting

- 1° Créez un rapport d'analyse statistique au format pdf contenant les différents résultats obtenus dans les exercices soit sous forme graphique, soit sous forme de tableau, avec une légende pour chaque partie, ainsi que vos commentaires/réponses aux questions posées. Que pouvez vous conclure des résultats obtenus ?
- 2° Un échantillon aléatoire de 5 patients vous est fourni sous la forme d'un fichier csv. Importez ce fichier dans SAS et créez un nouveau *dataset* Patient à partir de ce fichier.
- 3° Créez alors un fichier pdf par patient de Patient contenant les éléments suivants :

Etude Mentalor	PATIENT XXXXXXX	UPBRAIN
-----------------------	------------------------	----------------

Démographie

Sexe : Homme / Femme
 Age : xx ans
 Niveau d'éducation : Xx ans
 Groupe de traitement : Mentalor 50mg / Mentalor 120mg / Placebo

Signes Vitaux

Nom de visite (trié par date)	Température (Celsius)	Rythme Cardiaque (pulsation par min)	Tension Systolique (mmHg)	Tension Diastolique (mmHg)	Diagnostic de tension

Mini Mental Test

Question	Visit1	Visit2	Visit3	Visitn...
Indiquez le label de la question	Indiquez le score obtenu			

Insérez une représentation Graphique de l'évolution du score total du Mini Mental Test

Evènements Indésirables

SOC Term	AE Term	Date de Début (DD-MON-YYYY)	Date de fin (DD-MON-YYYY)	Sévérité
Si SOC term contient plusieurs AE term, ne pas répéter le SOC Term		Si AESTDTDD ou AESTDTMO vide, remplacé le par UNK	Si AEENDTDD ou AEENDTMO vide, remplacé le par UNK. Si évènement en cours (stop date complètement vide), indiquer « En cours »	Mild / Severe / Moderate

- 4° Le rendu se fera sous la forme d'un unique répertoire zippé contenant :
 - le rapport statistique décrit au 3.1
 - les rapports individuels des 5 patients (voir 3.3)
 - le(s) programme(s) SAS commenté(s) avec un Main program pour exécuter
 - Le plan de projet
 Ce fichier zippé sera déposé sur l'espace Moodle prévu à cet effet.