

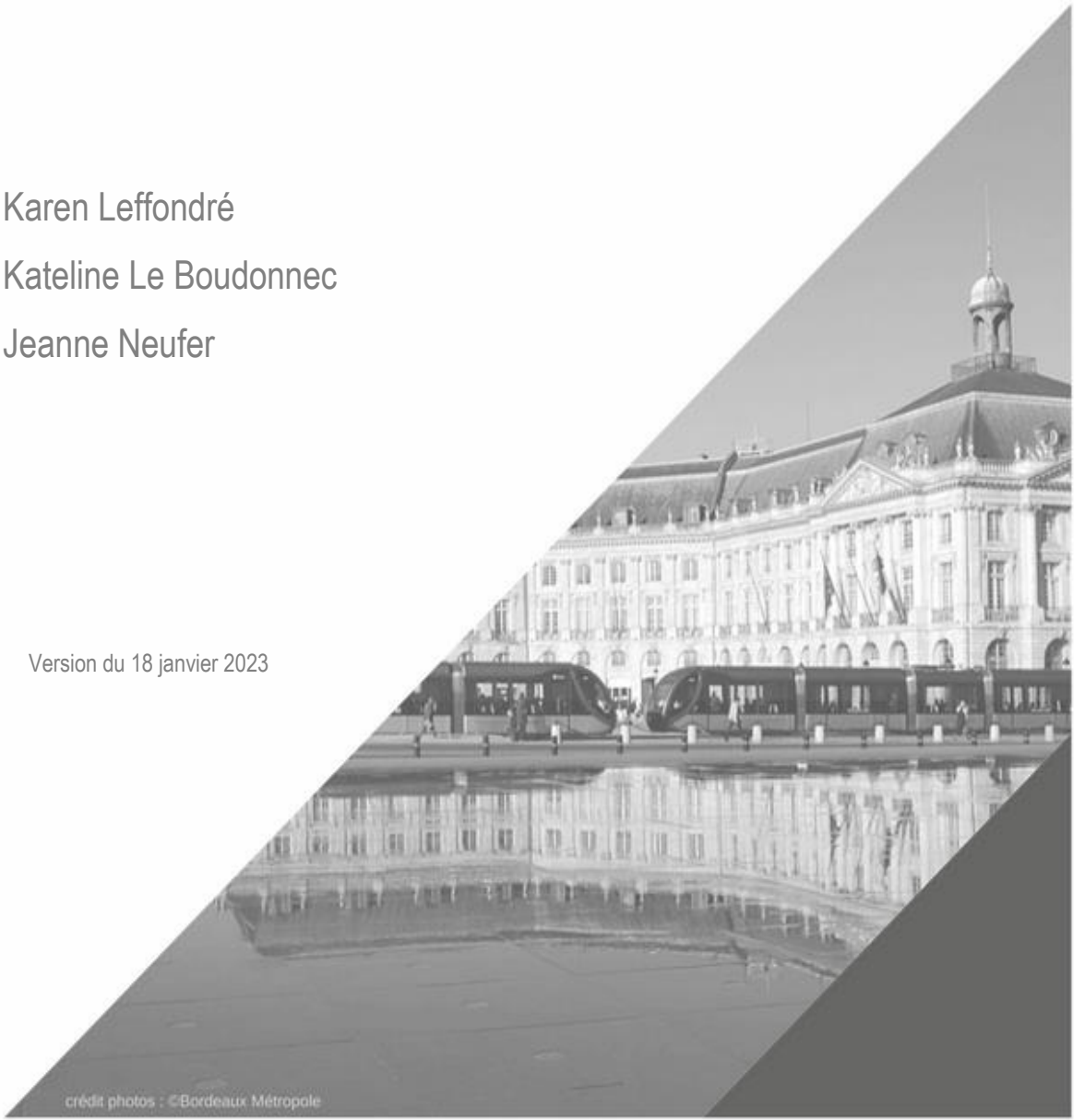
2022-2023

UE STA101

Enoncé des 6 séances de TD

Karen Leffondré
Kateline Le Boudonnec
Jeanne Neufer

Version du 18 janvier 2023



Vous avez 6 séances de TD de 2h chacune (donc 12h en classe) pour répondre à l'ensemble des questions énoncées dans la suite de ce document.

Les données de l'exercice proviennent d'un sous-échantillon de la cohorte Paquid. Cette cohorte a été constituée dans le but d'étudier le vieillissement cérébral normal et pathologique chez des sujets de plus de 65 ans, vivant à domicile en Gironde et en Dordogne. Les sujets ont été tirés au sort sur les listes électorales entre 1988 et 1990.

Les sujets inclus dans l'étude ont été interrogés à l'aide d'un questionnaire standardisé lors de la visite initiale (T0), et lors des suivis à 1 an, 3 ans, 5 ans et 8 ans après la visite initiale. A chaque visite, des informations concernant les caractéristiques sociodémographiques, les conditions de vie, l'état de santé et les performances intellectuelles des sujets ont été recueillies. Pour évaluer les performances cognitives, plusieurs tests psychométriques ont été utilisés, parmi lesquels : le Mini Mental State Examination (MMSE) mesurant la performance cognitive globale, le test de rétention visuelle de Benton, le test d'Isaac utilisé pour explorer l'évocation lexicale, le test de barrages de Zazzo et le test de raisonnement logique simple de Wechsler.

Dans cet exercice, nous nous intéressons au test de Wechsler. Il nécessite une bonne capacité d'attention et présente une composante vitesse. Il se présente sous la forme d'une table qui sert de modèle et décrit neuf paires constituées d'un chiffre et d'un symbole, et d'une seconde table qui ne comporte que des chiffres et des cases vides. En se servant du modèle qui figure en haut de la feuille, le sujet doit compléter le plus de cases possible dans la seconde table en écrivant le symbole correspondant sous chaque chiffre proposé. Le score total est le nombre de cases correctement remplies en 90 secondes ; le score maximum est 93.

L'objectif principal de cette étude est d'étudier la relation entre le score du test de Wechsler à la visite initiale et différentes caractéristiques des sujets.

Les données sont dans le fichier **cweschler.txt**. Il contient des informations sur 501 sujets :

- ayant effectués les cinq tests précités à T0, T1, T3 et T5,
- ayant été vus à T8 et non déments à T8,
- n'étant ni sourds, ni aveugles ;
- n'étant pas confinés au lit ;
- n'ayant pas eu d'AVC, ni de traumatismes crâniens pendant les 5 premières années de suivi.

Les variables disponibles sont les suivantes :

numero : numéro d'identification

age : âge à T0 (en années)

sexe : sexe

- 0=hommes
- 1=femmes

metier : profession

- 0=femme au foyer
- 1=ouvrier agricole
- 2=exploitant agricole
- 3=employé de service
- 4=ouvrier
- 5=artisan-commerçant
- 6=autres employés
- 7=profession intellectuelle

codew0 : score au test de Wechsler à T0

comvin : consommation de vin déclarée à T0

- 0=non buveur
- 1=1/4 de litre /jour
- 2=1/2 litre /jour ou plus

certif : niveau d'études

- 0=sans certificat d'études primaire
- 1=avec certificat d'études primaire

PARTIE 1 - Statistique descriptive (TD1) (à préparer pour le TD1)

- 1) Importer le fichier de données cweschler.txt.
- 2) Décrire la distribution de l'âge et du score au test de Wechsler à T0 à l'aide de graphiques appropriés et commentez.
- 3) Décrire toutes les variables disponibles à l'aide d'indices numériques, résumer ces indices dans un seul tableau pouvant être inclus dans un rapport, puis commenter.

PARTIE 2 - Tests d'hypothèses (TD1 et TD2)

- 1) Tester l'association entre le **score** au test de Wechsler à T0 et le **sexe** à l'aide d'un test approprié
 - a. Décrire les données illustrant cette association (à préparer pour le TD1).
 - b. Ecrire sur papier toutes les étapes du test à réaliser, en calculant la réalisation de la statistique de test vous-même à partir des données, et en déterminant la région critique et la p-valeur avec R (à préparer pour le TD1).
 - c. Utiliser la procédure SAS ou la fonction R permettant de réaliser le test, et comparer avec votre résultat à la question b.
- 2) Tester la même association mais **chez les plus de 80 ans** à l'inclusion.
 - a. Créer le dataset ne comprenant que les sujets de plus de 80 ans.
 - b. Décrire les données illustrant l'association dans cette population.
 - c. Ecrire sur papier toutes les étapes du test à réaliser, en calculant la réalisation de la statistique de test vous-même à partir des données, et en déterminant la région critique et la p-valeur avec R.
 - d. Utiliser la procédure SAS ou la fonction R permettant de réaliser le test, et comparer avec votre résultat à la question c.
- 3) Tester l'association entre le **score** au test de Wechsler à T0 et la **consommation de vin**
 - a. Décrire les données illustrant cette association.
 - b. Ecrire sur papier toutes les étapes du test à réaliser, en calculant la réalisation de la statistique de test vous-même à partir des données, et en déterminant la région critique et la p-valeur avec R.
 - c. Utiliser la procédure SAS ou la fonction R permettant de réaliser le test, et comparer avec votre résultat à la question b.
- 4) Tester l'association entre le **niveau d'éducation** et la **consommation de vin chez les plus de 80 ans** à l'inclusion.
 - a. Décrire les données illustrant cette association.
 - b. Ecrire sur papier toutes les étapes du test à réaliser.
 - c. Utiliser la procédure SAS ou la fonction R permettant de réaliser le test.

PARTIE 3 - Régression linéaire simple (TD3 et TD4)

- 1) Estimer l'association brute entre le score au test de Wechsler à T0 et l'**âge** à T0 (en quantitatif) à l'aide d'un modèle de régression
 - a. Sur papier
 - i. Ecrire l'équation du modèle.
 - ii. Décrire les conditions d'application de ce modèle de régression simple (hypothèses fondamentales), et ce qu'elles signifient concrètement pour ce modèle.
 - iii. Donner l'interprétation des coefficients de régression.
 - iv. Ecrire la formule de l'intervalle de confiance du coefficient de régression d'intérêt principal.
 - v. Décrire les étapes du test à réaliser pour obtenir la p-valeur pour ce coefficient de régression d'intérêt principal.
 - b. A partir des données
 - i. Tracez les graphiques nécessaires pour étudier les conditions d'application du modèle avant estimation, et commentez.
 - ii. Estimer le coefficient de corrélation de Pearson entre les deux variables.
 - iii. Estimer le modèle.
 - iv. Interpréter tous les résultats.
 1. Les coefficients de régression, les IC à 95%, et les p-valeurs.
 2. Le R².
 - v. Vérifier que le carré du coefficient de corrélation de Pearson est bien égal au R².
 - vi. Tracer la droite estimée sur un nuage de points.
 - vii. Calculer la différence estimée de score moyen à T0 entre deux sujets ayant 10 ans d'écart à T0, et l'intervalle de confiance correspondant.
 - viii. Proposez un modèle permettant d'obtenir un intercept interprétable, estimez-le, et interprétez les résultats.
- 2) Estimer l'association brute entre le score à T0 et le **sexe**
 - a. Ecrire sur papier l'équation du modèle.
 - b. Estimer le modèle et interpréter les résultats.
 - c. Comparer le résultat du test avec le test PARTIE 2.
- 3) Estimer l'association brute entre le score à T0 et le **niveau d'étude**
 - a. Ecrire sur papier l'équation du modèle.
 - b. Estimer le modèle et interpréter les résultats.
- 4) Estimer l'association brute entre le score à T0 et la **profession**
 - a. Regrouper judicieusement certaines catégories de la variable « métier » en créant une nouvelle variable binaire sur la profession.
 - b. Ecrire sur papier l'équation du modèle avec cette nouvelle variable binaire.
 - c. Estimer le modèle et interpréter les résultats.
- 5) Estimer l'association brute entre le score à T0 et la **consommation de vin**
 - a. Ecrire sur papier l'équation du modèle.
 - b. Estimer le modèle et interpréter les résultats.
 - c. Effectuer le test nécessaire en décrivant les étapes.
 - d. Comparer le résultat du test avec le test PARTIE 2.
- 6) Présenter tous les résultats des modèles précédant dans un seul tableau sur Word pouvant être inclus dans un rapport (coefficients de régression estimé, intervalle de confiance, et p-valeur pour chaque variable).

PARTIE 4 - Régression linéaire multiple (TD5 et TD6)

- 1) Estimer un modèle permettant d'estimer l'association entre la consommation de vin et le score à T0, ajustée sur l'âge et le niveau d'éducation. Interpréter les résultats.
- 2) Etudier l'adéquation du modèle précédent à partir des résidus.
- 3) Etudier si l'association entre la consommation de vin et le score à T0 diffère chez les hommes et chez les femmes de même âge et de même niveau d'éducation.
 - a. Faire des tableaux croisés entre la consommation de vin, le sexe, et le certificat d'étude, pour déterminer s'il y a suffisamment de sujets dans chaque case pour pouvoir répondre à cette question.
 - b. Ecrire sur papier l'équation du modèle, et indiquez les paramètres nécessaires pour répondre à la question.
 - c. Estimer le modèle avec l'interaction nécessaire, interpréter les résultats pour la consommation de vin
 - d. Réaliser le test permettant d'évaluer le degré de significativité statistique de cette interaction, en détaillant les étapes sur papier, et conclure.
 - e. Présenter les résultats du modèle avec interaction dans un tableau sur Word, même si l'interaction est non significative, et commenter les résultats.