

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique



Université de Constantine 03
Salah Bounider
Faculté de Médecine
Département de Médecine



Polycopié du Cours et Travaux Pratiques

Introduction à l'Analyse des Données Biomédicales

-Biostatistique-informatique-

Destiné aux Étudiants de 1ère Année de
Médecine

Rédigé par : Dr. Abdennour Boulesnane

Année Universitaire : 2025-2026

Avant-propos

Cher(e)s Étudiant(e)s,

Il me fait plaisir de vous présenter ce polycopié dédié au module de Biostatistique-informatique pour les étudiants de première année de médecine. Dans ce module, vous trouverez une introduction complète à l'analyse de données biomédicales et son traitement statistique à l'aide d'Excel et SPSS. Le contenu de ce module est hautement pertinent pour les défis scientifiques actuels en médecine et en médecine computationnelle. L'utilisation efficace des données médicales et l'application de méthodes statistiques et computationnelles sont essentielles pour avancer dans le diagnostic et le traitement des maladies. C'est pourquoi, ce polycopié est conçu pour vous fournir une base solide en statistique et informatique, qui est essentielle pour la recherche en médecine, afin de prendre des décisions éclairées et de mener des recherches de qualité.

Le but de ce polycopié est de vous offrir un support pédagogique clair et complet pour faciliter votre apprentissage. Chaque chapitre commence par un résumé des points clés à retenir, suivi d'une explication détaillée des concepts abordés et de nombreux exemples pour vous aider à bien comprendre les notions présentées. Les travaux pratiques proposés vous permettront de mettre en pratique ces notions et d'acquérir les compétences nécessaires pour réussir dans notre discipline.

Dans les sept chapitres suivants, vous apprendrez à collecter vos données, à les prétraiter en utilisant différentes techniques, à effectuer des analyses statistiques avancées telles que l'analyse de relations entre variables, et à utiliser les distributions de probabilité avec SPSS. Chaque chapitre est accompagné d'exemples pratiques pour vous aider à mieux comprendre les concepts et les techniques abordés.

Nous espérons que ce polycopié vous offrira une expérience d'apprentissage enrichissante et stimulante. Nous vous souhaitons à tous une excellente lecture et une réussite éclatante dans vos études!

Cordialement,

Dr. Abdenmour Boulesnane
Dernière mise à jour le : 10 mars 2026



Table des matières

| | |
|--|------------|
| Table des figures | v |
| Liste des tableaux | vii |
| 1 Analyse des Données Biomédicales | 1 |
| 1.1 Introduction | 1 |
| 1.2 Science des Données | 1 |
| 1.2.1 Définition | 1 |
| 1.2.2 Méthodologie de la Science des Données | 2 |
| 1.3 Médecine Computationnelle | 3 |
| 1.3.1 Définition | 3 |
| 1.3.2 Domaines d'application | 4 |
| 1.4 Science des données en médecine computationnelle | 5 |
| 1.4.1 Données Biomédicales | 5 |
| 1.4.2 Outils d'analyse de données biomédicales | 6 |
| 1.4.3 Excel | 7 |
| 1.4.4 Statistical Package for the Social Sciences : SPSS | 7 |
| 1.5 Conclusion | 8 |
| 2 Collection de Données avec Excel | 9 |
| 2.1 Introduction | 9 |
| 2.2 Démarrer Excel | 9 |
| 2.3 Terminologie | 10 |
| 2.4 Fenêtre EXCEL | 10 |
| 2.5 Fichier XLS ou XLSX? | 12 |
| 2.6 Gestion des Classeurs | 12 |
| 2.6.1 Créer classeur | 12 |
| 2.6.2 Ouvrir un classeur | 12 |
| 2.6.3 Enregistrer un classeur | 12 |
| 2.7 Cellule Active Et Plage De Cellules | 13 |
| 2.8 Saisie des Données | 13 |
| 2.9 Séries de Données | 14 |
| 2.10 Validation des Variables Statistiques | 15 |
| 2.10.1 Variables Quantitatives | 15 |
| 2.10.2 Variables Qualitatives | 16 |
| 2.11 Conclusion | 17 |
| 3 Organisation des Données Collectées dans SPSS | 18 |

TABLE DES MATIÈRES

| | | |
|----------|---|-----------|
| 3.1 | Introduction | 18 |
| 3.2 | L'Environnement SPSS | 18 |
| 3.3 | Manipulation de Données dans SPSS | 20 |
| 3.3.1 | Définition des métadonnées (metadata) | 21 |
| 3.4 | Saisie et affichage des éléments de données dans l'onglet «Data View» | 24 |
| 3.5 | Sauvegarde des données SPSS | 25 |
| 3.6 | Ouverture de fichiers de données SPSS | 25 |
| 3.7 | Transfert de données d'un fichier Excel vers SPSS | 25 |
| 3.8 | Conclusion | 27 |
| 4 | Prétraitement des données | 28 |
| 4.1 | Introduction | 28 |
| 4.2 | Navigation dans le SPSS Viewer | 28 |
| 4.3 | Travailler avec les données dans SPSS | 29 |
| 4.4 | Remplacement des valeurs manquantes | 30 |
| 4.5 | Tri des observations | 31 |
| 4.6 | Recodage des variables | 32 |
| 4.7 | Suppression d'une variable ou d'une observation | 33 |
| 4.8 | Fractionnement des données | 34 |
| 4.9 | Sélection des données | 35 |
| 4.9.1 | Condition logique simple | 35 |
| 4.9.2 | Condition logique complexe | 38 |
| 4.10 | Conclusion | 40 |
| 5 | Analyse des données | 41 |
| 5.1 | Introduction | 41 |
| 5.2 | Collecte des données dans SPSS | 41 |
| 5.3 | Prétraitement des données dans SPSS | 42 |
| 5.4 | Utilisation des statistiques descriptives | 43 |
| 5.4.1 | Fréquences pour les variables catégorielles (qualitatives) | 43 |
| 5.4.2 | Fréquences pour les variables continues | 46 |
| 5.4.3 | Résumé des variables continues avec la procédure Descriptives | 48 |
| 5.5 | Conclusion | 48 |
| 6 | Analyse des relations entre variables statistiques | 49 |
| 6.1 | Introduction | 49 |
| 6.2 | Collecte des données dans SPSS | 49 |
| 6.3 | Prétraitement des données dans SPSS | 50 |
| 6.4 | Distributions statistiques bivariées | 50 |
| 6.4.1 | Relations entre variables catégorielles (qualitatives) | 50 |
| 6.4.2 | Relations entre variables quantitatives | 53 |
| 6.5 | Représentation graphique des données | 56 |
| 6.5.1 | Création de graphiques avec Chart Builder | 57 |
| 6.5.2 | Affichage d'une relation linéaire | 59 |
| 6.6 | Conclusion | 60 |
| | Travaux Pratiques | 61 |

Références Bibliographiques

71

Table des figures

| | | |
|------|--|----|
| 1.1 | Organigramme de la méthodologie de la science des données . . . | 2 |
| 1.2 | Recherches en médecine computationnelle | 4 |
| 2.1 | Environnement Excel. | 10 |
| 2.2 | Utiliser l’apostrophe pour saisir du texte dans Excel. | 14 |
| 2.3 | Utiliser la poignée de recopie dans Excel pour compléter une série de données. | 14 |
| 2.4 | Validation des données quantitatives | 16 |
| 2.5 | Validation des données qualitatives | 16 |
| 2.6 | Résultat de la validation des données qualitatives | 17 |
| 3.1 | Environnement SPSS. | 19 |
| 3.2 | Données SPSS. | 20 |
| 3.3 | Vue des variables. | 21 |
| 3.4 | Type de variable. | 22 |
| 3.5 | Libellés de valeurs. | 23 |
| 3.6 | Valeurs manquantes. | 24 |
| 3.7 | Enregistrer les données dans SPSS. | 25 |
| 3.8 | Ouvrir les données dans SPSS. | 26 |
| 3.9 | Ouvrir la source de données Excel. | 26 |
| 4.1 | IBM SPSS statistics viewer. | 28 |
| 4.2 | SPSS data file. | 29 |
| 4.3 | Method Menu. | 30 |
| 4.4 | Replacing Missing Values. | 30 |
| 4.5 | Result after replacement. | 31 |
| 4.6 | Sort Cases. | 31 |
| 4.7 | Dialog box : Recode into Different Variables. | 32 |
| 4.8 | Dialog box : Old and New Values. | 33 |
| 4.9 | Result after Recoding. | 33 |
| 4.10 | Split file. | 34 |
| 4.11 | Frequencies result. | 35 |
| 4.12 | Select Cases. | 36 |
| 4.13 | Conditional expression. | 36 |
| 4.14 | Result after selection. | 37 |
| 4.15 | Frequencies result. | 37 |
| 4.16 | Conditional expression. | 39 |
| 4.17 | Result after selection. | 40 |

| | | |
|------|---|----|
| 5.1 | SPSS Data File. | 41 |
| 5.2 | Result after sorting. | 42 |
| 5.3 | Result after splitting. | 42 |
| 5.4 | Frequencies dialog box. | 43 |
| 5.5 | Frequencies : Statistics dialog box. | 44 |
| 5.6 | Frequencies : Charts dialog box. | 44 |
| 5.7 | Analysis result. | 45 |
| 5.8 | Frequencies dialog box. | 46 |
| 5.9 | Frequencies : Charts dialog box. | 47 |
| 5.10 | Analysis result. | 47 |
| 5.11 | Descriptive statistics result. | 48 |
| | | |
| 6.1 | SPSS Data File. | 50 |
| 6.2 | Result after sorting. | 50 |
| 6.3 | Crosstabs Dialog Box. | 52 |
| 6.4 | Crosstabs Output. | 52 |
| 6.5 | Corrélations bivariées. | 53 |
| 6.6 | Correlation Table | 54 |
| 6.7 | Boîte de dialogue Linear Regression. | 55 |
| 6.8 | Résultats de la régression linéaire. | 56 |
| 6.9 | Chart Builder. | 57 |
| 6.10 | Boxplot : 1D pour la variable Bloodsugar. | 58 |
| 6.11 | Diagramme de dispersion de deux variables quantitatives | 59 |
| 6.12 | Droite de régression linéaire | 60 |

Liste des tableaux

| | | |
|-----|---|----|
| 2.1 | XLS vs XLSX | 12 |
| 2.2 | Variables quantitatives et qualitatives | 15 |
| 4.1 | Conjonction et disjonction logiques. | 38 |
| 6.1 | Analyse des relations statistiques. | 51 |

Chapitre 1

Analyse des Données Biomédicales

1.1 Introduction

L'analyse des données biomédicales est importante, car elle permet aux chercheurs d'identifier les modèles de maladies, de développer des modèles prédictifs, de faciliter le développement de médicaments, de permettre une médecine personnalisée et d'intégrer des données provenant de plusieurs sources pour fournir une image plus complète de la santé des patients. Grâce à ces applications, l'analyse des données biomédicales joue un rôle crucial dans l'avancement de notre compréhension des maladies et dans le développement de nouveaux traitements et thérapies pour améliorer les résultats pour les patients.

1.2 Science des Données

1.2.1 Définition

La science des données est un domaine interdisciplinaire qui implique l'extraction, le traitement, l'analyse, la visualisation et l'interprétation d'ensembles de données volumineux et complexes. Il utilise une combinaison de méthodes statistiques et informatiques pour découvrir des modèles et des informations à partir de données qui peuvent éclairer la prise de décision et fournir un avantage concurrentiel.

La science des données implique une gamme de compétences, y compris l'analyse statistique, l'apprentissage automatique, la visualisation des données et la gestion des données. Cela implique souvent de travailler avec des ensembles de données volumineux et divers provenant de diverses sources, notamment les médias sociaux, les transactions de commerce électronique, les capteurs, les dossiers médicaux et les données financières.

L'objectif de la science des données est de transformer les données en informations exploitables qui peuvent éclairer la prise de décision, améliorer l'efficacité et fournir un avantage concurrentiel. Les scientifiques des données utilisent

une gamme d'outils et de techniques pour atteindre cet objectif, notamment le nettoyage et le prétraitement des données, l'analyse exploratoire des données, l'ingénierie des fonctionnalités, la sélection et la validation des modèles et la visualisation des données.

La science des données a un large éventail d'applications dans tous les secteurs, notamment la santé, la finance, le marketing, les médias sociaux et le commerce électronique. Parmi les exemples d'applications de science des données, citons la détection des fraudes, les systèmes de recommandation, le marketing personnalisé, la maintenance prédictive et le diagnostic médical.

1.2.2 Méthodologie de la Science des Données

La méthodologie de la science des données est une approche systématique de résolution de problèmes basés sur les données qui implique plusieurs étapes clés (voir Figure 1.1) :

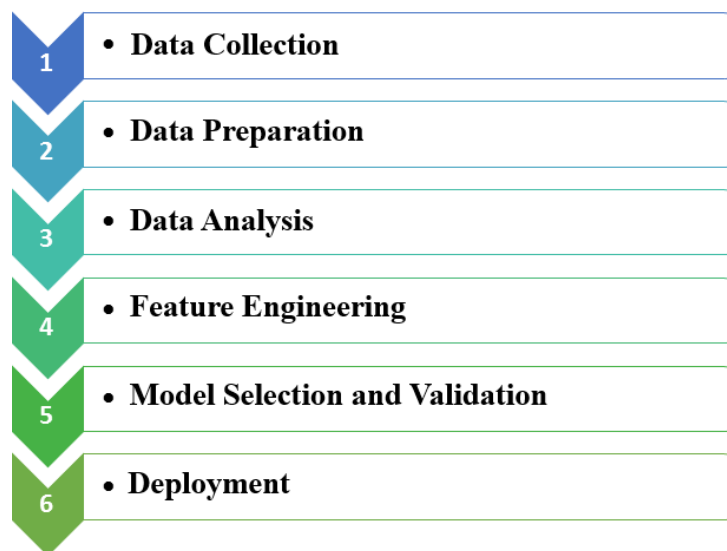


FIGURE 1.1 – Organigramme de la méthodologie de la science des données

1. **Collection de données** : Cela implique d'identifier les sources de données pertinentes, de collecter les données et de les stocker dans un format structuré.
2. **Préparation des données** : Cela implique le nettoyage, la transformation et le prétraitement des données pour les rendre aptes à l'analyse.
3. **L'analyse et L'analyse exploratoire des données** : L'analyse exploratoire des données (EDA) et l'analyse des données sont deux méthodes liées, mais distinctes d'analyse des données.

L'EDA est une étape préliminaire dans l'analyse des données qui implique l'utilisation de diverses techniques statistiques et de visualisation pour explorer et comprendre les données. L'objectif de l'EDA est de découvrir des modèles, des tendances et des anomalies dans les données qui peuvent éclairer le processus d'analyse des données. L'EDA implique des techniques telles que les histogrammes, les boîtes à moustaches, les diagrammes de dispersion et les matrices de corrélation pour explorer

visuellement les données et identifier les principales caractéristiques et relations.

L'analyse des données, en revanche, est un processus plus formel et structuré d'analyse des données qui implique l'application de techniques statistiques et d'apprentissage automatique aux données. Le but de l'analyse des données est d'obtenir des informations et de faire des prédictions à partir des données en utilisant une méthodologie bien définie. L'analyse des données implique des techniques telles que les tests d'hypothèses, l'analyse de régression, la classification et le regroupement pour modéliser et analyser les données.

Bien que l'EDA et l'analyse des données soient des processus distincts, ils sont souvent utilisés ensemble dans le flux de travail d'analyse des données.

4. **Ingénierie des caractéristiques** : Cela implique de sélectionner et de transformer les caractéristiques ou variables pertinentes dans les données pour créer de nouvelles caractéristiques qui amélioreront les performances des modèles.
5. **Sélection et validation des modèles** : cela implique de sélectionner les modèles d'apprentissage automatique ou statistiques appropriés, de les entraîner sur les données et d'évaluer leurs performances à l'aide de métriques et de techniques de validation appropriées.
6. **Déploiement** : il s'agit de déployer le modèle dans un environnement de production et de l'intégrer dans le flux de travail de l'entreprise.

La méthodologie de la science des données est un processus itératif qui consiste à faire des allers-retours entre ces étapes jusqu'à ce qu'une solution satisfaisante soit trouvée. Cela nécessite une combinaison de compétences techniques, de connaissances du domaine et de créativité pour développer des solutions efficaces à des problèmes complexes axés sur les données.

1.3 Médecine Computationnelle

1.3.1 Définition

La médecine computationnelle est un domaine interdisciplinaire qui combine des principes de l'informatique, des mathématiques et de l'ingénierie avec la médecine pour développer des outils et des modèles informatiques pour résoudre des problèmes de santé. Cela implique l'utilisation d'approches basées sur les données, la modélisation informatique et la simulation pour améliorer le diagnostic médical, le traitement et les soins aux patients.

La médecine computationnelle vise à fournir des soins médicaux personnalisés et précis grâce à l'utilisation de modèles et de simulations basés sur les données. Ces modèles peuvent aider à identifier les facteurs de risque, à prédire la progression de la maladie et à optimiser les plans de traitement. Par exemple, les algorithmes d'apprentissage automatique peuvent être utilisés pour analyser des images médicales afin de détecter les premiers signes de cancer ou d'identifier des modèles dans de grands ensembles de données qui peuvent être utilisés pour développer des plans de traitement plus efficaces.

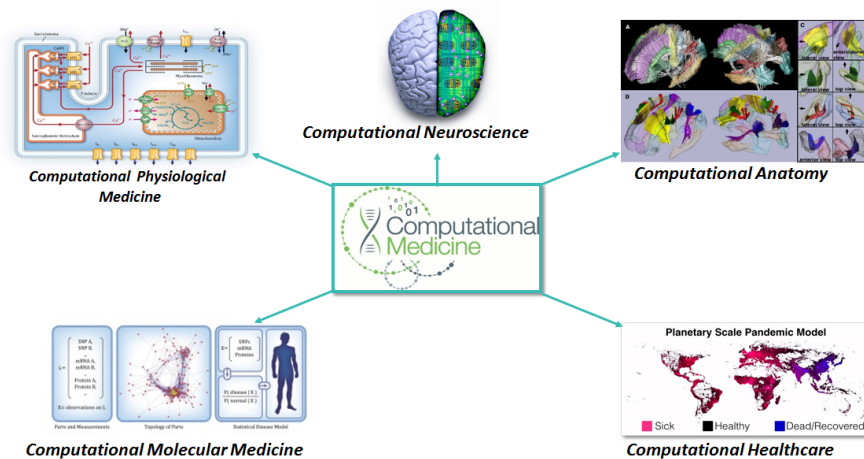


FIGURE 1.2 – Recherches en médecine computationnelle

Certaines des applications de la médecine computationnelle comprennent la découverte et le développement de médicaments, l'imagerie médicale et le diagnostic, la modélisation prédictive, la classification des maladies et la médecine personnalisée. La médecine computationnelle est un domaine en croissance rapide qui a le potentiel de transformer les soins de santé en fournissant des méthodes de diagnostic et de traitement plus précises et plus efficaces.

1.3.2 Domaines d'application

La médecine computationnelle a un large éventail d'applications dans divers domaines, notamment :

1. **Imagerie médicale** : La médecine computationnelle peut être utilisée pour analyser des images médicales telles que des IRM, des tomodensitogrammes et des rayons X pour faciliter le diagnostic, la planification du traitement et la surveillance des maladies.
2. **Génomique et médecine personnalisée** : La médecine computationnelle peut être utilisée pour analyser les données génomiques afin d'identifier les mutations pathogènes, de prédire le risque de maladie et d'élaborer des plans de traitement personnalisés.
3. **Dossiers de santé électroniques** : La médecine computationnelle peut être utilisée pour analyser les dossiers de santé électroniques (DSE) afin d'identifier les modèles de maladie, de prédire les résultats de la maladie et d'élaborer des plans de traitement personnalisés.
4. **Découverte et développement de médicaments** : La médecine computationnelle peut être utilisée pour identifier de nouvelles cibles médicamenteuses, concevoir et optimiser des molécules médicamenteuses et prédire la toxicité des médicaments.
5. **Essais cliniques** : La médecine computationnelle peut être utilisée pour concevoir et analyser des essais cliniques afin d'améliorer leur efficacité et leur efficacité.
6. **Santé publique** : la médecine computationnelle peut être utilisée pour analyser les données de santé publique afin d'identifier les épidémies, de

prévoir la propagation des maladies et d'élaborer des stratégies de prévention et de contrôle des maladies.

7. **Dispositifs médicaux et capteurs** : la médecine computationnelle peut être utilisée pour développer et optimiser des dispositifs médicaux et des capteurs pour le diagnostic, la surveillance et le traitement des maladies.
8. **Informatique de santé** : La médecine computationnelle peut être utilisée pour développer et analyser des systèmes informatiques de santé afin d'améliorer la prestation des soins de santé et les résultats pour les patients.

En résumé, la médecine computationnelle a diverses applications dans divers domaines, notamment l'imagerie médicale, la génomique et la médecine personnalisée, les dossiers de santé électroniques, la découverte et le développement de médicaments, les essais cliniques, la santé publique, les dispositifs et capteurs médicaux et l'informatique de la santé.

1.4 Science des données en médecine computationnelle

La science des données joue un rôle essentiel dans la médecine computationnelle. L'abondance de données médicales générées à partir de diverses sources, y compris les dossiers de santé électroniques, l'imagerie médicale, la génomique et les appareils portables, a créé un besoin d'approches axées sur les données pour analyser et interpréter ces données.

Les méthodes de la science des données, telles que l'apprentissage automatique, la modélisation statistique et l'exploration de données, peuvent être utilisées pour extraire des modèles et des informations à partir de données médicales qui peuvent éclairer le diagnostic médical, le traitement et les soins aux patients. Par exemple, les algorithmes d'apprentissage automatique peuvent être utilisés pour analyser des images médicales afin de détecter les premiers signes de cancer ou d'identifier des modèles dans de grands ensembles de données qui peuvent être utilisés pour développer des plans de traitement plus efficaces.

En médecine computationnelle, la science des données peut être utilisée pour développer des modèles prédictifs capables de prévoir la progression de la maladie ou de prédire les résultats pour les patients. La science des données peut également être utilisée pour identifier les facteurs de risque et personnaliser les plans de traitement en fonction de données spécifiques au patient, telles que les informations génétiques, les antécédents médicaux et les facteurs liés au mode de vie.

Dans l'ensemble, la science des données est un élément crucial de la médecine computationnelle, car elle permet aux chercheurs et aux cliniciens d'exploiter les vastes quantités de données médicales disponibles pour améliorer le diagnostic médical, le traitement et les soins aux patients.

1.4.1 Données Biomédicales

Les données biomédicales désignent tout type de données liées à la santé humaine et à la biologie. Il peut inclure un large éventail d'informations, telles que

1.4. Science des données en médecine computationnelle

les dossiers médicaux des patients, les données d'essais cliniques, les données génomiques, les données d'imagerie et de nombreux autres types de données liées à la santé.

Les données biomédicales sont généralement utilisées pour mieux comprendre les mécanismes de la maladie, identifier les traitements potentiels et améliorer les soins aux patients. Avec l'essor de l'analyse des mégadonnées et de l'intelligence artificielle, l'utilisation des données biomédicales suscite un intérêt croissant pour développer des modèles prédictifs et des approches de médecine personnalisée.

Cependant, les données biomédicales soulèvent également d'importantes préoccupations en matière d'éthique et de confidentialité, en particulier compte tenu de la nature sensible des données et du potentiel d'utilisation abusive. À ce titre, des réglementations strictes régissent la collecte, le stockage et l'utilisation des données biomédicales afin de garantir qu'elles sont traitées de manière responsable et éthique.

1.4.2 Outils d'analyse de données biomédicales

Il existe de nombreux outils disponibles pour analyser les données biomédicales, notamment :

1. **Logiciels statistiques** : les logiciels statistiques tels que R, SAS, Python et **SPSS** sont couramment utilisés pour analyser les données biomédicales, y compris les essais cliniques et les études épidémiologiques.
2. **Outils de visualisation de données** : des outils tels que Tableau, MATLAB et **Excel** peuvent être utilisés pour créer des visualisations de données biomédicales, telles que des graphiques et des diagrammes, afin d'identifier des modèles et des tendances.
3. **Outils d'apprentissage automatique et d'intelligence artificielle** : ces outils, tels que TensorFlow et Keras, peuvent être utilisés pour développer des modèles prédictifs et identifier des modèles dans de grands ensembles de données biomédicales.
4. **Outils d'analyse génomique** : Des outils tels que GATK, SAMtools et Picard peuvent être utilisés pour analyser les données génomiques, y compris les données de séquençage de l'ADN et les données d'expression génique.
5. **Outils d'analyse d'imagerie** : des logiciels tels que ImageJ et OsiriX peuvent être utilisés pour analyser les données d'imagerie médicale, y compris les tomodensitogrammes, les IRM et les radiographies.
6. **Outils d'analyse de réseau** : Ces outils, tels que Cytoscape et Gephi, peuvent être utilisés pour analyser des réseaux biologiques complexes et identifier les relations entre les différents composants.
7. **Outils d'exploration de texte et de traitement du langage naturel** : ces outils, tels que Pubmed et MetaMap, peuvent être utilisés pour extraire des informations de la littérature biomédicale et des dossiers de santé électroniques.

Ce ne sont là que quelques exemples des nombreux outils disponibles pour analyser les données biomédicales. Le choix de l'outil dépendra des besoins spécifiques du chercheur et de la nature des données analysées.

1.4.3 Excel

Excel est un progiciel de tableur couramment utilisé pour l'analyse et la gestion de données dans de nombreux domaines, y compris la recherche biomédicale. Bien qu'Excel puisse être utile pour certains types d'analyse de données biomédicales, ce n'est peut-être pas le progiciel le plus approprié pour tous les types de données biomédicales.

Une limitation d'Excel est qu'il a une limite maximale sur le nombre de lignes et de colonnes pouvant être traitées, ce qui peut être un problème pour les grands ensembles de données. De plus, Excel ne dispose pas de fonctionnalités intégrées pour l'analyse statistique avancée ou la visualisation des données, ce qui peut limiter son utilité pour certains types d'analyse de données biomédicales.

Cependant, Excel peut être utile pour les tâches de gestion de données de base telles que la saisie de données, le tri, le filtrage et les calculs de base. Il peut également être utile pour générer des tableaux et des graphiques simples pour visualiser les données. De plus, Excel peut être utilisé pour suivre des données expérimentales ou créer des feuilles de calcul de base pour calculer des mesures statistiques de base telles que la moyenne, la médiane et l'écart type.

1.4.4 Statistical Package for the Social Sciences : SPSS

SPSS signifie Statistical Package for the Social Sciences, qui est un progiciel utilisé pour l'analyse statistique dans les sciences sociales, y compris la psychologie, la sociologie et d'autres domaines connexes. C'est un outil populaire pour analyser les données dans les études de recherche et il est largement utilisé dans les milieux universitaires et commerciaux.

SPSS offre une interface conviviale pour effectuer des analyses statistiques, permettant aux chercheurs de saisir facilement des données, d'exécuter des analyses et de générer des tableaux et des graphiques pour présenter les résultats. Certains des tests statistiques pouvant être effectués à l'aide de SPSS comprennent les tests t, l'ANOVA, l'analyse de régression, l'analyse factorielle et l'analyse par grappes.

SPSS fournit également une large gamme d'outils de gestion et de transformation des données, tels que le tri, la fusion et le recodage des données, qui peuvent être utiles pour préparer les données à analyser.

En outre, SPSS peut être utile pour certains types d'analyse de données biomédicales, telles que l'analyse de données d'enquête ou l'exécution de tests statistiques sur des données d'essais cliniques. Il fournit une interface conviviale pour la saisie et l'analyse des données, et peut générer des tableaux et des graphiques pour présenter les résultats.

Dans l'ensemble, SPSS est un outil puissant et polyvalent pour l'analyse statistique, en particulier pour ceux qui travaillent dans les sciences des données. Il est conçu pour être facile à utiliser, même pour ceux qui ont des connaissances statistiques limitées, et peut produire des résultats précis et fiables.

1.5 Conclusion

L'analyse des données biomédicales et la médecine computationnelle sont des domaines en pleine expansion qui transforment la recherche médicale et les soins de santé. En exploitant des ensembles de données volumineux et complexes, des méthodes statistiques avancées, l'apprentissage automatique et des modèles computationnels, les chercheurs et cliniciens peuvent mieux comprendre les mécanismes des maladies, améliorer la précision des diagnostics, optimiser les plans de traitement et faire progresser la médecine personnalisée.

L'intégration des techniques de science des données dans la médecine computationnelle permet la modélisation prédictive, la reconnaissance de motifs et la prise de décisions basées sur les preuves, contribuant ainsi à l'amélioration des résultats pour les patients. Des outils tels que **Excel** et **SPSS** jouent un rôle clé dans ces processus, en offrant des plateformes accessibles pour la gestion des données, l'analyse statistique et la visualisation afin de soutenir la recherche et la prise de décision clinique.

Ainsi, la maîtrise de ces outils et des méthodologies d'analyse des données biomédicales est essentielle pour la recherche et la pratique médicale modernes. Ces compétences permettent aux chercheurs et cliniciens d'extraire des informations pertinentes à partir des données et de contribuer de manière significative à l'avancement des soins de santé.

Chapitre 2

Collection de Données avec Excel

2.1 Introduction

La collecte de données est un élément essentiel de la recherche biomédicale, car elle contribue à générer les informations et les connaissances nécessaires pour faire progresser notre compréhension des maladies et développer de nouveaux traitements et thérapies. Sans collecte de données, il serait difficile de faire progresser la recherche biomédicale et d'améliorer les résultats pour les patients.

Dans la recherche biomédicale, Excel est couramment utilisé pour la collecte de données. Excel est un outil polyvalent qui permet aux utilisateurs de créer des feuilles de calcul personnalisées pour stocker et organiser les données de manière structurée. Il offre également des fonctionnalités telles que la validation des données, des formules et des modèles qui peuvent aider à garantir l'exactitude des données, à rationaliser l'analyse des données et à standardiser la collecte des données.

2.2 Démarrer Excel

Il existe plusieurs façons de démarrer **Excel** :

- **Menu Démarrer (Windows 7)** : Cliquez sur le bouton Démarrer → Tous les programmes → Microsoft Office → Microsoft Office Excel 2013.
- **Menu Démarrer (Windows 11)** : Cliquez sur le bouton Démarrer → Tapez "Excel" dans la barre de recherche → Sélectionnez **Microsoft Excel** dans les résultats.
- **Raccourci** : Pour faciliter l'accès au logiciel, il est conseillé de créer un raccourci sur le Bureau. Ensuite, double-cliquez simplement sur ce raccourci pour lancer **Excel**.



2.3 Terminologie

Excel, Classeur (Workbook), Feuille de calcul (Worksheet), Colonne (Column), Ligne (Row), Cellule (Cell), Référence de cellule (Cell Reference), Plage de cellules (Cell Range) (voir Figure 2.1).

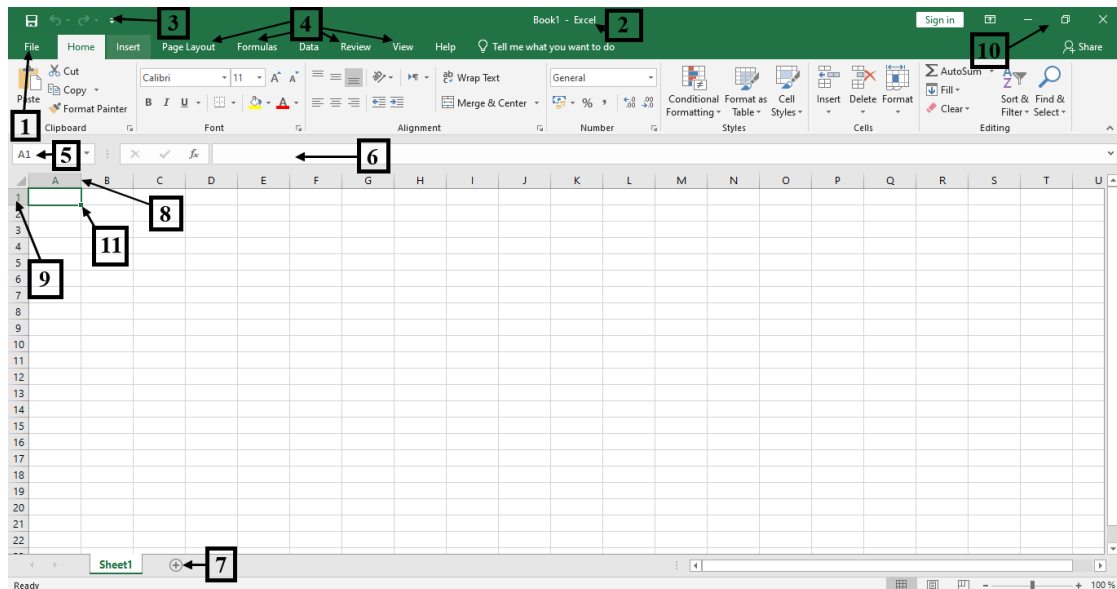


FIGURE 2.1 – Environnement Excel.

- Un fichier Excel est appelé un Workbook (Classeur). Il peut contenir plusieurs Worksheets (Feuilles de calcul) (7). Par défaut, dans Excel 2013, il contient une seule feuille : Sheet1. Chaque worksheet est composée de colonnes (8) et de lignes (9). L'intersection entre une colonne et une ligne forme une cellule.
- Dans Excel, chaque colonne est référencée par une ou plusieurs lettres (A, B, C, ..., XFD) — soit un total de 16 384 colonnes. Les lignes sont numérotées de 1 à 1 048 576. La référence d'une cellule est obtenue en combinant la lettre de la colonne et le numéro de la ligne (sans espace). Par exemple, l'intersection de la 4^e colonne (D) et de la 6^e ligne donne la cellule D6.
- Par défaut, les cellules sont vides, mais elles peuvent contenir des valeurs telles que du texte, des nombres ou des formules.

2.4 Fenêtre EXCEL

- File Tab (1) : Dans Excel 2013, le Office Button a été remplacé par l'onglet File, situé dans le coin supérieur gauche de la fenêtre. En cliquant dessus, on ouvre la Backstage view, qui contient deux panneaux principaux. Le panneau de droite affiche une liste des workbooks récemment utilisés, tandis que le panneau de gauche propose des commandes fréquemment utilisées telles que New, Open, Save, Print, etc.

■ Barre de titre **(2)** : Elle indique le nom du classeur en cours, suivi du nom de l'application utilisée (Microsoft Excel). À droite, il existe les trois boutons **(10)** : réduire, niveau inférieur (agrandir) et fermer.

■ Barre d'outils accès rapide **(3)** : Elle contient les boutons de commandes fréquemment utilisés (vous pouvez les utiliser sans passer par les onglets). Par défaut, elle affiche les trois boutons (enregistrer, annuler et de répéter). Mais on peut la personnaliser en y ajoutant d'autres outils tels que nouveau, ouvrir ou impression rapide, etc.

■ Ruban : il s'affiche sous la barre de titre. Le ruban est organisé hiérarchiquement ; il comporte plusieurs onglets **(4)**. Ils peuvent être des onglets fixes **(4)** : "Accueil, Insertion, Mise en page, Formules, Données, Révision et Affichage" ou des onglets contextuels. Ces derniers s'affichent lorsque vous sélectionnez un objet spécifique. Les onglets contextuels affichés dépendent de l'objet qui est sélectionné. Par exemple : la sélection d'un graphique fait apparaître les trois onglets contextuels Outils de graphique : création, disposition, mise en forme. Ensuite, chaque onglet est divisé en plusieurs groupes qui comportent des boutons de commande et des galeries. Par exemple : l'onglet Accueil comporte les groupes : "Presse-papiers, Police, Alignement, Nombre, Style, Cellules et Édition". Remarque : vous pouvez réduire le ruban (masquer les groupes) : double-cliquez sur l'onglet actif **([Ctrl] + [F1])** :

- Pour réafficher les groupes temporairement, faire un simple clic sur un onglet.
- Pour restaurer le ruban, double-cliquez de nouveau sur un onglet.

■ Zone de nom **(5)** : elle affiche l'adresse ou le nom de la cellule active (actuellement sélectionnée) ou de la sélection. Une cellule est active à la fois. Elle est entourée d'une bordure épaisse.

■ Barre de formule **(6)** : elle affiche le contenu de la cellule active. Elle peut être utilisée pour la saisie ou la modification du contenu de cette cellule.

■ Poignée de la recopie **(11)** : le petit carré vert dans le coin inférieur droit d'une cellule (lorsque vous pointez sur la poignée de recopie, le pointeur se transforme en croix noire).

2.5 Fichier XLS ou XLSX ?

Voir la Table 2.1.

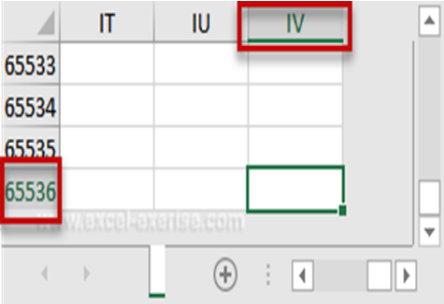
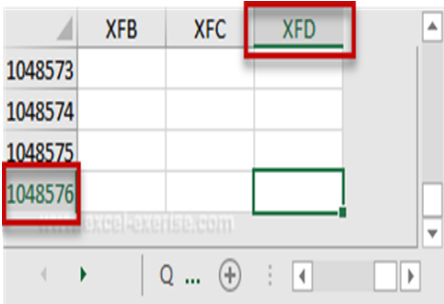
| Excel 97-2003 | Excel 2007 et versions ultérieures (2010, 2013, 2016, 2019, 2021, 365) |
|--|---|
| <p>Dans un classeur xls, les limites sont de 65 536 lignes et 256 colonnes, ce qui correspond à la colonne IV.</p> | <p>Maintenant, avec un classeurxlsx, les limites sont de 1 048 576 lignes et 16 384 colonnes, qui est la colonne XFD.</p> |
|  |  |

TABLE 2.1 – XLS vs XLSX

2.6 Gestion des Classeurs

2.6.1 Créer classeur


Pour créer un nouveau classeur **[Ctrl]+[N]** : **File Tab** ⇒ New ⇒ Blank Workbook.

2.6.2 Ouvrir un classeur

Pour ouvrir un classeur depuis Excel **[Ctrl]+[O]** : **File Tab** ⇒ Open.

2.6.3 Enregistrer un classeur

Excel offre plusieurs choix pour enregistrer un nouveau classeur (le créer physiquement) et le conserver sur votre disque dur :

1. Cliquer **File Tab** ⇒ Save (**[Ctrl]+[S]**)
2. Cliquer sur la disquette  de la barre d'outils accès rapide (**[ctrl]+[S]**)
3. Cliquer **File Tab** ⇒ Save As (**[F12]**)

Lorsque vous utilisez une de ces méthodes d'enregistrement (1, 2 ou 3) pour la première fois, la boîte de dialogue "enregistrer sous" s'affiche, dans laquelle vous spécifiez :

- ✓ Le nom du fichier
- ✓ Le disque et le dossier de sauvegarde
- ✓ Le type de fichier : le format de fichier souhaité pour l'enregistrement. Par défaut, les classeurs sont enregistrés au format (.xlsx) (Excel 2007 et versions ultérieures, y compris 2013)

2.7 Cellule Active Et Plage De Cellules

■ La cellule active (active cell) est celle dans laquelle la saisie sera enregistrée. Elle se distingue par une bordure plus marquée. Par défaut, A1 est la cellule active à l'ouverture du classeur. L'adresse (ou le nom) et le contenu de la cellule active sont affichés dans la zone de nom et la barre de formule respectivement.

■ Tout rectangle de cellules est appelé "plage de cellules" (range of cells), ou "plage". Pour désigner une plage de cellules, il est courant d'utiliser la référence de la première cellule en haut à gauche suivie d'un double-point et de la référence de la dernière cellule en bas à droite.

Exemple :

- A1 :B3 fait référence aux cellules : A1, B1, A2, B2, A3, B3
- C1 :E3 fait référence aux cellules : C1, D1, E1, C2, D2, E2, C3, D3, E3
- A1 :A8 fait référence aux cellules : A1, A2, A3, A4, A5, A6, A7, A8
- A1 :E1 fait référence aux cellules : A1, B1, C1, D1, E1

2.8 Saisie des Données

Avant saisir de données dans une cellule, il faut la sélectionner.

❖ Texte

- Le texte est automatiquement aligné à gauche.
- Le texte ne passe pas à la ligne même s'il est long et dépasse la largeur de la colonne.
- Pour passer à la ligne suivante : **[alt]+[entrée]** au sein d'une cellule
- Si le texte commence par " + ", " - " ou " = ", Excel affiche un message d'erreur "(#NOM?)" car il interprète ce texte comme une formule. Pour éviter cela, ajoutez une apostrophe « ' » avant le texte (exemple : '+Médecine).
- Ajouter une apostrophe « ' » avant un nombre pour que ce dernier soit interprété comme un texte (exemple : '2019 dans la cellule A2).

❖ Nombre

- Le nombre est automatiquement aligné à droite.
- Pour saisir un nombre négatif, il faut le précéder du signe " - " ou le mettre entre parenthèses.
- L'écriture 32e9 signifie $32 * 10^9$ c.-à-d. 32 suivis de neuf zéros.

❖ Date

- La date est automatiquement alignée à droite.
- Pour entrer une date dans une cellule, vous pouvez la saisir sous la forme : jj/mm/aaaa ou jj-mm-aaaa.
- Pour saisir la date d'aujourd'hui, tapez **[ctrl]+[;]** (cette date est constante).
- Pour qu'une date soit interprétée comme un texte, faites-la précéder par une apostrophe « ' » (exemple : '2019) (voir Figure 2.2).

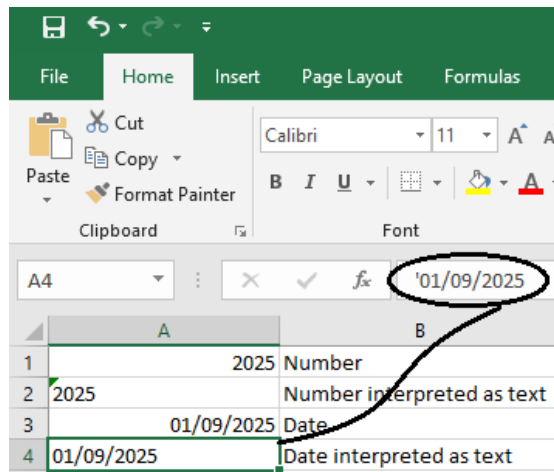


FIGURE 2.2 – Utiliser l’apostrophe pour saisir du texte dans Excel.

2.9 Séries de Données

- ▶ Taper Patient, Date, dans les cellules A1, B1 respectivement.
- ▶ Sélectionner la cellule A1, faire glisser la poignée de recopie sur quelques cellules (vers le bas) (voir Figure 2.3).
- ▶ Pour les dates vous avez la possibilité d’incrémentation par jour mois...

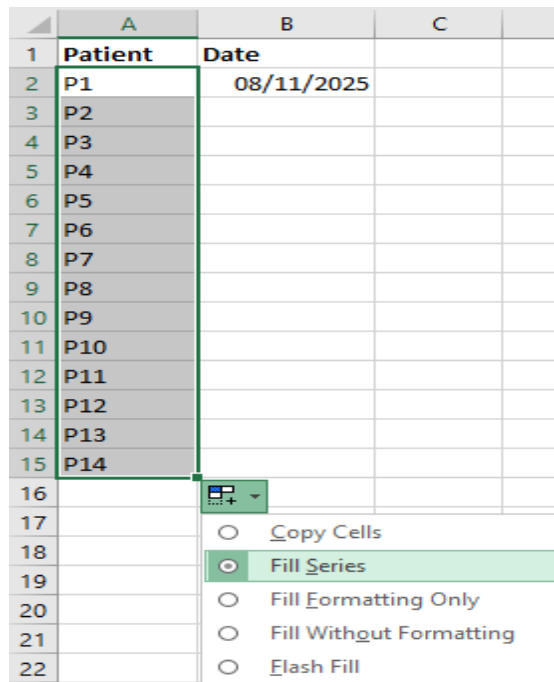


FIGURE 2.3 – Utiliser la poignée de recopie dans Excel pour compléter une série de données.

2.10 Validation des Variables Statistiques

Les variables sont utilisées pour décrire les individus d'une population. Chaque colonne correspond à une variable.

■ Une variable a un nom : "ID", "Age", "poids" ...

■ Une variable a une valeur à un certain moment : MED01 , 22 ans, 59 kg (chaque ligne représente un individu, unité statistique ...)

TABLE 2.2 – Variables quantitatives et qualitatives

| Types de variables | | | |
|--|---|--|---|
| Quantitative (Numérique) | | Qualitative (Catégorielle) | |
| Continue | Discrète | Ordinale | Nominale |
| Se compose de valeurs numériques qui peuvent être mesurées, mais pas comptées (infinie). | Se compose de valeurs numériques qui peuvent être mesurées et comptées (finie). | Se compose de texte ou d'étiquettes qui ont un ordre logique. | Se compose de texte ou d'étiquettes sans ordre logique. |
| Par exemple. Poids {56,06 kg, 87 kg} | Par exemple. Nombre d'enfants {0, 1, 2, 3,..., 10} | Par exemple. Taille de la tumeur {petite, moyenne, grande} | Par exemple. Sexe {Homme, Femme} |

Note :

Le fait qu'une variable soit représentée par des nombres ne signifie pas qu'elle soit automatiquement quantitative. Les valeurs numériques peuvent simplement servir d'étiquettes pour des catégories qualitatives. Par exemple, si nous codons « Homme » par 1 et « Femme » par 2, ces nombres servent uniquement à identifier des catégories distinctes, sans indiquer une relation de mesure entre elles.

2.10.1 Variables Quantitatives

■ En utilisant la fenêtre "Validation des données", il est possible de limiter une saisie en imposant des restrictions telles que la saisie d'un nombre entier.

Exemple : pour la variable age : Pour restreindre la saisie à des nombres entiers entre 0 et 100, commencez par sélectionner les cellules concernées. (avant la saisie de ces valeurs) dans notre exemple C2 :C6 (l'étape S1). Ensuite

Data → **Data Tools** → **Data Validation** (S2 et S3). Choisir **Allow : Whole number** (S4). Ensuite **Data : between** : Saisissez **0** dans le champ **Minimum** et **100** dans le champ **Maximum** (S5) (voir Figure 2.4).

Notes :

■ Pour les variables quantitatives continues, nous choisissons **Decimal** dans la liste "Allow".

■ Dans la liste "Data", vous avez plusieurs choix comme : **greater than, less than, between**, etc.

2.10. Validation des Variables Statistiques

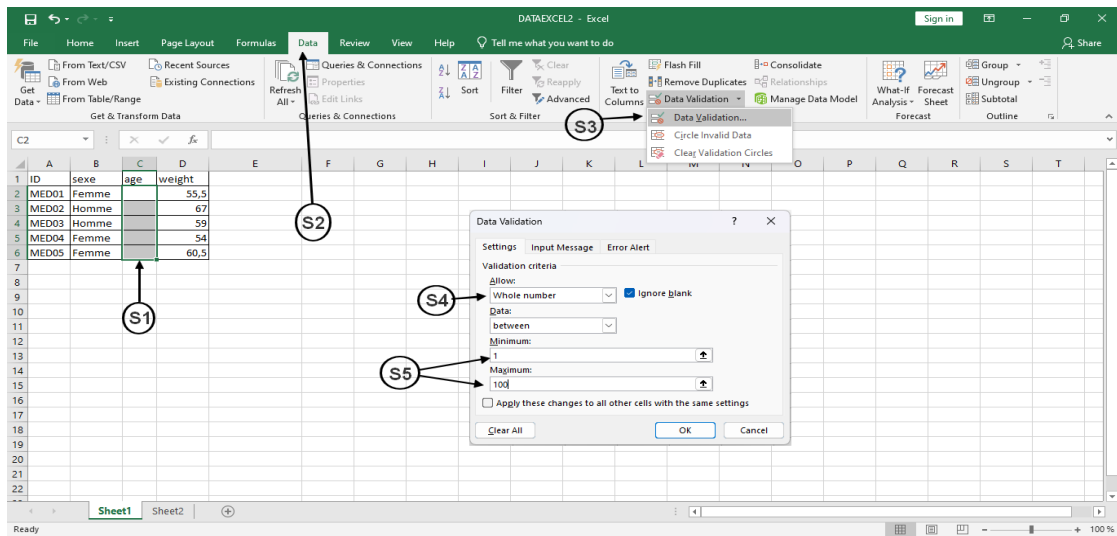


FIGURE 2.4 – Validation des données quantitatives

2.10.2 Variables Qualitatives

■ Sélectionner les cellules où vous allez saisir le sexe (l'étape E1 dans la figure suivante) (dans notre exemple B2 :B6) (voir Figure 2.5).

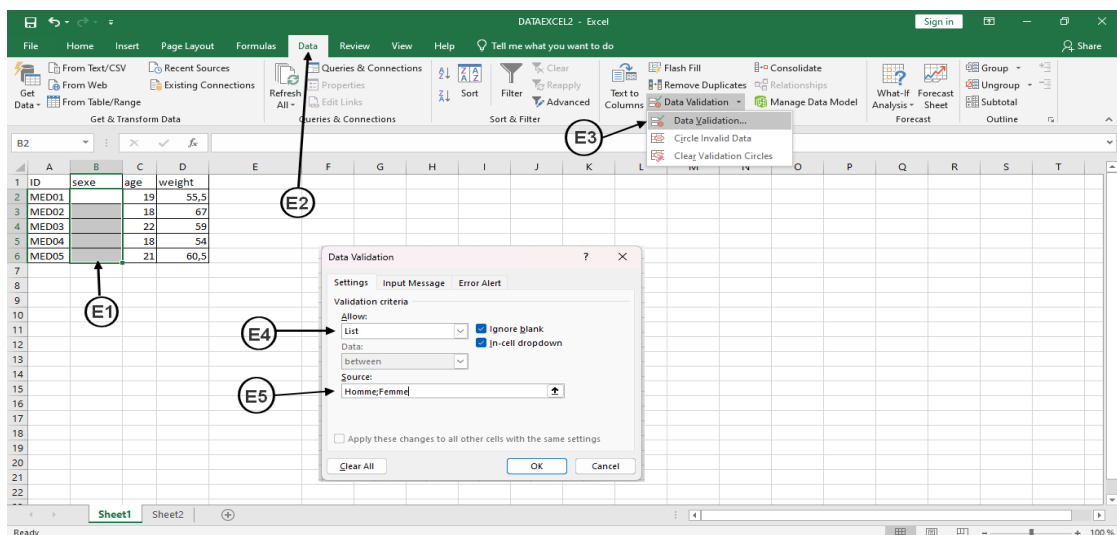
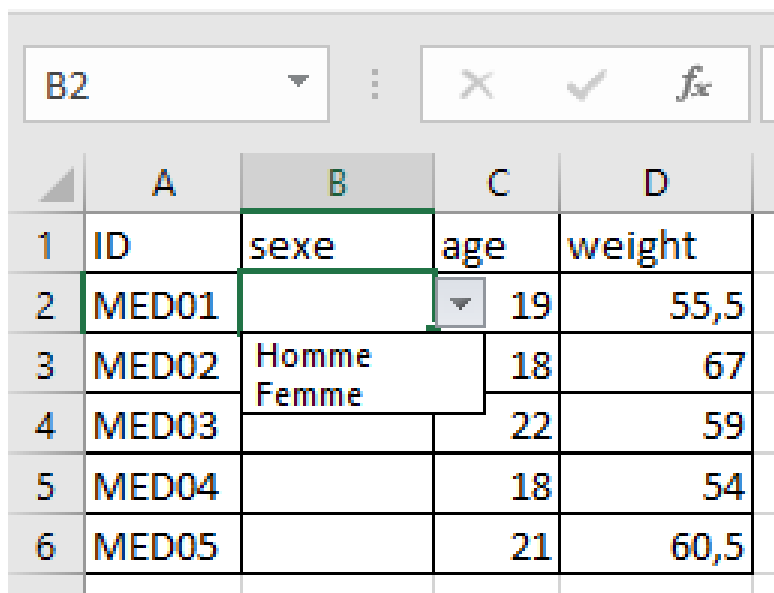


FIGURE 2.5 – Validation des données qualitatives

Ensuite, choisir autoriser **list** (E4) dans : **Data** → **Data Validation** → **Data Validation** (E2 et E3). Taper les éléments de la liste (**Homme ; Femme**) dans le champ source (E5) (séparés par point-virgule ;).

Le résultat sera :



| | A | B | C | D |
|---|-------|----------------|-----|--------|
| 1 | ID | sexe | age | weight |
| 2 | MED01 | | 19 | 55,5 |
| 3 | MED02 | Homme Femme | 18 | 67 |
| 4 | MED03 | | 22 | 59 |
| 5 | MED04 | | 18 | 54 |
| 6 | MED05 | | 21 | 60,5 |

FIGURE 2.6 – Résultat de la validation des données qualitatives

Note :

■ Veuillez trouver le fichier de données Excel de ce chapitre à partir de ce lien : <https://aboulesnane.net/wp-content/datafiles/DATAEXCEL2.xlsx>

2.11 Conclusion

Excel est un outil populaire pour la collecte de données en recherche biomédicale. Pour collecter des données de manière efficace avec Excel, il est important de planifier votre feuille de calcul, d'utiliser la validation des données pour garantir l'exactitude, d'appliquer des formules pour faciliter l'analyse et de sécuriser vos données. Par conséquent, Excel peut être un outil puissant pour organiser et analyser les données en recherche biomédicale.

En résumé, maîtriser Excel est une compétence essentielle pour les chercheurs et étudiants en biomédecine, car il permet une collecte de données précise, facilite l'exploration initiale des données et prépare le terrain pour des analyses avancées et une recherche reproductible.

Chapitre 3

Organisation des Données Collectées dans SPSS

3.1 Introduction

SPSS (Statistical Package for the Social Sciences) est un outil logiciel puissant largement utilisé pour organiser, analyser et visualiser des données dans la recherche sociale et biomédicale. Il offre une gamme d'options personnalisables pour l'organisation et la gestion des données, de puissants outils statistiques pour l'analyse des données et une documentation claire et transparente pour la reproductibilité.

3.2 L'Environnement SPSS

L'environnement SPSS se compose de plusieurs composants, comme le montre l'image ci-dessous (Figure 3.1) :

1. The Title Bar : affiche le nom du fichier actuel et de l'application.
2. The Menu Bar : Cette barre donne accès à différentes commandes qui sont regroupées selon leur fonction. SPSS a un certain nombre d'options de menu situées en haut de l'écran (comme tout autre programme informatique). Ouvrez SPSS et sélectionnez chacune des options de menu une par une.
 - ◆ **The 'File' menu (shortcut Alt + F)** : Essentiellement, ce menu vous permet d'ouvrir des fichiers existants, d'en créer de nouveaux et d'imprimer ou d'enregistrer tout ce sur quoi vous travaillez. Les listes Données récemment utilisées et Fichiers récemment utilisés sont utiles, car elles vous permettent d'accéder rapidement aux fichiers que vous avez récemment ouverts ou sur lesquels vous avez travaillé.
 - ◆ **The 'Edit' menu (shortcut Alt + E)** : Ce menu devrait vous être familier si vous avez déjà utilisé des traitements de texte. Annuler et Rétablir peuvent aider à rectifier les erreurs que vous faites. Couper, Copier et Coller vous permettent de déplacer des blocs de nombres d'une zone de la feuille de calcul à une autre. Chercher... et Aller à

Organisation des Données Collectées dans SPSS

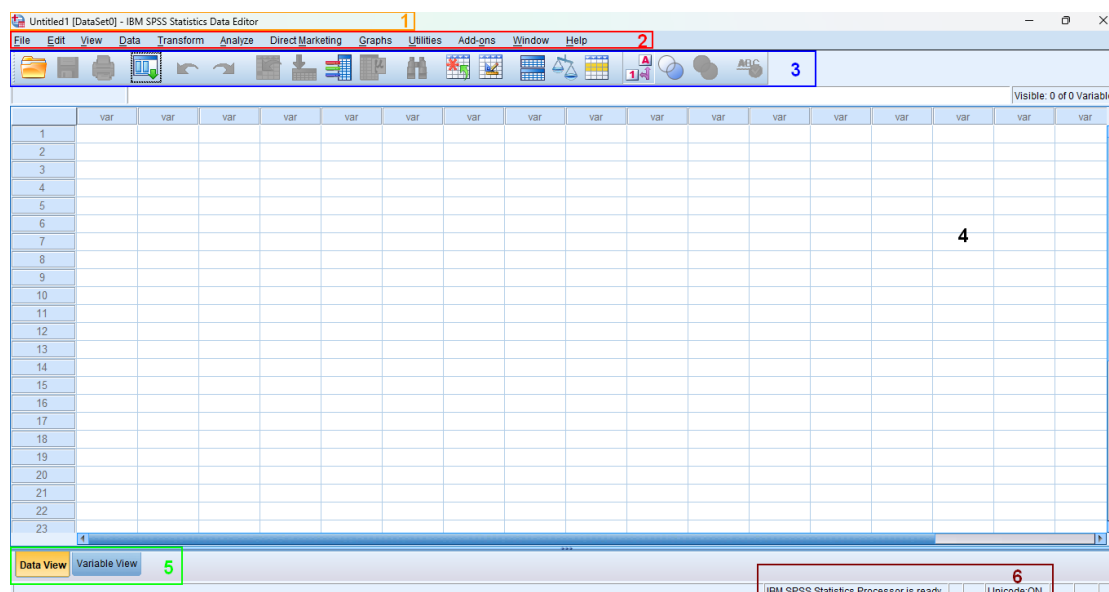


FIGURE 3.1 – Environnement SPSS.

l'observation... vous permet de localiser un score de données ou un participant particulier, ce qui est très pratique lorsque vous traitez un grand nombre de données.

- ◆ **The 'View' menu (shortcut Alt + V)** : Le menu Affichage traite des aspects visuels de la feuille de calcul, en particulier : quelles barres d'outils sont affichées, quelles polices sont utilisées, pour voir les lignes de la grille sur la feuille de calcul ou les étiquettes de valeur sont affichées pour vos variables. . . , etc.
- ◆ **The 'Data' menu (shortcut Alt + D)** : Ce menu vous permet d'organiser votre fichier de données. Il est peu probable que vous utilisiez initialement la plupart des options de ce menu ; cependant, quelques-unes des options peuvent être utiles. Par exemple, vous pouvez identifier certaines erreurs potentielles commises lors de la saisie de données en signalant d'éventuelles entrées de données en double à l'aide de l'outil **Identify Duplicate Cases**.
- ◆ **The 'Transform' menu (shortcut Alt + T)** : Ce menu vous permet de manipuler vos variables.
- ◆ **The 'Analyze' menu (shortcut Alt + A)** : C'est le menu que vous utiliserez probablement le plus et dont vous aurez initialement besoin : Statistiques descriptives, Comparer les moyennes, Modèle linéaire général, Corrélation et Régression, . . . , etc.
- ◆ **The 'Direct Marketing' menu (shortcut Alt + M)** : C'est plus pour les entreprises qui souhaitent réaliser des études de marché. Vous n'aurez pas besoin d'utiliser ce menu !
- ◆ **The 'Graphs' menu (shortcut Alt + G)** : Ce menu vous permet de présenter les données sous forme graphique, ce qui vous aidera à mieux comprendre vos données. Il existe plusieurs façons de créer des graphiques dans SPSS, mais c'est un bon point de départ.

- ◆ **The 'Utilities' menu (shortcut Alt + U)** : En pratique, il est utile pour créer des analyses personnalisées et automatisées. . . , mais n'hésitez pas à l'ignorer pour l'instant !
 - ◆ **The 'Window' menu (shortcut Alt + W)** : Ce menu vous permet d'accéder rapidement à d'autres fenêtres qui pourraient être masquées.
 - ◆ **The 'Help' menu (shortcut Alt + H)** : Ce menu peut être très utile, car il vous offre de l'aide et des informations à la fois sur le système du programme lui-même et sur les tests statistiques qu'il propose.
3. The Toolbar : fournit des raccourcis vers des commandes de menu couramment utilisés.
 4. The Data Editor Window : Le nom en tête de chaque colonne est le nom de la variable, c'est-à-dire le nom que vous utiliserez pour faire référence à une variable, tandis que, chaque ligne représente une observation (un cas).
 5. The Tabs : Data View et Variable View : Data View est l'endroit où nous inspectons nos données réelles et la Variable View est l'endroit où nous voyons des informations supplémentaires sur nos données.
 6. The Status Bar : en bas de chaque fenêtre, SPSS fournit plusieurs informations telles que : l'état de la commande, l'état du filtre, etc.

3.3 Manipulation de Données dans SPSS

■ Les données SPSS ont trois composants principaux : les observations, les variables et les métadonnées. Lorsque vous recevez des données, vous aurez rarement un problème avec les observations, occasionnellement un problème avec les variables, mais presque toujours un problème avec les métadonnées (metadata).

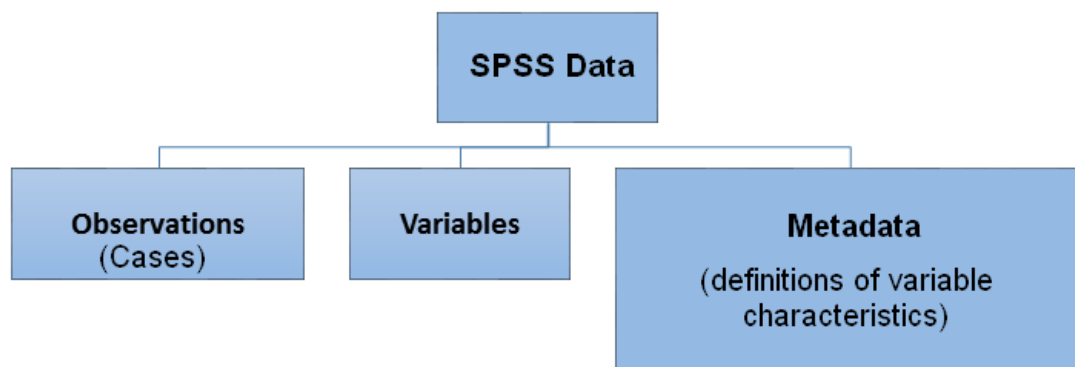


FIGURE 3.2 – Données SPSS.

Organisation des Données Collectées dans SPSS

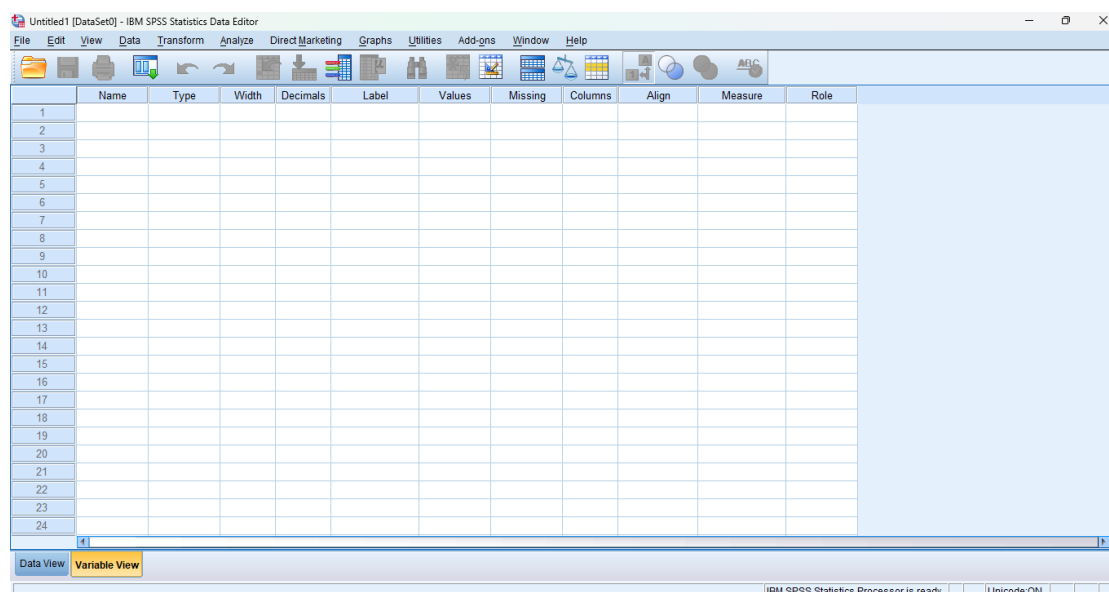


FIGURE 3.3 – Vue des variables.

■ SPSS peut lire des données à partir de divers formats, notamment des bases de données, des fichiers texte, Microsoft Excel, CSV. ... etc. Vous pouvez également taper directement dans SPSS ; et, vous pouvez même coller les données copiées dans SPSS.

3.3.1 Définition des métadonnées (metadata)

■ Dans SPSS, les données sont organisées sous forme d'observations (lignes) et chaque observation est constituée d'un ensemble de variables (colonnes). Tout d'abord, vous définissez les caractéristiques des variables qui composent une observation, puis vous entrez les données dans les variables qui composent le contenu des observations.

■ Pour saisir des données dans SPSS, utilisez l'onglet «Variable View». Comme vous pouvez le voir dans la figure ci-dessous (voir Figure 3.3), les attributs de la variable (tels que the Name, the Type et the Width) sont définis en haut de la fenêtre. Tout ce que vous avez à faire est d'entrer quelque chose dans chaque colonne pour chaque variable.

■ Les 11 caractéristiques sont les seuls nécessaires pour spécifier complètement tous les attributs d'une variable. Lorsque vous ajoutez une nouvelle variable, vous constaterez que des valeurs par défaut raisonnables apparaissent pour la plupart des caractéristiques. Les 11 caractéristiques d'une variable sont :

1. **Name** : Cliquez simplement sur la cellule et saisissez un court descriptif, tel que : Age, revenu, sexe, patient. ... Bien que vous puissiez saisir des noms plus longs ici, il est recommandé de les maintenir courts, car ils seront utilisés dans des listes nommées ainsi que pour des balises d'identification sur des graphiques de données et autres formats où l'espace peut être limité.

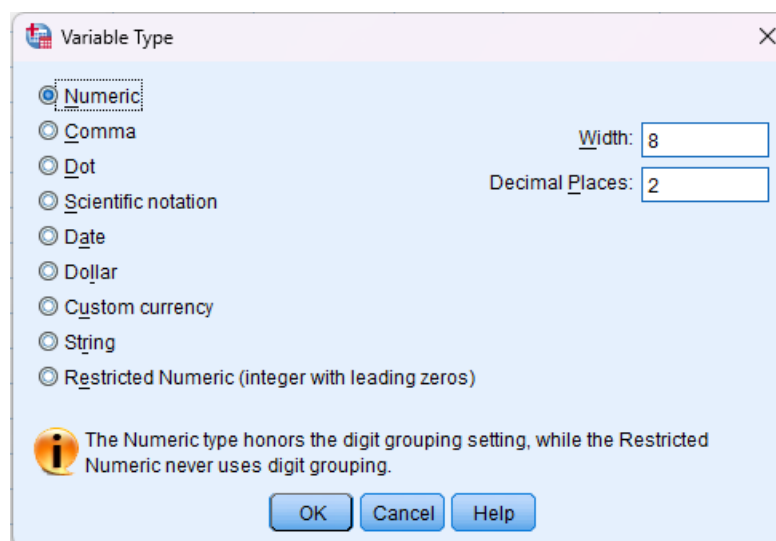


FIGURE 3.4 – Type de variable.

- Les noms de variables doivent commencer par une lettre (A-Z ou a-z).
 - Ils peuvent contenir des lettres, des chiffres (0-9) et des caractères de soulignement (_).
 - Aucun autre caractère spécial, espace ou accent n'est permis.
 - Le nom ne doit pas être un mot-clé réservé par SPSS, comme ALL, BY, AND, etc.
 - Chaque nom de variable dans le même fichier de données doit être unique.
 - Les noms de variables ne peuvent pas débuter par un chiffre.
2. **Type** : La plupart des données que vous saisirez ne seront que des nombres normaux. Cependant, les données telles que la devise doivent être affichées dans un format spécial, et les données telles que les dates nécessitent des procédures de calcul spéciales. Pour ce type de données, il vous suffit de spécifier le type dont vous disposez et SPSS s'occupe des détails pour vous. Pour afficher la boîte de dialogue illustrée à la Figure 3.4, sélectionnez une cellule dans la colonne Type, puis cliquez sur le bouton représenté par trois points qui apparaît.
 3. **Width** : La colonne Width dans la définition d'une variable détermine le nombre de caractères utilisés pour afficher la valeur. Si la valeur à afficher n'est pas assez grande pour remplir l'espace, la sortie sera complétée par des blancs. S'il est plus grand que vous ne l'avez spécifié, il sera reformaté pour s'adapter ou des astérisques seront affichés.
 4. **Decimals** : La colonne Decimals contient le nombre de chiffres qui apparaissent à droite de la virgule décimale lorsque la valeur apparaît à l'écran.
 5. **Label** : Le nom et le libellé ont le même objectif fondamental : ce sont des descripteurs qui identifient la variable. La différence est que le nom est l'identifiant court et l'étiquette est le long. Vous pouvez également ignorer la définition du libellé. Si vous n'avez pas d'étiquette définie pour une variable, SPSS utilisera le nom de variable que vous avez défini pour tout.

6. **Values** : La colonne Valeurs est l'emplacement où vous pouvez assigner des libellés à toutes les valeurs possibles d'une variable. Pour afficher la boîte de dialogue illustrée à la Figure 3.5, sélectionnez une cellule dans la colonne Valeurs, puis cliquez sur le bouton représenté par trois points qui apparaît.

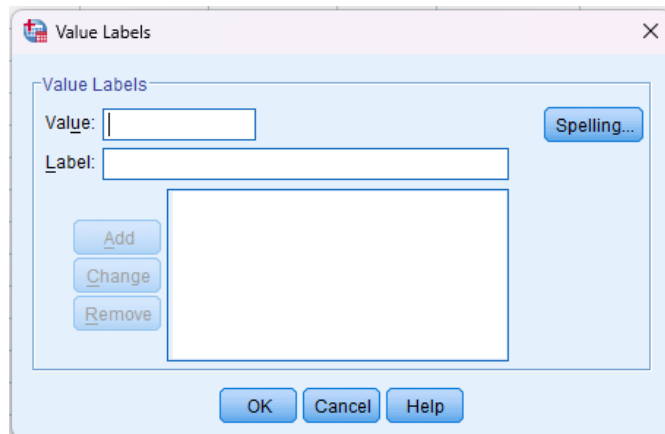


FIGURE 3.5 – Libellés de valeurs.

Normalement, vous faites une seule entrée pour chaque valeur possible qu'une variable peut prendre. Mais, par exemple, pour une variable nommée Sexe, vous pouvez affecter la valeur 1 à l'étiquette Homme et 2 à l'étiquette Femme. Si vous définissez des libellés, votre sortie peut afficher des étiquettes au lieu de valeurs.

Pour définir un libellé pour une valeur, procédez comme suit :

- (a) Dans la zone Value, entrez la valeur.
 - (b) Dans la zone Label, entrez un libellé.
 - (c) Cliquez sur le bouton Add. (La valeur et l'étiquette apparaissent dans le grand bloc de texte.)
 - (d) Pour modifier ou supprimer une définition, sélectionnez-la simplement dans le bloc de texte, apportez vos modifications, puis cliquez sur le bouton Change.
 - (e) Répétez les étapes (a) à (d) si nécessaire.
 - (f) Pour enregistrer les étiquettes de valeur et fermer la boîte de dialogue, cliquez sur OK.
7. **Missing** : Vous pouvez spécifier des codes pour les données manquantes. Pour afficher la boîte de dialogue illustrée à la Figure 3.6, sélectionnez une cellule dans la colonne Manquant. Un bouton représenté par trois points apparaîtra, cliquez dessus pour ouvrir la boîte de dialogue.

3.4. Saisie et affichage des éléments de données dans l'onglet «Data View»

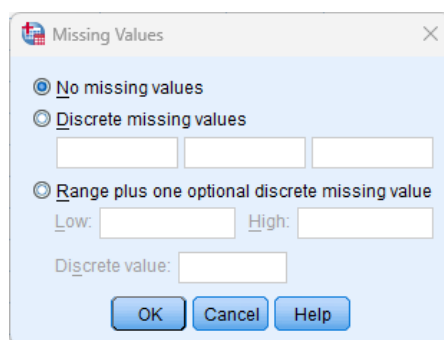


FIGURE 3.6 – Valeurs manquantes.

Par exemple, supposons que vous saisissiez des réponses à des questions, et que l'une des questions soit : " Combien d'enfants avez-vous?" La réponse normale à cette question est un nombre, vous définissez donc le type de variable comme un nombre. Vous pouvez définir -99 comme la valeur saisie lorsque la réponse est « Je ne me souviens pas » et -98 peut être utilisé lorsque la réponse est « Je ne peux pas dire ».

8. **Columns** : Dans l'attribut Colonnes, vous pouvez spécifier la largeur de la colonne à utiliser pour saisir les données.
9. **Align** : La colonne Align détermine la position des données dans son espace alloué.
10. **Measure** : Votre valeur pour l'attribut Measure spécifie le niveau de mesure de votre variable. Voici le niveau des options de mesure dans SPSS :
 - **Nominal** : Une valeur qui spécifie une catégorie ou un type de chose. Vous pouvez avoir 0 pour Désapprouver et 1 pour Approuver. Ou vous pouvez utiliser 1 pour signifier Femme et 2 pour signifier Homme.
 - **Ordinal** : Une valeur qui spécifie la position (ordre) de quelque chose dans une liste. Par exemple, premier, deuxième et troisième sont des nombres ordinaux.
 - **Scale** : Un nombre qui spécifie une magnitude. L'échelle peut être la distance, le poids, l'âge ou un décompte de quelque chose.
11. **Role** : Vous n'avez pas besoin de vous soucier de la colonne Rôle pour le moment.

3.4 Saisie et affichage des éléments de données dans l'onglet «Data View»

Après avoir défini les variables, vous pouvez commencer à saisir les données. Cliquez sur l'onglet **Data View** de la fenêtre de l'éditeur de données. En haut des colonnes, vous voyez les noms des variables. La saisie de données dans l'une de ces cellules est simple : il vous suffit de cliquer sur la cellule et de commencer à taper.

3.5 Sauvegarde des données SPSS

Tout ce que vous avez à faire est de choisir **File → Save or Save As (Ctrl + S)** sélectionnez votre type de fichier, puis entrez un nom de fichier. The SPSS Statistics File Format is «.sav» (voir Figure 3.7).

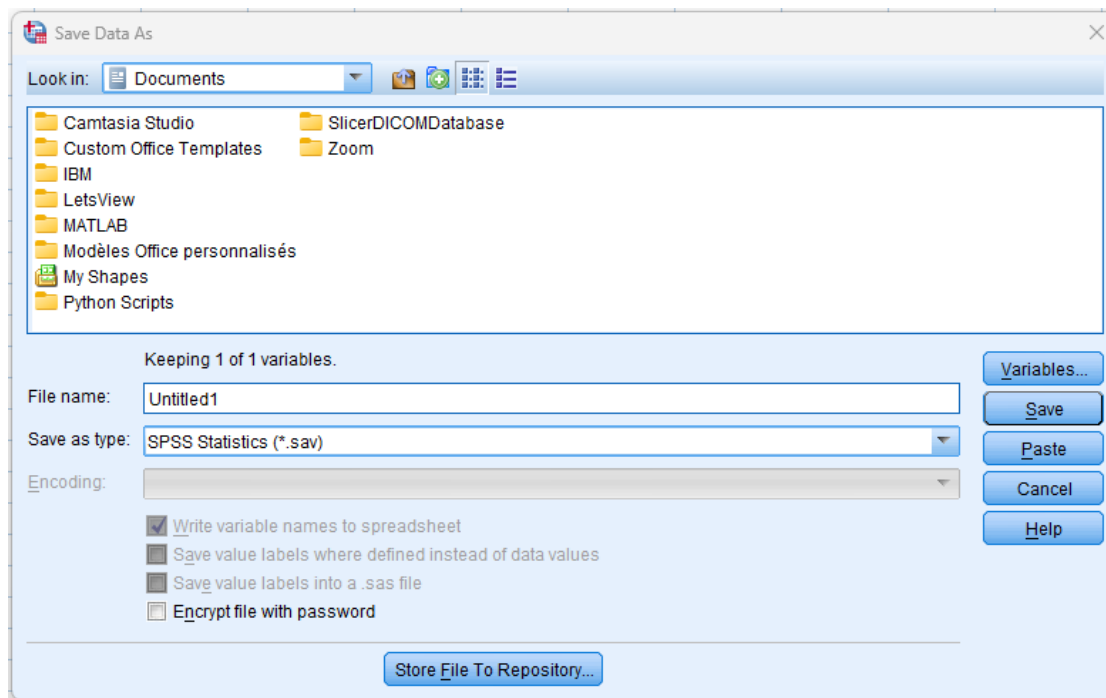


FIGURE 3.7 – Enregistrer les données dans SPSS.

Vous avez le choix entre de nombreux types de fichiers : formats de texte brut, formats de feuille de calcul Excel, formats Lotus, formats dBase, formats SAS, format SYLK, format Portable et 18 formats Stata.

3.6 Ouverture de fichiers de données SPSS

Pour ouvrir un fichier de données, choisissez **File → Open → Data (Ctrl + O)** et sélectionnez le fichier à charger. Lorsque vous le faites, les noms de variables et les données sont chargés dans SPSS.

3.7 Transfert de données d'un fichier Excel vers SPSS

Pour ouvrir votre fichier Excel dans SPSS :

1. **File → Open → Data (Ctrl + O)**, dans le menu SPSS.
2. Sélectionnez le type de fichier que vous souhaitez ouvrir, Excel (*.xls, *.xlsx, *.xlsm).

3.7. Transfert de données d'un fichier Excel vers SPSS

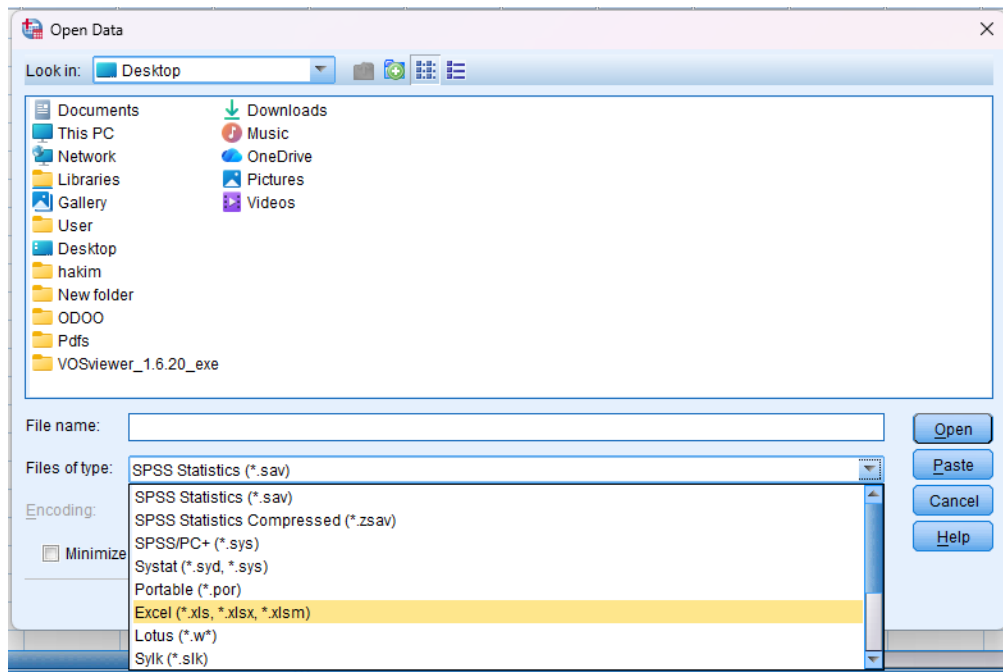


FIGURE 3.8 – Ouvrir les données dans SPSS.

3. Dans la boîte de dialogue "Open Data", sélectionnez le fichier que vous souhaitez ouvrir (voir Figure 3.8)..
4. Cliquez sur **Open**.
5. La boîte de dialogue suivante apparaît (Figure 3.9).

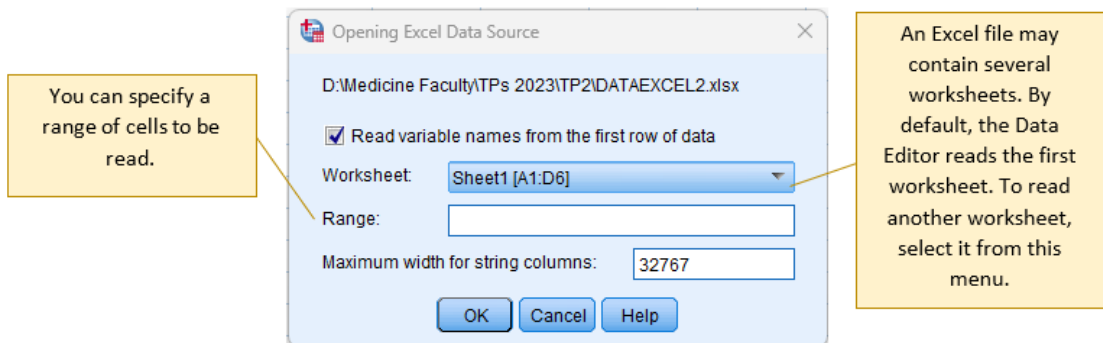


FIGURE 3.9 – Ouvrir la source de données Excel.

6. Cliquez sur **OK**.

Note :

■ Veuillez trouver le fichier de données SPSS de ce chapitre à partir de ce lien : <https://aboulesnane.net/wp-content/datafiles/DataSPSS3.sav>

3.8 Conclusion

L'organisation des données collectées dans SPSS est importante pour une gestion précise et efficace des données, la reproductibilité des résultats de recherche, la sécurité et la confidentialité des données de recherche et la collaboration entre les chercheurs. SPSS offre une interface conviviale et des options personnalisables pour organiser les données, telles que le recodage, les transformations de données et le nettoyage des données. SPSS permet une gestion des données rapide et efficace, comme la possibilité d'importer et d'exporter des données à partir de différentes sources, de fusionner des ensembles de données et de nettoyer des données. Dans l'ensemble, l'organisation des données collectées dans SPSS est essentielle pour mener des recherches fiables et valides en sciences sociales et biomédicales.

Chapitre 4

Prétraitement des données

4.1 Introduction

Le prétraitement des données est une étape cruciale de l'analyse des données qui consiste à nettoyer, transformer et préparer les données brutes avant l'analyse. Il permet d'améliorer la qualité, la précision et la compatibilité des données avec les techniques d'analyse et les outils logiciels. Le prétraitement des données peut également réduire le temps et les efforts nécessaires à l'analyse des données en simplifiant le processus de nettoyage et de transformation des données. Il permet aussi de créer des visualisations et des résumés facilitant la compréhension et l'interprétation des données.

4.2 Navigation dans le SPSS Viewer

Lorsque vous exécutez une analyse, produisez un graphique ou même enregistrez un fichier, **la fenêtre SPSS Statistics Viewer** apparaît automatiquement pour afficher ce que vous avez créé (Figure 4.1).

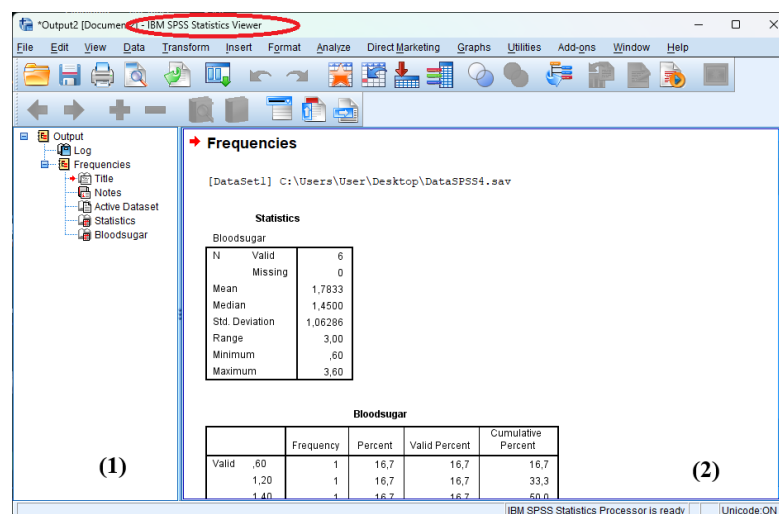


FIGURE 4.1 – IBM SPSS statistics viewer.


Prétraitement des données


■ Le volet d'organisation (outline pane) (1), à gauche, contient un aperçu de toutes les informations stockées dans le Viewer.

■ Le volet de contenu (content pane) (2), à droite, contient les tableaux statistiques, graphiques et textes.

■ Pour masquer un objet dans le volet de contenu, vous devez :

✓ Sélectionner l'objet (tableau, graphique, etc.) dans le volet d'organisation ou dans le volet de contenu.

✓ Cliquer sur l'icône livre dans la barre d'outils .

✓ Pour afficher à nouveau cet objet, cliquez sur l'icône .

Masquer des éléments sans les supprimer permet à l'utilisateur de se concentrer plus facilement sur les résultats d'intérêt tout en conservant l'ensemble des résultats.

4.3 Travailler avec les données dans SPSS

| | Patient | Weight | Bloodsugar | Sexe | Bloodgroup |
|---|---------|--------|------------|------|------------|
| 1 | P1 | 53,00 | 1,20 | 1 | 1 |
| 2 | P2 | -99,00 | 1,50 | 1 | 3 |
| 3 | P3 | 88,00 | 1,40 | 2 | 2 |
| 4 | P4 | 45,00 | ,60 | 1 | 4 |
| 5 | P5 | 90,00 | 2,40 | 2 | 2 |
| 6 | P6 | 175,00 | 3,60 | 2 | 1 |

FIGURE 4.2 – SPSS data file.

1. Télécharger le fichier de données SPSS appelé "DataSPSS4.sav" à partir de : <https://aboulesnane.net/wp-content/datafiles/DataSPSS4.sav>
2. Les données contiennent cinq variables nommées : Patient, Weight, Bloodsugar, Sexe et Bloodgroup (voir Figure 4.2).
 - (a) La variable "**Patient**" est une variable de type chaîne.
 - (b) La variable "**Weight**" est une variable quantitative continue de type numérique. (Pour la variable Weight, les valeurs manquantes sont représentées par le nombre -99 (voir Figure 3.6)).
 - (c) La variable "**Bloodsugar**" est une variable quantitative continue de type numérique.
 - (d) Les valeurs possibles pour la variable qualitative "**Sexe**" sont : 1=Homme et 2=Femme.
 - (e) Les valeurs possibles pour la variable qualitative "**Bloodgroup**" sont : 1=AB, 2=A, 3=B et 4=O.

4.4 Remplacement des valeurs manquantes

Afin de remplacer les valeurs manquantes par des valeurs acceptables, nous utiliserons des mesures de tendance centrale telles que la moyenne, la médiane, le mode, etc. Par exemple, dans la variable **Weight** de nos données, nous avons une valeur manquante représentée par -99. Afin de ne pas perturber la distribution de nos données, nous pouvons remplacer cette valeur par la moyenne de la série.

Pour ce faire :

1. Choisissez **Transform → Replace Missing Values** : la boîte de dialogue **Replace Missing Values**... apparaît.
2. Déplacez la variable **Weight** vers la zone "New Variables".
3. Dans le menu **Method** (voir Figure 4.3), vous pouvez sélectionner la meilleure méthode pour remplacer les valeurs manquantes (Linear Interpolation, Measures of Central Tendency, etc.). Dans notre cas, nous utiliserons **Series Mean**.

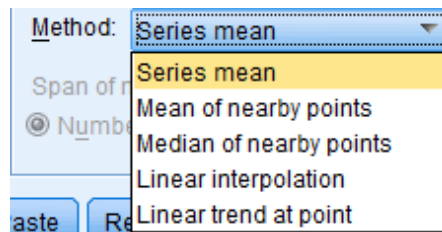


FIGURE 4.3 – Method Menu.

4. Cliquez sur **OK** (voir Figure 4.4).

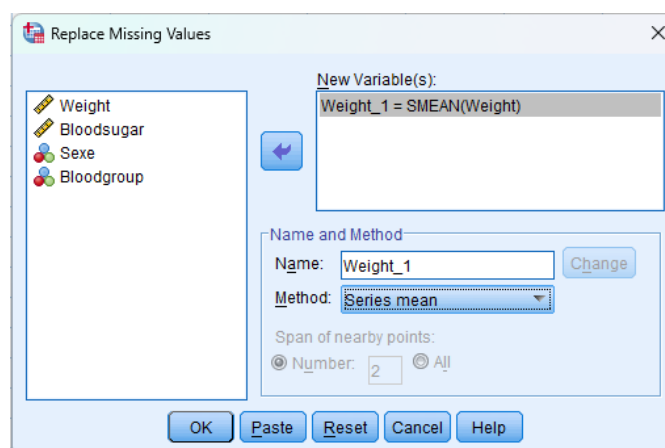


FIGURE 4.4 – Replacing Missing Values.

En conséquence, une nouvelle colonne apparaît avec le nom "**Weight_1**", où la valeur -99 est remplacée par la valeur moyenne 90.2 (voir Figure 4.5).

Prétraitement des données

| | Patient | Weight | Bloodsugar | Sexe | Bloodgroup | Weight_1 |
|---|---------|--------|------------|------|------------|----------|
| 1 | P1 | 53,0 | 1,2 | 1 | 1 | 53,0 |
| 2 | P2 | -99,0 | 1,5 | 1 | 3 | 90,2 |
| 3 | P3 | 88,0 | 1,4 | 2 | 2 | 88,0 |
| 4 | P4 | 45,0 | ,6 | 1 | 4 | 45,0 |
| 5 | P5 | 90,0 | 2,4 | 2 | 2 | 90,0 |
| 6 | P6 | 175,0 | 3,6 | 2 | 1 | 175,0 |

FIGURE 4.5 – Result after replacement.

4.5 Tri des observations

1. Choisissez **Data** → **Sort Cases** : la boîte de dialogue **Sort Cases** apparaît (Figure 4.6).

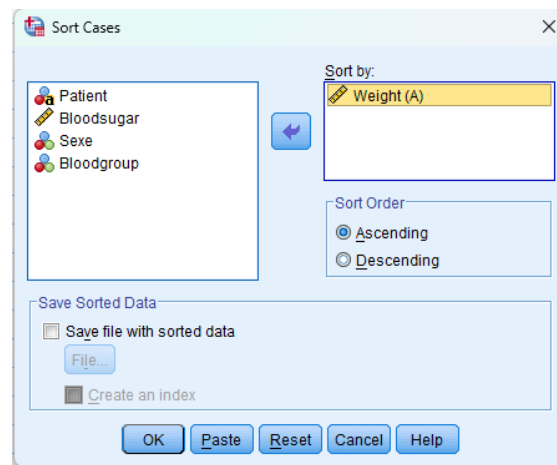


FIGURE 4.6 – Sort Cases.

2. Sélectionnez la variable **Weight**.
3. Cliquez sur **OK** pour trier les données.

■ L'ordre d'affichage des données n'affecte jamais l'analyse statistique. Le tri sert uniquement à mieux visualiser les informations dans le Data Editor.

■ Il est possible de trier les données soit en ordre croissant soit en ordre décroissant. Par défaut, les données sont triées en **ordre croissant**.

■ Le tri peut être basé sur un ou plusieurs critères.

Remarque :

■ Pour annuler le tri :

✓ Il n'est pas possible d'annuler directement le tri des données.

✓ Une astuce consiste à trier à nouveau les données selon la variable "**Patient**" (répétez la même procédure décrite ci-dessus en sélectionnant la variable **Patient** au lieu de **Weight**).

4.6 Recodage des variables

En recodant les variables, nous pouvons regrouper un ensemble de valeurs en certaines catégories prédéfinies selon le type de données. Ce processus permet d'économiser du temps et des efforts en restructurant les données collectées de manière plus efficace.

Par exemple, supposons que nous souhaitons recoder la variable **Weight**, où nous regrouperons les valeurs comme suit :

1. Featherweight : [0-50]
2. Middleweight :]50-90]
3. Heavyweight :]90-120]
4. Super Heavyweight : >120

Voici comment procéder dans SPSS :

1. Choisissez **Transform → Recode into Different Variables** : la boîte de dialogue **Recode into Different Variables** apparaît.
2. Cliquez sur le bouton flèche pour déplacer la variable **Weight** vers la zone de travail à droite.
3. Nommez la nouvelle variable de sortie dans la zone à droite par **WeightCat** → Cliquez sur le bouton **Change** pour enregistrer le nouveau nom de la variable, comme illustré dans la Figure 4.7.

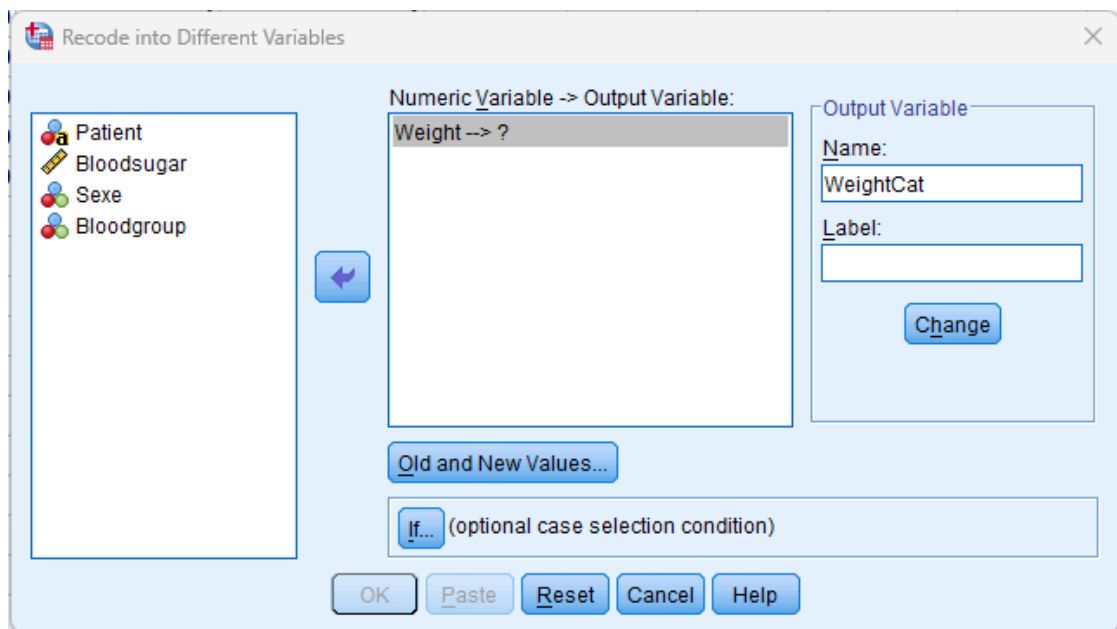


FIGURE 4.7 – Dialog box : Recode into Different Variables.

4. Cliquez sur le bouton **Old and New Values**.
5. Pour les catégories 1, 2 et 3 : sélectionnez le bouton radio **Range** → Entrez les valeurs Min et Max → À côté du bouton radio **Value** : entrez le numéro de la catégorie → Cliquez sur le bouton **Add** (voir Figure 4.8).

Prétraitement des données

- Pour la catégorie 4 : sélectionnez le bouton radio **Range, value through HIGHEST** → Entrez la valeur Min (c.-à-d. 120) → À côté du bouton radio **Value** : entrez le numéro de la catégorie (c.-à-d. 4) → Cliquez sur le bouton **Add** (voir Figure 4.8).

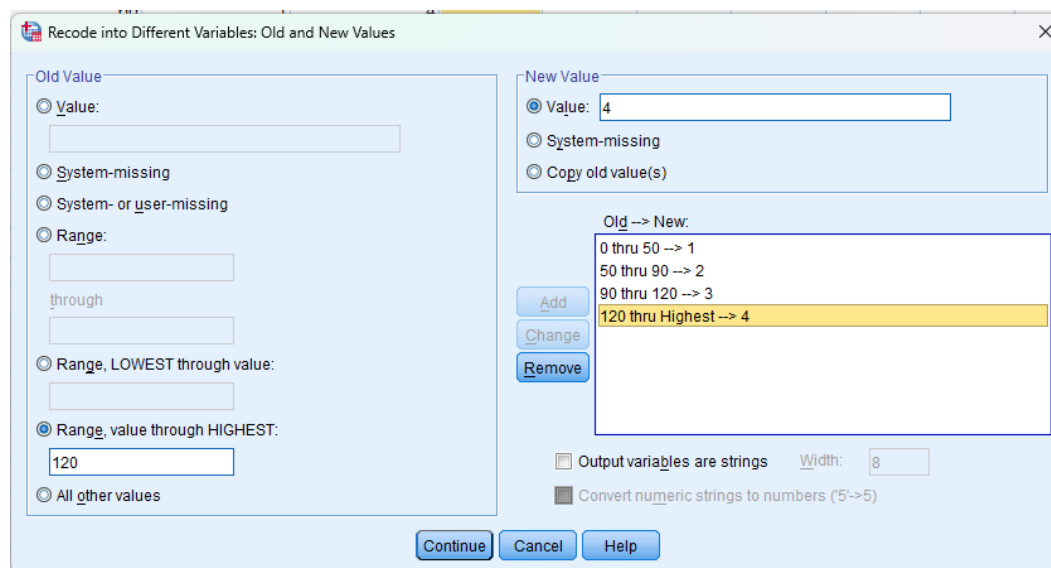


FIGURE 4.8 – Dialog box : Old and New Values.

- Cliquez sur **Continue**, puis sur **OK**. La Figure 4.9 ci-dessous montre le résultat final.

| | Patient | Weight | Bloodsugar | Sexe | Bloodgroup | WeightCat |
|---|---------|--------|------------|------|------------|-----------|
| 1 | P1 | 53,00 | 1,20 | 1 | 1 | 2,00 |
| 2 | P2 | 90,20 | 1,50 | 1 | 3 | 3,00 |
| 3 | P3 | 88,00 | 1,40 | 2 | 2 | 2,00 |
| 4 | P4 | 45,00 | ,60 | 1 | 4 | 1,00 |
| 5 | P5 | 90,00 | 2,40 | 2 | 2 | 2,00 |
| 6 | P6 | 175,00 | 3,60 | 2 | 1 | 4,00 |

FIGURE 4.9 – Result after Recoding.

Il est utile d'indiquer à toute personne consultant votre sortie ce que représentent ces valeurs recodées. Pour cela, cliquez sur l'onglet **Variable View** en bas de la feuille de calcul, puis cliquez dans le champ **Values** (pour la nouvelle variable **WeightCat**) et ajoutez les étiquettes de valeurs comme décrit précédemment (c.-à-d. 1=Featherweight, 2=Middleweight, 3=Heavyweight, 4=Super Heavyweight).

4.7 Suppression d'une variable ou d'une observation

► Supposons que nous souhaitons supprimer la nouvelle variable **WeightCat**. Pour supprimer une variable dans Data View, cliquez sur le nom de la variable

et appuyez sur la touche **Delete** du clavier, ou cliquez avec le bouton droit sur le nom de la variable puis cliquez sur **Clear**.

► Pour supprimer une observation (une ligne complète de données), suivez les étapes suivantes : cliquez sur le numéro de l'observation à gauche (toute la ligne sera sélectionnée), appuyez sur **Delete** sur le clavier, ou cliquez avec le bouton droit sur le numéro de l'observation puis cliquez sur **Clear**.

4.8 Fractionnement des données

Dans certaines études comparatives, nous divisons les données en plusieurs groupes selon certains critères, puis nous effectuons nos analyses sur chaque sous-groupe séparément. Le fractionnement des données permet de réaliser ce processus.

1. Choisissez **Data → Split File** : la boîte de dialogue **Split File** apparaît.
2. Sélectionnez le bouton radio **Compare groups**.
3. Choisissez **Sexe** comme variable de comparaison puis cliquez sur **OK** (voir Figure 4.10).

Remarque : après ce processus, nous constatons qu'il n'y a pas de changement significatif, sauf que les données ont été réorganisées selon le sexe des patients.

4. Choisissez **Analyze → Descriptive Statistics → Frequencies**.
5. Choisissez **Bloodgroup** et placez-le dans la zone Variable(s).
6. Cliquez sur **OK** : le résultat obtenu, illustré dans la Figure 4.11, est regroupé par **Sexe**.

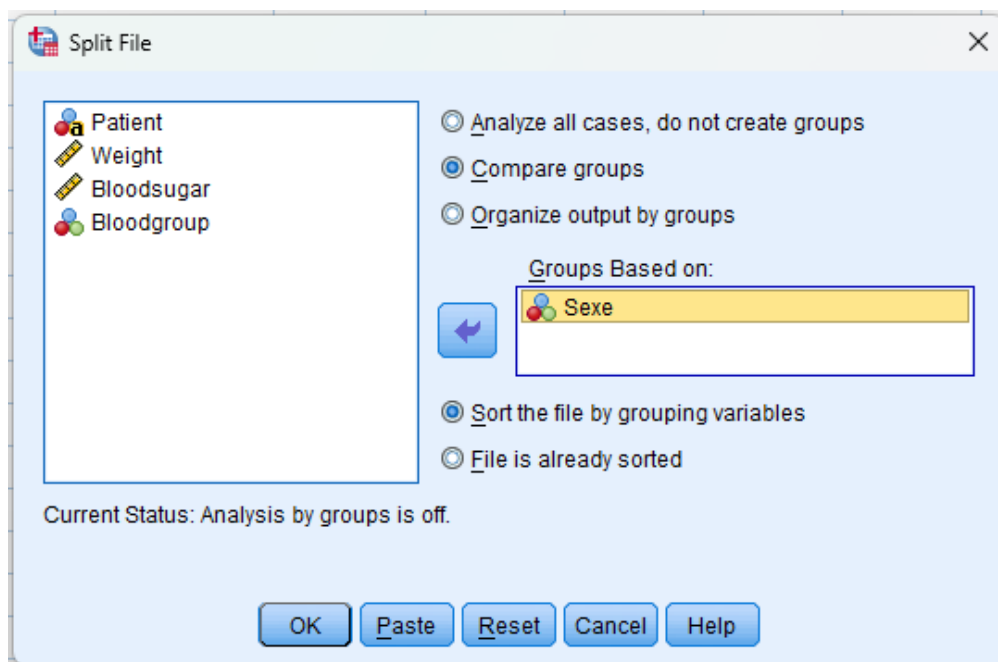


FIGURE 4.10 – Split file.

➔ **Frequencies**

Statistics

Bloodgroup

| | | | |
|-------|---|---------|---|
| Homme | N | Valid | 3 |
| | | Missing | 0 |
| Femme | N | Valid | 3 |
| | | Missing | 0 |

Bloodgroup

| Sexe | | | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------|-------|-----------|---------|---------------|--------------------|
| Homme | Valid | AB | 1 | 33,3 | 33,3 | 33,3 |
| | | B | 1 | 33,3 | 33,3 | 66,7 |
| | | O | 1 | 33,3 | 33,3 | 100,0 |
| | | Total | 3 | 100,0 | 100,0 | |
| Femme | Valid | AB | 1 | 33,3 | 33,3 | 33,3 |
| | | A | 2 | 66,7 | 66,7 | 100,0 |
| | | Total | 3 | 100,0 | 100,0 | |

FIGURE 4.11 – Frequencies result.

Remarques :

- Le fractionnement peut être basé sur un ou plusieurs critères.
- Pour annuler le fractionnement :
 1. Choisissez **Data → Split File**.
 2. Sélectionnez le bouton radio **Analyze all cases, do not create groups**, puis cliquez sur le bouton **OK** (ou appuyez sur **Reset → OK**).

4.9 Sélection des données

4.9.1 Condition logique simple

1. Choisissez **Data → Select Cases** : la boîte de dialogue **Select Cases** apparaît, comme illustré dans la Figure 4.12.
2. Sélectionnez le bouton radio **If condition is satisfied**, puis cliquez sur le bouton **If...** : vous pouvez maintenant spécifier les critères de sélection (voir Figure 4.13).
3. Déplacez la variable **Sexe** de la liste à gauche vers la zone d'expression (en haut à gauche) : vous pouvez déplacer la variable soit en la faisant glisser, soit en la sélectionnant puis en cliquant sur le bouton flèche.
4. Utilisez votre clavier ou le pavé numérique à l'écran pour saisir **=2** dans la zone d'expression.

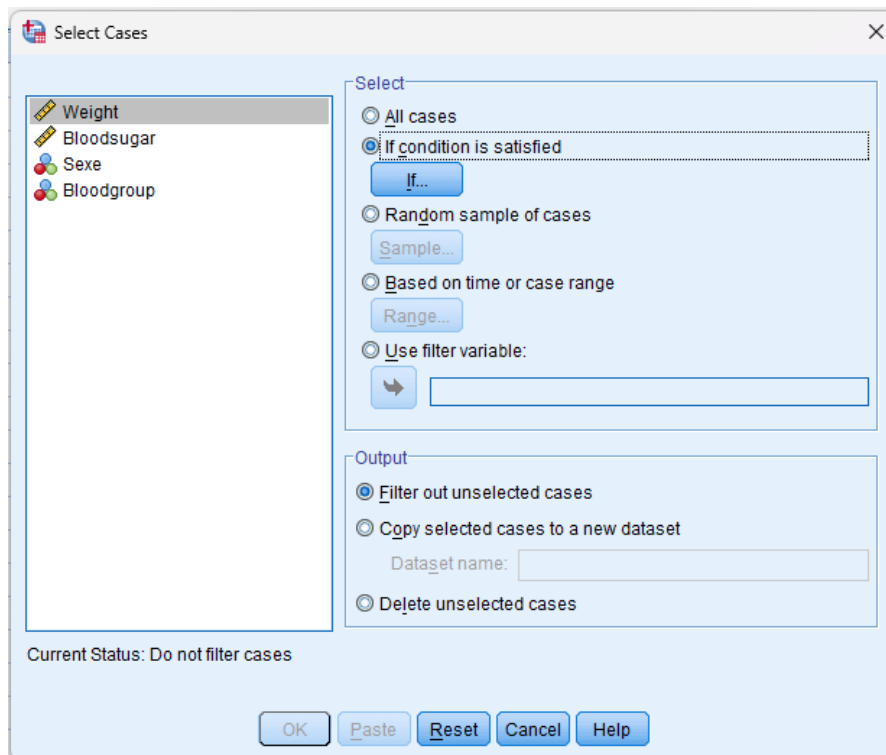


FIGURE 4.12 – Select Cases.

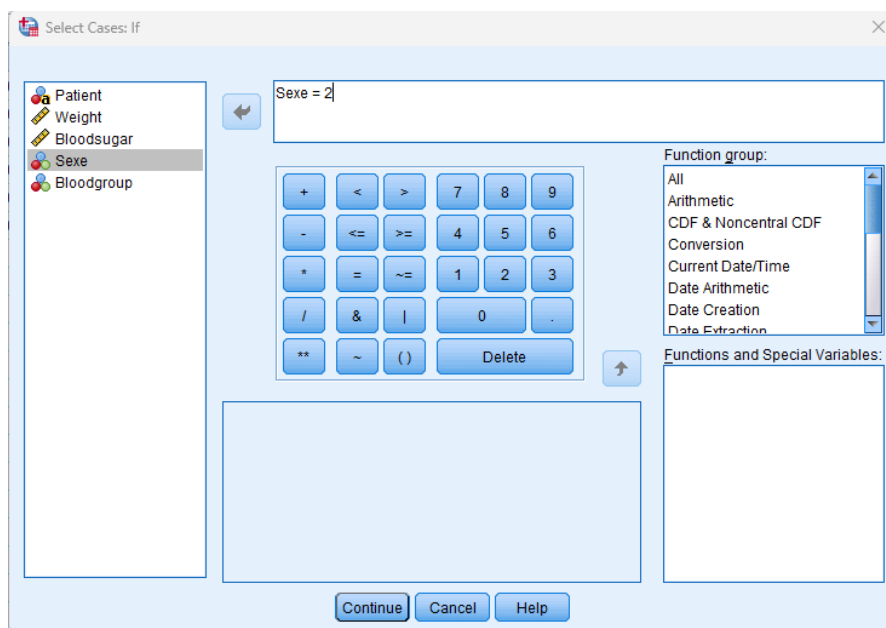


FIGURE 4.13 – Conditional expression.

5. Cliquez sur **Continue**, puis sur **OK**. La figure ci-dessous montre le résultat final. Les barres diagonales sur certains identifiants de lignes (dans la première colonne) indiquent que les patients (**Homme**) sont ignorés (pour le moment) et que seuls les patients femmes (**Femme**) sont analysés.

La variable *filter_\$* est également créée et contient les valeurs 0 et 1 cor-

Prétraitement des données

| | Patient | Weight | Bloodsugar | Sexe | Bloodgroup | filter_\$ |
|--------------|---------|--------|------------|------|------------|-----------|
| 1 | P1 | 53,00 | 1,20 | 1 | 1 | 0 |
| 2 | P2 | 90,20 | 1,50 | 1 | 3 | 0 |
| 3 | P3 | 88,00 | 1,40 | 2 | 2 | 1 |
| 4 | P4 | 45,00 | ,60 | 1 | 4 | 0 |
| 5 | P5 | 90,00 | 2,40 | 2 | 2 | 1 |
| 6 | P6 | 175,00 | 3,60 | 2 | 1 | 1 |

FIGURE 4.14 – Result after selection.

respondant respectivement aux cas non sélectionnés et aux cas sélectionnés (voir Figure 4.14).

Pour observer les effets de la sélection :

- Choisissez **Analyze → Descriptive Statistics → Frequencies**.
- Choisissez **Bloodgroup** et placez-le dans la zone Variable(s).
- Cliquez sur **OK** : le résultat obtenu, illustré dans la figure ci-dessous (Figure 4.15), montre que SPSS affiche uniquement les résultats des groupes sanguins pour les patientes.

→ Frequencies

Statistics

Bloodgroup

| | | |
|---|---------|---|
| N | Valid | 3 |
| | Missing | 0 |

Bloodgroup

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------|-----------|---------|---------------|--------------------|
| Valid | AB | 1 | 33,3 | 33,3 | 33,3 |
| | A | 2 | 66,7 | 66,7 | 100,0 |
| | Total | 3 | 100,0 | 100,0 | |

FIGURE 4.15 – Frequencies result.

4.9.2 Condition logique complexe

Les conditions logiques sont des concepts fondamentaux en logique et en programmation. Elles permettent de prendre des décisions basées sur l'évaluation d'expressions qui produisent soit vrai (True : T), soit faux (False : F).

| Conjonction | | | Disjonction | | |
|---|----------|--------------|--|----------|--------------|
| Table de vérité de la conjonction (& : AND) (T : True, F : False) | | | Table de vérité de la disjonction (: OR) (T : True, F : False) | | |
| P | Q | P ∧ Q | P | Q | P ∨ Q |
| T | T | T | T | T | T |
| T | F | F | T | F | T |
| F | T | F | F | T | T |
| F | F | F | F | F | F |

TABLE 4.1 – Conjonction et disjonction logiques.

1. Opérateurs logiques

Les opérateurs logiques suivants permettent de combiner ou de modifier des expressions :

- **AND** (\wedge , &)
 - Une expression est vraie si **toutes** les conditions sont vraies.
 - Exemple : $A \wedge B$ est vrai si A **et** B sont vrais.
- **OR** (\vee , |)
 - Une expression est vraie si **au moins une** condition est vraie.
 - Exemple : $A \vee B$ est vrai si A ou B (ou les deux) sont vrais.
- **NOT** (\neg , ~)
 - Inverse une condition : vrai devient faux et faux devient vrai.
 - Exemple : $\neg A$ est vrai si A est faux.

2. Propriétés des opérateurs logiques

- **Propriété commutative**

$$A \wedge B = B \wedge A$$

$$A \vee B = B \vee A$$

Prétraitement des données

— Propriété associative

$$(A \wedge B) \wedge C = A \wedge (B \wedge C)$$

$$(A \vee B) \vee C = A \vee (B \vee C)$$

— Propriété distributive

$$A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$$

$$A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$$

► Essayons maintenant de sélectionner les hommes ayant un poids supérieur ou égal à 50 ou les patients ayant le groupe sanguin A.

► Entrez la condition suivante :

(Sexe = 1 & Weight >= 50) | Bloodgroup = 2

(voir Figure 4.16). Les parenthèses entre les conditions sont très importantes.

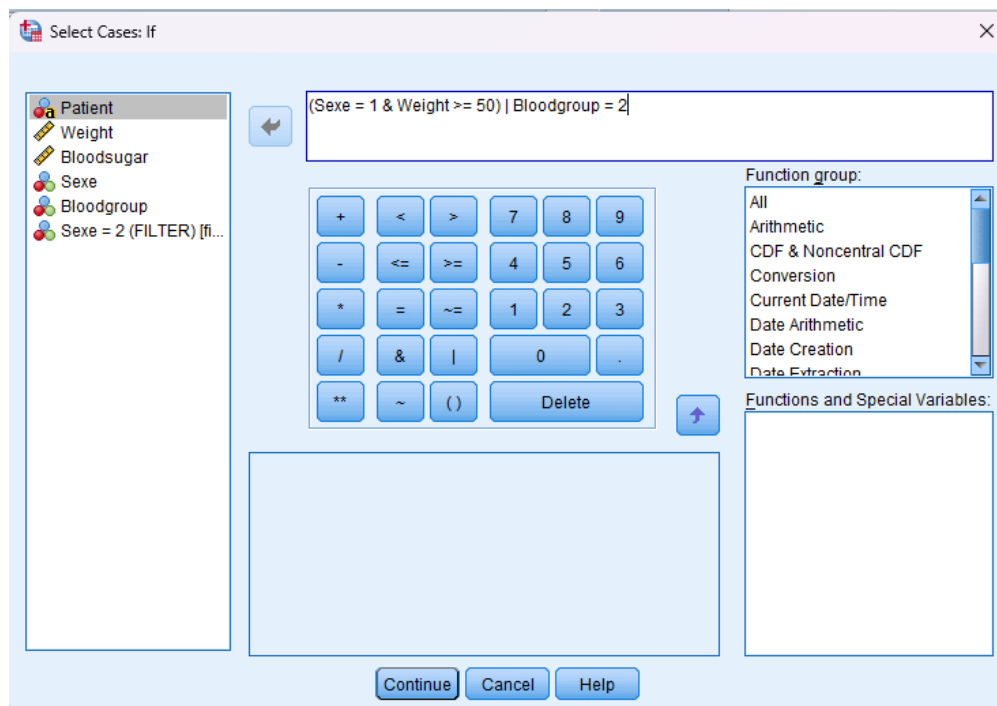


FIGURE 4.16 – Conditional expression.

Le résultat de la sélection sera :

| | Patient | Weight | Bloodsugar | Sexe | Bloodgroup | filter_\$ |
|--------------|---------------|-------------------|-----------------|------------------|---------------|-------------------------|
| 1 | P1 | 53,00 | 1,20 | Homme | AB | Selected |
| 2 | P2 | 90,20 | 1,50 | Homme | B | Selected |
| 3 | P3 | 88,00 | 1,40 | Femme | A | Selected |
| 4 | P4 | 45,00 | ,60 | Homme | O | Not Selected |
| 5 | P5 | 90,00 | 2,40 | Femme | A | Selected |
| 6 | P6 | 175,00 | 3,60 | Femme | AB | Not Selected |

FIGURE 4.17 – Result after selection.

Remarques

■ Pour annuler la sélection :

1. Choisissez **Data → Select Cases**.
2. Sélectionnez le bouton radio **All cases**, puis cliquez sur **OK** (ou appuyez sur **Reset → OK**).

■ La sélection peut être basée sur un ou plusieurs critères.

4.10 Conclusion

Le prétraitement des données constitue une étape fondamentale de l'analyse des données, car il influence directement la fiabilité et la validité des résultats obtenus. Dans ce chapitre, nous avons exploré plusieurs opérations essentielles de prétraitement dans SPSS, notamment la gestion des valeurs manquantes, le tri et le recodage des variables, la sélection et le fractionnement des données ainsi que l'application de conditions logiques.

Ces techniques permettent aux analystes de nettoyer, structurer et organiser efficacement les données avant d'effectuer des analyses statistiques. En utilisant correctement les outils de prétraitement de SPSS, les utilisateurs peuvent améliorer la qualité des données, assurer une meilleure compatibilité avec les méthodes d'analyse et extraire des informations plus précises et pertinentes à partir de leurs données.

Chapitre 5

Analyse des données

5.1 Introduction

Dans l'analyse des données biomédicales, les variables qualitatives et continues sont souvent analysées conjointement afin de mieux comprendre des phénomènes complexes, tels que les facteurs de risque d'une maladie particulière ou l'efficacité d'un traitement médical. Il est donc important d'avoir une compréhension approfondie des méthodes statistiques applicables à ces deux types de variables afin d'interpréter correctement les données biomédicales et de prendre des décisions éclairées.

5.2 Collecte des données dans SPSS

| | Patient | Weight | Bloodsugar | Sexe | Bloodgroup |
|---|---------|--------|------------|------|------------|
| 1 | P1 | 63,00 | 1,20 | 1 | 1 |
| 2 | P2 | 90,20 | 1,70 | 1 | 3 |
| 3 | P3 | 88,00 | 1,40 | 2 | 2 |
| 4 | P4 | 45,00 | ,60 | 1 | 4 |
| 5 | P5 | 90,00 | 2,40 | 2 | 2 |
| 6 | P6 | 175,00 | 3,60 | 2 | 1 |
| 7 | P7 | 63,00 | 2,10 | 1 | 2 |

FIGURE 5.1 – SPSS Data File.

1. Télécharger le fichier de données SPSS appelé "**DataSPSS5.sav**" à partir de : <https://aboulesnane.net/wp-content/datafiles/DataSPSS5.sav>
2. Les données contiennent cinq variables nommées : Patient, Weight, Bloodsugar, Sexe et Bloodgroup (voir Figure 5.1).
 - (a) La variable "**Patient**" est une variable de type **String**.
 - (b) La variable "**Weight**" est une variable quantitative continue de type Numeric.

5.3. Prétraitement des données dans SPSS

- (c) La variable "**Bloodsugar**" est une variable quantitative continue de type Numeric.
- (d) Les valeurs possibles pour la variable qualitative "**Sexe**" sont : 1=Homme et 2=Femme.
- (e) Les valeurs possibles pour la variable qualitative "**Bloodgroup**" sont : 1=AB, 2=A, 3=B et 4=O.

5.3 Prétraitement des données dans SPSS

■ Les données des patients ont été triées selon leurs valeurs de glycémie (**Bloodsugar**) et de poids (**Weight**) (voir section 4.5).

Le résultat du tri est :

| | Patient | Weight | Bloodsugar | Sexe | Bloodgroup |
|---|---------|--------|------------|------|------------|
| 1 | P4 | 45,00 | ,60 | 1 | 4 |
| 2 | P7 | 60,00 | 1,20 | 1 | 2 |
| 3 | P1 | 63,00 | 1,20 | 1 | 1 |
| 4 | P3 | 88,00 | 1,40 | 2 | 2 |
| 5 | P2 | 90,20 | 1,70 | 1 | 3 |
| 6 | P5 | 90,00 | 2,40 | 2 | 2 |
| 7 | P6 | 175,00 | 3,60 | 2 | 1 |

FIGURE 5.2 – Result after sorting.

■ Les données des patients ont été fractionnées selon leurs valeurs de groupe sanguin (**Bloodgroup**) (voir section 4.8).

Le résultat du fractionnement est :

| | Patient | Weight | Bloodsugar | Sexe | Bloodgroup |
|---|---------|--------|------------|------|------------|
| 1 | P1 | 63,00 | 1,20 | 1 | 1 |
| 2 | P6 | 175,00 | 3,60 | 2 | 1 |
| 3 | P7 | 60,00 | 1,20 | 1 | 2 |
| 4 | P3 | 88,00 | 1,40 | 2 | 2 |
| 5 | P5 | 90,00 | 2,40 | 2 | 2 |
| 6 | P2 | 90,20 | 1,70 | 1 | 3 |
| 7 | P4 | 45,00 | ,60 | 1 | 4 |

FIGURE 5.3 – Result after splitting.

5.4 Utilisation des statistiques descriptives

5.4.1 Fréquences pour les variables catégorielles (qualitatives)

■ La technique la plus courante pour décrire les données catégorielles — niveaux de mesure nominal et ordinal — consiste à demander **une table de fréquences**, qui fournit un résumé indiquant le nombre et le pourcentage de cas appartenant à chaque catégorie d'une variable. Les utilisateurs peuvent également demander des statistiques supplémentaires telles que le mode ou la médiane, entre autres.

■ Voici comment exécuter la procédure Frequencies afin de créer une table de fréquences permettant d'obtenir des statistiques descriptives pour les variables qualitatives :

1. Choisissez **Analyze → Descriptive Statistics → Frequencies** : la boîte de dialogue Frequencies apparaît. Dans cet exemple, et sur la base du **fractionnement précédent des données**, vous souhaitez étudier la distribution de la variable **Sexe** (Homme, Femme) pour chaque valeur de **Bloodgroup** (AB, A, B, O).
2. Sélectionnez la variable **Sexe** et placez-la dans la zone Variable(s), comme illustré dans la Figure 5.4.

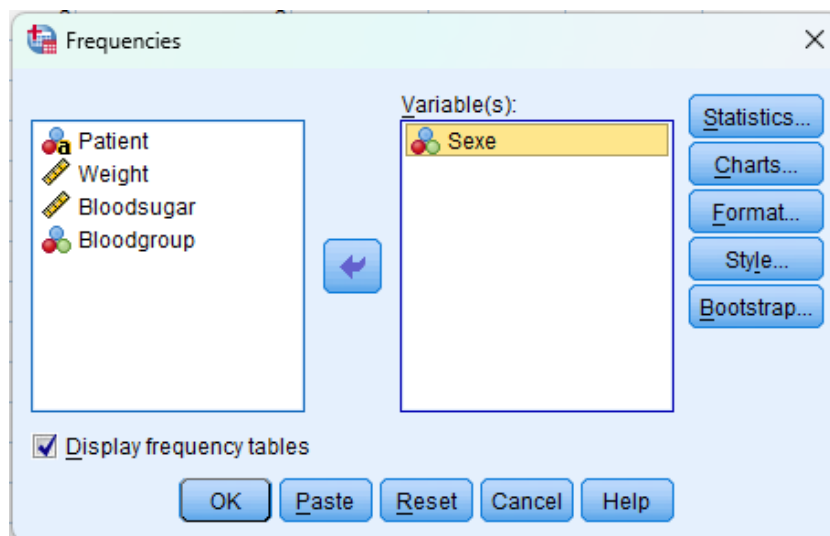


FIGURE 5.4 – Frequencies dialog box.

3. Cliquez sur le bouton **Statistics** : la boîte de dialogue **Frequencies : Statistics** apparaît (voir Figure 5.5).
4. Dans la section **Central Tendency**, cochez la case **Mode**, comme illustré ci-dessous.

Cette boîte de dialogue propose de nombreuses statistiques, mais il est essentiel de ne sélectionner que celles qui correspondent au niveau de mesure des variables placées dans la zone Variable(s). Pour les variables nominales, la mesure de tendance centrale la plus appropriée est le **mode**,

5.4. Utilisation des statistiques descriptives

car ces variables n'ont ni ordre ni hiérarchie. Pour les variables ordinales, la **médiane** est préférable car elle tient compte de l'ordre des catégories, mais le mode peut également être utilisé si l'on souhaite identifier la catégorie la plus fréquente.

5. Cliquez sur **Continuer**.

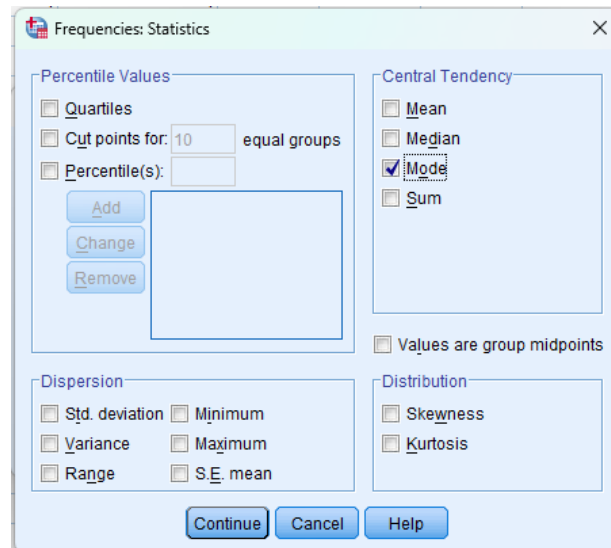


FIGURE 5.5 – Frequencies : Statistics dialog box.

6. Cliquez sur le bouton **Charts** : la boîte de dialogue **Frequencies : Charts** apparaît.
7. Dans la section **Chart Type**, sélectionnez le bouton radio **Bar charts** ; dans la section Chart Values, sélectionnez le bouton radio **Percentages** (voir Figure 5.6).

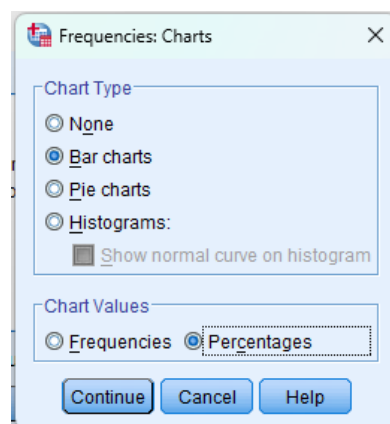


FIGURE 5.6 – Frequencies : Charts dialog box.

8. Cliquez sur **Continuer**, puis sur **OK** : SPSS exécute la procédure Frequencies et calcule les statistiques descriptives, la table de fréquences et le diagramme en barres demandés.
9. Le résultat obtenu, illustré dans les figures ci-dessous, est regroupé par **Bloodgroup** (voir Figure 5.7).

Analyse des données

→ Frequencies

[Jeu_de_données1] C:\Users\User\Desktop\DataSPSS5.sav

Statistics

Sexe

| Sexe | N | Valid | Missing | Mode |
|------|---|-------|---------|----------------|
| AB | 2 | 2 | 0 | 1 ^a |
| A | 3 | 3 | 0 | 2 |
| B | 1 | 1 | 0 | 1 |
| O | 1 | 1 | 0 | 1 |

a. Multiple modes exist. The smallest value is shown

(Effectif)
Absolute frequencies

(Fréquences)
Relative frequencies (%)

Cumulative frequencies (%)

| Bloodgroup | Frequency | Percent | Valid Percent | Cumulative Percent |
|----------------|-----------|---------|---------------|--------------------|
| AB Valid Homme | 1 | 50,0 | 50,0 | 50,0 |
| AB Valid Femme | 1 | 50,0 | 50,0 | 100,0 |
| AB Total | 2 | 100,0 | 100,0 | |
| A Valid Homme | 1 | 33,3 | 33,3 | 33,3 |
| A Valid Femme | 2 | 66,7 | 66,7 | 100,0 |
| A Total | 3 | 100,0 | 100,0 | |
| B Valid Homme | 1 | 100,0 | 100,0 | 100,0 |
| O Valid Homme | 1 | 100,0 | 100,0 | 100,0 |

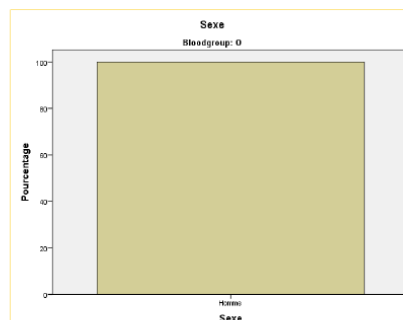
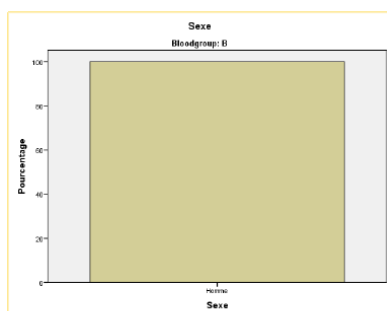
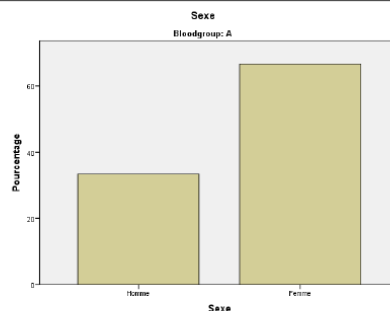
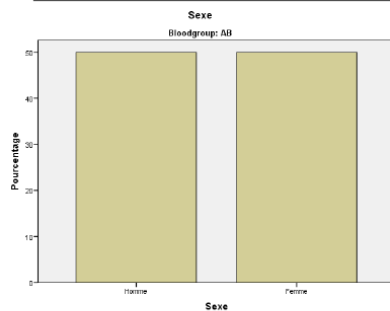


FIGURE 5.7 – Analysis result.

Remarque :

Dans IBM SPSS Statistics Viewer, vous pouvez modifier le style des graphiques en utilisant le Chart Editor :

1. Double-cliquez sur le graphique dans le Statistics Viewer pour ouvrir le Chart Editor.
2. Une fois le **Chart Editor** ouvert, vous pouvez modifier le style des éléments en double-cliquant sur l'élément à modifier (par exemple une barre, un axe ou une légende). Une fenêtre **Properties** s'ouvre alors, vous permettant d'ajuster le style, la couleur, la taille et d'autres paramètres.

5.4.2 Fréquences pour les variables continues

■ Comme vous l'avez vu, les **tables de fréquences** affichent les effectifs et les pourcentages, ce qui est extrêmement utile lorsque l'on travaille avec des variables qualitatives. Cependant, pour les variables continues, qui peuvent prendre un grand nombre de valeurs, les tables de fréquences deviennent moins utiles.

■ Pour exécuter la procédure Fréquences pour des variables continues, procédez comme suit :

1. **Annuler le fractionnement des données.**
2. Choisissez **Analyze → Descriptive Statistics → Fréquences.**
3. Sélectionnez les variables **Weight** et **Bloodsugar**, puis placez-les dans la zone Variable(s).
4. Décochez l'option **Display frequency tables**, comme illustré dans la Figure 5.8 ci-dessous.

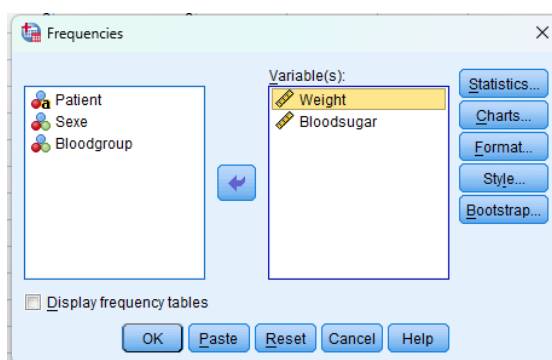


FIGURE 5.8 – Fréquences dialog box.

5. Cliquez sur le bouton **Statistics.**
6. Dans la section **Central Tendency**, cochez les cases **Mean**, **Median** et **Mode**. Dans la section **Dispersion**, sélectionnez **Standard Deviation**, **Variance**, **Minimum** et **Maximum**. Vous pouvez également sélectionner d'autres indicateurs comme : Quartiles, Sum, etc.

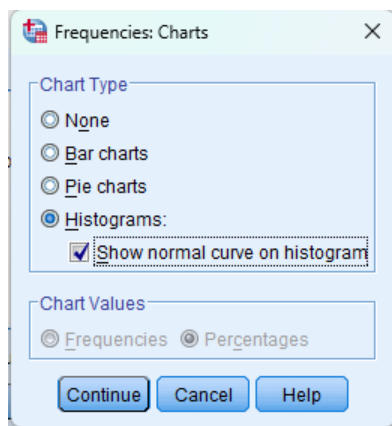


FIGURE 5.9 – Frequencies : Charts dialog box.

7. Cliquez sur **Continue**.

➔ **Frequencies**

| Statistics | | | |
|----------------|---------|--------------------|------------|
| | | Weight | Bloodsugar |
| N | Valid | 7 | 7 |
| | Missing | 0 | 0 |
| Mean | | 87,3143 | 1,7286 |
| Median | | 88,0000 | 1,4000 |
| Mode | | 45,00 ^a | 1,20 |
| Std. Deviation | | 42,49029 | ,99115 |
| Variance | | 1805,425 | ,982 |
| Minimum | | 45,00 | ,60 |
| Maximum | | 175,00 | 3,60 |
| Sum | | 611,20 | 12,10 |
| Percentiles | 25 | 60,0000 | 1,2000 |
| | 50 | 88,0000 | 1,4000 |
| | 75 | 90,2000 | 2,4000 |

a. Multiple modes exist. The smallest value is shown

Histogramme

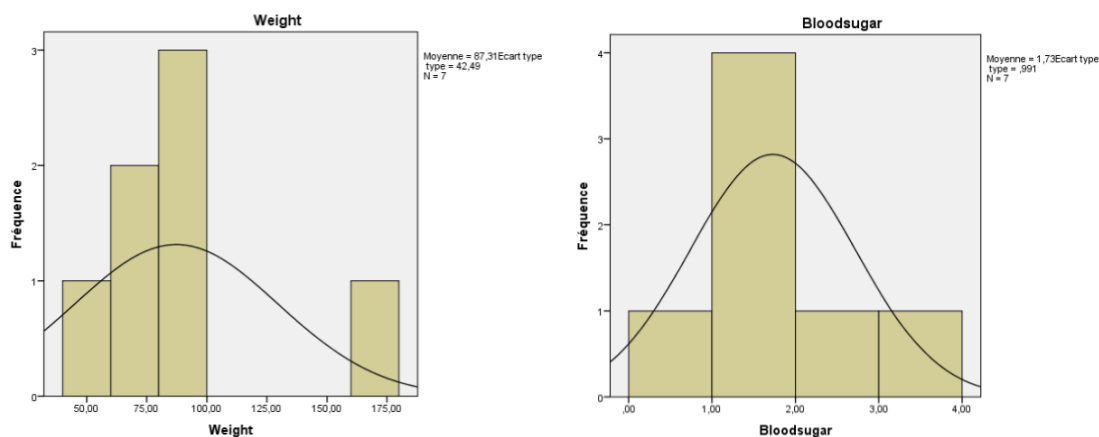


FIGURE 5.10 – Analysis result.

8. Cliquez sur le bouton **Charts**.
9. Sélectionnez le bouton radio **Histograms** et cochez l'option **Display normal curve on histogram**, comme illustré dans la Figure 5.9.
10. Cliquez sur **Continue**, puis sur **OK** : SPSS exécute la procédure Frequencies et calcule les statistiques descriptives ainsi que l'histogramme demandé.
11. Le résultat obtenu est illustré dans la Figure 5.10.

5.4.3 Résumé des variables continues avec la procédure Descriptives

■ La procédure Descriptives fournit un résumé concis de différentes statistiques ainsi que le nombre de cas ayant des valeurs valides pour chaque variable incluse dans le tableau.

■ Pour utiliser la procédure Descriptives, procédez comme suit :

1. Choisissez **Analyze → Descriptive Statistics → Descriptives** : la boîte de dialogue Descriptives apparaît.
2. Sélectionnez les variables **Weight** et **Bloodsugar**, puis placez-les dans la zone Variable(s).
3. Cliquez sur **OK** : SPSS exécute la procédure Descriptives et calcule les statistiques descriptives comme illustré ci-dessous (Figure 5.11).

| | N | Minimum | Maximum | Mean | Std. Deviation |
|--------------------|---|---------|---------|---------|----------------|
| Weight | 7 | 45,00 | 175,00 | 87,3143 | 42,49029 |
| Bloodsugar | 7 | ,60 | 3,60 | 1,7286 | ,99115 |
| Valid N (listwise) | 7 | | | | |

FIGURE 5.11 – Descriptive statistics result.

5.5 Conclusion

Dans ce chapitre, nous avons exploré différentes méthodes d'analyse des données biomédicales à l'aide de SPSS, en mettant l'accent sur les variables qualitatives et continues. Ces approches permettent de résumer les données de manière systématique, tout en tenant compte des caractéristiques spécifiques des variables étudiées.

L'utilisation des statistiques descriptives illustre la puissance de SPSS pour organiser et analyser efficacement les données biomédicales. En soulignant l'importance d'une méthodologie rigoureuse et adaptée, ce chapitre fournit une base solide pour réaliser des analyses exploratoires, essentielles pour identifier des tendances, des corrélations ou des différences significatives dans les données.

Chapitre 6

Analyse des relations entre variables statistiques

6.1 Introduction

L'analyse des données biomédicales peut être utilisée pour étudier les relations entre les variables dans différents contextes, tels que l'étude des facteurs de risque d'une maladie ou l'évaluation de l'efficacité d'un traitement médical. L'analyse des relations entre variables consiste à examiner l'association ou la corrélation entre différentes variables, qui peuvent être catégorielles ou continues.

Des méthodes statistiques telles que les tableaux croisés et l'analyse de régression peuvent être utilisées pour identifier et quantifier les relations entre les variables qualitatives et continues, respectivement. De plus, l'analyse de corrélation peut être utilisée pour déterminer la force et la direction de la relation entre deux variables continues.

En comprenant les relations entre les variables, les chercheurs peuvent mieux appréhender les facteurs sous-jacents contribuant à un résultat ou à un phénomène particulier et prendre des décisions éclairées basées sur leur analyse.

6.2 Collecte des données dans SPSS

1. Télécharger le fichier de données SPSS nommé "**DataSPSS6.sav**" depuis : <https://aboulesnane.net/wp-content/datafiles/DataSPSS6.sav>
2. Les données contiennent cinq variables nommées : Patient, Weight, Bloodsugar, LungCancer et Smoking (voir Figure 6.1).
 - (a) La variable "**Patient**" est une variable de type **String**.
 - (b) La variable "**Weight**" est une variable quantitative continue de type Numeric.
 - (c) La variable "**Bloodsugar**" est une variable quantitative continue de type Numeric.
 - (d) Les valeurs possibles pour la variable qualitative "**LungCancer**" sont : 0=Non et 1=Oui.

6.3. Prétraitement des données dans SPSS

| | Patient | Weight | Bloodsugar | LungCancer | Smoking |
|----|---------|--------|------------|------------|---------|
| 1 | P1 | 63,00 | 1,20 | 1 | 2 |
| 2 | P2 | 90,20 | 1,70 | 1 | 3 |
| 3 | P3 | 88,00 | 1,40 | 0 | 2 |
| 4 | P4 | 45,00 | ,60 | 1 | 3 |
| 5 | P5 | 90,00 | 2,40 | 0 | 2 |
| 6 | P6 | 175,00 | 3,60 | 0 | 1 |
| 7 | P7 | 60,00 | 1,20 | 1 | 2 |
| 8 | P8 | 120,00 | 1,92 | 0 | 1 |
| 9 | P9 | 55,00 | ,70 | 0 | 1 |
| 10 | P10 | 160,00 | 4,62 | 1 | 3 |

FIGURE 6.1 – SPSS Data File.

(e) Les valeurs possibles pour la variable qualitative "**Smoking**" sont :
1=Non, 2=Parfois, 3=Beaucoup.

6.3 Prétraitement des données dans SPSS

■ Les données des patients ont été triées selon leurs valeurs de poids (**Weight**) (voir Section 4.5).

Le résultat du tri est :

| | Patient | Weight | Bloodsugar | LungCancer | Smoking |
|----|---------|--------|------------|------------|---------|
| 1 | P4 | 45,00 | ,60 | 1 | 3 |
| 2 | P9 | 55,00 | ,70 | 0 | 1 |
| 3 | P7 | 60,00 | 1,20 | 1 | 2 |
| 4 | P1 | 63,00 | 1,20 | 1 | 2 |
| 5 | P3 | 88,00 | 1,40 | 0 | 2 |
| 6 | P5 | 90,00 | 2,40 | 0 | 2 |
| 7 | P2 | 90,20 | 1,70 | 1 | 3 |
| 8 | P8 | 120,00 | 1,92 | 0 | 1 |
| 9 | P10 | 160,00 | 4,62 | 1 | 3 |
| 10 | P6 | 175,00 | 3,60 | 0 | 1 |

FIGURE 6.2 – Result after sorting.

6.4 Distributions statistiques bivariées

6.4.1 Relations entre variables catégorielles (qualitatives)

L'une des méthodes les plus courantes pour analyser les données consiste à utiliser des tableaux croisés (cross-tabulations). Comme mentionné précédem-

Analyse des relations entre variables statistiques

| | | Variables indépendantes | |
|-----------------------|---------------|--|---|
| Variables dépendantes | Variables | Qualitatives | Quantitatives |
| | Qualitatives | Tableaux croisés, Tests non paramétriques | Régression logistique, Analyse discriminante |
| | Quantitatives | Test t, ANOVA | Corrélation, Régression linéaire |

TABLE 6.1 – Analyse des relations statistiques.

ment, on utilise un tableau croisé lorsque l'on souhaite étudier la relation entre deux ou plusieurs variables catégorielles.

Par exemple, on peut examiner l'impact (la relation) du tabagisme sur le cancer du poumon.

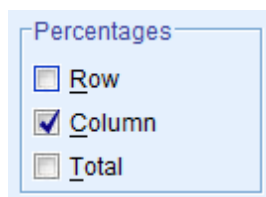
Voici comment effectuer un tableau croisé avec nos données (entre les variables **LungCancer** et **Smoking**) :

1. Choisissez **Analyze → Descriptive Statistics → Crosstabs** : la boîte de dialogue Crosstabs apparaît.

Bien que les variables puissent être placées dans Rows ou Columns, **il est habituel de placer la variable indépendante dans les colonnes du tableau croisé**. Dans les analyses bivariées, la variable indépendante est celle qui influence théoriquement l'autre variable appelée variable dépendante.

Dans cette étude, la variable indépendante est le comportement de tabagisme, car on suppose que le tabagisme a un impact direct sur le développement du cancer du poumon.

2. Sélectionnez la variable **LungCancer** et placez-la dans la zone Row(s).
3. Sélectionnez la variable **Smoking** et placez-la dans la zone Column(s), comme illustré dans la Figure 6.3.
4. Cliquez sur le bouton **Cells** : sélectionnez les pourcentages par ligne, par colonne, ou les deux.



5. Cliquez sur **Continue**, puis sur **OK**.
6. Le résultat obtenu est illustré dans la Figure 6.4.

6.4. Distributions statistiques bivariées

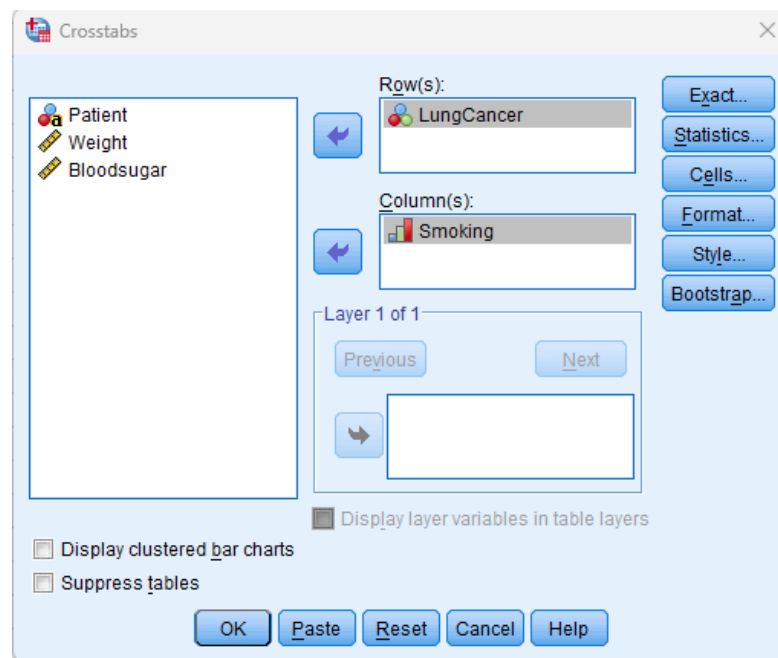


FIGURE 6.3 – Crosstabs Dialog Box.

Case Processing Summary

| | Cases | | | | | |
|----------------------|-------|---------|---------|---------|-------|---------|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| LungCancer * Smoking | 10 | 100,0% | 0 | 0,0% | 10 | 100,0% |

LungCancer ^ Smoking Crosstabulation

| | | | Smoking | | | Total |
|------------|-----|------------------|---------|---------|----------|--------|
| | | | Non | Parfois | Beaucoup | |
| LungCancer | Non | Count | 3 | 2 | 0 | 5 |
| | | % within Smoking | 100,0% | 50,0% | 0,0% | 50,0% |
| | Oui | Count | 0 | 2 | 3 | 5 |
| | | % within Smoking | 0,0% | 50,0% | 100,0% | 50,0% |
| Total | | Count | 3 | 4 | 3 | 10 |
| | | % within Smoking | 100,0% | 100,0% | 100,0% | 100,0% |

FIGURE 6.4 – Crosstabs Output.

Attention :

Correlation does not imply causation signifie que même si deux variables évoluent ensemble (par exemple les ventes de glaces et les attaques de requins), cela ne signifie pas que l'une cause directement l'autre. Elles peuvent être liées par un facteur caché (comme la saison ou la température), une coïncidence temporelle ou simplement le hasard.

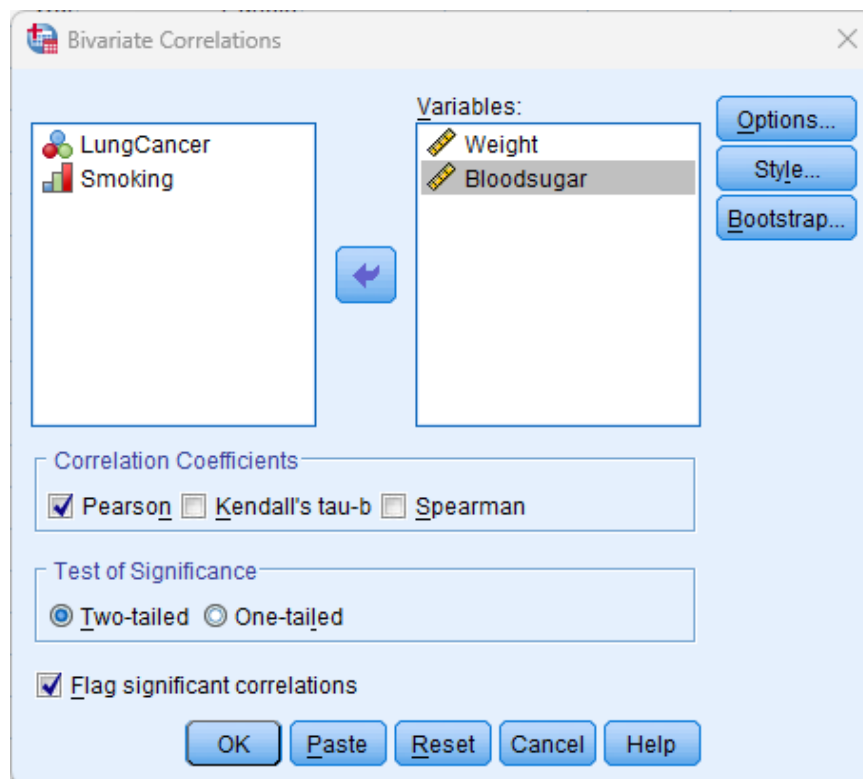


FIGURE 6.5 – Corrélations bivariées.

6.4.2 Relations entre variables quantitatives

Les deux techniques statistiques les plus utilisées pour analyser les relations entre variables continues sont la corrélation de Pearson et la régression linéaire.

La corrélation permet de déterminer si deux variables sont liées entre elles. La régression linéaire va plus loin en tentant de prédire la valeur d'une variable à partir d'une autre variable.

❖ Exécution de la procédure Bivariate

Le coefficient de corrélation de Pearson est une mesure qui permet d'évaluer l'intensité d'une relation linéaire (droite) entre deux variables. Il prend des valeurs comprises entre -1 et $+1$, de sorte que plus la valeur absolue est élevée, plus la corrélation est forte. Par exemple, une corrélation de $+1$ indique que les données se situent parfaitement sur une droite croissante (relation positive), tandis qu'une corrélation de -1 correspond à des données formant une droite parfaitement décroissante (relation négative). Une corrélation égale à 0 indique qu'il n'existe aucune relation linéaire entre les variables. Pour tester une corrélation, procédez comme suit :

1. Choisissez **Analyse → Correlate → Bivariate**.
2. Sélectionnez les variables **Weight** et **Bloodsugar** et placez-les dans la zone Variables.
3. Cliquez sur **Options** et cochez **Cross-product deviations and covariances**.

6.4. Distributions statistiques bivariées

4. Cliquez sur **Continue**, puis sur **OK**.
5. Le résultat obtenu est illustré dans la Figure 6.6.

→ Correlations

Covariance using Bessel's correction, that is, dividing by N - 1.

| | | Weight | Bloodsugar |
|------------|-----------------------------------|-----------|------------|
| Weight | Pearson Correlation | 1 | ,925** |
| | Sig. (2-tailed) | | ,000 |
| | Sum of Squares and Cross-products | 17694,596 | 475,289 |
| | Covariance | 1966,066 | 52,810 |
| | N | 10 | 10 |
| Bloodsugar | Pearson Correlation | ,925** | 1 |
| | Sig. (2-tailed) | ,000 | |
| | Sum of Squares and Cross-products | 475,289 | 14,927 |
| | Covariance | 52,810 | 1,659 |
| | N | 10 | 10 |

** . Correlation is significant at the 0.01 level (2-tailed).

FIGURE 6.6 – Correlation Table

Dans cet exemple, on observe une très forte corrélation positive (0,925) statistiquement significative entre la glycémie (**Bloodsugar**) et le poids (**Weight**). D'une autre manière, à partir du même tableau, nous pouvons également calculer le coefficient de corrélation à partir de la matrice de covariance, comme suit :

$$r = \frac{XY \text{ covariance}}{\sqrt{X \text{ variance}} \sqrt{Y \text{ variance}}}$$

$$r = \frac{\left(\frac{\sum (X - \bar{X})(Y - \bar{Y})}{N} \right)}{\sqrt{\frac{\sum (X - \bar{X})(X - \bar{X})}{N}} * \sqrt{\frac{\sum (Y - \bar{Y})(Y - \bar{Y})}{N}}}$$

$$r = \frac{\left(\frac{475,289}{10} \right)}{\sqrt{\frac{17694,596}{10}} * \sqrt{\frac{14,927}{10}}} = 0.925$$

❖ Exécution de la procédure de régression linéaire simple

Les corrélations permettent de déterminer si deux variables continues sont liées de manière linéaire. L'analyse de régression consiste à prédire le futur (l'inconnu) à partir de données collectées dans le passé (le connu). La régression permet de quantifier davantage les relations en développant une **équation** afin de pouvoir prédire, par exemple, le niveau de glycémie à partir du poids corporel du patient.

La régression linéaire est une technique statistique utilisée pour prédire une **variable dépendante continue** à partir d'une ou plusieurs **variables indépendantes continues**.

Puisque nous avons une forte corrélation linéaire entre la glycémie et le poids, nous pouvons effectuer une régression linéaire comme suit :

1. Sélectionnez **Analyze → Regression → Linear**.

La boîte de dialogue **Linear Regression** apparaît. Dans cet exemple, vous souhaitez prédire le niveau de glycémie à partir du poids. Vous pouvez placer la variable dépendante, la glycémie (**Bloodsugar**), dans la zone **Dependent** ; il s'agit de la variable pour laquelle vous souhaitez définir une équation de prédiction. Vous pouvez placer la variable prédictive **Weight** dans la zone **Independent(s)** ; c'est la variable qui sera utilisée pour prédire la variable dépendante.

2. Sélectionnez la variable **Bloodsugar** et déplacez-la dans la zone Dependent.
3. Sélectionnez la variable **Weight** et déplacez-la dans la zone Independent(s), comme illustré dans la Figure 6.7.

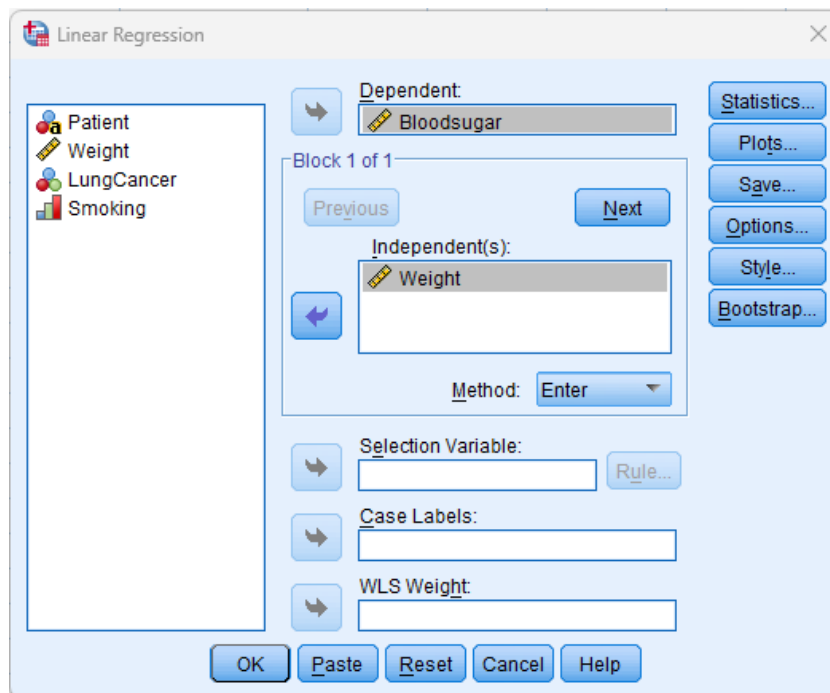


FIGURE 6.7 – Boîte de dialogue Linear Regression.

6.5. Représentation graphique des données

4. Cliquez sur **OK** : SPSS effectue la régression linéaire (voir Figure 6.8).

→ Regression

Variables Entered/Removed^a

| Model | Variables Entered | Variables Removed | Method |
|-------|---------------------|-------------------|--------|
| 1 | Weight ^b | . | Enter |

a. Dependent Variable: Bloodsugar

b. All requested variables entered.

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | ,925 ^a | ,855 | ,837 | ,51969 |

a. Predictors: (Constant), Weight

ANOVA^a

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|--------|-------------------|
| 1 | Regression | 12,767 | 1 | 12,767 | 47,270 | ,000 ^b |
| | Residual | 2,161 | 8 | ,270 | | |
| | Total | 14,927 | 9 | | | |

a. Dependent Variable: Bloodsugar

b. Predictors: (Constant), Weight

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|--------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | -,608 | ,405 | | -1,502 | ,172 |
| | Weight | ,027 | ,004 | ,925 | 6,875 | ,000 |

a. Dependent Variable: Bloodsugar

FIGURE 6.8 – Résultats de la régression linéaire.

La colonne B contient les coefficients de régression (a : la pente, b : l'ordonnée à l'origine) qui sont utilisés dans l'équation de prédiction. Dans cet exemple, le niveau de glycémie peut être prédit à l'aide de l'équation suivante :

$$\text{Bloodsugar} = a * \text{Weight} + b$$

$$\text{Bloodsugar} = 0.027 * \text{Weight} - 0.608$$

6.5 Représentation graphique des données

Le menu **Graphs** dans SPSS contient trois options principales : **Chart Builder**, **Graphboard Template Chooser** et **Legacy Dialogs**. Ces options repré-

Analyse des relations entre variables statistiques

sentent différentes façons d'effectuer la même tâche. Les **Legacy Dialogs** correspondent aux graphiques originaux de SPSS et sont principalement utilisés par les personnes qui les emploient depuis longtemps et trouvent difficile de passer à une autre option.

Graphboard Template Chooser et **Chart Builder** permettent de créer des graphiques de différentes manières. Dans le **Graphboard Template Chooser**, vous sélectionnez d'abord les variables que vous souhaitez afficher. Sur la base de ces informations, différentes options de graphiques sont proposées. **Chart Builder** commence par présenter différents types de graphiques ; après avoir sélectionné un graphique, vous spécifiez ensuite les variables à utiliser.

6.5.1 Création de graphiques avec Chart Builder

SPSS inclut **Chart Builder**, qui utilise une interface graphique pour vous guider dans les étapes de création de graphiques. Le programme vérifie continuellement vos actions et empêche l'utilisation de fonctionnalités incompatibles ou qui ne fonctionneraient pas correctement.

À travers un exemple, nous allons voir comment générer un graphique. Supposons que nous souhaitons tracer le **boxplot** de la variable **Bloodsugar**.

1. Choisissez **Graphs** → **Chart Builder**. Un avertissement apparaît indiquant qu'avant d'utiliser cette boîte de dialogue, le niveau de mesure doit être correctement défini pour chaque variable du graphique. (Nous avons déjà défini le niveau de mesure correct, vous pouvez donc continuer.)
2. Cliquez sur **OK** : la boîte de dialogue **Chart Builder** apparaît.
3. Assurez-vous que l'onglet **Gallery** est sélectionné : dans la liste "**Choose from**", sélectionnez "**Boxplot**" comme type de graphique.
4. Différents types de boxplots apparaissent dans la galerie à droite de la liste, comme illustré dans la Figure 6.9.

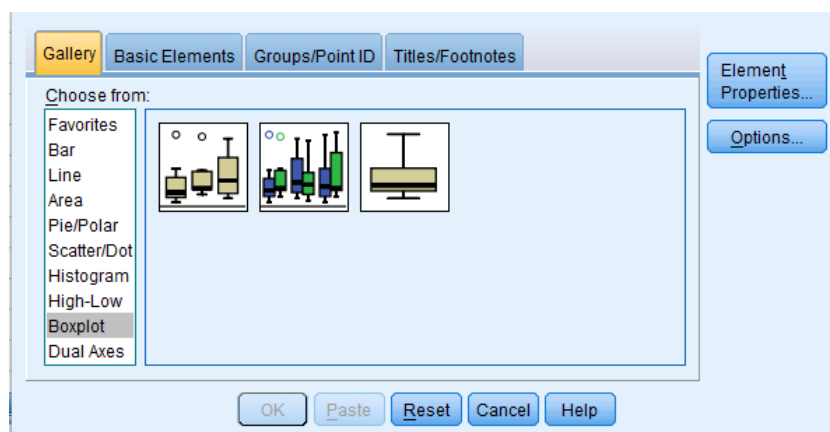


FIGURE 6.9 – Chart Builder.

5. Sélectionnez et faites glisser "**ID : Boxplot**" dans la zone du graphique.

L'onglet **Element Properties** apparaît alors à droite de la zone de prévisualisation. Cet onglet indique quelles caractéristiques de l'élément peuvent

6.5. Représentation graphique des données

être modifiées. Par exemple, vous pouvez modifier la statistique affichée ou le style du graphique. Dans cet exemple, nous n'utiliserons pas cet onglet ; vous pouvez simplement le fermer.

6. Dans la liste des variables, sélectionnez la variable **Bloodsugar** et faites-la glisser vers l'axe Y du graphique.

L'affichage graphique dans la fenêtre de prévisualisation ne représente jamais vos données réelles, même après l'insertion des noms des variables.

7. Cliquez sur **OK** pour produire le graphique : le résultat est illustré dans la Figure 6.10.

→ GGraph

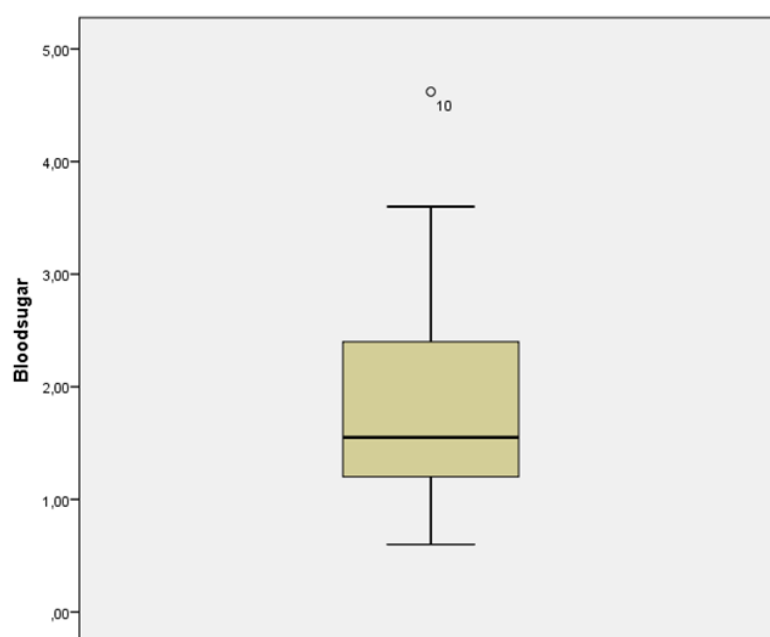


FIGURE 6.10 – Boxplot : 1D pour la variable Bloodsugar.

6.5.2 Affichage d'une relation linéaire

Les étapes suivantes montrent comment construire un diagramme de dispersion simple :

1. Choisissez **Graphs → Chart Builder**.
2. Cliquez sur **OK** puis sur le bouton **Reset**.
3. Dans la liste "**Choose from**", sélectionnez **Scatter/Dot**.
4. Sélectionnez le premier graphique (Simple Scatter) et faites-le glisser dans la zone supérieure.
5. Dans la liste des variables, sélectionnez **Weight** et faites-le glisser vers la zone X-Axis du graphique.
6. Dans la liste des variables, sélectionnez **Bloodsugar** et faites-le glisser vers la zone Y-Axis du graphique.
7. Cliquez sur **OK** : le graphique illustré dans la Figure 6.11 apparaît.

→ **GGraph**

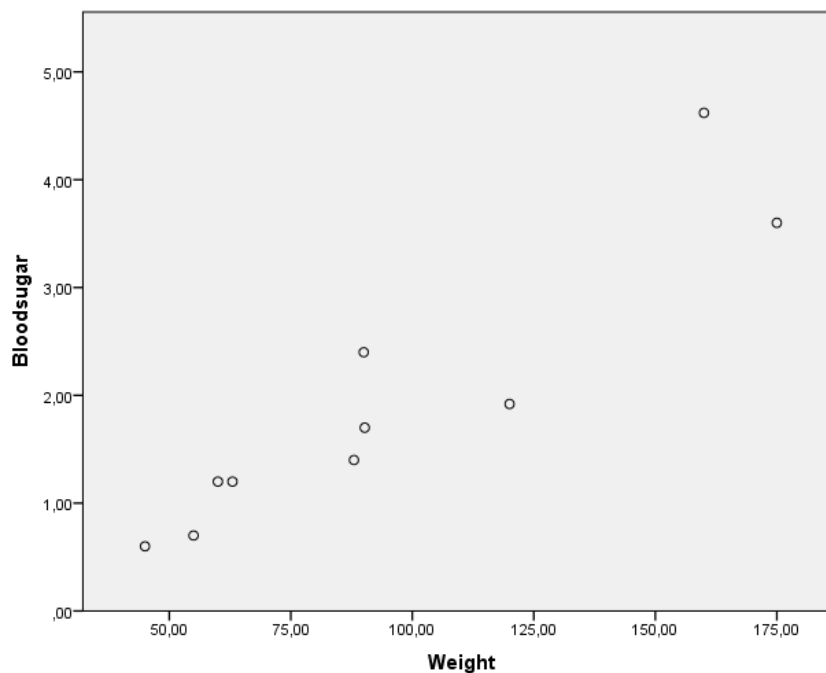


FIGURE 6.11 – Diagramme de dispersion de deux variables quantitatives

8. Double-cliquez sur le graphique produit : la boîte de dialogue **Chart Editor** apparaît.
9. Cliquez sur l'icône **Add Fit Line at Total**, puis cliquez sur **Close** et fermez la fenêtre **Chart Editor**.

10. Le résultat obtenu est illustré dans la Figure 6.12.

GGraph

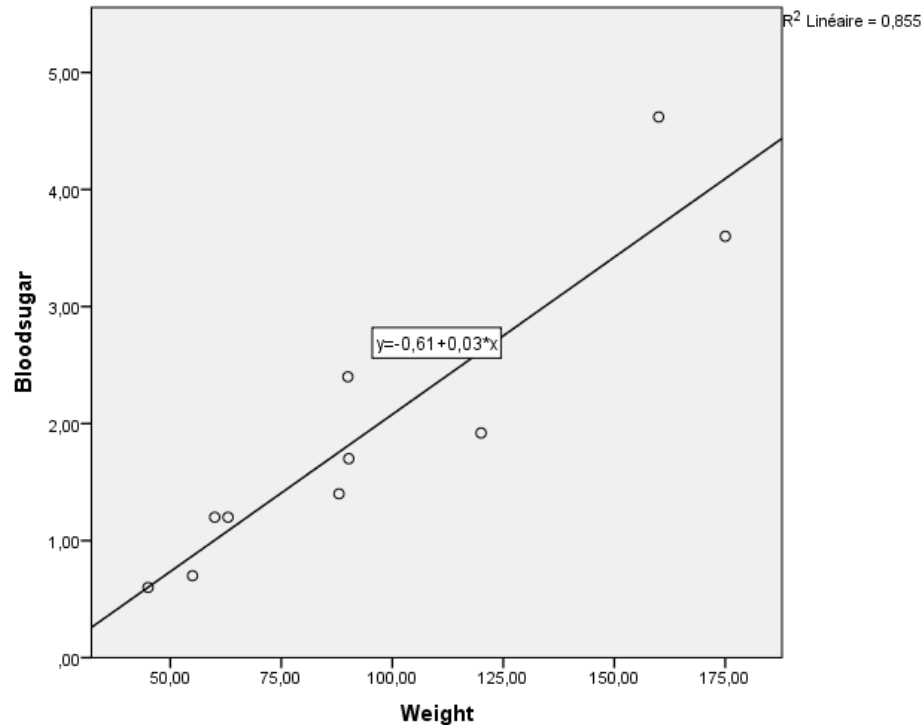


FIGURE 6.12 – Droite de régression linéaire

La droite superposée au diagramme de dispersion est la droite d'ajustement qui décrit la relation linéaire donnée par l'équation : $y = 0.03 * x - 0.61$, où y représente **Bloodsugar** et x représente **Weight**.

6.6 Conclusion

Les chercheurs peuvent utiliser SPSS pour étudier les relations entre variables en analysant à la fois des variables qualitatives et continues. SPSS propose différents outils et fonctions, tels que les tableaux croisés et l'analyse de corrélation, pour examiner l'association ou la corrélation entre les variables.

L'analyse par tableaux croisés peut être utilisée pour identifier les relations entre deux ou plusieurs variables qualitatives, tandis que l'analyse de corrélation permet de déterminer la force et la direction de la relation entre deux variables continues. De plus, l'analyse de régression peut être utilisée pour examiner la relation entre une variable dépendante et une ou plusieurs variables indépendantes, et pour créer un modèle permettant de prédire les valeurs de la variable dépendante à partir des variables indépendantes.

En utilisant ces outils et ces fonctions, les chercheurs peuvent étudier les relations entre les variables dans leurs données biomédicales et prendre des décisions éclairées basées sur leurs analyses.

Travaux Pratiques

TP 1

But

Dans ce TP, nous apprendrons à créer un fichier Excel contenant plusieurs variables quantitatives et qualitatives. Nous visons également à apprendre à valider les données saisies.

Exercice

| | A | B | C | D | E |
|---|---------|-----|-------|-------|-------------|
| 1 | Patient | Age | Sexe | Blood | Temperature |
| 2 | P1 | 77 | Femme | B | 37,4 |
| 3 | P2 | 64 | Homme | A | 38 |
| 4 | P3 | 87 | Homme | AB | 38,5 |
| 5 | P4 | 70 | Femme | A | 38,2 |
| 6 | P5 | 85 | Femme | O | 38,1 |

1. Créer un fichier de données EXCEL avec différents types de variables (qualitatives et quantitatives) :
 - (a) Créer un nouveau fichier de données EXCEL.
 - (b) Les données doivent contenir cinq variables nommées : Patient, Age, Sexe, Blood, et Temperature (voir la Figure ci-dessus).
 - (c) Saisir les données de la variable "Patient" de P1 à P5 par glissement ?
2. Avant de saisir les données des variables Age, Sexe, Blood, et Temperature, vous devez vous assurer que chaque variable est valide.
 - (a) Vérifier la validité des données en variables quantitatives (Age, Temperature) ?
 - (b) Vérifier la validité des données en variables qualitatives (Sexe, Blood) ?
3. Enregistrez le fichier Excel obtenu sous le nom « Tp1.xlsx ».
4. Supprimez la variable « Température » et enregistrez les données dans un autre fichier appelé « Tp1_2.xlsx ».

Solution

1. (a) Ouvrir Microsoft Excel.
(b) Dans la première ligne, saisir les noms des cinq variables dans chaque colonne : A1 : Patient, B1 : Age, C1 : Sex, D1 : Blood, et E1 : Temperature.
(c) Sélectionner la cellule A2 → Saisir P1 → faire glisser le fill handle vers le bas jusqu'à A6 pour remplir les identifiants des patients.
2. (a) Pour la variable quantitative discrète Age : sélectionner les cellules B2 :B6 → Data → Data Validation → Choisir Allow : Whole number → Data : between → Minimum : 1 → Maximum : 120 → OK.
(b) Pour la variable quantitative continue Temperature : sélectionner les cellules E2 :E6 → Data → Data Validation → Choisir Allow : Decimal → Data : between → Minimum : 29.5 → Maximum : 42.3 → OK.
(c) Pour la variable qualitative Sex : sélectionner les cellules C2 :C6 → Data → Data Validation → Choisir Allow : List → Source : Homme ;Femme → OK.
(d) Pour la variable qualitative Blood : sélectionner les cellules D2 :D6 → Data → Data Validation → Choisir Allow : List → Source : AB ;A ;B ;O → OK.
(e) Compléter les données comme indiqué dans la figure ci-dessus.
3. File Tab → Save → File name : Tp1.xlsx → Save.
4. Sélectionner la colonne E → appuyer sur la touche "Delete" → File Tab → Save As → File name : Tp1_2.xlsx → Save.

Pour plus de détails, reportez-vous au chapitre 2

TP 2

But

Dans ce TP, nous apprendrons à saisir des données et à définir correctement des variables à l'aide du programme SPSS. De plus, il est important de savoir comment gérer les fichiers SPSS et comment transférer des données d'Excel vers SPSS.

Exercice

| | Patient | Sexe | Temperature | Bloodsugar | Hypertension |
|---|---------|------|-------------|------------|--------------|
| 1 | P1 | 1 | 38,10 | 1,20 | 1 |
| 2 | P2 | 1 | 37,10 | 1,70 | 2 |
| 3 | P3 | 2 | 38,00 | 1,40 | 1 |
| 4 | P4 | 1 | 38,10 | ,60 | 1 |

1. Créer un fichier de données SPSS avec différents types de variables (qualitatives et quantitatives) :
 - (a) Créer un nouveau fichier de données SPSS appelé « Tp2.sav ».
 - (b) Les données doivent contenir cinq variables nommées : Patient, Sexe, Temperature, Bloodsugar et Hypertension (voir la figure ci-dessus).
 - (c) La variable « Patient » est une variable de type **String**.
 - (d) Entrer les valeurs possibles pour la variable qualitative « Sexe » : 1 = Homme et 2 = Femme.
 - (e) La variable « Temperature » est une variable quantitative continue de type **Numeric**.
 - (f) La variable « Bloodsugar » est une variable quantitative continue de type **Numeric**.
 - (g) Entrer les valeurs possibles pour la variable qualitative « Hypertension » : 1 = Yes, 2 = No.
2. Enregistrer le fichier SPSS obtenu.
3. Créer un autre fichier Excel et essayer de l'ouvrir dans SPSS.

Solution

1. Pour saisir des données dans SPSS, vous pouvez suivre ces étapes :
 - (a) Ouvrir SPSS et cliquer sur l'onglet **Variable View**.
 - (b) Saisir les noms des variables dans la première colonne « **Name** ». Chaque nom de variable doit être unique et descriptif.
 - (c) Définir les propriétés des variables (metadata) dans les colonnes : **Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure**.

Travaux Pratiques

- (d) Allez dans **Data View** et saisissez les données des patients.
2. Une fois toutes les données définies, enregistrer le fichier via **File → Save**.
3. Vous pouvez aussi importer des données dans SPSS depuis Excel. Pour cela : **File → Open > Data → Excel File Type (*.xls, *.xlsx, *.xlsm)** → sélectionner le fichier Excel → cliquer sur **Open**.

Après l'importation, il peut être nécessaire d'ajuster les types de variables ou de recoder certaines données selon les besoins d'analyse.

Pour plus de détails, reportez-vous au chapitre 3.

TP 3

But

1. Aider l'étudiant à comprendre les techniques appropriées pour collecter et saisir des données avec précision à l'aide du logiciel SPSS.
2. Le TP vise à mettre en évidence les meilleures pratiques pour le traitement des données dans SPSS, y compris les stratégies de traitement des données manquantes, de tri des données, de recodage des variables, de fractionnement des données et de sélection des observations.

Exercice 1

| | Patient | Weight | Bloodsugar | Sexe | Bloodgroup |
|---|---------|--------|------------|------|------------|
| 1 | P1 | 53,00 | 1,20 | 1 | 1 |
| 2 | P2 | 44,00 | -99,00 | 1 | 3 |
| 3 | P3 | 88,00 | 1,40 | 2 | 2 |
| 4 | P4 | 45,00 | ,60 | 1 | 4 |
| 5 | P5 | 90,00 | 2,40 | 2 | 2 |
| 6 | P6 | 175,00 | 3,60 | 2 | 1 |

1. Créer un nouveau fichier de données SPSS appelé « Tp3.sav ».
2. Les données doivent contenir cinq variables nommées : Patient, Weight, Bloodsugar, Sexe, et Bloodgroup (voir la Figure ci-dessus).
3. La Variable « Patient » est une variable de type Chaîne.
4. La Variable « Weight » est une variable quantitative continue de type Numérique.
5. La Variable « Bloodsugar » est une variable quantitative continue de type Numérique. (Pour la variable Bloodsugar, on représente les valeurs manquantes par le nombre -99).
6. Entrez les valeurs possibles pour la variable qualitative « Sexe » : 1=Homme et 2=Femme.
7. Entrez les valeurs possibles pour la variable qualitative « Bloodgroup » : 1=AB , 2=A, 3=B et 4=O.

Exercice 2

1. Dans la variable Bloodsugar, remplacez les valeurs manquantes par une valeur moyenne de la série ?
2. Triez les données des patients en fonction de leurs valeurs de glycémie (Bloodsugar) ?
3. Dans une nouvelle variable nommée "BloodsugarCat" : essayez de recoder les valeurs de la variable "Bloodsugar" comme suit : Classe 1 : Normal $\leq 0,99$; Classe 2 : Prédiabète $]0,99 - 1,25]$; Classe 3 : Diabète $> 1,25$.

4. Supprimer la nouvelle variable "BloodsugarCat" ?
5. Comparez séparément les données des patients en fonction de leurs valeurs de Bloodgroup ?
6. Annuler le fractionnement des données ?
7. Sélectionnez des patients avec un poids ≥ 70 et une glycémie < 1 ou des patients hommes avec un groupe sanguin égal à AB ?

Solution

1. Choisissez Transformer → Remplacer les valeurs manquantes → Passez la variable Bloodsugar vers la zone «Nouvelles variables» → OK → Remplacer manuellement la valeur -99 dans la colonne de Bloodsugar par 1,84 → Supprimer la colonne Bloodsugar_1.
2. Choisissez Données → Trier les observations → Trier par : Bloodsugar → OK.
3. Choisissez Transformer → Création de variables → Déplacer la variable Bloodsugar vers la zone de travail à droite → Dans Nom : Nommez la nouvelle variable comme BloodsugarCat → Changer → Anciennes et nouvelles valeurs → Pour la catégorie 1 : on sélectionne le bouton radio Plage, du MINIMUM à la valeur → Entrer la valeur Max (c.a.d. 0,99) → A côté du bouton radio Valeur : on entre le numéro de la catégorie (c.a.d. 1) → Cliquer sur le bouton Ajouter → Pour la catégorie 2 : on sélectionne le bouton radio Plage → Entrer les valeurs Min(c.a.d. 0,99) et Max(c.a.d. 1,25) → A côté du bouton radio Valeur : on entre le numéro de la catégorie (c.a.d. 2) → Cliquer sur le bouton Ajouter → Pour la catégorie 3 : on sélectionne le bouton radio Plage, de la valeur au MAXIMUM → Entrer la valeur Min (c.a.d. 1,25) → A côté du bouton radio Valeur : on entre le numéro de la catégorie (c.a.d. 3) → Cliquer sur le bouton Ajouter → Poursuivre → OK.
4. Dans l'onglet vue des données, cliquez sur le nom de la variable "BloodsugarCat" → Appuyez sur la touche Suppr du clavier.
5. Choisissez Données → Scinder un fichier → Sélectionnez le bouton radio "Comparer les groupes" → Choisissez Bloodgroup comme variable de comparaison → OK → Choisissez Analyse → Statistiques descriptives → Fréquences → Choisissez Sexe et placez-la dans la zone Variable(s) → OK.
6. Choisissez Données → Scinder un fichier → Réinitialiser → OK.
7. Choisissez Données → Sélectionner des observations → Sélectionnez le bouton radio "Selon une condition logique" → Cliquez sur le bouton Si... → Dans la zone d'expression, Écrivez "(Weight ≥ 70 & Bloodsugar <1) | (Sexe = 1 & Bloodgroup = 1)" → Poursuivre → OK.

Pour plus de détails, reportez-vous au chapitre 4

TP 4

But

1. Aider l'étudiant à comprendre les techniques appropriées pour collecter et saisir des données avec précision à l'aide du logiciel SPSS.
2. Réviser certaines des méthodes utilisées dans le traitement des données SPSS.
3. Présenter les techniques utilisées pour analyser les données qualitatives et quantitatives à l'aide de procédures de fréquence et descriptives spécifiquement dans SPSS.

Exercice 1

| | Patient | Weight | Bloodsugar | Sexe | Bloodgroup |
|---|---------|--------|------------|------|------------|
| 1 | P1 | 63,00 | 1,20 | 1 | 1 |
| 2 | P2 | 90,20 | 1,70 | 1 | 3 |
| 3 | P3 | 88,00 | 1,40 | 2 | 2 |
| 4 | P4 | 45,00 | ,60 | 1 | 4 |
| 5 | P5 | 90,00 | 2,40 | 2 | 2 |
| 6 | P6 | 175,00 | 3,60 | 2 | 1 |
| 7 | P7 | 63,00 | 2,10 | 1 | 2 |

1. Créer un nouveau fichier de données SPSS appelé « Tp4.sav ».
2. Les données doivent contenir cinq variables nommées : Patient, Weight, Bloodsugar, Sexe, et Bloodgroup (voir la Figure ci-dessus).
3. La Variable « Patient » est une variable de type Chaîne.
4. La Variable « Weight » est une variable quantitative continue de type Numérique.
5. La Variable « Bloodsugar » est une variable quantitative continue de type Numérique.
6. Entrez les valeurs possibles pour la variable qualitative « Sexe » : 1=Homme et 2=Femme.
7. Entrez les valeurs possibles pour la variable qualitative « Bloodgroup » : 1=AB , 2=A, 3=B et 4=O.

Exercice 2

1. Triez les données des patients en fonction de leurs valeurs de poids (Weight) et de glycémie (Bloodsugar). Que remarquez-vous ? (Focus sur les patients P1 et P7).
2. Comparez séparément les données des patients en fonction de leurs valeurs de sexe.

Travaux Pratiques

3. Basant sur le fractionnement des données précédent, nous souhaitons étudier la distribution de variable Bloodgroup (AB, A, B, O) pour chaque valeur de Sexe (Homme, Femme). Analyser statistiquement et graphiquement la variable qualitative Bloodgroup avec la procédure de fréquence ?
4. Annuler le fractionnement des données ?
5. Analyser statistiquement et graphiquement la variable quantitative Weight avec la procédure de fréquence ?
6. Résumer statistiquement les variables continues avec la procédure descriptive ?

Solution

1. Choisissez Data → Sort Cases → Trier par : Weight puis Bloodsugar, respectivement → OK. (On observe que P1 est positionné avant P7.)
2. Choisissez Data → Split File → Sélectionnez le bouton radio "Compare groups" → Choisissez Sexe comme variable de regroupement → OK.
3. Choisissez Analyze → Descriptive Statistics → Frequencies → Placez Bloodgroup dans la zone Variable(s) → Statistics → Cochez la case Mode → Continue → Charts → Sélectionnez le bouton radio Bar charts → Sélectionnez le bouton radio Percentages → Continue → OK.
4. Choisissez Data → Split File → Reset → OK.
5. Choisissez Analyze → Descriptive Statistics → Frequencies → Placez la variable Weight dans la zone Variable(s) → Statistics → Cochez les cases suivantes : Mean, Median, Mode, Standard deviation, Variance, Minimum et Maximum → Continue → Charts → Sélectionnez le bouton radio Histogram → Cochez la case "Show normal curve on histogram" → Continue → Décochez la case "Display frequency tables" → OK.
6. Choisissez Analyze → Descriptive Statistics → Descriptives → Placez les variables Weight et Bloodsugar dans la zone Variable(s) → OK.

Pour plus de détails, reportez-vous au chapitre 5.

TP 5

But

1. Aider l'étudiant à comprendre les techniques appropriées pour collecter et saisir des données avec précision à l'aide du logiciel SPSS.
2. Pour montrer comment analyser les relations entre les variables qualitatives à l'aide de tableaux croisés : Les tableaux croisés sont utilisés pour créer des tableaux de contingence qui résument la relation entre deux ou plusieurs variables catégorielles. Le TP vise à montrer comment créer et interpréter ces tableaux dans SPSS.
3. Démontrer comment analyser les relations entre des variables quantitatives à l'aide de la corrélation et de la régression linéaire : La corrélation et la régression linéaire sont des méthodes statistiques utilisées pour modéliser la relation entre deux ou plusieurs variables continues. Le TP vise à montrer comment utiliser ces méthodes pour créer et interpréter des modèles de régression dans SPSS.
4. Démontrer comment créer des tracés et des graphiques dans SPSS : Le TP vise à montrer comment utiliser SPSS pour créer des visualisations de données afin de mieux comprendre les relations entre les variables ou de communiquer les résultats aux autres.

Exercice 1

| | Patient | Weight | Bloodsugar | Sexe | Bloodgroup |
|----|---------|--------|------------|------|------------|
| 1 | P1 | 63,00 | 1,20 | 1 | 1 |
| 2 | P2 | 90,20 | 1,70 | 1 | 3 |
| 3 | P3 | 88,00 | 1,40 | 2 | 2 |
| 4 | P4 | 45,00 | ,60 | 1 | 4 |
| 5 | P5 | 90,00 | 2,40 | 2 | 2 |
| 6 | P6 | 175,00 | 3,60 | 2 | 1 |
| 7 | P7 | 60,00 | 1,20 | 1 | 2 |
| 8 | P8 | 120,00 | 1,92 | 1 | 1 |
| 9 | P9 | 55,00 | ,70 | 2 | 4 |
| 10 | P10 | 160,00 | 4,62 | 2 | 3 |

1. Créer un nouveau fichier de données SPSS appelé "Tp5.sav".
2. Les données doivent contenir cinq variables nommées : Patient, Weight, Bloodsugar, Sexe, et Bloodgroup (voir la Figure ci-dessus).
3. La Variable "Patient" est une variable de type Chaîne.
4. La Variable "Weight" est une variable quantitative continue de type Numérique.
5. La Variable "Bloodsugar" est une variable quantitative continue de type Numérique.

6. Entrez les valeurs possibles pour la variable qualitative "Sexe" : 1=Homme et 2=Femme.
7. Entrez les valeurs possibles pour la variable qualitative "Bloodgroup" : 1=AB , 2=A, 3=B et 4=O.

Exercice 2

1. Triez les données des patients en fonction de leurs valeurs de poids (Weight) ?
2. En utilisant les tableaux croisés, y a-t-il une relation entre les variables qualitatives Sexe et Bloodgroup ?
3. Prouver qu'il existe une corrélation linéaire entre la variable Bloodsugar et Weight ?; Si oui, essayez d'extraire cette équation linéaire en utilisant la technique de régression linéaire.
4. représenter graphiquement la relation linéaire entre la variable Bloodsugar et Weight ?

Solution

1. Choisissez Data → Sort Cases → Trier par : Weight → OK.
2. Choisissez Analyze → Descriptive Statistics → Crosstabs → Placez la variable Sexe dans la zone Row(s) et la variable Bloodgroup dans la zone Column(s) → Cells → Sélectionnez Row percentages → Continue → OK.
3. (a) Choisissez Analyze → Correlate → Bivariate → Placez les variables Weight et Bloodsugar dans la zone Variables → Options → Cochez l'option "Cross-product deviations and covariances" → Continue → OK.
(b) Sélectionnez Analyze → Regression → Linear → Placez la variable Bloodsugar dans la zone Dependent et la variable Weight dans la zone Independent(s) → OK.
4. Choisissez Graphs → Chart Builder → OK → Reset → Scatter/Dot → Simple Scatter → Sélectionnez Weight et faites-le glisser vers le rectangle intitulé X-Axis dans le graphique → Sélectionnez Bloodsugar et faites-le glisser vers le rectangle intitulé Y-Axis dans le graphique → OK → Double-cliquez sur le graphique généré → Cliquez sur l'icône "Add Fit Line at Total" → Close → Fermez la fenêtre "Chart Editor".

Pour plus de détails, reportez-vous au chapitre 6.

Références Bibliographiques

- [Cronk, 2019] CRONK, B. C. (2019). *How to use SPSS® : A step-by-step guide to analysis and interpretation*. Routledge.
- [Denis, 2018] DENIS, D. J. (2018). *SPSS data analysis for univariate, bivariate, and multivariate statistics*. John Wiley & Sons.
- [Jean-Pierre, 2016] JEAN-PIERRE, L. (2016). *Statistiques et probabilités : Cours et exercices corrigés*.
- [Microsoft Corporation,] MICROSOFT CORPORATION. Microsoft excel. <https://support.microsoft.com/en-us/excel>. Accessed : 2019-09-01.
- [Nasir et al., 2022] NASIR, M. A., BAKOUCH, H. S. et JAMAL, F. (2022). *Introductory Statistical Procedures with SPSS*. Bentham Science Publishers.
- [Salcedo et McCormick, 2020] SALCEDO, J. et MCCORMICK, K. (2020). *SPSS Statistics for Dummies*. John Wiley & Sons.
- [Twigg, 2010] TWIGG, M. (2010). *Discovering statistics using spss*.
- [uiowa, 2024] UIOWA (2024). Normal distribution applet/calculator. <https://homepage.divms.uiowa.edu/~mbognar/applets/normal.html>. Accessed : 2024-02-01.