

## **Statistiques en santé : exploitation de données et rédaction scientifique**

### **Questionnaire à choix multiples**

**1. Il est essentiel de détailler le mieux possible dans le paragraphe « matériels et méthodes » :**

- a. Quelles données sont étudiées et comment les mesures sont récoltées
- b. Le(s) logiciel(s) utilisé(s) pour effectuer les analyses statistiques ainsi que leurs versions
- c. Les types de représentations graphiques qui serviront d'illustrations dans la publication

**2. Dans le paragraphe « Résultats », il est important de décrire :**

- a. Les caractéristiques de la population étudiée
- b. Les résultats significatifs seulement, de manière littérale, sans nécessairement spécifier la valeur de la p-value ou de l'estimation de l'effet calculé
- c. Les résultats qui répondent aux objectifs de l'étude avec leur p-value respective, avec le cas échéant les valeurs des estimations (avec leurs intervalles de confiance)

**3. Sur quoi peut impacter une mauvaise gestion des données (exemple : variables ou données manquantes, nombre de patients insuffisants) :**

- a. La validité des résultats et donc la conclusion
- b. La reproductibilité des découvertes dans la vraie vie ou par d'autres équipes qui réalisent le même protocole ou analyses
- c. La mise en pratique d'une méthodologie statistique convenable

**4. La rédaction d'une méthodologie statistique basique (mais très complète) comme complexe (si les analyses sont très poussées) permet :**

- a. D'avoir une publication garantie dans un journal avec un très gros *impact factor*
- b. D'évaluer le sérieux et la robustesse de l'analyse des données de l'étude en question, et aussi le niveau des auteurs en matière de biostat
- c. De pouvoir reproduire l'expérience sur une autre cohorte ou avec d'autres données en l'utilisant telle un mode d'emploi

**5. Pour savoir manipuler les logiciels utilisés pour faire des biostatistiques, il est nécessaire en premier lieu d'avoir des connaissances en matière de :**

- a. Biologie d'une manière générale, parfois pointues dans certains domaines, et de mathématiques (à un niveau très poussé)
- b. Base de données/data-management, statistiques (tests, modélisations) et programmation
- c. De rédaction de rapport ou publication scientifique, de création de tables, etc.

**6. Les logiciels de statistiques comme SAS, RStudio, ou même Python, permettent entre autres de :**

- a. Coder des graphiques de haut-niveau
- b. Accéder à des bases SQL
- c. Générer des tables en format Excel ou pdf

**7. Grâce aux nombreuses autres fonctionnalités des logiciels de statistiques, il est aisément et également possible de :**

- a. Créer des diaporamas en format Powerpoint
- b. Générer des fichiers audios
- c. Rédiger des réponses automatiques avec des algorithmes de machine learning comme ChatGPT

**8. Cocher la ou les bonnes(s) pratique(s) parmi les suivantes :**

- a. On peut toujours intégrer dans une publication des graphiques et ses propres calculs directement produits sur Excel depuis la base de données brute
- b. C'est très utile de refaire des graphiques sur Paint ou de les améliorer avec des logiciels de retouche (type Photoshop)
- c. Utiliser le test ou la modélisation statistique qui rendra mon objectif significatif (avec une p-value inférieure à 5%)

Réponses :

1. a et b

2. a et c

3. a, b et c

4. b et c

5. b

6. a, b et c

7. aucune bonne réponse

8. aucune bonne réponse