

TP 2 : Tests de comparaison à deux échantillons et test du chi-deux
Date limite de rendu du compte-rendu : dimanche 16 décembre

Instructions : Un compte-rendu du TP rédigé sous la forme d'un fichier pdf est à rendre par binôme d'étudiants. Vous devrez déposer votre compte rendu sous moodle. Chaque binôme remettra un document pdf ayant pour nom *nomsetudiants_tptest.pdf*. Le document à remettre devra comporter **au maximum** 10 pages. Des consignes de rédaction se trouvent à la fin du document.

Exercice 1

Une étude sur l'insuffisance pondérale des nouveaux nés à été menée en 1986 aux États-Unis (Baystate Medical Center, Springfield, Massachusetts) auprès de 189 mères. Les données sont disponibles sur moodle sous le nom de `birthwt.dat`. On qualifie dans la suite un bébé d'hypotrophe si son poids de naissance est inférieur à 2 500 grammes.

Les variables de la base sont les suivantes :

- low : indicateur d'insuffisance pondérale (1 si poids du bébé inférieur à 2 500 g, 0 sinon)
- age : âge de la mère (en années)
- smoke : tabacologie de la mère durant la grossesse (1 si fumeur, 0 si non fumeur)
- lwt : poids de la mère au début de la grossesse (en livre)
- ht : hypertension pendant la grossesse (1 si hypertension, 0 sinon)
- ui : infection urinaire au cours de la grossesse (1 si infection, 0 sinon)
- bwt : poids du bébé à la naissance (en grammes).

1. On s'intéresse au poids du bébé à la naissance.
 - (a) Mettre en place une méthodologie descriptive pour étudier le lien entre le poids des bébés et la présence ou non d'infection urinaire au cours de la grossesse.
 - (b) Mettre en place un test pour répondre à la question précédente en précisant, si il y a lieu, les hypothèses de modélisation nécessaires à sa réalisation et en vérifiant ces hypothèses.
 - (c) Étudier le lien entre la variable poids du bébé à la naissance et toutes les autres variables de la base de données. Pour chaque variable, proposer une étude descriptive puis effectuer un test pour conclure. Présenter les résultats de tous les tests dans un tableau récapitulatif.
 - (d) D'après vos différentes analyses, quelles sont les variables qui influent sur le poids du bébé à la naissance ?
2. Etudier le lien entre la variable indicateur d'insuffisance pondérale (low) et toutes les autres variables de la base de données. Après avoir effectué une étude descriptive, présenter les résultats des tests dans un tableau récapitulatif.
3. D'après vos différentes analyses, quels facteurs de risque proposeriez-vous pour l'insuffisance pondérale à la naissance ?

Exercice 2

On s'intéresse à la base de données `employes.csv` disponible sous moodle. On veut étudier l'influence du sexe sur la catégorie d'employé dans cette base de données.

1. Quelle méthodologie permet d'étudier le lien entre sexe et catégorie d'employés ?
2. Donner le tableau de contingence de la catégorie et du sexe.
3. Quelle est la proportion d'employés qui sont des hommes parmi les cadres ? Parmi les secrétaires ?
4. Quelle est la proportion d'employés qui sont des secrétaires parmi les hommes ? Parmi les femmes ?
5. Proposer un graphique permettant de visualiser le lien entre le sexe et la catégorie d'employés. Mettre en oeuvre le test permettant d'étudier le lien entre ces 2 variables. Quelle est la statistique de ce test, sa loi sous l'hypothèse (H_0), sa valeur produite par R ?
6. La catégorie des employés est-elle liée à leur sexe ? Justifier votre réponse à partir de vos analyses.

Consignes pour la rédaction du compte-rendu

Votre compte-rendu devra être paginé et contenir pour chaque exercice :

1. Une introduction présentant la base de données et les objectifs de l'étude.
2. Un corps de texte dans lequel vous présenterez les détails de votre analyse et commenterez les résultats obtenus. Vous présenterez les résultats pertinents de statistique descriptive univariée et bivariée (sous forme de tableaux et/ou graphiques commentés). Vous présenterez ensuite les résultats de l'étude du lien entre les variables de la base (résultats descriptifs et tests), en commentant les résultats obtenus par rapport à l'objectif annoncé de l'étude.
3. Une conclusion générale sur l'étude incluant un retour sur les objectifs annoncés.