

SCI 1018 Statistiques appliquées en sciences

Consignes optionnelles sur le Laboratoire 3

Consignes

Le cours Statistiques appliquées aux sciences vous permet à terme d'élaborer et d'optimiser des plans d'expérience. L'enseignement de ces compétences est prévu à la Leçon 9 du cours. Cette leçon comprend une partie théorique ainsi qu'un laboratoire sur un sujet de votre choix. Afin de favoriser le succès de ce laboratoire, il est conseillé de présenter à la personne tutrice :

- Le sujet que vous avez retenu;
- Les hypothèses scientifiques et statistiques choisies;
- Le protocole d'échantillonnage;
- Le ou les tests statistiques que vous jugez appropriés.

Vous devriez remettre ces informations avant la Leçon 9 afin de démarrer votre laboratoire à temps et sur une base plus solide. Prenez note que cette préparation au Laboratoire 3 n'est pas évaluée.

Vous trouverez dans les pages qui suivent les consignes officielles et obligatoires du Laboratoire 3.

Laboratoire 3 – Élaboration et optimisation de plans d'échantillonnage

Ce laboratoire est obligatoire et s'échelonne sur deux leçons. Ce laboratoire devra être remis selon les échéanciers de votre feuille de route. Le but de ce laboratoire est d'élaborer un plan d'échantillonnage afin de tester une hypothèse scientifique développée à partir de l'un des thèmes que vous trouverez ci-dessous. Vous devrez choisir un thème, formuler une hypothèse à tester, élaborer un plan d'échantillonnage pour la collecte de données réelles, choisir l'analyse statistique appropriée à partir de celles présentées dans le cours et réaliser cette analyse statistique sur les données.

Votre laboratoire devra obligatoirement inclure les éléments suivants :

1. Introduction : quelques lignes sur l'hypothèse scientifique que vous désirez tester et l'hypothèse statistique que vous utiliserez.
2. Méthodes : une description détaillée de votre plan d'échantillonnage permettant de répéter l'expérience exactement comme vous l'avez exécutée (matériel utilisé, identification des unités expérimentales, nombre de répétitions, identification de la variable réponse et de son unité de mesure, le ou les facteurs de l'expérience).
3. Analyse : une justification de l'analyse que vous avez choisie pour analyser les données et une énumération des suppositions de cette analyse.
4. Un tableau présentant les données brutes que vous avez récoltées.
5. Le jeu de données en format texte sauvegardé dans un fichier séparé, tel que vous avez importé dans R pour réaliser votre analyse.
6. Résultats : l'exécution de l'analyse, la vérification des suppositions, le `summary()` de votre analyse si pertinent.
7. Graphiques : un ou deux graphiques illustrant les résultats.

8. Discussion : quelques lignes sur l'interprétation des résultats en relation avec votre hypothèse scientifique de départ.

Votre laboratoire doit comprendre **tous** les éléments identifiés plus haut, ainsi que le code R utilisé pour importer les données, exécuter l'analyse, vérifier les hypothèse, et réaliser les graphiques tels qu'ils apparaissent dans votre laboratoire. Afin de bien distinguer le code R et les sorties (*output*) du reste du texte, vous devez utiliser une police monospace comme celle-ci qui attribue le même espace à tous les caractères (p. ex., police de type Courier New) – ceci préserve l'alignement des éléments des sorties. La structure du laboratoire à remettre doit s'inspirer de la forme des solutions des autoévaluations.

Suggestions de thèmes à étudier

1. La germination de graines (quelques possibilités ou d'autres si vous êtes inspirés) ¹ :
 - le temps de germination de trois espèces (p. ex., haricots, lentilles, luzerne, oignons, radis) ;
 - le succès de germination d'une espèce à pH neutre (eau distillée) et à pH acide (gouttes de jus de citron) ;
 - le succès de germination à différentes températures (4°C) et température de la pièce (20°C) ;
 - le succès de germination dans différents substrats (p. ex., terre noire, papier absorbant, aucun substrat) ;
 - la longueur des racines de graines germées dans différents substrats. Veuillez noter que
2. La moisissure (parmi les exemples suivants) ² :

1. Le temps de germination de certaines espèces peut être relativement long. Assurez-vous que le calendrier du cours vous permet de réaliser votre expérience dans les délais prescrits.

2. Le temps d'apparition de moisissures peut être relativement long pour certains substrats. Assurez-vous que le calendrier du cours vous permet de réaliser votre expérience dans les délais prescrits.

- la présence de moisissure sur des carrés de pain de trois types (pain biologique, pain blanc générique, pain blé entier) ;
 - le temps qu'il faut pour qu'apparaissent des moisissures sur des carrés de pain de trois types (pain biologique, pain blanc générique, pain blé entier) ;
 - le temps qu'il faut pour qu'apparaissent des moisissures sur des carrés de pain d'un type de pain à différentes températures ;
 - le temps qu'il faut pour qu'apparaissent des moisissures sur des pêches soumises à différentes températures.
3. La vitesse de frappe au clavier :
 - la vitesse d'écriture entre hommes et femmes pour le même texte ;
 - la vitesse d'écriture entre hommes et femmes selon le type de texte (nouvelle du sport *vs* nouvelle des arts et spectacles).
 4. L'occurrence de la pratique de jeux vidéos chez les hommes et les femmes de différentes classes d'âge (15 – 20 ans, 21 – 25 ans, > 25 ans).
 5. Le nombre de minutes passées à écouter des émissions en baladodiffusion par semaine chez les hommes et les femmes de différentes classes d'âge.
 6. nombre de minutes passées sur internet par semaine entre les hommes et les femmes de différentes classes d'âge.
 7. La fréquence d'utilisation de sacs réutilisables par les clients de différents types de commerces (épicerie, SAQ, quincaillerie, boutique de vêtements).
 8. La proportion de véhicules domestiques (GM, Chrysler, Ford) par rapport aux véhicules d'origine étrangère dans deux quartiers de votre ville.
 9. La capacité à distinguer des bananes biologiques *vs* non biologiques en dégustation en aveugle chez un groupe de gens.
 10. La préférence entre le vin et la bière selon le sexe et la classe d'âge.
 11. La fréquence de certaines couleurs de bonbons dans des sacs de bonbons (p. ex., jujubes,

- fèves en gelée) selon différentes compagnies (Dare, Choix du Président, Nos Compliments, etc.).
12. La différence entre le volume moyen de raisins secs de différentes compagnie (Sun-Maid, Choix du Président, Nos Compliments, Sans nom, etc.).
 13. Le nombre moyen de pépins dans les poivrons verts biologiques et les poivrons verts non biologiques.
 14. La préférence de chats pour trois types de nourriture (e.g., croquettes sèches, nourriture en boîte de conserve, nourriture de table).
 15. Le nombre d'hommes et de femmes qui accompagnent leur chien au parc.
 16. Le temps que mettent différents types de boissons gazeuses dans un verre (avec ou sans glace) à perdre tout leur CO₂ à la température de la pièce.
 17. La longueur du texte de nouvelles internationales, d'environnement et de chroniques culinaires dans différents journaux.
 18. La fréquence d'un mot ou d'un groupe de mots dans différents journaux selon le sujet abordé ou depuis l'avènement d'un événement important (p. ex., nomination d'un nouveau premier ministre)³.
 19. Le nombre de publicités dans différents types de quotidiens en semaine *vs* en fin de semaine.
 20. Le prix de l'essence à différentes stations-service au Québec, en Ontario et au Manitoba (<http://www.gasbuddy.com>).
 21. La fréquence de personnes portant des écouteurs dans différents types de transport en commun (autobus, métro, train de banlieue).

3. La fréquence d'un mot ou d'un groupe de mots peut être évaluée grâce à des logiciels d'analyse de texte, tels que TextSTAT (<http://neon.niederlandistik.fu-berlin.de/en/textstat/>) et Textalyser (<http://textalyser.net/>).