

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. Abdenmour 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 : 8 novembre 2025



Table des matières

Table des figures	iv
Liste des tableaux	v
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
Travaux Pratiques	18

Références Bibliographiques	20
------------------------------------	-----------

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

Liste des tableaux

2.1 XLS vs XLSX	12
2.2 Variables quantitatives et qualitatives	15

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) :

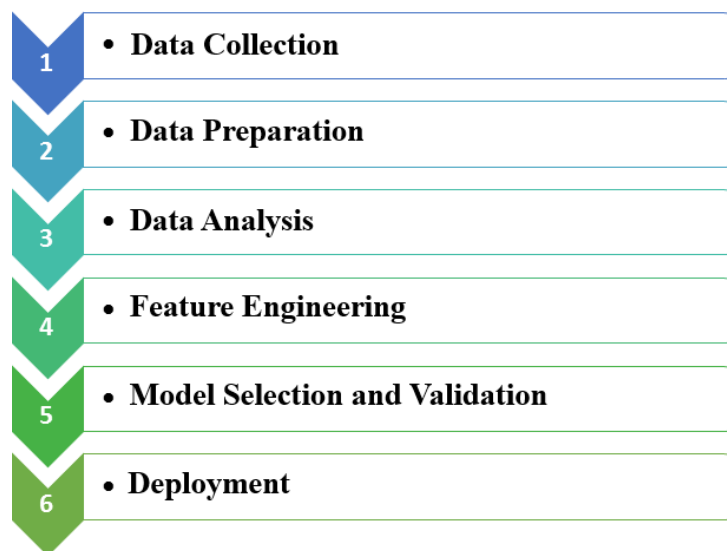


FIGURE 1.1 – Organigramme de la méthodologie de la science des données

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.

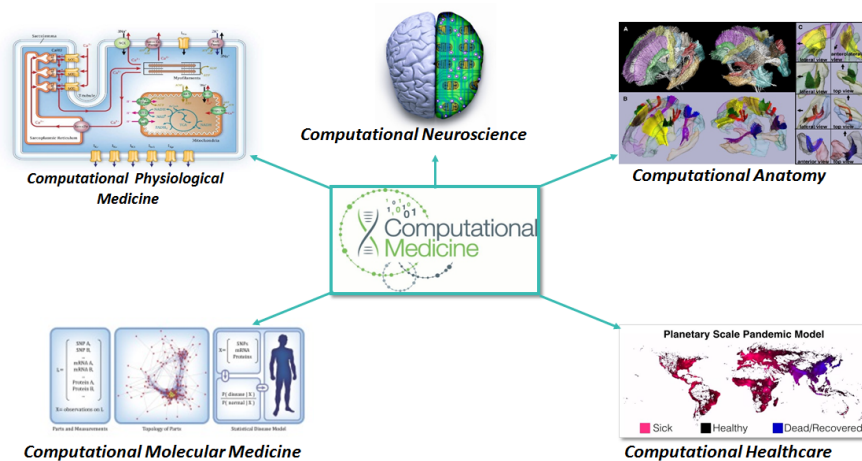


FIGURE 1.2 – Recherches en médecine computationnelle

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

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. **Progiciels 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).

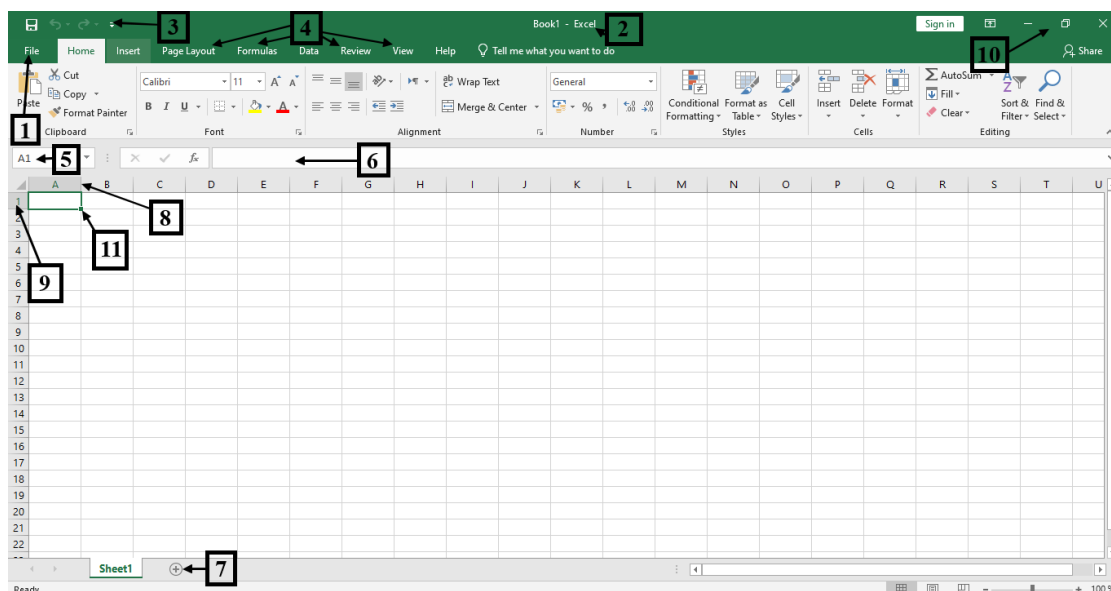


FIGURE 2.1 – Environnement Excel.

- 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.

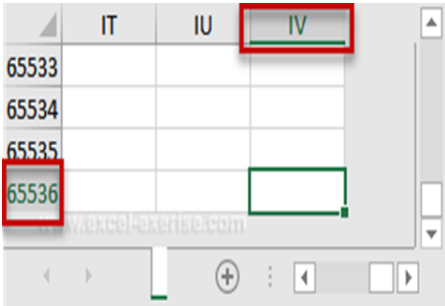
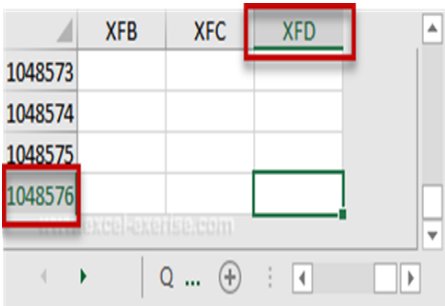
Excel 97-2003	Excel 2007 et versions ultérieures (2010, 2013, 2016, 2019, 2021, 365)
Dans un classeur xls, les limites sont de 65 536 lignes et 256 colonnes, ce qui correspond à la colonne IV.	Maintenant, avec un classeurxlsx, les limites sont de 1 048 576 lignes et 16 384 colonnes, qui est la colonne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

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).

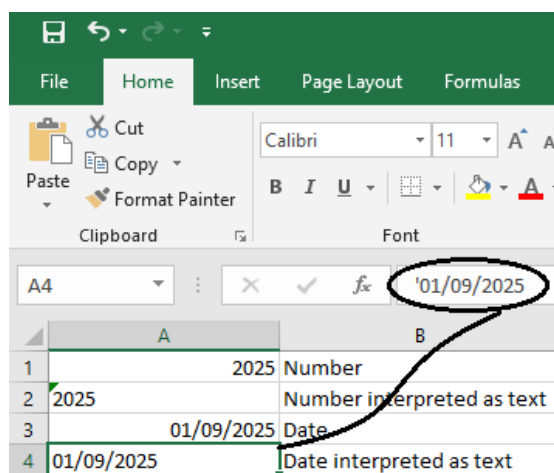


FIGURE 2.2 – Utiliser l’apostrophe pour saisir du texte dans Excel.

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...

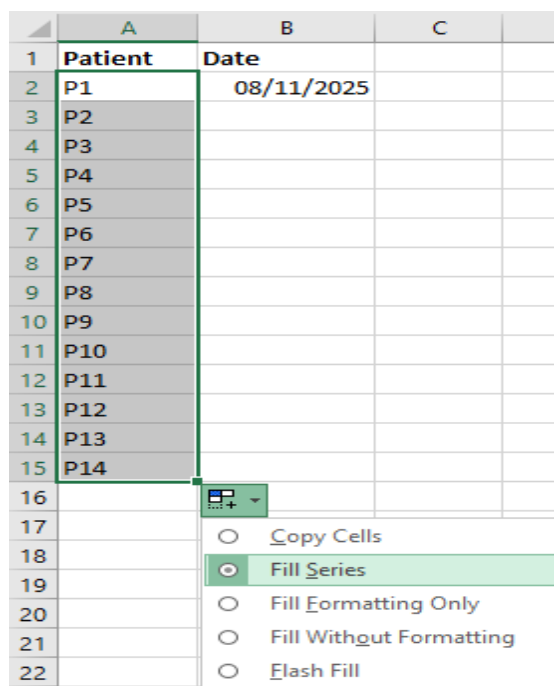


FIGURE 2.3 – Utiliser la poignée de recopie dans Excel pour compléter une série de données.

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

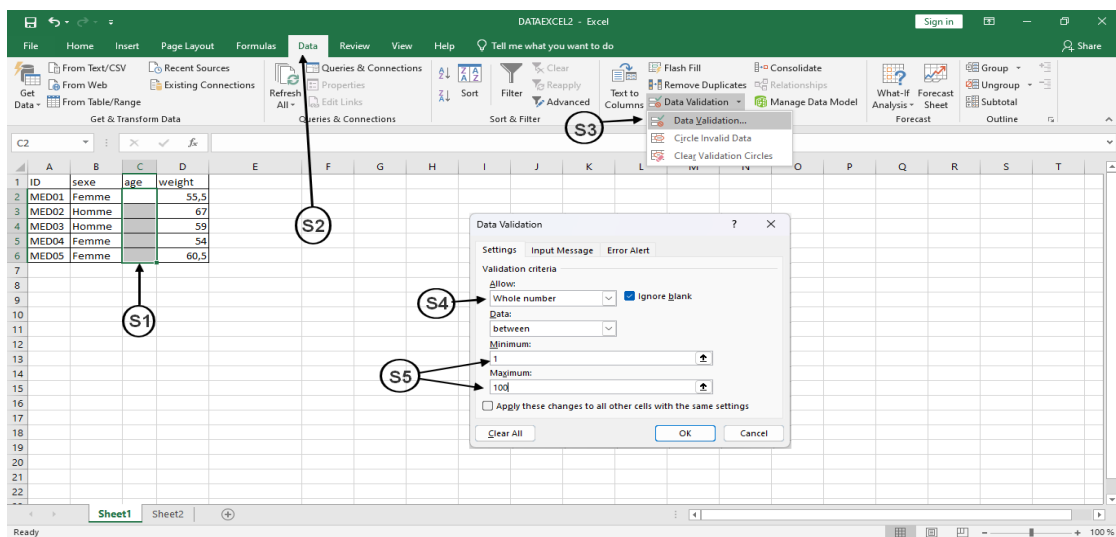


FIGURE 2.4 – Validation des données quantitatives

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).

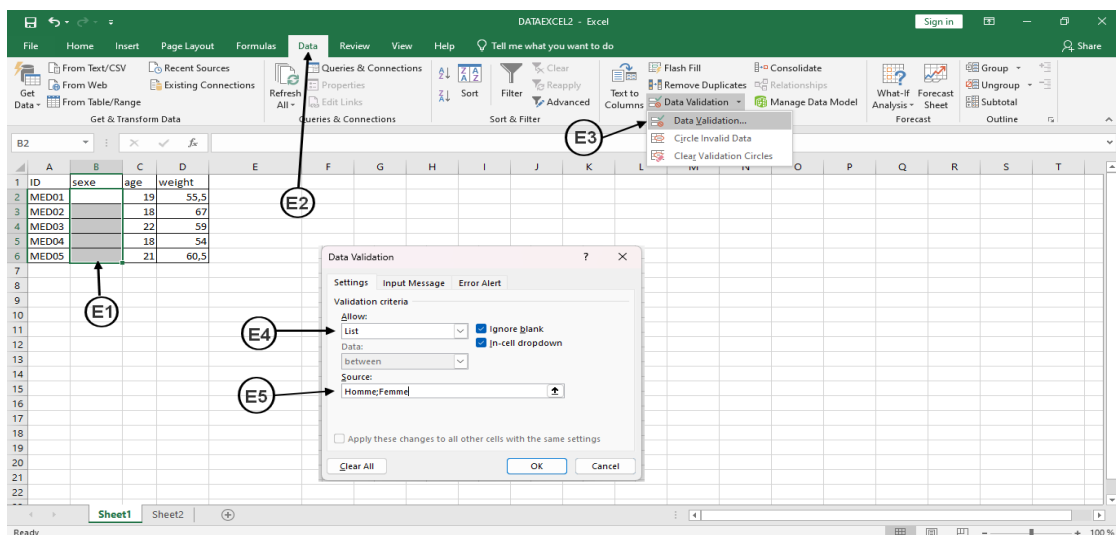
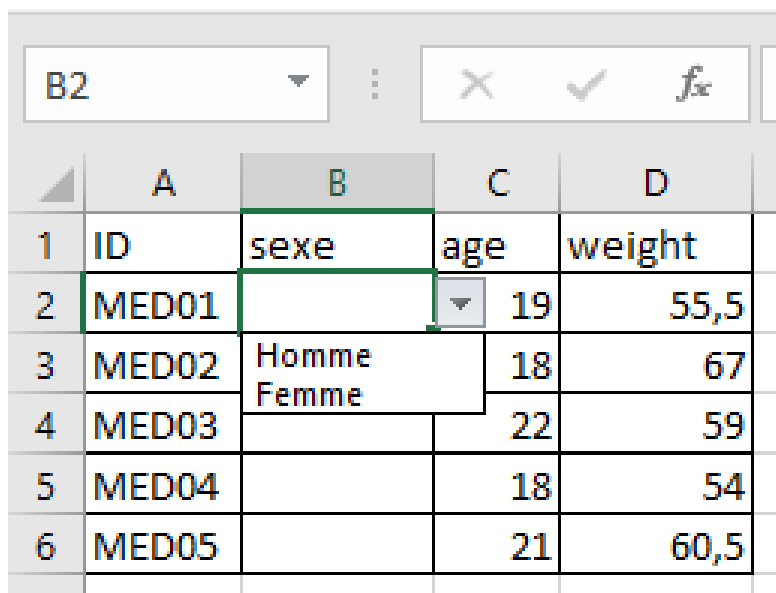


FIGURE 2.5 – Validation des données qualitatives

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.

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

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.