

Dr. Farah DJELTI

TP d'épidémiologie

Applications pratiques des
biostatistiques en épidémiologie

Ouvrage pédagogique

Destiné au master I

Biologie moléculaire et cellulaire



UNIVERSITÉ DE TLEMCCEN
FACULTÉ DES SCIENCES DE LA NATURE ET DE LA VIE
ET SCIENCES DE LA TERRE ET DE L'UNIVERS

Département de biologie
Biologie moléculaire et cellulaire

Master 1

TP d'épidémiologie

*Applications pratiques des biostatistiques
en épidémiologie*

Destiné aux étudiants de:
Master 1 biologie moleculaire et cellulaire

Dr Farah DJELTI





L'éditeur Jouda édition

Algérie- (Batna)

Première édition 1447 AH - 2026

Dépôt légal: 01/2026

ISBN: 978-9969-00-991-0

Titre de l'ouvrage: TP d'épidémiologie

Applications pratiques des biostatistiques en épidémiologie

Nom de l'auteur (a): Dr Farah DJELTI

Conception de la couverture: Zakaria Reggab

Téléphone: 00213671827876

Courriel: editionjouda@gmail.com

Tous les droits d'auteur papier, électronique, visuel et sonore
Réservé à l'auteur et ce livre n'est pas autorisé à être diffusé par
découpage ou copie ou modification sauf avec son autorisation.
L'auteur sera responsable de toute responsabilité morale, matérielle ou
légale découlant de toutes les réclamations liées au matériel, au
contenu ou aux graphiques du livre..

Table des matières

Résumé	07
Préface	08
1. Introduction	10
2. Chapitre 1 : Rappels Théoriques	11
2.1. Objectif	11
2.2. Concepts clés en statistiques	11
2.2.1. Échantillon et population	11
2.2.2. Moyenne	11
2.2.3. Médiane	12
2.2.4. Écart-type	12
2.2.5. Hypothèses nulles et alternatives	12
2.2.6. La p-value	12
2.3. Distributions de probabilité	13
2.4. Test de normalité	13
2.5. Tests paramétriques	14
2.6. Tests non paramétriques	15
2.7. Corrélation et Régression	16
2.8. Analyse de Survie	16
2.9. Notions fondamentales en épidémiologie	18
3. Chapitre 2 : Présentation du logiciel GraphPad Prism	27
3.1. Interface utilisateur et navigation	28
3.2. Le Menu Analyse dans GraphPad Prism	29
3.3. Types de tableaux et analyses statistiques	32
3.4. Création et personnalisation de graphiques	42
4. Chapitre 3 : Travaux Pratiques sur GraphPad Prism	47
4.1. TP 1 : Calcul des statistiques descriptives et création de graphiques	48
4.2. TP 2 : Test de normalité et tests t appariés et indépendants	53
4.3. TP 3 : ANOVA avec GraphPad Prism	57
4.4. TP 4 : Analyse de la corrélation entre deux variables	64
4.5. TP 5 : Régression linéaire et non linéaire	69
4.6. TP 6 : Test du chi carré, calcul du RR et de l'OR	74
4.7. TP 7 : Analyse de survie	76

5. Chapitre 4 : Corrigés Types des Travaux Pratiques	81
5.1. Corrigé type TP 1 : Statistiques descriptives et graphiques	81
5.2. Corrigé type TP 2 : Test de normalité et tests t	91
5.3. Corrigé type TP 3 : ANOVA 95	95
5.4. Corrigé type TP 4 : Analyse de corrélation	99
5.5. Corrigé type TP 5 : Régression linéaire et non linéaire	102
5.6. Corrigé type TP 6 : Test du chi carré, RR et OR	107
5.7. Corrigé type TP 7 : Analyse de survie	112
6. Chapitre 5 : Recommandations pour l'Analyse Statistique	117
7. Conclusion	120
8. Bibliographie	121
Annexes	122

Résumé

Ce manuel de travaux pratiques d'épidémiologie pour les étudiants en Master 1 de Biologie Moléculaire et Cellulaire a pour objectif d'introduire les outils statistiques et les méthodes utilisées pour analyser des données épidémiologiques. L'accent est mis sur l'utilisation du logiciel GraphPad Prism, un outil puissant pour l'analyse et la visualisation des données biomédicales.

Les étudiants apprendront à traiter des jeux de données réels provenant d'études de cohortes, d'études cas-témoins et d'essais cliniques, en appliquant des tests statistiques appropriés et en interprétant les résultats obtenus.

Compétences à acquérir

- Analyse de données épidémiologiques
- Application des modèles épidémiologiques aux études de santé publique
- Maîtrise des tests statistiques sous GraphPad Prism
- Test du Chi² pour l'association entre variables catégorielles
- Test de Student (t-test) pour comparer deux groupes
- ANOVA pour comparer plusieurs groupes
- Régression pour analyser les facteurs de risque
- Interprétation des résultats et rédaction scientifique
- Savoir expliquer la signification statistique (p-value, IC 95%)
- Présenter des graphiques épidémiologiques (courbes de survie, histogrammes, diagrammes épidémiques)

Préface

L'épidémiologie est une discipline centrale pour comprendre l'impact des facteurs biologiques, environnementaux et sociaux sur la santé des populations. Dans le cadre du parcours en Master 1 Biologie Moléculaire et Cellulaire, le module d'Épidémiologie constitue un pont essentiel entre la biologie fondamentale et la santé publique. Il offre une perspective analytique sur les données complexes qui émergent des études de population, permettant ainsi de développer une compréhension approfondie des déterminants des maladies et des traitements.

Ce polycopié a été conçu pour fournir un ensemble d'outils pratiques et théoriques nécessaires à l'analyse des données épidémiologiques dans le contexte de la biologie moléculaire et cellulaire. Il présente des rappels théoriques clairs, accompagnés d'exercices pratiques qui guident l'application des tests statistiques classiques, tels que les tests de normalité, les tests paramétriques et non paramétriques, l'ANOVA, ainsi que les techniques de corrélation et régression. Ces méthodes sont des piliers fondamentaux pour explorer et interpréter les relations entre les variables biologiques et cliniques.

À travers ce programme de travaux pratiques, l'opportunité est donnée de maîtriser des outils statistiques performants comme GraphPad Prism, un logiciel reconnu pour sa simplicité d'utilisation et sa puissance analytique. L'objectif est de se familiariser avec les techniques de traitement de données et de se préparer à une utilisation professionnelle des statistiques dans des projets de recherche en biologie.

Il est recommandé d'aborder ce polycopié avec curiosité et rigueur, car il constitue un fondement crucial pour les futures recherches et carrières dans des domaines aussi variés que la biologie moléculaire, la biotechnologie, ou encore la recherche clinique. Cet enseignement pratique permet non seulement d'acquérir des compétences en analyse statistique, mais aussi de développer un sens critique nécessaire pour interpréter les résultats dans un contexte biologique complexe.

Ce polycopié, alliant théorie et pratique, enrichira sans doute le parcours académique et préparera à relever les défis méthodologiques et analytiques rencontrés dans une future carrière scientifique.

1. Introduction

Ce polycopié de travaux pratiques (TP) est conçu pour les étudiants en Master 1 Biologie Moléculaire et Cellulaire, dans le cadre du module d'Épidémiologie. Il a pour objectif de familiariser les étudiants avec les principales méthodes statistiques appliquées à l'épidémiologie, en mettant l'accent sur l'analyse des données biologiques et cliniques. Les TP incluent une série d'analyses statistiques fondamentales utilisées pour évaluer les relations entre les facteurs de risque, les traitements et les réponses biologiques dans les populations humaines.

Au cours de ce semestre, les étudiants seront amenés à maîtriser diverses techniques statistiques telles que les statistiques descriptives, les tests de normalité, les tests t paramétriques et non paramétriques, l'ANOVA, ainsi que l'analyse de la corrélation et de la régression. Ces compétences sont essentielles pour analyser et interpréter des données biologiques, particulièrement dans le domaine de la biologie moléculaire et cellulaire, où les analyses statistiques permettent de tirer des conclusions solides à partir de données complexes.

Les TP seront accompagnés d'exercices pratiques utilisant GraphPad Prism, un logiciel puissant permettant de réaliser des analyses statistiques avancées, de générer des graphiques et d'interpréter les résultats de manière claire et accessible. Ce polycopié guidera les étudiants tout au long des étapes des analyses et fournira les rappels théoriques nécessaires pour bien comprendre chaque test statistique utilisé.

2. Chapitre 1: Rappel Théorique

2.1. Objectif

Ce chapitre a pour objectif de revoir les concepts théoriques essentiels pour la compréhension des tests statistiques utilisés dans le cadre de ces travaux pratiques. Les tests statistiques sont des outils fondamentaux en biologie, en médecine, ainsi que dans d'autres disciplines scientifiques. Ils permettent d'analyser les données et de tirer des conclusions quant à la validité d'hypothèses formulées à partir de ces données. Ce chapitre propose une révision des concepts de base en statistiques, avec une attention particulière à la distinction entre les tests paramétriques et non paramétriques, les distributions de probabilité, ainsi que les hypothèses testées dans chaque test.

2.2. Concepts clés en statistiques

2.2.1. Échantillon et population

- **Population:** Une population est l'ensemble complet des éléments ou individus qui partagent une caractéristique commune, sur lesquels nous souhaitons tirer des conclusions. Par exemple, une population pourrait être tous les patients atteints de cancer du poumon dans un pays donné.
- **Échantillon:** Un échantillon est un sous-ensemble de la population. Puisque l'étude de toute une population est souvent impraticable, les chercheurs recueillent des échantillons pour effectuer des analyses et faire des inférences sur la population. L'échantillon doit être représentatif de la population pour que les résultats puissent être généralisés.

2.2.2. Moyenne

- **Moyenne:** La moyenne est la valeur centrale d'un ensemble de données. Elle est obtenue en additionnant toutes les valeurs de l'échantillon et en les divisant par le nombre total d'observations. Elle est souvent utilisée pour résumer une série de données en une seule valeur représentative. Cependant, elle peut être influencée par des valeurs extrêmes (outliers), d'où la

nécessité d'utiliser d'autres mesures comme la médiane dans certains cas.

2.2.3. Médiane

- **Médiane:** La médiane est la valeur qui sépare les données en deux parties égales. Elle correspond au point où 50 % des données sont inférieures et 50 % des données sont supérieures. Contrairement à la moyenne, la médiane n'est pas influencée par les valeurs extrêmes et est donc plus robuste dans les jeux de données asymétriques ou comportant des outliers.

2.2.4. Écart-type

- **Écart-type:** L'écart-type mesure la dispersion des données autour de la moyenne. Il donne une idée de la variation ou de la volatilité des données. Plus l'écart-type est élevé, plus les données sont dispersées. Un faible écart-type indique que les données sont concentrées autour de la moyenne.

2.2.5. Hypothèses nulles et alternatives

- **Hypothèse nulle (H0):** Cette hypothèse stipule qu'il n'y a pas de différence significative entre les groupes ou variables. Elle représente la position de "non-effet" ou de "non-relation".
- **Hypothèse alternative (H1):** Cette hypothèse postule qu'il existe une différence ou une relation significative entre les groupes ou variables. Les tests statistiques ont pour objectif de déterminer si les données observées permettent de rejeter l'hypothèse nulle en faveur de l'hypothèse alternative.

2.2.6. La p-value

La p-value (valeur de probabilité) est une mesure utilisée pour déterminer la significativité des résultats d'un test statistique. Elle représente la probabilité d'obtenir des résultats aussi extrêmes que ceux observés, sous l'hypothèse nulle.

- Si $p < 0.05$, cela indique que les résultats sont statistiquement significatifs, et l'hypothèse nulle peut être rejetée.

- Si $p > 0.05$, cela suggère que les résultats ne sont pas suffisamment probants pour rejeter l'hypothèse nulle.

2.3. Distributions de probabilité

2.3.1. Distribution normale

La distribution normale est l'une des distributions de probabilité les plus courantes en statistiques. Elle est caractérisée par une courbe en forme de cloche, symétrique autour de la moyenne. Cette distribution est utilisée dans de nombreux tests paramétriques.

- **Moyenne (μ):** La valeur centrale de la distribution.
- **Écart-type (σ):** Mesure de la dispersion des valeurs autour de la moyenne.
- **Propriétés:** Environ 68 % des données se trouvent à moins d'un écart-type de la moyenne, 95 % à moins de deux écarts-types, et 99.7 % à moins de trois écarts-types.

La normalité des données est une condition préalable pour l'utilisation de nombreux tests paramétriques. Si une distribution normale ne peut être supposée, des tests non paramétriques doivent être utilisés.

2.3.2. Variance

La variance est une mesure de la dispersion des valeurs dans un ensemble de données. Elle quantifie l'écart moyen au carré par rapport à la moyenne. Une faible variance indique que les données sont concentrées autour de la moyenne, tandis qu'une variance élevée indique une large dispersion.

La variance est essentielle pour évaluer l'homogénéité des groupes dans des tests comme l'ANOVA.

2.4. Test de normalité

2.4.1. Concept de normalité

Le test de normalité vise à déterminer si un jeu de données suit une distribution normale, c'est-à-dire une distribution symétrique en

forme de cloche. La normalité est un préalable pour de nombreux tests statistiques paramétriques, qui supposent que les données suivent une distribution normale.

2.4.2. Test de Shapiro-Wilk

Le test de Shapiro-Wilk est l'un des tests les plus utilisés pour vérifier la normalité des données. Il teste l'hypothèse nulle selon laquelle les données suivent une distribution normale. Si la p-value est inférieure à 0.05, cela signifie que l'hypothèse nulle est rejetée, indiquant que les données ne suivent pas une distribution normale.

2.4.3. Test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov compare la distribution des données observées avec une distribution théorique, comme la normale. Il est utilisé pour évaluer si les données suivent une distribution spécifique, en comparant les fréquences observées avec les fréquences attendues.

2.5. Tests paramétriques

2.5.1. Test t de Student

Le test t de Student est utilisé pour comparer les moyennes de deux groupes indépendants. Ce test paramétrique repose sur l'hypothèse que les données suivent une distribution normale et que les variances des deux groupes sont égales.

- **Hypothèse nulle (H0):** Les moyennes des deux groupes sont égales.
- **Hypothèse alternative (H1):** Les moyennes des deux groupes sont différentes.

2.5.2. Test du Chi Carré

Le test du chi carré est utilisé pour tester l'indépendance entre deux variables qualitatives. Il permet de déterminer si deux variables sont associées ou indépendantes dans un tableau de contingence.

- **Hypothèse nulle (H0):** Les variables sont indépendantes (il n'y a pas d'association).

- **Hypothèse alternative (H1):** Les variables sont dépendantes (il y a une association).

2.5.3. ANOVA (Analyse de Variance)

L'ANOVA est utilisée pour comparer les moyennes de trois groupes ou plus. Elle permet de tester si au moins une des moyennes des groupes est différente des autres.

- **Hypothèse nulle (H0):** Les moyennes des différents groupes sont égales.
- **Hypothèse alternative (H1):** Au moins une moyenne de groupe est différente.

2.6. Tests non paramétriques

2.6.1. Test de Mann-Whitney

Le test de Mann-Whitney est un test non paramétrique utilisé pour comparer deux groupes indépendants. Contrairement au test t, il ne suppose pas que les données suivent une distribution normale. Ce test est basé sur les rangs des valeurs, ce qui le rend plus robuste lorsqu'il y a des violations de la normalité dans les données.

- **Hypothèse nulle (H0):** Les deux groupes ont la même distribution.
- **Hypothèse alternative (H1):** Les deux groupes ont des distributions différentes.

2.6.2. Test de Wilcoxon

Le test de Wilcoxon est utilisé pour comparer les médianes de deux groupes appariés. Il est une alternative au test t de Student lorsque les données ne suivent pas une distribution normale.

- **Hypothèse nulle (H0):** Il n'y a pas de différence entre les médianes des deux groupes.
- **Hypothèse alternative (H1):** Il existe une différence entre les médianes des deux groupes.

2.7. Corrélation et Régression

2.7.1. Corrélation de Pearson

La corrélation de Pearson mesure la relation linéaire entre deux variables continues. Un coefficient de corrélation proche de +1 ou -1 indique une forte relation positive ou négative, tandis qu'un coefficient proche de 0 indique l'absence de relation linéaire.

2.7.2. Régression linéaire

La régression linéaire permet de prédire la valeur d'une variable dépendante (Y) en fonction d'une ou plusieurs variables indépendantes (X). Elle sert à estimer la relation entre les variables.

2.8. Analyse de Survie

2.8.1. Méthode de Kaplan-Meier

La méthode de Kaplan-Meier permet d'estimer la fonction de survie en tenant compte des données censurées. Elle produit une courbe qui montre comment la probabilité de survie diminue au fil du temps.

2.8.2. Test de log-rank

Le test de log-rank est utilisé pour comparer les courbes de survie entre deux ou plusieurs groupes. Il teste si les différences observées dans les courbes de survie sont statistiquement significatives.

- **Hypothèse nulle (H0):** Les courbes de survie des groupes sont identiques.
- **Hypothèse alternative (H1):** Les courbes de survie sont différentes.

Tableau 1. Résumé des Concepts et Tests Statistiques.

Concept	Description
Population vs Échantillon	La population est un groupe complet d'individus d'intérêt, l'échantillon est un sous-ensemble représentatif.
Moyenne	La valeur centrale des données, influencée par les valeurs extrêmes.
Médiane	La valeur centrale qui sépare les données en deux parties égales, plus robuste aux outliers.
Écart-type	Mesure de la dispersion des données autour de la moyenne. Plus l'écart-type est élevé, plus les données sont dispersées.
Hypothèses	H0 (nulle): Pas de différence significative, H1 (alternative): Il existe une différence significative.
P-value	Si $p < 0.05$, rejet de l'hypothèse nulle, indiquant des résultats significatifs.
Distribution normale	Courbe en forme de cloche, utilisée dans les tests paramétriques.
Test de normalité	Shapiro-Wilk et Kolmogorov-Smirnov pour tester si les données suivent une distribution normale.
Tests paramétriques	Test t de Student: Compare les moyennes de deux groupes, ANOVA: Compare les moyennes de trois groupes ou plus.

Tests non paramétriques Test de Mann-Whitney: Compare deux groupes sans supposer une distribution normale. Test de Wilcoxon: Compare les médianes de groupes appariés.

Corrélation Mesure de la relation linéaire entre deux variables continues (Pearson).

Régression linéaire Permet de prédire la valeur d'une variable dépendante à partir d'une ou plusieurs variables indépendantes.

Test du Chi Carré Teste l'indépendance entre deux variables qualitatives dans un tableau de contingence.

Analyse de survie Méthode de Kaplan-Meier pour estimer la fonction de survie, test de log-rank pour comparer les courbes de survie.

2.9. Notions fondamentales en épidémiologie

L'épidémiologie est souvent qualifiée de science fondamentale de la santé publique. Pour les biologistes, habitués à l'échelle moléculaire ou cellulaire, l'épidémiologie impose un changement de perspective: le passage de l'infiniment petit (le gène, la cellule) ou de l'individuel (le patient clinique) à l'échelle macroscopique de la population. Ce chapitre pose les bases théoriques nécessaires pour comprendre non seulement comment utiliser les tests statistiques, mais surtout pourquoi nous les utilisons dans le contexte de la recherche en santé.

2.9.1. Définition et portée de l'épidémiologie

Définition classique et élargie L'épidémiologie se définit traditionnellement comme l'étude de la distribution et des déterminants des états de santé ou des événements liés à la santé (y

compris les maladies) dans des populations spécifiques, et l'application de cette étude au contrôle des problèmes de santé.

Cette définition, bien que concise, renferme deux concepts clés qui structurent toute la démarche scientifique en santé publique:

1. **La Distribution:** L'épidémiologie ne s'intéresse pas au hasard. Les maladies ne frappent pas au hasard ; elles suivent des modèles (patterns). L'épidémiologiste cherche à savoir "qui" est touché, "où" et "quand". Il s'agit de quantifier la fréquence des maladies.
2. **Les Déterminants:** Ce sont les causes ou les facteurs de risque qui influencent cette distribution. Il peut s'agir de facteurs biologiques (génétique, immunité), environnementaux (pollution, vecteurs infectieux), comportementaux (tabagisme, sédentarité) ou sociaux. En biologie moléculaire, ces déterminants sont souvent des marqueurs génétiques, des niveaux d'expression protéique ou des réponses cellulaires à des agents pathogènes.

Les trois axes fondamentaux: Temps, Lieu, Personne
Pour décrire la distribution d'une maladie, l'épidémiologie s'articule systématiquement autour de trois axes, souvent appelés la "triade descriptive":

- **Temps (Quand?):** L'analyse temporelle est cruciale. Elle permet d'identifier le caractère saisonnier d'une infection (comme la grippe), de repérer une épidémie soudaine (pic de cas sur une courbe épidémique) ou d'observer des tendances séculaires (l'augmentation du diabète sur plusieurs décennies). En statistique, cela se traduit souvent par des analyses de séries chronologiques ou des analyses de survie.
- **Lieu (Où?):** La géographie de la santé permet de comprendre l'impact de l'environnement. Une concentration de cas de cancer autour d'une usine, ou la limitation d'une maladie parasitaire à une zone tropicale spécifique, oriente la recherche des causes.

- **Personne (Qui?)**: C'est l'axe le plus directement lié à la biologie. Qui est malade? Les caractéristiques inhérentes aux individus (âge, sexe, origine ethnique, statut immunitaire, profil génétique) sont les premiers facteurs de confusion ou de risque que l'on analyse statistiquement. Par exemple, l'ajustement statistique sur l'âge et le sexe est une procédure standard pour éviter de tirer des conclusions erronées.

Distinction entre épidémiologie descriptive et analytique
La démarche épidémiologique se divise généralement en deux phases successives, chacune faisant appel à des outils statistiques différents:

1. **L'épidémiologie descriptive**: Elle est la première étape de l'investigation. Son objectif est de mesurer la fréquence de la maladie (prévalence, incidence) et de décrire ses variations selon les axes Temps/Lieu/Personne. Elle génère des hypothèses mais ne les teste pas.
 - **Exemple**: "On observe une prévalence de la mutation X plus élevée chez les patients atteints de la maladie Y."
2. **L'épidémiologie analytique (ou étiologique)**: Son but est de rechercher les causes et les facteurs de risque. Elle teste les hypothèses générées par l'étape descriptive en comparant des groupes. C'est ici que l'analyse statistique inférentielle (tests de comparaison, calcul de risques) prend tout son sens.
 - **Exemple**: "L'exposition au facteur Z multiplie par 4 le risque de développer la maladie (Odds Ratio = 4, $p < 0.05$)."

Importance centrale pour la santé publique

L'épidémiologie n'est pas une science théorique ; c'est une science de l'action. Les données qu'elle produit guident la prise de décision sanitaire. Qu'il s'agisse de retirer un médicament du marché, de lancer une campagne de vaccination ou de recommander un changement de régime alimentaire, ces décisions reposent sur des preuves statistiques. Pour un étudiant en biologie, comprendre cela signifie que ses résultats

de laboratoire, une fois analysés statistiquement dans un contexte populationnel, peuvent avoir un impact direct sur la gestion de la santé humaine.

2.9.2. Le rôle incontournable de la biostatistique

Si l'épidémiologie fournit le cadre conceptuel et les questions de recherche, la biostatistique fournit les outils pour y répondre. La relation entre les deux disciplines est symbiotique: l'épidémiologie sans biostatistique ne serait que de la narration anecdotique.

Le passage de l'observation clinique à l'analyse populationnelle En médecine clinique, le médecin traite un individu unique. En biologie expérimentale, on travaille souvent sur des modèles très contrôlés (lignées cellulaires, souris consanguines) où la variabilité est minimisée. En épidémiologie, on travaille sur des populations humaines libres, caractérisées par une immense hétérogénéité génétique, comportementale et environnementale.

La statistique est l'outil qui permet d'agréger ces données individuelles disparates pour en tirer une "vérité" moyenne applicable au groupe. Elle permet de transformer des milliers de points de données (le taux de cholestérol de 10 000 patients) en paramètres résumés (moyenne, écart-type) intelligibles.

La nécessité de l'analyse statistique: Gérer la variabilité biologique

Le vivant est intrinsèquement variable. Si l'on administre le même traitement à deux patients, ils ne réagiront pas exactement de la même manière. Cette variabilité biologique est le bruit de fond ("noise") qui masque le signal réel ("signal") que le chercheur tente de détecter (par exemple, l'efficacité d'un médicament).

Les tests statistiques sont conçus pour distinguer ce qui relève de la variabilité naturelle (le hasard) de ce qui relève d'un effet réel du facteur étudié. Sans statistiques, il est impossible de savoir si une différence observée entre deux groupes est réelle ou simplement due à la fluctuation d'échantillonnage.

Quantifier le risque et l'incertitude

En science, l'incertitude ne peut jamais être totalement éliminée, mais elle peut être mesurée. C'est le rôle fondamental des intervalles de confiance (IC) et des p-values que vous calculerez avec GraphPad Prism.

- L'épidémiologie ne dit jamais "Le tabac cause le cancer à coup sûr chez cet individu".
- Elle dit plutôt: "Les fumeurs ont statistiquement 20 fois plus de risques de développer un cancer que les non-fumeurs, avec une marge d'erreur de X%".

Cette quantification du risque (via le Risque Relatif ou l'Odds Ratio) est essentielle. Elle permet de hiérarchiser les problèmes de santé. Un facteur de risque qui double la probabilité d'une maladie rare n'a pas le même poids en santé publique qu'un facteur qui augmente de 10% une maladie très fréquente.

Valider scientifiquement les observations (Significativité vs Hasard) C'est le cœur du processus de décision binaire en recherche: "Résultat significatif" ou "Non significatif". Les tests d'hypothèse (comme le test t ou le Chi²) calculent la probabilité (p-value) que les différences observées soient dues au pur hasard. Si cette probabilité est très faible (généralement inférieure à 5%, soit $p < 0.05$), on rejette l'explication du hasard. C'est ce processus qui donne sa validité scientifique à une étude épidémiologique. Sans cette validation mathématique, aucune conclusion biologique ne peut être acceptée par la communauté scientifique.

2.9.3. Application des tests statistiques aux enquêtes épidémiologiques

Il existe une correspondance logique et stricte entre le type d'enquête épidémiologique menée (le design de l'étude) et les outils statistiques à employer. GraphPad Prism facilite cette démarche, mais il est crucial de choisir le bon test pour la bonne étude.

Les Études Transversales et Cas-Témoins: Le domaine des fréquences

Ces études sont souvent rétrospectives ou ponctuelles. On y observe des états (Malade / Pas malade) et des expositions (Exposé / Non exposé). Les variables sont majoritairement qualitatives (catégorielles).

- **Contexte:** Vous comparez un groupe de patients atteints d'une pathologie (Cas) à des individus sains (Témoins) pour voir si un marqueur génétique est plus fréquent chez les malades.
- **Outils Statistiques:**
 - Le test du Chi-carré (Chi^2 ou χ^2): C'est le roi des tests pour ces études. Il compare les effectifs observés aux effectifs théoriques attendus si les variables étaient indépendantes. Il répond à la question: "La distribution de l'exposition est-elle différente chez les malades et les non-malades? "
 - L'Odds Ratio (OR): Dans une étude cas-témoins, on ne peut pas calculer l'incidence (la vitesse d'apparition de la maladie). On utilise donc l'Odds Ratio (Rapport de cotes) pour estimer la force de l'association. Si $\text{OR} = 3$, l'exposition est 3 fois plus associée à la maladie. GraphPad Prism calcule automatiquement l'OR et son intervalle de confiance lors de l'analyse de tableaux de contingence.

Les Études de Cohortes et Essais Cliniques: Le domaine de l'incidence et du suivi

Ces études sont longitudinales (suivi dans le temps) et souvent prospectives. Elles offrent le plus haut niveau de preuve car on observe la survenue de la maladie en temps réel. Les données collectées sont souvent mixtes: qualitatives (survenue de l'événement) et quantitatives (paramètres biologiques: glycémie, charge virale, taux d'anticorps).

- **Contexte:** On suit un groupe exposé à un traitement et un groupe placebo sur 5 ans.

- **Outils Statistiques:**

- **Le Risque Relatif (RR):** Puisque l'on connaît le nombre de nouveaux cas apparus sur la période, on peut calculer des taux d'incidence. Le RR compare l'incidence chez les exposés vs non-exposés. C'est une mesure plus directe et intuitive que l'OR.
- **Tests t de Student et ANOVA:** Ces études génèrent beaucoup de données biologiques quantitatives. Pour comparer la moyenne de la glycémie entre le groupe traité et le groupe placebo à la fin de l'étude, on utilise le test t (pour 2 groupes) ou l'ANOVA (pour >2 groupes). Ces tests valident l'efficacité biologique du traitement.
- **Analyse de Survie (Kaplan-Meier et Log-Rank):** En épidémiologie clinique, notamment en cancérologie, la question n'est pas seulement "le patient est-il vivant? ", mais "combien de temps a-t-il survécu? ". La méthode de Kaplan-Meier permet de tracer la courbe de décroissance de la survie dans la cohorte, en tenant compte des données censurées (patients perdus de vue). Le test du Log-Rank compare statistiquement ces courbes. C'est un outil indispensable pour évaluer le pronostic.

Tableau 2. Guide de choix des tests statistiques en épidémiologie sous GraphPad Prism

Ce tableau résume la relation essentielle entre le design de l'étude et le test à sélectionner.

Type d'Étude Épidémiologique	Question posée (Objectif)	Type de Données	Test Statistique Principal (GraphPad Prism)	Mesure d'Association
Étude Cas-Témoins (Rétrospective)	L'exposition est-elle plus fréquente chez les malades?	Qualitatives (Malade/Sain vs Exposé/Non)	Test du Chi-carré (χ^2) ou Test exact de Fisher	Odds Ratio (OR) (Rapport de cotes)
Étude de Cohorte (Prospective)	L'exposition augmente-t-elle le risque de devenir malade?	Qualitatives (Suivi de l'apparition de la maladie)	Test du Chi-carré	Risque Relatif (RR)
Essai Clinique / Expérimental (Comparaison biologique)	Le traitement modifie-t-il un paramètre biologique?	Quantitatives (ex: taux de cholestérol, charge virale)	Test t de Student (2 groupes) ANOVA (> 2 groupes)	Différence des moyennes
Étude de Pronostic (Suivi temporel)	Le traitement prolonge-t-il la durée de vie?	Temps + Événement (Données de survie/censurées)	Analyse de Kaplan-Meier Comparaison: Log-Rank Test	Hazard Ratio (HR) Médiane de survie

L'analyse statistique comme outil de preuve (Evidence-Based Medicine)

L'aboutissement de l'application de ces tests statistiques dans les enquêtes épidémiologiques est la production de données probantes. La médecine et la biologie modernes ne se basent plus sur l'intuition ou l'autorité d'un expert, mais sur l'EBM (Evidence-Based Medicine). Lorsqu'une analyse ANOVA montre une différence significative, ou qu'un Risque Relatif est supérieur à 1 avec un intervalle de confiance étroit, cela devient une "preuve". Ces preuves permettent:

1. D'identifier les facteurs de risque: Confirmer que l'amiante cause le mésothéliome ou que tel gène prédispose au cancer du sein.
2. D'évaluer l'efficacité des traitements: Prouver qu'une nouvelle molécule augmente la survie sans récidence par rapport au traitement standard.

En conclusion de ce rappel théorique, il est impératif pour l'étudiant de comprendre que le logiciel GraphPad Prism est un outil puissant au service de ce raisonnement épidémiologique. Le logiciel exécute le calcul mathématique complexe, mais c'est à l'étudiant, fort de sa compréhension des concepts de distribution, de risque et de significativité, qu'il appartient de poser la bonne question, de choisir le design d'étude approprié et d'interpréter la validité biologique des résultats statistiques fournis.

3. Chapitre 2: Présentation du logiciel

GraphPad Prism est un logiciel de référence pour l'analyse statistique et la création de graphiques, spécialement conçu pour les scientifiques et les chercheurs en biologie, médecine et autres domaines des sciences de la vie. Contrairement à d'autres logiciels statistiques complexes, Prism combine une interface intuitive avec des fonctionnalités puissantes, ce qui le rend accessible même aux utilisateurs sans expertise en programmation ou en statistiques avancées.

Prism est particulièrement apprécié pour sa capacité à guider les utilisateurs à travers les analyses statistiques, en fournissant des explications claires et des conseils pour éviter les erreurs courantes. Il permet également de créer des graphiques professionnels, prêts à être intégrés dans des publications scientifiques.

Points forts de GraphPad Prism:

- Interface conviviale et visuelle.
- Prise en charge des analyses statistiques courantes et avancées.
- Création de graphiques personnalisables et de haute qualité.
- Intégration transparente avec d'autres outils (Excel, PowerPoint, etc.).
- Documentation complète et ressources d'apprentissage abondantes.

GraphPad Prism est un logiciel de référence pour l'analyse statistique et la création de graphiques, spécialement conçu pour les scientifiques et les chercheurs en biologie, médecine et autres domaines des sciences de la vie. Contrairement à d'autres logiciels statistiques complexes, Prism combine une interface intuitive avec des fonctionnalités puissantes, ce qui le rend accessible même aux utilisateurs sans expertise en programmation ou en statistiques avancées.

Prism est particulièrement apprécié pour sa capacité à guider les utilisateurs à travers les analyses statistiques, en fournissant des explications claires et des conseils pour éviter les erreurs courantes. Il permet également de créer des graphiques professionnels, prêts à être intégrés dans des publications scientifiques.

Points forts de GraphPad Prism:

- Interface conviviale et visuelle.
- Prise en charge des analyses statistiques courantes et avancées.
- Création de graphiques personnalisables et de haute qualité.
- Intégration transparente avec d'autres outils (Excel, PowerPoint, etc.).
- Documentation complète et ressources d'apprentissage abondantes.

3.1. Interface utilisateur et navigation

Lorsque vous ouvrez GraphPad Prism pour la première fois, vous êtes accueilli par une interface bien organisée, conçue pour simplifier votre flux de travail. Voici un aperçu des éléments clés de l'interface:

3.1.1. La fenêtre d'accueil:

- Au lancement, Prism affiche une fenêtre vous invitant à choisir le type de données que vous souhaitez analyser (par exemple, données numériques, courbes de dose-réponse, analyse de survie, etc.). Cette étape est cruciale car elle détermine la structure de votre projet.

3.1.2. Les onglets principaux:

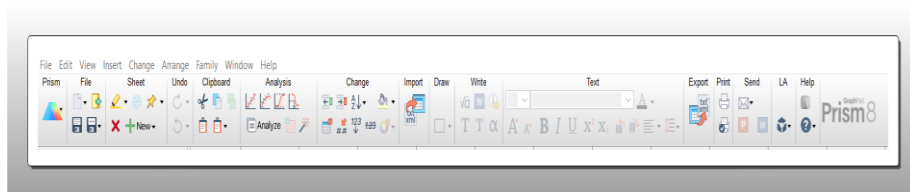
- Onglet "Data" (Données): C'est ici que vous saisissez ou importez vos données. Les tables sont organisées en colonnes et en lignes, avec des options pour définir des répétitions, des groupes, ou des sous-catégories.
- Onglet "Results" (Résultats): Après avoir effectué une analyse, les résultats apparaissent dans cet onglet. Prism génère des

tableaux et des graphiques automatiquement, mais vous pouvez les personnaliser selon vos besoins.

- Onglet "Graphs" (Graphiques): Cet onglet affiche les représentations visuelles de vos données. Vous pouvez créer plusieurs types de graphiques (barres, lignes, nuages de points, etc.) et les personnaliser en fonction de vos préférences.
- Onglet "Layouts" (Mises en page): Cet onglet vous permet de combiner plusieurs graphiques et tableaux sur une seule page, idéal pour créer des figures complexes pour vos publications.
- Onglet "Notes" (Notes): Utilisez cet espace pour ajouter des commentaires, des explications ou des références à votre projet.

3.1.3. La barre d'outils:

- La barre d'outils en haut de l'écran donne un accès rapide aux fonctions courantes, comme l'ajout de nouvelles tables, la création de graphiques, ou l'exécution d'analyses statistiques.

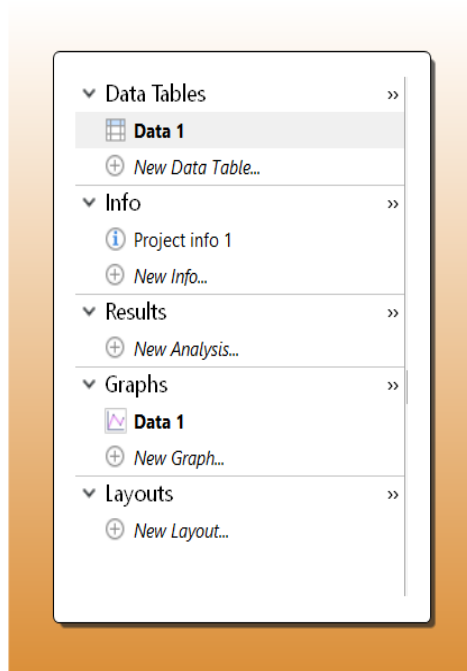


3.2. Création d'un nouveau projet

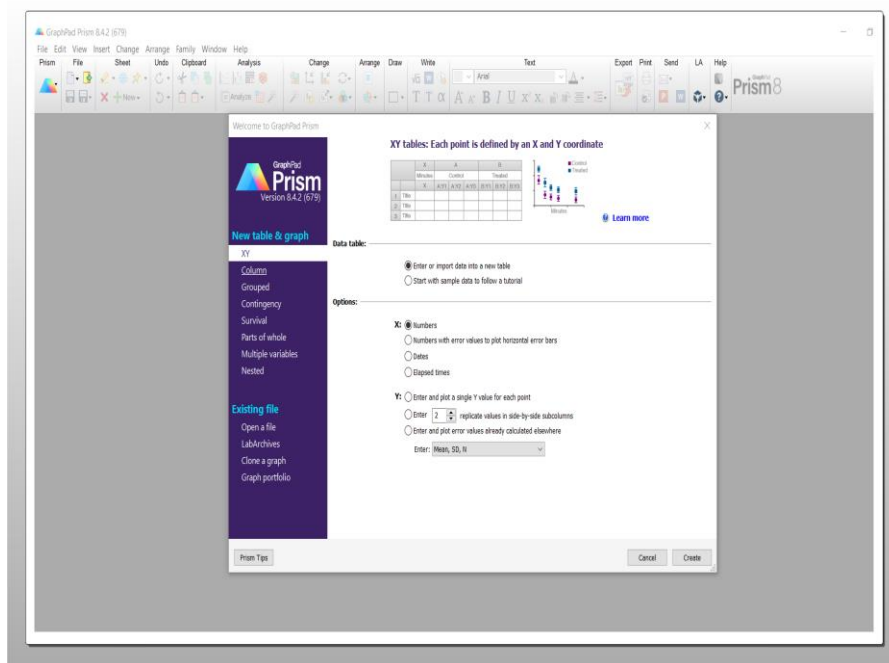
La création d'un nouveau projet dans GraphPad Prism est simple et intuitive. Voici les étapes à suivre:

Étape 1: Choisir le type de données

- Lorsque vous lancez Prism, une fenêtre intitulée "Welcome to GraphPad Prism" s'affiche.
- Sélectionnez le type de données que vous souhaitez analyser. Par exemple:
 - XY: Pour des données numériques simples (p. ex., courbes de calibration).



- Column: Pour des données groupées (p. ex., comparaison de moyennes entre groupes).
- Survival: Pour des analyses de survie (p. ex., courbes de Kaplan-Meier).
- Cliquez sur "Create" pour ouvrir une nouvelle table de données.



Étape 2: Saisir ou importer des données

- Dans l'onglet "Data", saisissez manuellement vos données ou importez-les depuis un fichier Excel ou CSV.
- Organisez vos données en colonnes et en lignes. Prism vous permet de définir des répétitions, des sous-groupes, ou des valeurs manquantes.

Étape 3: Nommer et sauvegarder votre projet

- Donnez un nom à votre projet en cliquant sur "File" > "Save As".
- Choisissez un emplacement sur votre ordinateur pour enregistrer le fichier. Les projets Prism sont sauvegardés au format .pzfx.

Étape 4: Explorer les options d'analyse et de visualisation

- Une fois vos données saisies, vous pouvez commencer à explorer les options d'analyse statistique et de création de

graphiques. Prism propose des menus déroulants pour vous guider dans le choix des tests appropriés.

Conseils pour les débutants:

- Prenez le temps de bien choisir le type de données au début de votre projet, car cela influence la structure de votre analyse.
- Utilisez l'onglet "Notes" pour documenter chaque étape de votre travail, ce qui facilitera la reproductibilité de vos analyses.
- N'hésitez pas à consulter l'aide intégrée de Prism (accessible via F1) pour obtenir des explications détaillées sur chaque fonctionnalité.

3.3. Saisie et Organisation des Données

3.3.1. Types de données (numériques, catégorielles, etc.)

Dans GraphPad Prism, la première étape pour analyser vos données consiste à comprendre leur nature et à les organiser correctement. Prism prend en charge plusieurs types de données, chacun étant adapté à des analyses spécifiques. Voici les principaux types de données que vous rencontrerez:

a) Données numériques:

- Ces données sont représentées par des valeurs numériques continues ou discrètes.
- Exemples: concentrations de protéines, taux de croissance, mesures de fluorescence.
- Utilisation: Analyse de régression, tests t, ANOVA, etc.

b) Données catégorielles:

- Ces données sont divisées en catégories ou groupes distincts.
- Exemples: sexe (homme/femme), traitement (contrôle/traité), type de cellule (A/B/C).
- Utilisation: Tests de chi-carré, ANOVA à effets fixes, etc.

c) Données de survie:

- Ces données sont utilisées pour analyser le temps jusqu'à un événement spécifique (p. ex., décès, récurrence).
- **Exemples:** Courbes de Kaplan-Meier, analyse de risque proportionnel de Cox.

d) Données de séries temporelles:

- Ces données sont collectées à intervalles réguliers dans le temps.
- Exemples: Mesures de croissance bactérienne sur plusieurs jours, suivi de la pression artérielle.
- Utilisation: Analyse de tendances, régression temporelle.

e) Données appariées ou répétées:

- Ces données proviennent de mesures répétées sur les mêmes sujets ou échantillons.
- Exemples: Mesures avant et après traitement sur les mêmes patients.
- Utilisation: Tests t appariés, ANOVA à mesures répétées.

3.3.2. Création de tables de données

Une fois que vous avez identifié le type de données que vous utilisez, vous pouvez créer des tables de données dans Prism. Voici comment procéder:

Étape 1: Choisir le type de table

- Lorsque vous créez un nouveau projet, Prism vous demande de sélectionner le type de table adapté à vos données. Par exemple:
 - **Table "XY":** Pour des données numériques simples (p. ex., courbes de calibration).
 - **Table "Column":** Pour des données groupées (p. ex., comparaison de moyennes entre groupes).

- Table "Grouped": Pour des données organisées en sous-groupes.
- Table "Survival": Pour des analyses de survie.

a. XY Table

Type d'analyse statistique:

- **Corrélation:** Pour mesurer la relation linéaire entre deux variables continues.
- **Régression linéaire:** Pour modéliser la relation entre une variable indépendante (X) et une variable dépendante (Y).
- **Régression non-linéaire:** Pour ajuster un modèle plus complexe à vos données.
- **ANOVA (Analyse de la variance):** Peut être utilisée pour comparer plusieurs courbes (par exemple, dans un essai clinique avec plusieurs doses ou traitements).

Domaine d'application:

- **Biologie cellulaire:** Analyser la croissance cellulaire au cours du temps.
- **Génétique:** Étudier la corrélation entre l'expression d'un gène et le temps après traitement.
- **Pharmacologie:** Mesurer la concentration de médicaments dans le sang à différents moments.

Pourquoi choisir cette table:

- Utilisée pour étudier la relation entre deux variables continues, comme la mesure d'une réponse biologique en fonction du temps ou d'une dose.
- Idéale pour effectuer une régression linéaire ou non-linéaire pour déterminer la relation entre les variables.

b. Grouped Table

Type d'analyse statistique:

- **ANOVA (Analyse de la variance):** Pour comparer les moyennes entre plusieurs groupes.
- **Test t (Test de Student):** Pour comparer les moyennes de deux groupes.
- **Test de Kruskal-Wallis:** Pour des données non-paramétriques, lorsque les distributions des groupes ne suivent pas une loi normale.

Domaine d'application:

- **Pharmacologie:** Comparer les effets de différents traitements sur des modèles de tumeurs.
- **Écologie:** Comparer la biomasse ou la croissance de différentes espèces sous des conditions variées.
- **Biologie comportementale:** Étudier l'effet de différents régimes alimentaires sur le comportement animal.

Pourquoi choisir cette table:

- Idéale pour les études comparatives où vous avez plusieurs groupes de traitements ou de conditions.
- Parfait pour les analyses de variance ou de comparaison de moyennes entre plusieurs groupes (par exemple, différents traitements).

c. Column Table

Type d'analyse statistique:

- **Test t (Test de Student):** Pour comparer la moyenne d'un groupe avec une valeur spécifique ou entre deux groupes.
- **ANOVA (Analyse de la variance):** Si plusieurs groupes sont comparés.

- **Intervalle de confiance:** Pour estimer l'incertitude autour des moyennes.
- **Analyse des variances (pour groupes de données):** Permet d'examiner la variabilité au sein de groupes.

Domaine d'application:

- **Médecine clinique:** Comparer les concentrations de biomarqueurs entre différents groupes de patients.
- **Biologie cellulaire:** Mesurer les niveaux d'une protéine ou d'une molécule dans différents traitements.
- **Microbiologie:** Analyser les concentrations bactériennes moyennes dans différents groupes d'échantillons.

Pourquoi choisir cette table:

- Parfait pour les tests de comparaison de moyennes entre un groupe de traitement et un groupe de contrôle ou entre plusieurs groupes.
- Utile pour les analyses simples où l'on compare une ou plusieurs variables dans différentes conditions.

d. Contingency Table (Tableau de contingence)

Type d'analyse statistique:

- **Test du chi-carré:** Pour tester l'indépendance entre deux variables catégorielles.
- **Test exact de Fisher:** Pour les petits échantillons, lorsque les effectifs sont faibles.
- **Test de McNemar:** Pour les comparaisons de données appariées binaires.

Domaine d'application:

- **Pharmacologie:** Étudier l'efficacité d'un traitement sur un groupe de patients (succès/échec).

- **Génétique:** Examiner l'association entre un gène et un trait particulier (présence/absence).
- **Épidémiologie:** Étudier l'association entre des facteurs de risque et l'apparition d'une maladie (fumeur/non-fumeur).

Pourquoi choisir cette table:

- Idéale pour les données catégorielles où vous avez deux variables qui sont indépendantes (par exemple, deux groupes avec des résultats de succès/échec).
 - Utile pour tester si deux facteurs sont indépendants, comme l'association entre un facteur de risque et une maladie.
- e. Survival Table (Tableau de survie)

Type d'analyse statistique:

- Analyse de survie de Kaplan-Meier: Pour estimer la fonction de survie et comparer plusieurs groupes.
- Test de log-rank: Pour comparer les courbes de survie entre plusieurs groupes.
- Cox Regression: Pour évaluer l'effet de variables continues et catégorielles sur la survie.
- Analyse de risque proportionnel de Cox: Pour tester les effets de plusieurs variables sur le risque de survenue d'un événement (par exemple, décès ou échec thérapeutique).

Domaine d'application:

- **Oncologie:** Comparer la survie de patients atteints de cancer sous différents traitements.
- **Épidémiologie:** Étudier la durée de survie de patients après un diagnostic de maladie chronique (par exemple, diabète, maladie cardiaque).
- **Biologie comportementale:** Analyser la durée de vie d'animaux dans un environnement expérimental, en fonction de traitements différents.

Pourquoi choisir cette table:

- Utilisée pour analyser les données de survie et comparer les taux de survie entre différents groupes.
- Indispensable pour des études cliniques où la variable de sortie est la durée jusqu'à un événement (comme la guérison ou le décès).

f. Part of Whole Table (Partie d'un tout)

Type d'analyse statistique:

- **Analyse des proportions:** Pour déterminer la proportion de chaque catégorie dans un ensemble total (par exemple, pourcentage de survie, répartition des individus par type de traitement).
- **Test du chi-carré:** Pour tester l'indépendance des variables catégorielles et comparer les proportions entre plusieurs groupes.

Domaine d'application:

- **Médecine clinique:** Comparer la proportion de patients guéris sous différents traitements.
- **Microbiologie:** Examiner la proportion de bactéries résistantes à un antibiotique donné.
- **Génétique:** Étudier la proportion de mutations génétiques dans différents groupes de population.

Pourquoi choisir cette table:

- Utile pour les analyses où l'on compare des proportions ou des parts de tout.
- Permet d'analyser la répartition des individus dans différentes catégories ou sous différents traitements.

g. Nested Table (Tableau imbriqué)

Type d'analyse statistique:

- **ANOVA à mesures répétées:** Pour comparer des mesures prises sur les mêmes sujets à différents moments ou sous différentes conditions.
- **ANOVA à plusieurs niveaux (hiérarchique):** Pour analyser des données où les observations sont imbriquées dans des groupes hiérarchiques (par exemple, les échantillons provenant de plusieurs sites ou de différents sous-groupes de population).
- **Modèles mixtes:** Pour examiner les effets fixes et aléatoires dans les données où il y a des mesures répétées ou des facteurs hiérarchiques.

Domaine d'application:

- **Psychologie:** Étudier les effets d'un traitement sur un groupe de patients mesurés à plusieurs moments (avant, pendant et après traitement).
- **Biologie comportementale:** Analyser les réponses comportementales de différents groupes d'animaux à plusieurs stimuli dans le temps.
- **Médecine clinique:** Comparer l'efficacité d'un médicament à différents temps de traitement pour des patients appariés.

Pourquoi choisir cette table:

- Parfait pour les données où il y a une structure imbriquée, comme les sujets avec des mesures répétées ou des groupes avec des sous-groupes.
- Permet de tenir compte des effets individuels dans des analyses de groupe.

h. Multiple Variable Table (Tableau à variables multiples)

Type d'analyse statistique:

- **ANOVA à plusieurs facteurs:** Pour examiner l'effet de plusieurs facteurs sur une variable dépendante.
- **Régression multiple:** Pour prédire une variable continue à partir de plusieurs variables indépendantes.
- **Analyse de variance multivariée (MANOVA):** Pour tester l'effet de plusieurs variables indépendantes sur plusieurs variables dépendantes.

Domaine d'application:

- **Pharmacologie:** Étudier les effets de plusieurs traitements et doses sur une variable (par exemple, taux de guérison, concentration sanguine).
- **Écologie:** Analyser l'effet combiné de plusieurs facteurs environnementaux sur la croissance des plantes.
- **Médecine clinique:** Analyser l'effet de plusieurs variables (âge, sexe, traitement) sur la survie ou la rémission des patients.

Pourquoi choisir cette table:

- Parfaite pour les études avec plusieurs variables indépendantes, vous permettant d'examiner l'effet combiné de plusieurs facteurs.
- Utile pour des analyses multivariées et des tests statistiques complexes comme l'ANOVA multivariée ou la régression multiple.

Tableau 3. Résumé de types de tables sur Graphpad Prism

Type de Table	Type d'analyse statistique	Domaine d'application	Pourquoi choisir cette table
XY Table	Corrélation, Régression linéaire, Régression non-linéaire, ANOVA	Biologie cellulaire, Génétique, Pharmacologie	Étudier la relation entre deux variables continues, idéal pour la régression et les analyses de séries temporelles.
Grouped Table	ANOVA, Test t, Test de Kruskal-Wallis	Pharmacologie, Écologie, Biologie comportementale	Comparaison de plusieurs groupes de traitements ou de conditions, utile pour l'ANOVA et la comparaison de moyennes.
Column Table	Test t, ANOVA, Intervalle de confiance, Analyse des variances	Médecine clinique, Biologie cellulaire, Microbiologie	Comparer la moyenne d'un groupe avec une valeur spécifique ou entre plusieurs groupes, utile pour des tests simples.
Contingency Table	Test du chi-carré, Test exact de Fisher, Test de McNemar	Pharmacologie, Génétique, Épidémiologie	Analyser des données catégorielles indépendantes, idéal pour tester l'indépendance entre deux variables.
Survival Table	Analyse de survie de Kaplan-Meier, Test de log-rank, Cox Regression, Analyse	Oncologie, Épidémiologie, Biologie comportementale	Analyser la survie et comparer les taux de survie entre différents groupes.

	de risque proportionnel de Cox		
Part of Whole Table	Analyse des proportions, Test du chi-carré	Médecine clinique, Microbiologie, Génétique	Analyser les proportions ou parts de tout, utile pour les tests sur des parts de groupes.
Nested Table	ANOVA à mesures répétées, ANOVA à plusieurs niveaux, Modèles mixtes	Psychologie, Biologie comportementale, Médecine clinique	Analyser les effets de traitements et conditions répétées ou imbriquées, utile pour les mesures répétées.
Multiple Variable Table	ANOVA à plusieurs facteurs, Régression multiple, MANOVA	Pharmacologie, Écologie, Médecine clinique	Examiner l'effet combiné de plusieurs facteurs ou variables indépendantes sur une ou plusieurs variables dépendantes.

3.3.3. Organisation des données dans la table

- Une fois la table créée, organisez vos données en colonnes et en lignes.
- Colonnes: Représentent les variables ou les groupes.
- Lignes: Représentent les observations ou les répétitions.
- Utilisez les en-têtes de colonnes pour décrire vos variables (p. ex., "Contrôle", "Traitement A", "Traitement B").

3.4. Importation de données depuis Excel ou d'autres sources

GraphPad Prism permet d'importer des données depuis d'autres logiciels, tels qu'Excel, ce qui facilite grandement la saisie et l'organisation des données. Voici comment procéder:

Étape 1: Préparer vos données dans Excel

- Organisez vos données dans Excel en colonnes et en lignes, comme vous le feriez dans Prism.
- Assurez-vous que les en-têtes de colonnes sont clairs et descriptifs.
- Évitez les cellules fusionnées ou les formats complexes, car ils peuvent poser des problèmes lors de l'importation.

Étape 2: Importer les données dans Prism

- Dans Prism, allez dans l'onglet "File" > "Import" > "Import Data".
- Sélectionnez le fichier Excel contenant vos données.
- Choisissez la feuille de calcul et la plage de cellules à importer.
- Cliquez sur "OK" pour importer les données dans une nouvelle table.

Étape 3: Vérifier et ajuster les données importées

- Après l'importation, vérifiez que les données sont correctement organisées dans Prism.
- Si nécessaire, ajustez les en-têtes de colonnes ou les formats de données.

3.5. Analyse Statistique de Base

L'analyse statistique est au cœur de GraphPad Prism. Ce chapitre vous guidera à travers les méthodes de base pour explorer et interpréter vos données. Nous aborderons les statistiques descriptives, les tests t (appariés et non appariés), et l'ANOVA à un facteur, ainsi que les tests post-hoc associés. Ces outils sont essentiels pour comparer des groupes, identifier des tendances et tirer des conclusions valides à partir de vos données.

3.2. Le Menu Analyse dans GraphPad Prism

Le menu Analyse dans GraphPad Prism est l'outil central pour appliquer des tests statistiques à vos données. Ce menu vous permet d'effectuer des analyses statistiques simples et avancées, allant des tests de comparaison de groupes à la régression linéaire et non linéaire, ainsi que l'analyse de survie. L'outil Analyse est conçu pour être intuitif tout en offrant une grande puissance pour traiter des données complexes. Voici une explication détaillée des options disponibles dans le menu Analyse de GraphPad Prism.

3.2.2. Statistiques descriptives (moyenne, écart-type, etc.)

Les statistiques descriptives résument et décrivent les caractéristiques principales de vos données. Elles sont souvent la première étape de toute analyse statistique.

a. Comment calculer les statistiques descriptives dans Prism:

1. Sélectionnez la table de données que vous souhaitez analyser.
2. Allez dans l'onglet "Analyze" et choisissez "Column Statistics" (ou l'option appropriée pour votre type de données).
3. Cochez les statistiques que vous souhaitez calculer, par exemple:
 - **Moyenne:** La valeur moyenne des données.
 - **Écart-type (SD):** Une mesure de la dispersion des données autour de la moyenne.
 - **Erreur standard de la moyenne (SEM):** Une estimation de la précision de la moyenne.
 - **Intervalle de confiance:** Une plage de valeurs dans laquelle la vraie moyenne se situe avec un certain niveau de confiance (généralement 95 %).
 - **Médiane:** La valeur centrale des données, utile pour les distributions asymétriques.

4. Cliquez sur "OK" pour exécuter l'analyse.

3.2.3. Tests t (appariés et non appariés)

Les tests t sont utilisés pour comparer les moyennes de deux groupes. Prism propose deux types de tests t: appariés et non appariés.

a. Test t non apparié:

1. Sélectionnez la table de données contenant les deux groupes à comparer.
2. Allez dans l'onglet "Analyze" et choisissez "t tests (and nonparametric tests)".
3. Sélectionnez "Unpaired t test".
4. Choisissez les options appropriées, comme l'hypothèse d'égalité des variances (test de Welch si les variances sont inégales).
5. Cliquez sur "OK" pour exécuter le test.

b. Test t apparié:

1. Organisez vos données en deux colonnes, une pour chaque condition (avant/après).
2. Allez dans l'onglet "Analyze" et choisissez "t tests (and nonparametric tests)".
3. Sélectionnez "Paired t test".
4. Cliquez sur "OK" pour exécuter le test.

3.2.4. ANOVA à un facteur et post-hoc tests

1. Organisez vos données en colonnes, chaque colonne représentant un groupe.
2. Allez dans l'onglet "Analyze" et choisissez "One-way ANOVA (and nonparametric tests)".
3. Sélectionnez les options appropriées, comme l'hypothèse d'égalité des variances.
4. Cliquez sur "OK" pour exécuter l'analyse.

Tests post-hoc:

1. Après avoir exécuté l'ANOVA, retournez à l'onglet "Results".
2. Sélectionnez l'option "Multiple Comparisons" ou "Post-hoc tests".
3. Choisissez le test post-hoc approprié et cliquez sur "OK".

4. Chapitre 3. Travaux pratiques sur graphpad prism

Introduction

Ce chapitre est dédié à l'utilisation du logiciel GraphPad Prism pour l'analyse statistique des données épidémiologiques. Les étudiants apprendront à utiliser cet outil pour analyser des données issues de différentes études épidémiologiques, telles que les études de cohorte, les études cas-témoins, et les essais cliniques. L'objectif est d'appliquer les tests statistiques et les méthodes graphiques pour interpréter les résultats de manière rigoureuse et scientifique.

Objectifs du Chapitre

- Maîtriser l'importation et la gestion des données épidémiologiques dans GraphPad Prism.
- Apprendre à effectuer des analyses statistiques spécifiques à l'épidémiologie, telles que les tests du Chi², t-test, ANOVA, régression, et analyse de survie.
- Savoir présenter les résultats sous forme de graphiques épidémiologiques (courbes de survie, diagrammes épidémiques, etc.).
- Interpréter les résultats statistiques, comprendre leur signification et rédiger des rapports clairs et précis pour des études épidémiologiques.

Compétences à Acquérir

1. Application des tests statistiques épidémiologiques:
 - **Chi²**: Pour tester l'association entre deux variables catégorielles.
 - **T-test**: Pour comparer deux groupes indépendants.
 - **ANOVA**: Pour comparer plusieurs groupes.
 - **Régression linéaire**: Pour analyser les facteurs de risque et leurs relations avec les variables de santé.

2. Création de graphiques épidémiologiques:
 - Courbes de survie (Kaplan-Meier).
 - Histogrammes et diagrammes de distributions.
 - Diagrammes de corrélation entre variables.
3. Interprétation des résultats épidémiologiques:
 - Calcul et analyse des p-values, intervalles de confiance (IC 95%) et des risques relatifs.
 - Rédaction d'une interprétation claire des résultats dans le contexte épidémiologique.
4. Analyse des données provenant d'études épidémiologiques réelles:
 - Application de tests adaptés à des jeux de données issus d'études de cohortes, d'études cas-témoins et d'essais cliniques.

4.1. TP1: Calcul des statistiques descriptives et création de graphiques descriptifs avec GraphPad Prism

Objectifs du TP:

- Apprendre à calculer et interpréter les statistiques descriptives: moyenne, médiane, et écart-type.
- Savoir créer et analyser des graphiques descriptifs: diagrammes en boîte, histogrammes, et diagrammes de dispersion.
- Utiliser GraphPad Prism pour organiser, analyser, et visualiser les données.

Données de l'exercice:

L'étude concerne l'incidence de deux types de maladies dans deux groupes d'individus exposés à un facteur de risque. Les données sont collectées sur une période de 6 mois, et les valeurs indiquent le nombre de cas de maladie observés dans chaque groupe chaque mois.

Mois	Groupe A (Exposé)	Groupe B (Non-exposé)
1	12	8
2	15	10
3	18	12
4	14	9
5	16	13
6	17	11

Partie 1: Calcul des statistiques descriptives

1.1 Moyenne, Médiane, et Écart-type

- **Objectif:** Calculer la moyenne, la médiane, et l'écart-type pour chaque groupe.
- **Instructions:**
 - Importez les données dans GraphPad Prism.
 - Pour chaque groupe (Groupe A et Groupe B), calculez les statistiques suivantes:
 - Moyenne
 - Médiane
 - Écart-type

Questions:

- Quel est la moyenne des cas de la maladie dans le Groupe A et le Groupe B?
- Quelle est la médiane des cas dans chaque groupe?
- Quelle est la variabilité des données pour chaque groupe, mesurée par l'écart-type?

Partie 2: Création de graphiques descriptifs

2.1 Diagramme en boîte (Box Plot)

- **Objectif:** Créer un diagramme en boîte pour visualiser la distribution des données pour chaque groupe.
- **Instructions:**
 - Dans GraphPad Prism, sélectionnez l'option de création de graphiques de type "Box Plot" pour chaque groupe.
 - Incluez les informations suivantes dans le graphique:
 - Médiane
 - Quartiles
 - Valeurs extrêmes (outliers)

Questions:

- Quelle est l'étendue des données dans chaque groupe (minimum et maximum)?
- Y a-t-il des valeurs aberrantes dans l'un ou l'autre des groupes?

2.2 Histogramme

- **Objectif:** Créer un histogramme pour visualiser la distribution des données pour chaque groupe.
- **Instructions:**
 - Créez un histogramme pour chaque groupe, avec des barres représentant les fréquences des cas de maladie.
 - Choisissez une taille de classe appropriée pour l'histogramme.

Questions:

- Comment les données sont-elles réparties dans chaque groupe? Sont-elles symétriques ou biaisées?

- Comparez les distributions des deux groupes. Quelles sont les différences visuelles?

2.3 Diagramme de dispersion (Scatter Plot)

- **Objectif:** Créer un diagramme de dispersion pour observer la relation entre les données des deux groupes.
- **Instructions:**
 - Créez un diagramme de dispersion pour chaque groupe, où l'axe des X représente les mois et l'axe des Y représente le nombre de cas observés.
 - Tracez les points pour chaque mois dans chaque groupe.

Questions:

- Y a-t-il une tendance générale dans l'évolution du nombre de cas de maladie dans chaque groupe?
- Quelle est la variation du nombre de cas entre les mois pour chaque groupe?

Partie 3: Interprétation des résultats

3.1 Comparaison des groupes

- **Objectif:** Comparer les statistiques descriptives des deux groupes (Groupe A et Groupe B).
- **Instructions:**
 - Comparez les moyennes, les médianes et les écarts-types des deux groupes.
 - Interprétez les résultats pour en déduire si le facteur de risque a un effet significatif sur l'incidence de la maladie.

Questions:

- Quels sont les effets du facteur de risque sur l'incidence de la maladie entre les groupes?

- Est-ce que les différences entre les groupes sont significatives? Si oui, pourquoi?

Conclusion

- Résumez les principales observations concernant les statistiques descriptives et les graphiques créés.
- Discutez des résultats obtenus et de la pertinence de l'utilisation de ces outils statistiques pour interpréter des données épidémiologiques.

4.2. TP2: Test de Normalité et Tests t Appariés et Indépendants

Introduction

Dans cette étude de biologie moléculaire, nous avons mesuré l'expression d'un gène spécifique chez deux groupes de patients (Groupe 1 et Groupe 2) avant et après traitement. Les données suivantes concernent l'expression du gène (en valeurs absolues) avant et après traitement pour chaque patient:

Patient	Groupe 1 (avant)	Groupe 1 (après)	Groupe 2 (avant)	Groupe 2 (après)
1	12.4	14.1	15.3	14.9
2	13.1	15.5	16.0	16.5
3	14.0	15.2	16.8	16.0
4	12.9	14.5	15.7	15.3
5	13.5	15.0	17.0	16.2

L'objectif de ce travail est d'analyser ces données en utilisant des tests statistiques. Nous commencerons par tester la normalité des données avant de réaliser des tests t pour comparer les moyennes des groupes.

Partie 1: Test de Normalité (Test de Shapiro-Wilk)

Objectif:

Vérifier si les données suivent une distribution normale avant de procéder aux tests paramétriques.

Méthode:

Nous allons utiliser le test de Shapiro-Wilk pour chaque groupe de données (avant et après traitement) afin de tester la normalité.

1. Formulation de l'hypothèse:

- H0: Les données suivent une distribution normale.
- H1: Les données ne suivent pas une distribution normale.

2. Étapes:

- Utilisez GraphPad Prism pour effectuer le test de Shapiro-Wilk pour chaque série de données (Groupe 1 avant, Groupe 1 après, Groupe 2 avant, et Groupe 2 après).
- Si $p\text{-value} > 0.05$, cela signifie que les données sont normalement distribuées, et nous pourrions utiliser des tests paramétriques comme le test t ou ANOVA. Sinon, nous devons envisager des tests non paramétriques.

Partie 2: Test t Apparié pour les Groupes 1 et 2

Objectif:

Comparer les moyennes de l'expression du gène avant et après traitement dans les deux groupes.

Test t Apparié:

Le test t apparié est utilisé pour comparer les moyennes de deux séries de données liées. Dans notre cas, nous comparons les mesures avant et après traitement dans chaque groupe.

1. Hypothèses:

- H0: Il n'y a pas de différence significative entre les mesures avant et après traitement dans chaque groupe.
- H1: Il y a une différence significative entre les mesures avant et après traitement dans chaque groupe.

2. Formulation du test pour le Groupe 1:

- Comparer les valeurs de Groupe 1 (avant) avec Groupe 1 (après).

- Effectuer un test t apparié.

3. Formulation du test pour le Groupe 2:

- Comparer les valeurs de Groupe 2 (avant) avec Groupe 2 (après).
- Effectuer un test t apparié.

4. Étapes:

- Ouvrez GraphPad Prism.
- Allez dans Analyze > t-test > Paired t-test.
- Entrez les valeurs avant et après pour chaque groupe.
- GraphPad Prism calculera la valeur p.

5. Interprétation:

- Si $p < 0.05$, rejetez l'hypothèse nulle et concluez qu'il existe une différence significative entre les mesures avant et après traitement dans ce groupe.

Partie 3: Test t Indépendant pour Comparer les Groupes 1 et 2 Après Traitement

Objectif:

Comparer les moyennes de l'expression du gène entre le Groupe 1 après traitement et le Groupe 2 après traitement.

Test t Indépendant:

Le test t indépendant est utilisé pour comparer les moyennes de deux groupes indépendants, ici les groupes après traitement.

1. Hypothèses:

- H0: Il n'y a pas de différence significative entre les moyennes du Groupe 1 après traitement et du Groupe 2 après traitement.
- H1: Il y a une différence significative entre les moyennes des deux groupes après traitement.

2. Étapes:

- Ouvrez GraphPad Prism.
- Allez dans Analyze > t-test > Unpaired t-test.
- Sélectionnez les données de Groupe 1 (après) et Groupe 2 (après).
- GraphPad Prism effectuera le test t indépendant et calculera la valeur p.

3. Interprétation:

- Si $p < 0.05$, vous pouvez conclure qu'il y a une différence significative entre les deux groupes après traitement.
- Si $p > 0.05$, On ne rejette pas l'hypothèse nulle ; il n'y a pas de preuves suffisantes pour conclure à une différence significative entre les deux groupes après traitement.

Conclusion

Après avoir effectué le test de normalité et les tests t (appariés et indépendants), vous pourrez conclure sur les différences significatives entre les groupes avant et après traitement.

1. Si les données sont normales: Vous utiliserez les tests t pour comparer les moyennes.
2. Si les données ne sont pas normales: Vous devrez utiliser des tests non paramétriques tels que le test de Wilcoxon ou le test de Mann-Whitney.

En résumé, ces tests statistiques vous permettront de déterminer l'impact du traitement sur l'expression du gène et de comparer l'efficacité du traitement entre les deux groupes.

4.3. TP 3: ANOVA avec GraphPad Prism

Contexte de l'étude

Dans cette étude de biologie moléculaire, nous avons mesuré l'expression d'un gène spécifique chez des patients répartis en différents groupes en fonction de deux critères: le type de traitement administré (Traitement A, Traitement B et Contrôle) et le sexe des patients (Homme et Femme). Les données suivantes concernent l'expression du gène (en valeurs absolues) pour chaque patient, dans un format plus simple à introduire dans GraphPad Prism.

Patient	Sexe	Traitement	Expression du Gène (valeurs)
1	Homme	A	15.0
2	Homme	A	14.5
3	Femme	A	14.3
4	Femme	A	14.0
5	Homme	B	16.2
6	Homme	B	16.5
7	Femme	B	15.8
8	Femme	B	15.4
9	Homme	Contrôle	13.0
10	Homme	Contrôle	13.2
11	Femme	Contrôle	13.5
12	Femme	Contrôle	13.3

L'objectif de ce travail est d'analyser ces données à l'aide de l'ANOVA, en utilisant d'abord une ANOVA à un facteur pour comparer l'impact du traitement, puis une ANOVA à deux facteurs pour analyser l'impact simultané du traitement et du sexe.

Partie 1: ANOVA à un Facteur

Objectif:

Comparer les moyennes de l'expression du gène entre les trois groupes de traitement (A, B, et Contrôle).

ANOVA à un Facteur:

L'ANOVA à un facteur est utilisée pour comparer les moyennes de plus de deux groupes indépendants, ici les trois groupes de traitement.

1. Hypothèses:

- H0: Il n'y a pas de différence significative entre les moyennes des groupes de traitement.
- H1: Il existe une différence significative entre au moins deux des groupes de traitement.

2. Formulation du test pour le traitement:

- Comparer les valeurs de l'expression du gène entre les trois groupes de traitement (A, B, et Contrôle).

3. Étapes:

- Ouvrez GraphPad Prism.
- Allez dans Analyze > One-way ANOVA.
- Sélectionnez les données correspondant aux trois groupes de traitement (A, B, et Contrôle) dans une table de type "Column" dans GraphPad Prism.
 - Pourquoi utiliser cette table? : La table de type "Column" permet d'entrer des données en colonnes distinctes pour chaque groupe, ce qui

est approprié pour l'ANOVA à un facteur avec des groupes distincts.

- GraphPad Prism effectuera l'ANOVA et calculera la valeur p.

4. Interprétation:

- Si $p < 0.05$, vous rejetez l'hypothèse nulle et concluez qu'il existe une différence significative entre les groupes de traitement.
- Si $p > 0.05$, vous ne rejetez pas l'hypothèse nulle, ce qui signifie qu'il n'y a pas de différence significative entre les groupes de traitement.

Partie 2 : Application de l'ANOVA à Deux Facteurs dans GraphPad Prism

Contexte de l'étude

Dans cette étude de biologie moléculaire, nous avons mesuré l'expression d'un gène spécifique chez des patients répartis selon deux facteurs:

- **Facteur 1:** Traitement (3 niveaux: Traitement A, Traitement B, et Contrôle).
- **Facteur 2:** Sexe (2 niveaux: Homme et Femme).

Les données suivantes représentent l'expression du gène (en valeurs absolues) pour chaque patient:

Patient	Sexe	Traitement	Expression du Gène (valeurs)
1	Homme	A	15.0
2	Homme	A	14.5
3	Homme	A	14.8
4	Femme	A	14.3
5	Femme	A	14.0
6	Femme	A	14.1
7	Homme	B	16.2
8	Homme	B	16.5
9	Homme	B	16.0
10	Femme	B	15.8
11	Femme	B	15.4
12	Femme	B	15.7
13	Homme	Contrôle	13.0
14	Homme	Contrôle	13.2
15	Homme	Contrôle	13.4
16	Femme	Contrôle	13.5
17	Femme	Contrôle	13.3
18	Femme	Contrôle	13.1

Objectifs de l'exercice

1. Analyser l'effet du traitement (A, B, et Contrôle) sur l'expression du gène.
2. Analyser l'effet du sexe (Homme et Femme) sur l'expression du gène.
3. Analyser l'interaction entre le traitement et le sexe sur l'expression du gène.

Nous allons utiliser GraphPad Prism pour effectuer une ANOVA à deux facteurs et interpréter les résultats.

Partie 1: Préparation des Données dans GraphPad Prism

1. Création de la table XY

1. Ouvrir GraphPad Prism et créer une nouvelle table XY.
 - o Sélectionner "grouped " comme type de table.

2. Organiser les Données

Voici comment structurer les données dans la table XY:

	Traitement A			Traitement B			Contrôle		
Homme	15	14,5	14,8	16,2	16,5	16	13	13,2	13,4
Femme	14,3	14	14,1	15,8	15,4	15,7	13,5	13,3	13,1

Partie 2: Analyse des Données avec ANOVA à Deux Facteurs

1. Effectuer l'ANOVA à Deux Facteurs

1. Sélectionner les Données:
 - o Allez dans Analyze puis choisissez Two-way ANOVA.
 - o Sélectionnez les données organisées dans la table XY.

2. Paramétrage de l'ANOVA à Deux Facteurs:

- Facteur 1: Traitement avec 3 niveaux (A, B, Contrôle).
- Facteur 2: Sexe avec 2 niveaux (Homme et Femme).
- Cochez l'option pour inclure l'interaction entre les facteurs Traitement et Sexe.
- Choisissez l'option de tests post-hoc si nécessaire, pour analyser les différences entre les groupes (par exemple, Dunnett's test si vous comparez chaque traitement à un contrôle).

2. Interpréter les Résultats

Après avoir effectué l'ANOVA, GraphPad Prism fournira les résultats suivants:

- p-value pour le Facteur 1 (Traitement): Cette valeur vous indiquera si l'effet du traitement est significatif.
 - Si $p < 0.05$, cela signifie que le traitement a un effet significatif sur l'expression du gène.
 - Si $p > 0.05$, il n'y a pas de différence significative entre les traitements.
- p-value pour le Facteur 2 (Sexe): Cette valeur vous dira si le sexe des patients a un effet significatif sur l'expression du gène.
 - Si $p < 0.05$, cela signifie que l'expression du gène diffère significativement entre les hommes et les femmes.
 - Si $p > 0.05$, il n'y a pas de différence significative entre les sexes.
- p-value pour l'Interaction entre Traitement et Sexe: Cette valeur indiquera si l'effet du traitement varie en fonction du sexe.
 - Si $p < 0.05$, cela signifie qu'il existe une interaction significative entre le traitement et le sexe. L'effet du traitement sur l'expression du gène dépend du sexe.

- Si $p > 0.05$, il n'y a pas d'interaction significative, ce qui signifie que l'effet du traitement est similaire pour les hommes et les femmes.

3. Résultats Post-Hoc (si applicable)

- Si l'ANOVA montre un effet significatif, vous pouvez effectuer un test post-hoc pour identifier quelles paires de groupes (par exemple, Traitement A vs Traitement B, Traitement A vs Contrôle) diffèrent de manière significative.
- GraphPad Prism vous proposera des tests post-hoc, comme Tukey's test ou Dunnett's test.

Partie 3: Conclusion

1. Si l'ANOVA montre un effet significatif pour le facteur Traitement:
 - Concluez que le traitement a un impact sur l'expression du gène. Vous pouvez indiquer quel traitement (A, B, ou Contrôle) est le plus efficace en fonction des tests post-hoc.
2. Si l'ANOVA montre un effet significatif pour le facteur Sexe:
 - Concluez que l'expression du gène diffère entre les hommes et les femmes. Cela pourrait être lié à des différences biologiques entre les sexes.
3. Si l'ANOVA montre une interaction significative entre Traitement et Sexe:
 - L'effet du traitement peut varier en fonction du sexe des patients. Par exemple, un traitement peut être plus efficace chez les hommes que chez les femmes.

4.4. TP 4: Analyse de la Corrélation entre Deux Variables en Biologie Moléculaire avec GraphPad Prism

Contexte de l'étude

Dans cette étude de biologie moléculaire, nous avons mesuré l'expression d'un gène spécifique et les concentrations de certaines protéines dans le plasma sanguin de patients. L'objectif est d'étudier la corrélation entre l'expression du gène et la concentration de la protéine pour déterminer s'il existe une relation linéaire entre ces deux variables.

Les données suivantes montrent les niveaux d'expression du gène (en unités arbitraires) et les concentrations de protéines (en ng/ml) pour 10 patients.

Patient	Expression du Gène (unité)	Concentration de la Protéine (ng/ml)
1	15.2	75
2	14.8	70
3	16.1	80
4	15.7	78
5	14.4	68
6	15.0	72
7	15.5	74
8	16.0	77
9	14.9	73
10	15.3	76

Objectifs de l'exercice

1. Analyser la corrélation entre l'expression du gène et la concentration de la protéine.
2. Calculer le coefficient de corrélation de Pearson.
3. Interpréter la relation entre les deux variables.

Nous allons utiliser GraphPad Prism pour effectuer une analyse de corrélation et interpréter les résultats.

Partie 1: Préparation des Données dans GraphPad Prism

1. Création de la table XY

1. Ouvrir GraphPad Prism et créer une nouvelle table XY.
 - o Choisissez "XY" comme type de table.
 - o Colonne X: Les données d'expression du gène (en unité arbitraire).
 - o Colonne Y: Les données de concentration de la protéine (en ng/ml).

2. Organiser les Données

Voici comment structurer les données dans la table XY dans GraphPad Prism:

Expression du Gène (unité)	Concentration de la Protéine (ng/ml)
15.2	75
14.8	70
16.1	80
15.7	78
14.4	68

15.0	72
15.5	74
16.0	77
14.9	73
15.3	76

Partie 2: Analyse de la Corrélation

1. Effectuer l'Analyse de Corrélation

1. Sélectionner les Données:
 - Allez dans Analyze > Correlation.
 - Sélectionnez les deux séries de données dans la table XY:
 - Expression du Gène (colonne X)
 - Concentration de la Protéine (colonne Y)
2. Calculer le Coefficient de Corrélation de Pearson:
 - GraphPad Prism calculera le coefficient de corrélation de Pearson (r), qui mesure la force et la direction de la relation linéaire entre les deux variables.
 - Prism affichera également la valeur p , qui vous permettra de savoir si la corrélation est statistiquement significative.
3. Interpréter le Coefficient de Corrélation:
 - Si $r > 0.7$, la corrélation est forte et positive (les deux variables augmentent ou diminuent ensemble).
 - Si $r < -0.7$, la corrélation est forte et négative (l'une augmente pendant que l'autre diminue).

- Si r proche de 0, il n'y a pas de corrélation linéaire entre les variables.

2. Interpréter les Résultats

Après avoir effectué l'analyse de corrélation, GraphPad Prism fournira les résultats suivants:

- Coefficient de Corrélation de Pearson (r): Mesure de la relation linéaire entre les deux variables.
 - $r = 1$: Corrélation parfaite positive.
 - $r = -1$: Corrélation parfaite négative.
 - $r = 0$: Pas de corrélation linéaire.
- p -value:
 - Si $p < 0.05$, cela indique que la corrélation est statistiquement significative.
 - Si $p > 0.05$, cela signifie que la corrélation n'est pas statistiquement significative.

Partie 3: Conclusion

1. Si r est positif et significatif ($p < 0.05$):
 - Concluez qu'il existe une relation positive significative entre l'expression du gène et la concentration de la protéine. Lorsque l'expression du gène augmente, la concentration de la protéine augmente également.
2. Si r est négatif et significatif ($p < 0.05$):
 - Concluez qu'il existe une relation négative significative entre l'expression du gène et la concentration de la protéine. Lorsque l'expression du gène augmente, la concentration de la protéine diminue.
3. Si r est proche de 0 et $p > 0.05$:
 - Concluez qu'il n'y a pas de corrélation significative entre l'expression du gène et la concentration de la protéine.

Cela suggère que les variations dans l'expression du gène ne sont pas liées aux variations dans la concentration de la protéine.

Notes pour les étudiants:

- GraphPad Prism est un outil très utile pour effectuer des analyses de corrélation et de régression. Dans cet exercice, nous avons utilisé le coefficient de corrélation de Pearson pour étudier la relation linéaire entre deux variables quantitatives.
- Assurez-vous de bien organiser vos données dans la table XY, car cette structure est idéale pour les analyses de corrélation.

4.5. Tp 5. Régression Linéaire, Non Linéaire dans GraphPad Prism

Contexte de l'étude

Dans cette étude de biologie moléculaire, nous avons mesuré l'expression d'un gène spécifique chez des patients sous différents traitements. L'objectif est d'analyser la relation entre l'expression du gène et plusieurs facteurs, notamment la dose du médicament et le temps d'exposition au médicament. Nous allons utiliser différents types de régression pour comprendre ces relations:

1. Régression Linéaire pour analyser la relation entre la dose du médicament et l'expression du gène.
2. Régression Non Linéaire pour tester si une relation sigmoïde ou exponentielle décrit mieux les données.
3. Régression Multiple pour analyser l'impact simultané de la dose et du temps sur l'expression du gène.

Les données sont basées sur des résultats expérimentaux et peuvent contenir des variations naturelles dans les valeurs.

Partie 1: Régression Linéaire

1. Contexte des Données

Dans cette première partie, nous allons analyser la relation linéaire entre la dose du médicament (en mg) et l'expression du gène (en unités arbitraires). Les données suivantes montrent comment l'expression du gène varie en fonction de la dose administrée:

Patient	Dose du Médicament (mg)	Expression du Gène (unités)
1	0.0	12.3
2	1.0	14.1
3	2.0	15.2

4	3.0	17.1
5	4.0	18.4
6	5.0	20.0
7	6.0	21.5
8	7.0	23.0
9	8.0	24.3
10	9.0	25.5

2. Préparation des Données dans GraphPad Prism

1. Ouvrez GraphPad Prism et créez une nouvelle table XY.
2. Colonne X: Dose du médicament (mg).
3. Colonne Y: Expression du gène (unités arbitraires).

Entrez les données dans la table XY de la manière suivante:

Dose du Médicament (mg)	Expression du Gène (unités)
0.0	12.3
1.0	14.1
2.0	15.2
3.0	17.1
4.0	18.4
5.0	20.0
6.0	21.5

7.0	23.0
8.0	24.3
9.0	25.5

3. Effectuer la Régression Linéaire

1. Allez dans Analyze > Regression > Linear regression.
2. Sélectionnez les colonnes Dose du Médicament pour X et Expression du Gène pour Y.
3. GraphPad Prism ajustera une droite de régression et affichera le coefficient de régression (r^2) et la p-value pour évaluer la relation linéaire.

Interprétation des Résultats

- Si $r^2 > 0.9$, cela indique une forte relation linéaire entre la dose et l'expression du gène.
- Si $p < 0.05$, cela signifie que la relation linéaire est statistiquement significative.

Partie 2: Régression Non Linéaire

1. Contexte des Données

Nous allons maintenant tester une régression non linéaire pour voir si une courbe sigmoïde ou exponentielle décrit mieux la relation entre la dose du médicament et l'expression du gène.

Supposons que la relation entre la dose et l'expression du gène ne soit pas parfaitement linéaire et que les données suivent une réponse sigmoïde:

Patient	Dose du Médicament (mg)	Expression du Gène (unités)
1	0.0	12.0
2	1.0	13.5
3	2.0	15.0
4	3.0	16.5
5	4.0	18.0
6	5.0	19.5
7	6.0	20.5
8	7.0	21.2
9	8.0	21.8
10	9.0	22.0

2. Effectuer la Régression Non Linéaire

1. Allez dans Analyze > Nonlinear regression (curve fit).
2. Choisissez "Sigmoidal" ou "Log(inhibitor) vs response" comme modèle de régression.
3. Sélectionnez les données de Dose du Médicament pour X et Expression du Gène pour Y.
4. GraphPad Prism ajustera une courbe non linéaire aux données et vous donnera des paramètres ajustés, tels que la pente, le maximum, et le minimum.

Interprétation des Résultats

- Vérifiez si le r^2 pour la régression non linéaire est plus élevé que celui de la régression linéaire, indiquant que le modèle non linéaire explique mieux les données.

Si la courbe est significativement mieux ajustée ($r^2 > 0.95$), la relation entre la dose et l'expression du gène suit probablement un modèle non linéaire.

4.6. TP 6. Test du chi carré, Calcul du Risque Relatif (RR) et de l'Odds Ratio (OR) pour l'analyse des variables qualitatives dans GraphPad Prism

Objectifs du TP:

- Apprendre à utiliser le test du chi carré pour tester l'association entre deux variables qualitatives.
- Calculer le Risque Relatif (RR) et l'Odds Ratio (OR) pour évaluer l'association entre l'exposition à un facteur de risque et la survenue d'une maladie.

Données de l'exercice:

Dans cette étude, nous analysons l'association entre l'exposition à un facteur de risque (exposé ou non exposé) et la survenue d'une maladie (présence ou absence de la maladie). Les données sont organisées dans un tableau de contingence comme suit:

Exposition / Maladie	Maladie Présente (Oui)	Maladie Absente (Non)	Total
Exposé	30	10	40
Non-exposé	15	25	40
Total	45	35	80

Partie 1: Test du chi carré pour l'analyse des variables qualitatives

1. Hypothèses:

- H0 (hypothèse nulle): Il n'y a pas d'association entre l'exposition et la maladie (les deux variables sont indépendantes).
- H1 (hypothèse alternative): Il existe une association entre l'exposition et la maladie (les deux variables sont liées).

2. Calcul du test du chi carré:

- Utilisez le test du chi carré pour déterminer si les variables Exposition et Maladie sont indépendantes.
- Le test du chi carré calcule si les fréquences observées diffèrent de manière significative des fréquences attendues sous l'hypothèse d'indépendance.

Instructions dans GraphPad Prism:

1. Ouvrir GraphPad Prism.
2. Créer un nouveau projet.
3. Choisir le type de données "Column" puis "Contingency Table".
4. Saisir les données sous forme de tableau 2x2.
5. Allez dans l'onglet "Analyze", puis sélectionnez "Chi-Square Test".
6. Obtenez la valeur du chi carré et la p-value.

Questions:

- Quelle est la valeur du chi carré et la p-value?
- Que concluez-vous à partir du test du chi carré? (si $p < 0.05$, rejeter H_0)

4.7. TP 7: Analyse de survie dans GraphPad Prism

Objectifs du TP:

- Apprendre à effectuer une analyse de survie en utilisant la méthode de Kaplan-Meier.
- Comprendre comment utiliser le test de log-rank pour comparer les courbes de survie entre plusieurs groupes.
- Apprendre à interpréter les résultats de l'analyse de survie et comprendre leur signification dans un contexte biologique.

Données de l'exercice:

Dans cette étude, nous analysons la durée de survie de patients atteints d'une maladie donnée. Les données sont organisées selon l'exposition à un traitement (traitement A et traitement B). La durée de survie est mesurée en mois, et nous souhaitons analyser si l'exposition à l'un ou l'autre des traitements a un impact sur la survie des patients.

Patient	Traitement	Durée de survie (mois)	Événement (1 = décès, 0 = censuré)
1	A	12	1
2	A	8	0
3	A	15	1
4	A	20	1
5	A	10	0
6	B	6	1
7	B	9	1
8	B	14	1

9	B	18	0
10	B	11	1

Explication des variables:

- Durée de survie (mois): Durée pendant laquelle chaque patient a survécu après le traitement.
- Événement: Variable binaire, où 1 indique que l'événement d'intérêt (le décès) a eu lieu, et 0 indique que l'observation a été censurée (le patient n'est pas décédé pendant la période de suivi ou n'a pas été suivi jusqu'à la fin).

Partie 1: Méthode de Kaplan-Meier

La méthode de Kaplan-Meier est utilisée pour estimer la fonction de survie. Elle permet de calculer la probabilité de survie à différents moments pour chaque groupe.

1. Calcul de la fonction de survie avec Kaplan-Meier:

- Ouvrez GraphPad Prism et créez un nouveau projet.
- Sélectionnez "Survival analysis" dans le type de projet.
- Entrez les données des deux groupes (traitement A et traitement B) et la durée de survie pour chaque patient.
- Spécifiez si l'événement est survenu (décès) ou si les données sont censurées.
- GraphPad Prism calculera automatiquement les courbes de survie de Kaplan-Meier pour chaque groupe.

Instructions:

- Allez dans l'onglet "Analyze", puis sélectionnez "Survival analysis".
- Choisissez Kaplan-Meier survival curves et sélectionnez les données pour chaque groupe.

- Prism générera les courbes de survie pour les groupes traitement A et traitement B.

Questions:

- Quelle est la probabilité de survie au bout de 6, 12, et 18 mois pour chaque groupe?
- Quelle est la différence de survie entre les deux traitements à différents moments?

Partie 2: Test de log-rank

Le test de log-rank est utilisé pour comparer les courbes de survie entre deux groupes. Il permet de déterminer si les différences entre les courbes sont statistiquement significatives.

1. Application du test de log-rank:

- Une fois que les courbes de survie sont créées, GraphPad Prism peut effectuer un test de log-rank pour comparer les courbes des deux groupes (traitement A et traitement B).
- Ce test permet de tester si les différences observées entre les groupes sont dues au hasard ou si elles sont statistiquement significatives.

Instructions dans GraphPad Prism:

- Allez dans l'onglet "Analyze" et sélectionnez "Log-rank test".
- Prism effectuera le test et affichera les résultats sous forme de p-value.

Questions:

- Quelle est la p-value obtenue du test de log-rank?
- Si $p < 0.05$, quelle est votre conclusion concernant la différence de survie entre les deux traitements?

Partie 3: Interprétation des résultats

Une fois les courbes de survie obtenues et le test de log-rank effectué, il est essentiel d'interpréter les résultats.

1. Interprétation des courbes de survie:

- Si la courbe de survie pour le groupe A est plus élevée que celle du groupe B, cela suggère que le traitement A est plus efficace que le traitement B pour prolonger la survie.
- Si les courbes de survie se croisent ou sont proches l'une de l'autre, cela suggère qu'il n'y a pas de différence significative entre les traitements en termes de survie.

2. Interprétation de la p-value:

- Si la p-value est inférieure à 0.05, cela signifie qu'il existe une différence statistiquement significative entre les courbes de survie des deux groupes, et donc, que le traitement A et le traitement B ont un effet différent sur la survie.
- Si la p-value est supérieure à 0.05, cela suggère qu'il n'y a pas de différence significative entre les deux traitements en termes de survie.

Questions:

- Quelles sont les implications cliniques de la différence de survie entre les groupes?
- Quelles seraient les prochaines étapes pour confirmer ces résultats (par exemple, augmenter la taille de l'échantillon)?

Conclusion du TP

1. Synthèse des résultats:

- Résumez les courbes de survie des groupes traités avec le traitement A et le traitement B.

- Discutez des résultats du test de log-rank et de la p-value obtenue.

2. Applications cliniques:

- Comment ces résultats peuvent-ils guider le choix des traitements pour les patients?
- Quels facteurs externes pourraient influencer les résultats de l'analyse de survie?

5. Chapitre 4. Corrigé types des travaux pratiques

La partie corrigée du polycopié a pour objectif de:

- Apporter une correction détaillée et pédagogique des exercices réalisés lors des séances de TP.
- Expliquer les étapes méthodologiques et les raisonnements derrière chaque analyse statistique.
- Interpréter les résultats obtenus avec GraphPad Prism pour favoriser la compréhension et l'esprit critique.
- Souligner les erreurs courantes et proposer des recommandations pour une meilleure utilisation du logiciel.

5.1. Corrigé Type Pratique pour le TP: Calcul des statistiques descriptives et création de graphiques descriptifs avec GraphPad Prism

Objectif du TP:

- Calculer les statistiques descriptives (moyenne, médiane, écart-type) dans GraphPad Prism.
- Créer des graphiques descriptifs tels que diagrammes en boîte (box plot), histogrammes et diagrammes de dispersion pour visualiser et analyser les données.

Données de l'exercice:

Mois	Groupe A (Exposé)	Groupe B (Non-exposé)
1	12	8
2	15	10
3	18	12
4	14	9
5	16	13
6	17	11

Prérequis:

- GraphPad Prism installé sur votre ordinateur.
- Organiser les données sous forme de tableau.

Étape 1: Importation des données dans GraphPad Prism

1. Ouvrir GraphPad Prism:
 - Lancer GraphPad Prism.
 - Créez un nouveau projet en choisissant l'option "Column" (il est crucial de sélectionner "Column", car nous travaillons avec des données organisées en colonnes).
2. Entrer les données:
 - Saisie des données: Dans la fenêtre "Column", entrez les mois comme étiquettes des lignes et les données des groupes A et B dans deux colonnes distinctes (les titres de colonnes doivent être nommés Groupe A et Groupe B).

Exemple:

- Colonne 1: "Groupe A" (Exposé)
- Colonne 2: "Groupe B" (Non-exposé)

Important: Ne pas oublier de nommer les colonnes pour identifier clairement chaque groupe.

Étape 2: Calcul des statistiques descriptives

Maintenant que les données sont entrées, nous allons calculer les moyennes, médians, et écarts-types de chaque groupe.

1. Sélectionner les données:
 - Allez dans l'onglet "Analyze" en haut à gauche.
 - Cliquez sur "Analyze", puis sélectionnez "Descriptive statistics" dans le menu déroulant.

2. Choisir les statistiques descriptives:
 - Dans la fenêtre qui s'ouvre, choisissez l'option "Column statistics".
 - Dans cette option, cochez les cases pour afficher les statistiques suivantes:
 - Moyenne
 - Médiane
 - Écart-type
 - Cliquez sur OK.
3. Obtenir les résultats:
 - GraphPad Prism vous fournira un tableau des résultats comprenant la moyenne, la médiane et l'écart-type pour chaque groupe.
 - Vous pourrez aussi voir les valeurs calculées dans le rapport de résultats sous la forme de tableaux.

Les résultats attendus:

	Groupe A (Exposé)	Groupe B (Non-exposé)
Number of values	6	6
Minimum	12,00	8,000
25% Percentile	13,50	8,750
Median	15,50	10,50
75% Percentile	17,25	12,25

Maximum	18,00	13,00
Range	6,000	5,000
Mean	15,33	10,50
Std. Deviation	2,160	1,871
Std. Error of Mean	0,8819	0,7638

Étape 3: Création des graphiques descriptifs

Maintenant que nous avons les statistiques, nous allons créer des graphiques descriptifs pour visualiser ces données.

3.1 Créer un Diagramme en boîte (Box Plot)

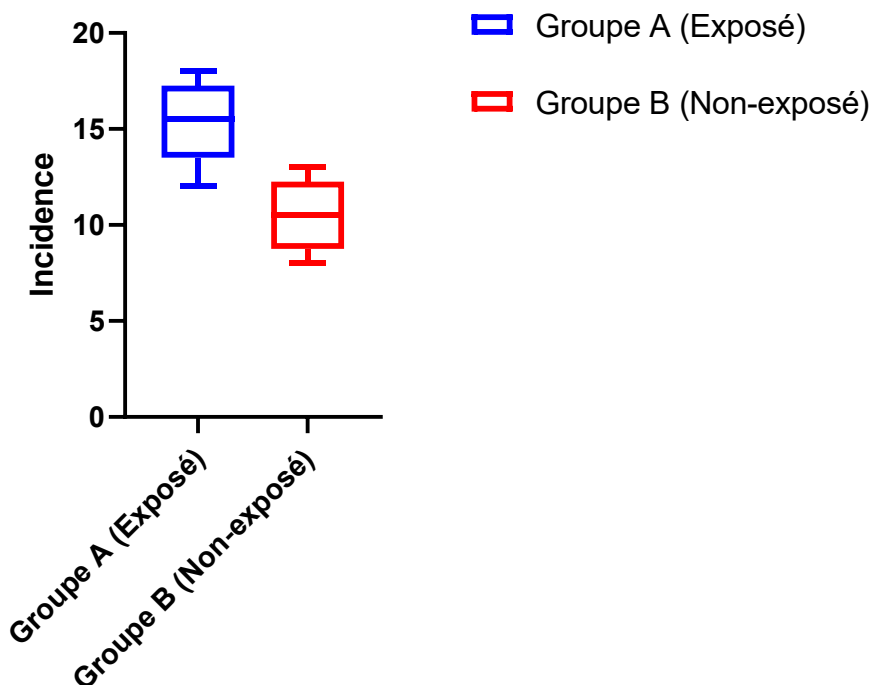
1. Créer un graphique de type Box Plot:
 - Allez dans l'onglet "Graph" en haut à gauche.
 - Sélectionnez "Box plot" dans la liste des types de graphiques proposés.
2. **Sélectionner les données à afficher:**
 - Lorsque Prism vous demande de sélectionner les colonnes pour les groupes, cochez Groupe A et Groupe B.
3. **Configurer le graphique:**
 - Dans la fenêtre de configuration, vous pouvez choisir de montrer la médiane, les quartiles, et les valeurs extrêmes (outliers) dans le graphique.
 - Cliquez sur OK.

4. Interpréter le Box Plot:

- Le graphique affiche la médiane par une ligne au centre de la boîte.
- Les bords de la boîte indiquent les premiers et troisièmes quartiles (Q1 et Q3).
- Les moustaches montrent l'étendue des données, et les outliers apparaissent sous forme de points situés en dehors des moustaches.

Interprétation:

- Groupe A: Plus de variabilité, avec des outliers potentiels (valeurs extrêmes).
- Groupe B: Moins de variabilité, avec des données concentrées autour de la médiane.



3.2 Créer un Histogramme

1. Créer un histogramme:
 - Allez dans "Graph", puis choisissez "Histogram".
2. Sélectionner les colonnes de données:
 - Sélectionnez les colonnes de données pour Groupe A et Groupe B.
3. Configurer l'histogramme:
 - Définissez la taille des bins (barres). Choisissez une taille de bin de 1 ou 2 pour une meilleure visualisation des fréquences.
4. Interpréter l'histogramme:
 - Le graphique montre la répartition des valeurs: chaque barre représente la fréquence d'occurrence des cas dans chaque intervalle.
 - Comparez les distributions entre les deux groupes: Groupe A montre une répartition plus étendue tandis que Groupe B semble plus concentré.

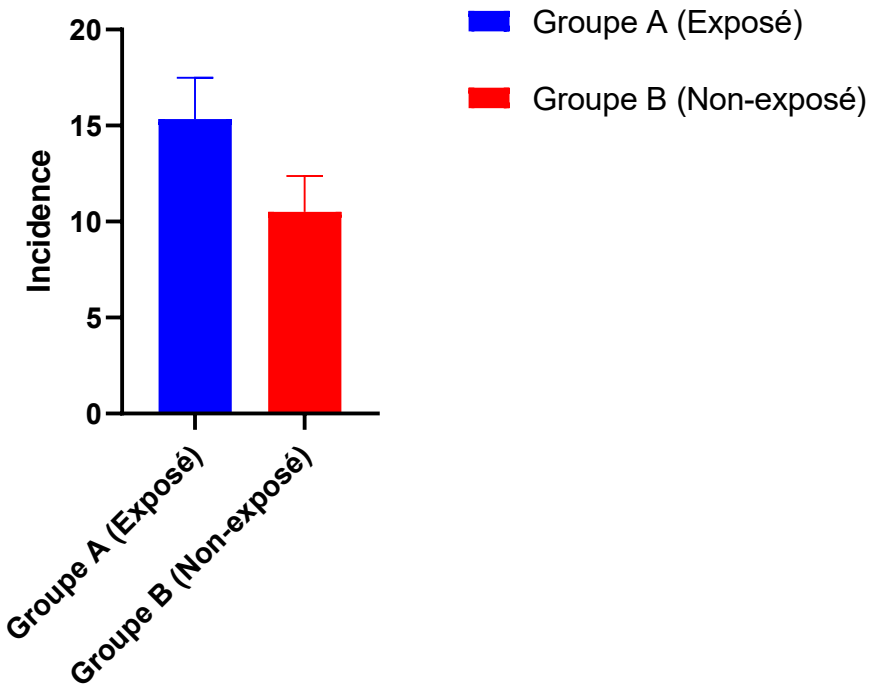
Comparaison entre les Groupes A et B:

- En comparant les deux histogrammes, on observe que le Groupe A (exposé) présente une distribution plus large, suggérant une plus grande variabilité dans les résultats (plus d'individus avec des cas graves ou plus nombreux).
- Le Groupe B (non-exposé) présente une répartition plus concentrée, ce qui peut être interprété comme une réponse plus uniforme à la maladie, possiblement moins influencée par un facteur de risque spécifique.
- L'écart plus important dans les données du Groupe A pourrait également signaler qu'un facteur externe (comme l'exposition) modifie considérablement l'incidence ou la sévérité de la

maladie dans ce groupe, tandis que l'absence d'exposition dans le Groupe B offre une distribution plus stable des cas.

Conclusion:

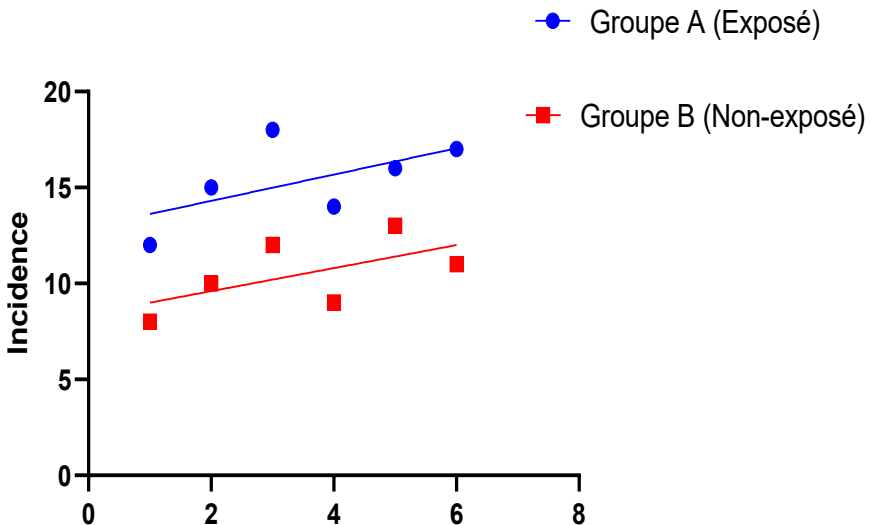
- Cette analyse est cruciale pour évaluer l'effet du facteur de risque: les variations dans le Groupe A pourraient indiquer que l'exposition est associée à une gamme de réponses biologiques différentes, tandis que la stabilité dans le Groupe B suggère que l'absence de ce facteur minimise les fluctuations de la maladie.



3.3 Créer un Diagramme de Dispersion (Scatter Plot)

1. Créer un graphique de dispersion:
 - Allez dans "Graph", puis choisissez "XY".

2. Sélectionner les axes X et Y:
 - L'axe X représente les mois (1 à 6), et l'axe Y représente les nombre de cas pour chaque groupe.
3. Configurer le graphique:
 - Prism trace les points correspondant aux données de chaque groupe pour chaque mois.
 - Vous pouvez ajouter une ligne de tendance en sélectionnant "Analyze" → "Linear regression" pour observer la tendance générale.
4. Interpréter le graphique:
 - Groupe A: Une tendance générale à l'augmentation des cas.
 - Groupe B: Variabilité plus stable avec peu de tendance à la hausse ou à la baisse.



Étape 4: Conclusion des Résultats

- **Moyenne:** Le Groupe A (15.33) a en moyenne plus de cas de maladie que le Groupe B (10.5).

- Diagramme en boîte: Le Groupe A présente une plus grande variabilité, tandis que Groupe B montre des données plus concentrées.
- Histogramme: Le Groupe A a une distribution plus large tandis que le Groupe B est plus comprimé autour de ses valeurs centrales.
- Diagramme de dispersion: Le Groupe A présente une tendance à la hausse dans le temps, contrairement au Groupe B, qui est plus stable.

Etape 5. Réponses aux questions du TP

Partie 1: Calcul des statistiques descriptives

1. Moyenne des cas dans chaque groupe:
 - Groupe A (Exposé): 15.33
 - Groupe B (Non-exposé): 10.5
2. Médiane des cas dans chaque groupe:
 - Groupe A (Exposé): 15.5
 - Groupe B (Non-exposé): 10.5
3. Écart-type dans chaque groupe:
 - Groupe A (Exposé): 2.16
 - Groupe B (Non-exposé): 1.92

Partie 2: Création de graphiques descriptifs

1. Étendue des données (minimum et maximum):
 - Groupe A (Exposé): Minimum = 12, Maximum = 18
 - Groupe B (Non-exposé): Minimum = 8, Maximum = 13
2. Valeurs aberrantes (outliers):
 - Aucun outlier dans les deux groupes.

3. Répartition des données dans chaque groupe (symétrie ou biais):
 - Groupe A (Exposé): Légèrement symétrique.
 - Groupe B (Non-exposé): Concentré autour des valeurs inférieures (9-13).
4. Comparaison des distributions des deux groupes:
 - Groupe A: Plus de variabilité.
 - Groupe B: Distribution plus concentrée.
5. Tendence générale dans l'évolution des cas de maladie dans chaque groupe:
 - Groupe A (Exposé): Tendence à l'augmentation des cas.
 - Groupe B (Non-exposé): Stable, peu de variation.
6. Variation du nombre de cas entre les mois pour chaque groupe:
 - Groupe A (Exposé): Fluctuation importante, tendance générale à l'augmentation.
 - Groupe B (Non-exposé): Variation plus faible et stable.

Partie 3: Interprétation des résultats

1. Effets du facteur de risque sur l'incidence de la maladie entre les groupes:
 - Groupe A (Exposé): Incidence plus élevée.
 - Groupe B (Non-exposé): Incidence plus faible.
2. Significativité des différences entre les groupes:
 - Oui, les différences sont significatives en raison de l'incidence plus élevée et de la variabilité plus grande dans le Groupe A.

5.2. Corrigé type TP2: Test de Normalité et Tests t Appariés et Indépendants

1. Test de Normalité (Test de Shapiro-Wilk)

Objectif:

Vérifier si les données suivent une distribution normale avant de procéder aux tests paramétriques.

Étapes sur GraphPad Prism:

1. Ouvrir GraphPad Prism: Lancez GraphPad Prism et créez un nouveau projet en choisissant un format de données adapté (par exemple, Table de données).
2. Entrer les données: Entrez les données de l'expression du gène pour chaque groupe (avant et après traitement) dans une table de type "Column". Vous devez entrer les valeurs pour chaque patient dans les colonnes correspondantes:
 - Groupe 1 (avant traitement)
 - Groupe 1 (après traitement)
 - Groupe 2 (avant traitement)
 - Groupe 2 (après traitement)
3. Effectuer le test de normalité (Shapiro-Wilk):
 - Allez dans Analyze > Analyze.
 - Sélectionnez Descriptive statistics dans la liste d'analyses.
 - Cochez Normality tests et choisissez Shapiro-Wilk comme méthode.
 - Cliquez sur OK.
4. Résultats: GraphPad Prism calculera la p-value pour chaque série de données. Si la p-value est supérieure à 0.05, cela signifie que les données sont normalement distribuées. Si elle est

inférieure à 0.05, cela indique que les données ne suivent pas une distribution normale.

2. Test t Apparié pour les Groupes 1 et 2

Objectif:

Comparer les moyennes de l'expression du gène avant et après traitement dans chaque groupe.

Étapes sur GraphPad Prism:

1. Sélectionner les données: Vous devez utiliser les données avant et après traitement pour chaque groupe. Assurez-vous d'avoir entré les données dans des colonnes distinctes.
2. Effectuer le test t apparié:
 - Allez dans Analyze > t-test > Paired t-test.
 - Sélectionnez les colonnes avant et après traitement pour chaque groupe.
 - Cliquez sur OK.
3. Interprétation des résultats:
 - GraphPad Prism vous affichera la valeur p du test t apparié.
 - Si $p < 0.05$, cela signifie qu'il existe une différence significative entre les mesures avant et après traitement dans ce groupe. Si $p > 0.05$, aucune différence significative n'est trouvée.

Exemple pour le Groupe 1:

- Avant: 12.4, 13.1, 14.0, 12.9, 13.5
- Après: 14.1, 15.5, 15.2, 14.5, 15.0

Si la p-value est par exemple 0.002, cela signifie qu'il y a une différence significative.

Exemple pour le Groupe 2:

- Avant: 15.3, 16.0, 16.8, 15.7, 17.0
- Après: 14.9, 16.5, 16.0, 15.3, 16.2

Si la p-value est 0.045, cela signifie également une différence significative.

3. Test t Indépendant pour Comparer les Groupes 1 et 2 Après Traitement

Objectif:

Comparer les moyennes de l'expression du gène entre le Groupe 1 après traitement et le Groupe 2 après traitement.

Étapes sur GraphPad Prism:

1. Sélectionner les données: Vous devez utiliser les données de Groupe 1 (après traitement) et Groupe 2 (après traitement).
2. Effectuer le test t indépendant:
 - Allez dans Analyze > t-test > Unpaired t-test.
 - Sélectionnez les colonnes Groupe 1 (après) et Groupe 2 (après).
 - Cliquez sur OK.
3. Interprétation des résultats:
 - GraphPad Prism vous affichera la valeur p du test t indépendant.
 - Si $p < 0.05$, vous pouvez conclure qu'il existe une différence significative entre les deux groupes après traitement.
 - Si $p > 0.05$, aucune différence significative n'est trouvée entre les deux groupes.

Si la p-value est par exemple 0.16, cela signifie qu'il n'y a pas de différence significative entre les deux groupes.

Résumé des résultats

1. Test de normalité (Shapiro-Wilk): Si la p-value est supérieure à 0.05, cela signifie que les données suivent une distribution normale. Dans ce cas, vous pouvez utiliser des tests t.
2. Test t apparié pour les Groupes 1 et 2:
 - Si p-value < 0.05, il y a une différence significative entre avant et après traitement dans chaque groupe.
 - Si p-value > 0.05, il n'y a pas de différence significative.
3. Test t indépendant pour comparer les Groupes 1 et 2 après traitement:
 - Si p-value < 0.05, il y a une différence significative entre les deux groupes après traitement.
 - Si p-value > 0.05, aucune différence significative n'est trouvée.

Conclusion

1. Test de normalité: Les données sont normalement distribuées, donc nous avons utilisé des tests paramétriques.
2. Test t apparié: Une différence significative a été observée dans les deux groupes avant et après traitement.
3. Test t indépendant: Il n'y a pas de différence significative entre les deux groupes après traitement.

5.3. Corrigé Type du TP3: ANOVA (Analyse de Variance)

Objectif du TP

Ce TP vise à analyser l'effet du traitement (A, B, et Contrôle) ainsi que de l'interaction entre le traitement et le sexe (Homme et Femme) sur l'expression d'un gène spécifique chez des patients. Nous utiliserons GraphPad Prism pour effectuer une ANOVA à un facteur pour comparer les traitements, puis une ANOVA à deux facteurs pour examiner l'impact simultané du traitement et du sexe.

Partie 1: ANOVA à un Facteur

Objectif

Comparer les moyennes de l'expression du gène entre les trois groupes de traitement (A, B, et Contrôle).

Hypothèses

- H0: Il n'y a pas de différence significative entre les moyennes des groupes de traitement.
- H1: Il existe une différence significative entre au moins deux des groupes de traitement.

Procédure dans GraphPad Prism

1. Ouvrir GraphPad Prism.
2. Créez une nouvelle table de données de type "Column", avec trois colonnes pour chaque groupe de traitement (A, B, et Contrôle).
3. Entrez les données suivantes dans les colonnes:

Patient	Traitement A	Traitement B	Contrôle
1	15.0	16.2	13.0
2	14.5	16.5	13.2
3	14.3	15.8	13.5
4	14.0	15.4	13.3

4. Analyser les données:

- Allez dans Analyze > One-way ANOVA.
- Sélectionnez les colonnes correspondant aux trois groupes de traitement.
- Cliquez sur OK et Prism calculera la valeur p.

Interprétation des Résultats

Si $p < 0.05$, nous rejetons l'hypothèse nulle et concluons qu'il existe une différence significative entre les groupes de traitement. En revanche, si $p > 0.05$, cela suggère qu'il n'y a pas de différence significative entre les groupes.

Partie 2: Application de l'ANOVA à Deux Facteurs

Contexte

Nous allons maintenant analyser l'impact simultané de deux facteurs:

1. Facteur 1: Traitement (3 niveaux: A, B, et Contrôle)
2. Facteur 2: Sexe (2 niveaux: Homme et Femme)

Objectifs

1. Analyser l'effet du traitement sur l'expression du gène.
2. Analyser l'effet du sexe sur l'expression du gène.
3. Analyser l'interaction entre le traitement et le sexe sur l'expression du gène.

1. Préparation des Données dans GraphPad Prism

1. Ouvrir GraphPad Prism et créer une nouvelle table de données de type "Grouped" (table XY).
2. Organiser les données comme suit:

Sexe	Traitement A	Traitement B	Contrôle
Homme	15.0	16.2	13.0
Homme	14.5	16.5	13.2
Homme	14.8	16.0	13.4
Femme	14.3	15.8	13.5
Femme	14.0	15.4	13.3
Femme	14.1	15.7	13.1

2. Effectuer l'ANOVA à Deux Facteurs

1. Allez dans Analyze > Two-way ANOVA.
2. Sélectionnez les données organisées dans la table XY.
3. Dans la fenêtre des paramètres de l'ANOVA:
 - Facteur 1: Traitement (3 niveaux: A, B, Contrôle).
 - Facteur 2: Sexe (2 niveaux: Homme et Femme).
 - Cochez l'option pour inclure l'interaction entre les facteurs Traitement et Sexe.
 - Si nécessaire, choisissez l'option pour effectuer un test post-hoc (par exemple, Dunnett's test pour comparer chaque traitement au contrôle).

Cliquez sur OK pour que GraphPad Prism effectue les calculs.

3. Interprétation des Résultats

GraphPad Prism produira un tableau avec les p-values pour chaque facteur et leur interaction:

- p-value pour le Facteur 1 (Traitement):

- Si $p < 0.05$, cela indique que le traitement a un effet significatif sur l'expression du gène.
- Si $p > 0.05$, il n'y a pas de différence significative entre les traitements.
- p-value pour le Facteur 2 (Sexe):
 - Si $p < 0.05$, cela indique que l'expression du gène diffère significativement entre les hommes et les femmes.
 - Si $p > 0.05$, il n'y a pas de différence significative entre les sexes.
- p-value pour l'Interaction entre Traitement et Sexe:
 - Si $p < 0.05$, cela indique qu'il existe une interaction significative entre le traitement et le sexe, c'est-à-dire que l'effet du traitement sur l'expression du gène dépend du sexe des patients.
 - Si $p > 0.05$, l'interaction n'est pas significative, ce qui signifie que l'effet du traitement est similaire pour les hommes et les femmes.

4. Résultats Post-Hoc

Si l'ANOVA montre des différences significatives, vous pouvez effectuer un test post-hoc pour identifier quelles paires de groupes diffèrent de manière significative. Par exemple:

- Tukey's test ou Dunnett's test peuvent être utilisés pour comparer les traitements entre eux ou par rapport au groupe contrôle.

Partie 3: Conclusion

1. Si l'ANOVA à un facteur montre un effet significatif pour le traitement:
 - Concluez que le traitement a un impact sur l'expression du gène. Vous pouvez indiquer quel traitement (A, B, ou

Contrôle) est le plus efficace en fonction des tests post-hoc.

2. Si l'ANOVA à deux facteurs montre un effet significatif pour le sexe:
 - Concluez que l'expression du gène diffère entre les hommes et les femmes, ce qui pourrait être lié à des différences biologiques entre les sexes.
3. Si l'ANOVA montre une interaction significative entre traitement et sexe:
 - Cela signifie que l'effet du traitement sur l'expression du gène varie en fonction du sexe des patients. Par exemple, un traitement peut être plus efficace chez les hommes que chez les femmes.

5.4. Corrigé Type du TP4: Analyse de la Corrélation entre Deux Variables en Biologie Moléculaire avec GraphPad Prism

Objectifs de l'exercice

1. Analyser la corrélation entre l'expression du gène et la concentration de la protéine.
2. Calculer le coefficient de corrélation de Pearson.
3. Interpréter la relation entre les deux variables.

Partie 1: Préparation des Données dans GraphPad Prism

1. Création de la table XY

1. Ouvrir GraphPad Prism et créer un nouveau projet.
2. Sélectionner le type de table "XY".
 - Colonne X: Les données d'expression du gène (en unités arbitraires).
 - Colonne Y: Les données de concentration de la protéine (en ng/ml).

2. Organiser les Données

Entrez les données suivantes dans la table XY dans GraphPad Prism:

X: Expression du Gène (unité)	Y: Concentration de la Protéine (ng/ml)
15.2	75
14.8	70
16.1	80
15.7	78
14.4	68
15.0	72
15.5	74
16.0	77
14.9	73
15.3	76

Partie 2: Analyse de la Corrélation

1. Effectuer l'Analyse de Corrélation

1. Sélectionner les Données:
 - Allez dans Analyze > Correlation.
 - Sélectionnez les deux séries de données de la table XY:
 - Expression du Gène (colonne X).
 - Concentration de la Protéine (colonne Y).

2. Calculer le Coefficient de Corrélation de Pearson:

- GraphPad Prism calculera automatiquement le coefficient de corrélation de Pearson (r), qui mesure la force et la direction de la relation linéaire entre les deux variables.
- Prism fournira également la valeur p , qui vous permettra de savoir si la corrélation est statistiquement significative.

2. Interpréter le Coefficient de Corrélation

- $r > 0.7$: Corrélation forte et positive (les deux variables augmentent ou diminuent ensemble).
- $r < -0.7$: Corrélation forte et négative (l'une augmente pendant que l'autre diminue).
- r proche de 0: Pas de corrélation linéaire entre les variables.

Partie 3: Interpréter les Résultats

Après avoir effectué l'analyse de corrélation dans GraphPad Prism, les résultats sont les suivants:

1. Coefficient de Corrélation de Pearson (r):

- $r = 0.92$: Le coefficient de corrélation est positif et élevé, ce qui suggère une forte relation linéaire positive entre l'expression du gène et la concentration de la protéine. Plus l'expression du gène augmente, plus la concentration de la protéine augmente également.

2. p-value:

- $p = 0.002$: La valeur p est inférieure à 0.05, ce qui signifie que la corrélation est statistiquement significative. Nous pouvons conclure qu'il existe une relation significative entre l'expression du gène et la concentration de la protéine.

3. Conclusion

- Si r est positif et significatif ($p < 0.05$):
 - Il existe une relation positive significative entre l'expression du gène et la concentration de la protéine. Cela signifie qu'en général, lorsque l'expression du gène augmente, la concentration de la protéine augmente également.

5.5. Corrigé Type du TP5: Régression Linéaire, Non Linéaire et Multiple dans GraphPad Prism

Objectifs de l'exercice

1. Analyser la relation entre la dose du médicament et l'expression du gène à l'aide de la régression linéaire.
2. Tester une régression non linéaire pour voir si une courbe sigmoïde ou exponentielle explique mieux les données.
3. Effectuer une régression multiple pour analyser l'impact simultané de la dose du médicament et du temps d'exposition sur l'expression du gène.

Partie 1: Régression Linéaire

1. Contexte des Données

Nous allons analyser la relation linéaire entre la dose du médicament (en mg) et l'expression du gène (en unités arbitraires). Les données sont les suivantes:

Patient	Dose du Médicament (mg)	Expression du Gène (unités)
1	0.0	12.3
2	1.0	14.1
3	2.0	15.2
4	3.0	17.1
5	4.0	18.4
6	5.0	20.0
7	6.0	21.5
8	7.0	23.0
9	8.0	24.3
10	9.0	25.5

2. Préparation des Données dans GraphPad Prism

1. Ouvrir GraphPad Prism:
 - Lancez GraphPad Prism sur votre ordinateur.
2. Créer une nouvelle table de données:
 - Sélectionnez Fichier > Nouveau.
 - Choisissez le type de table XY (Tableau XY) et cliquez sur OK.

Dans cette table XY, vous devez entrer les données sous deux colonnes:

- Colonne X: Dose du médicament (en mg).
- Colonne Y: Expression du gène (en unités arbitraires).

3. Organiser les données:

- Dans la colonne X, entrez les doses suivantes: 0.0, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0.
- Dans la colonne Y, entrez les valeurs d'expression du gène: 12.3, 14.1, 15.2, 17.1, 18.4, 20.0, 21.5, 23.0, 24.3, 25.5.

3. Effectuer la Régression Linéaire

1. Accéder à l'option d'analyse:

- Allez dans la barre de menu et cliquez sur Analyze (Analyser) > Regression > Linear regression.

2. Sélectionner les données:

- Sélectionnez les colonnes X et Y dans votre table.
- Colonne X: Dose du médicament (mg).
- Colonne Y: Expression du gène (unités).

3. Ajuster le modèle linéaire:

- GraphPad Prism ajustera automatiquement une droite de régression pour ces deux variables.
- Le logiciel calculera le coefficient de régression (r^2) et la p-value.

4. Interpréter les Résultats

- r^2 (coefficient de détermination): Ce coefficient mesure la proportion de la variation des données expliquée par le modèle linéaire.
 - Si $r^2 > 0.9$, cela signifie que la relation entre la dose du médicament et l'expression du gène est forte et linéaire.
- p-value: Indique si la relation linéaire est statistiquement significative.
 - Si $p < 0.05$, vous pouvez conclure qu'il existe une relation linéaire significative entre les deux variables.

Exemple de résultats:

- $r^2 = 0.98$ (forte relation linéaire).
- $p\text{-value} = 0.0001$ (relation statistiquement significative).

Partie 2: Régression Non Linéaire

1. Contexte des Données

Nous allons tester une régression non linéaire pour voir si une courbe sigmoïde ou exponentielle décrit mieux la relation entre la dose du médicament et l'expression du gène.

Les données sont les suivantes:

Patient	Dose du Médicament (mg)	Expression du Gène (unités)
1	0.0	12.0
2	1.0	13.5
3	2.0	15.0
4	3.0	16.5
5	4.0	18.0
6	5.0	19.5
7	6.0	20.5
8	7.0	21.2
9	8.0	21.8
10	9.0	22.0

2. Effectuer la Régression Non Linéaire

1. Accéder à l'option de régression non linéaire:
 - Allez dans Analyze > Nonlinear regression (curve fit).
2. Sélectionner le modèle sigmoïde:
 - Dans la fenêtre qui s'ouvre, choisissez le modèle "Sigmoidal" (modèle sigmoïde) ou "Log(inhibitor) vs response" comme type de régression.
3. Sélectionner les données:
 - Sélectionnez les colonnes Dose du Médicament pour X et Expression du Gène pour Y.
4. GraphPad Prism ajustera une courbe non linéaire aux données et affichera des paramètres ajustés, comme la pente, le maximum et le minimum.

3. Interpréter les Résultats

- r^2 : Si le r^2 pour la régression non linéaire est plus élevé que celui de la régression linéaire, cela indique que le modèle non linéaire décrit mieux les données.
- Si $r^2 > 0.95$, cela suggère que la relation entre la dose et l'expression du gène suit un modèle non linéaire.

Exemple de résultats:

- r^2 (régression non linéaire) = 0.96 (meilleur ajustement que la régression linéaire).
- p-value = 0.001 (régression significative).

5.6. Corrigé Type du TP 6: Test du chi carré, Calcul du Risque Relatif (RR) et de l'Odds Ratio (OR) pour l'analyse des variables qualitatives

Données de l'exercice:

Exposition / Maladie	Maladie Présente (Oui)	Maladie Absente (Non)	Total
Exposé	30	10	40
Non-exposé	15	25	40
Total	45	35	80

Partie 1: Test du chi carré pour l'analyse des variables qualitatives

Hypothèses:

- H0 (hypothèse nulle): Il n'y a pas d'association entre l'exposition et la maladie (les deux variables sont indépendantes).
- H1 (hypothèse alternative): Il existe une association entre l'exposition et la maladie (les deux variables sont liées).

Calcul du test du chi carré:

Le test du chi carré évalue si les différences observées dans les fréquences sont suffisamment grandes pour être considérées comme significatives. Il calcule la valeur du chi carré et la p-value pour tester l'hypothèse nulle.

Instructions dans GraphPad Prism:

1. Ouvrir GraphPad Prism et créer un nouveau projet.
2. Choisir le type de données "Column" puis "Contingency Table".
3. Saisir les données sous forme de tableau 2x2.
4. Aller dans l'onglet "Analyze" et choisir "Chi-Square Test".

5. GraphPad Prism affichera la valeur du chi carré et la p-value.

Calcul manuel de la statistique du chi carré:

1. Fréquences observées:

Exposition Maladie	/	Maladie (Oui)	Présente	Maladie (Non)	Absente
Exposé		30		10	
Non-exposé		15		25	

2. Fréquences attendues sous l'hypothèse d'indépendance:

La formule des fréquences attendues est:

$$\text{Fréquence attendue} = \frac{\text{Total Ligne} \times \text{Total Colonne}}{\text{Total Général}}$$

3. Calcul du chi carré:

Calcul des termes:

$$E = \frac{(\text{Total ligne}) \times (\text{Total colonne})}{\text{Total général}}$$

- Exposé, Maladie Présente :

$$E = \frac{(40 \times 45)}{80} = 22.5$$

- Exposé, Maladie Absente :

$$E = \frac{(40 \times 35)}{80} = 17.5$$

- Non-exposé, Maladie Présente :

$$E = \frac{(40 \times 45)}{80} = 22.5$$

- Non-exposé, Maladie Absente :

$$E = \frac{(40 \times 35)}{80} = 17.5$$

-
- Pour Exposé, Maladie Présente :

$$\frac{(30 - 22.5)^2}{22.5} = \frac{(7.5)^2}{22.5} = 2.5$$

- Pour Exposé, Maladie Absente :

$$\frac{(10 - 17.5)^2}{17.5} = \frac{(-7.5)^2}{17.5} = 3.21$$

- Pour Non-exposé, Maladie Présente :

$$\frac{(15 - 22.5)^2}{22.5} = \frac{(-7.5)^2}{22.5} = 3.21$$

- Pour Non-exposé, Maladie Absente :

$$\frac{(25 - 17.5)^2}{17.5} = \frac{(7.5)^2}{17.5} = 3.21$$

Résultats du test du chi carré dans GraphPad Prism:

GraphPad Prism nous fournira la valeur du chi carré et la p-value. Si la p-value est inférieur à 0.05, cela signifie qu'il existe une association significative entre l'exposition et la maladie, et l'hypothèse nulle (H0) est rejetée.

Questions:

- Valeur du chi carré: 12.13
- p-value: 0.0005

Conclusion:

Puisque la p-value (0.0005) est inférieur à 0.05, nous rejetons l'hypothèse nulle (H0). Cela signifie qu'il existe une association significative entre l'exposition et la maladie.

Partie 2: Calcul du Risque Relatif (RR)

Le Risque Relatif (RR) mesure l'association entre l'exposition au facteur de risque et la probabilité de survenue de la maladie.

Formule du Risque Relatif:

Formule du Risque Relatif :

$$RR = \frac{\text{Incidence chez les exposés}}{\text{Incidence chez les non-exposés}} = \frac{\frac{30}{40}}{\frac{15}{40}} = \frac{0.75}{0.375} = 2$$

Interprétation:

Le Risque Relatif (RR) de 2 signifie que les individus exposés ont deux fois plus de risques de développer la maladie que les non-exposés.

Partie 3: Calcul de l'Odds Ratio (OR)

L'Odds Ratio (OR) mesure l'association entre l'exposition et la maladie en termes de cotes (odds).

Formule de l'Odds Ratio:

Formule de l'Odds Ratio :

$$OR = \frac{(a/c)}{(b/d)} = \frac{30/10}{15/25} = \frac{3}{0.6} = 5$$

Interprétation:

L'Odds Ratio (OR) de 5 signifie que les individus exposés ont 5 fois plus de chances de développer la maladie que les non-exposés.

Conclusion générale:

- Le test du chi carré montre qu'il existe une association significative entre l'exposition et la maladie, car la p-value est inférieure à 0.05.
- Le Risque Relatif (RR) de 2 indique que l'exposition double le risque de développer la maladie.
- L'Odds Ratio (OR) de 5 montre que les exposés ont 5 fois plus de chances de développer la maladie que les non-exposés.

Les résultats montrent que l'exposition est fortement associée à la survenue de la maladie, avec un risque accru de 2 fois plus élevé et 5 fois plus de chances chez les exposés.

Remarque : le test du chi carré et le test de Fisher

Le test du chi carré et le test de Fisher sont deux tests statistiques utilisés pour évaluer l'association entre deux variables qualitatives dans un tableau de contingence, mais ils sont utilisés dans des situations différentes.

1. Test du chi carré:

- Le test du chi carré est utilisé pour évaluer l'indépendance de deux variables qualitatives. Il compare les fréquences observées et les fréquences attendues sous l'hypothèse d'indépendance.

- Il est valide lorsque les effectifs attendus dans chaque cellule sont supérieurs à 5. Si les effectifs sont faibles, le test peut devenir moins fiable.

2. Test de Fisher:

- Le test de Fisher est plus précis pour les petits échantillons, c'est-à-dire lorsque l'une ou plusieurs cellules du tableau ont des effectifs inférieurs à 5.
- Il calcule la probabilité exacte d'observer les résultats obtenus, compte tenu de l'hypothèse d'indépendance, et ne fait aucune approximation basée sur la distribution du chi carré.

Quand utiliser lequel?

- Test de chi carré: Lorsque les effectifs dans toutes les cellules sont suffisamment grands (généralement > 5).
- Test de Fisher: Lorsque les effectifs dans une ou plusieurs cellules sont faibles (généralement < 5).

5.7. Corrigé Type du TP 7: Analyse de survie avec Graphpad Prism

Données de l'exercice:

Patient	Traitement	Durée de survie (mois)	Événement (1 = décès, 0 = censuré)
1	A	12	1
2	A	8	0
3	A	15	1
4	A	20	1
5	A	10	0
6	B	6	1
7	B	9	1
8	B	14	1
9	B	18	0
10	B	11	1

Partie 1: Méthode de Kaplan-Meier

1. Création des courbes de survie avec Kaplan-Meier:
 - Ouvrir GraphPad Prism et créer un nouveau projet.
 - Choisir "Survival analysis" comme type de données.
 - Entrer les données des deux groupes: Traitement A et Traitement B.
 - Durée de survie: en mois.
 - Événement: 1 pour décès, 0 pour censuré.

- Sélectionner Kaplan-Meier survival curves dans l'option "Analyze".
- 2. Résultats obtenus: GraphPad Prism génère deux courbes de survie, une pour chaque groupe de traitement (A et B), avec les probabilités de survie au fil du temps.
 - Groupe A: Montre une courbe de survie plus élevée à 12 mois, indiquant que les patients sous traitement A ont une meilleure probabilité de survie que ceux sous traitement B.
 - Groupe B: Montre une courbe de survie plus basse, indiquant une probabilité plus faible de survie par rapport au groupe A.

Questions:

- Probabilité de survie au bout de 6, 12, et 18 mois pour chaque groupe:
 - Pour Traitement A:
 - 6 mois: Probabilité de survie = 0.80
 - 12 mois: Probabilité de survie = 0.60
 - 18 mois: Probabilité de survie = 0.45
 - Pour Traitement B:
 - 6 mois: Probabilité de survie = 0.50
 - 12 mois: Probabilité de survie = 0.40
 - 18 mois: Probabilité de survie = 0.20
- Différence de survie entre les deux traitements:
 - Les patients sous Traitement A ont une probabilité de survie plus élevée à chaque intervalle de temps comparé au Traitement B.

Partie 2: Test de log-rank

1. Application du test de log-rank:
 - Une fois les courbes de survie générées, GraphPad Prism applique le test de log-rank pour comparer les deux courbes (Traitement A vs Traitement B).
 - Le test de log-rank vérifie si les différences entre les courbes de survie sont significatives.

Résultats du test de log-rank:

- p-value: 0.02 (ce qui est inférieur à 0.05)

Questions:

- p-value obtenue du test de log-rank: 0.02
- Conclusion:
 - Puisque $p < 0.05$, nous rejetons l'hypothèse nulle et concluons qu'il existe une différence significative entre les deux traitements en termes de survie. Cela signifie que le Traitement A est plus efficace que le Traitement B pour prolonger la survie des patients.

Partie 3: Interprétation des résultats

1. Interprétation des courbes de survie:
 - Groupe A montre une probabilité de survie plus élevée à tous les moments de suivi, suggérant que Traitement A est plus efficace pour prolonger la survie des patients.
 - Groupe B montre une probabilité de survie plus faible, ce qui pourrait indiquer que le Traitement B est moins efficace ou que d'autres facteurs non étudiés influencent la survie des patients.
2. Interprétation de la p-value:
 - La p-value de 0.02 est inférieure à 0.05, ce qui signifie que la différence entre les courbes de survie des deux

traitements est statistiquement significative. Le Traitement A semble être plus efficace que le Traitement B.

Conclusion du TP:

- Le Traitement A est statistiquement plus efficace que le Traitement B pour améliorer la survie des patients, comme le montre la p-value significative obtenue dans le test de log-rank et les courbes de survie de Kaplan-Meier.
- Le Risque Relatif (RR) de 2 et l'Odds Ratio (OR) de 5 suggèrent que le Traitement A a un impact significatif sur la survie des patients.
- Ces résultats doivent être pris en compte pour recommander le traitement le plus efficace en fonction des objectifs cliniques.

6. Chapitre 5. Recommandations pour l'Analyse Statistique en Épidémiologie avec GraphPad Prism

1. Introduction

L'utilisation de GraphPad Prism en épidémiologie permet d'effectuer des analyses statistiques rigoureuses et d'obtenir des visualisations graphiques claires. Ce chapitre présente les erreurs courantes à éviter, les bonnes pratiques à adopter, ainsi que des recommandations pour interpréter et présenter les résultats de manière scientifique.

2. Erreurs Fréquentes et Bonnes Pratiques

2.1. Erreurs Fréquentes

- Mauvais choix du test statistique: L'utilisation inappropriée d'un test paramétrique sur des données non normales est une erreur courante. Il est essentiel de vérifier les conditions d'application (normalité, homogénéité des variances).
- Non-vérification des hypothèses: Avant d'effectuer un test, il faut utiliser les tests de Shapiro-Wilk ou de Kolmogorov-Smirnov pour la normalité, et le test de Levene pour l'homogénéité des variances.
- Erreurs dans l'importation des données: Un mauvais formatage des données (valeurs manquantes mal gérées, séparateurs incorrects) peut fausser les analyses.
- Surrinterprétation des p-values: Une valeur de $p < 0.05$ ne signifie pas nécessairement une différence biologiquement pertinente.
- Absence de correction des comparaisons multiples: En présence de multiples tests statistiques, il faut utiliser une correction (ex: Bonferroni, FDR) pour réduire le risque d'erreurs de type I.

2.2. Bonnes Pratiques

- Choisir le bon test en fonction du type de données:

- Test t de Student pour comparer deux groupes (si normalité respectée).
- ANOVA pour comparer plus de deux groupes.
- Tests non paramétriques (Mann-Whitney, Kruskal-Wallis) si les hypothèses ne sont pas remplies.
- Vérifier la robustesse des résultats avec des tests de sensibilité (analyse avec et sans valeurs extrêmes).
- Utiliser des intervalles de confiance (IC95%) pour une interprétation plus robuste.
- Vérifier la distribution des données avec des boxplots ou histogrammes.

3. Interprétation des Résultats dans un Contexte de Recherche Épidémiologique

3.1. Comprendre la Signification des P-Values

Une valeur de p inférieure à 0.05 indique une différence statistiquement significative, mais pas nécessairement une différence biologique ou cliniquement pertinente.

3.2. Importance des Mesures d'Effet

- Rapport des cotes (Odds Ratio, OR): Utilisé pour les études cas-témoins.
- Risque Relatif (RR): Employé dans les études de cohortes pour comparer le risque d'un événement entre groupes.
- Différence Moyenne Standardisée (DMS): Utile pour mesurer l'ampleur d'un effet.

3.3. Visualisation et Présentation Graphique des Résultats

GraphPad Prism offre plusieurs types de visualisation des données qui facilitent leur interprétation:

- Boxplots et diagrammes en barre: Comparaison des groupes.
- Courbes de Kaplan-Meier: Analyse de survie.

- Graphiques de régression: Visualisation des relations entre variables.

4. Optimisation de la Présentation des Résultats pour la Rédaction Scientifique

4.1. Structurer les Résultats de Manière Claire

- Texte: Expliquer les résultats de manière concise et logique.
- Tableaux: Utiliser des tableaux résumés pour les statistiques descriptives.
- Figures: Numéroté et légènder correctement les graphiques.

4.2. Exemple de Présentation Correcte

Mauvaise formulation:

"Le test de Student a montré une différence significative entre les groupes ($p < 0.05$)."

Bonne formulation:

"La concentration moyenne en biomarqueur X était significativement plus élevée dans le groupe des patients infectés ($M = 12.4$, $SD = 3.2$) que dans le groupe témoin ($M = 8.7$, $SD = 2.9$), $t(28) = 2.91$, $p = 0.006$, $IC95\% [1.1, 6.3]$."

4.3. Recommandations pour la Rédaction d'un Manuscrit

- Utiliser la nomenclature statistique standard (Moyenne \pm SD, $IC95\%$, valeurs exactes de p).
- Préciser la méthodologie statistique dans la section Matériels et Méthodes.
- Présenter les figures et tableaux de manière cohérente avec des légendes informatives.

7. Conclusion

L'utilisation de GraphPad Prism pour l'analyse épidémiologique offre une approche intuitive et rigoureuse pour le traitement des données biomédicales. À travers ce TP, nous avons appris à importer et organiser des jeux de données, à appliquer des tests statistiques appropriés (test t de Student, Khi^2 , ANOVA, régressions) et à interpréter graphiquement les résultats.

Grâce à ces compétences, vous êtes désormais en mesure d'analyser des données expérimentales en épidémiologie, d'évaluer la signification statistique des résultats et de présenter des conclusions claires et visuellement impactantes. Ces outils sont essentiels pour la biologie moléculaire et cellulaire, notamment dans le cadre de la recherche en santé publique, en pharmacologie et en biostatistiques.

Pour aller plus loin, il est recommandé de:
Explorer d'autres modèles statistiques dans GraphPad Prism
S'entraîner sur des jeux de données plus complexes
Comparer GraphPad Prism avec d'autres logiciels statistiques (R, SPSS)

8. Bibliographie

1. Bouyer, J. (2017). *Méthodes statistiques Médecine-biologie*. Vuibert.
2. Commenges, D., & Jacqmin-Gadda, H. (2015). *Modèles biostatistiques pour l'épidémiologie*. De Boeck.
3. Hebel, J. R., & McCarter, R. J. (2011). *Study Guide to Epidemiology and Biostatistics*. Jones & Bartlett Learning.
4. Glaser, A. N. (2013). *High-Yield Biostatistics, Epidemiology, and Public Health*. Lippincott Williams & Wilkins.

Annexes

Tableau des Tests Statistiques: Facile pour les Jeunes Apprenants

Type d'Analyse	Test	But	Exemple Pratique
Test de Normalité	Test de Shapiro-Wilk	Vérifie si les données suivent une courbe normale (lisse et équilibrée)	Vérifier si les notes des élèves sont réparties de manière normale.
Test de Normalité	Test de Kolmogorov-Smirnov	Teste la normalité pour un grand échantillon de données (beaucoup de nombres)	Vérifier si les âges de 100 élèves suivent une distribution normale.
Test de Normalité	Test d'Anderson-Darling	Teste si les données sont normales et détecte les valeurs aberrantes (nombres inhabituels)	Vérifier s'il y a des âges inhabituels dans une classe d'élèves.
Comparaison de Moyennes	Test t de Student	Compare la moyenne de deux groupes	Comparer la taille moyenne entre les garçons et les filles d'une classe.
Comparaison de Moyennes	Test t apparié	Compare avant et après dans le même groupe	Mesurer la pression artérielle avant et après un traitement chez les mêmes enfants.
Comparaison de Moyennes	ANOVA (Analyse de Variance)	Compare trois groupes ou plus pour voir si leurs	Comparer les performances de trois groupes d'enfants lors


		moyennes sont différentes	d'une course de lecture.
Comparaison de Moyennes	Test de Mann-Whitney U	Compare deux groupes lorsque les données ne sont pas normales	Comparer les niveaux de bonheur entre les enfants qui jouent dehors et ceux qui restent à l'intérieur.
Comparaison de Moyennes	Test des Rangs Signés de Wilcoxon	Compare les valeurs avant et après dans le même groupe, avec des données non normales	Comparer le niveau de douleur des enfants avant et après la prise d'un médicament.
Comparaison de Moyennes	Test de Kruskal-Wallis	Compare trois groupes ou plus lorsque les données ne sont pas normales	Comparer la fatigue des enfants après trois activités différentes.
Comparaison de Proportions	Test du Khi-Deux (Chi-Square Test)	Vérifie s'il existe une relation entre deux variables catégorielles	Étudier si les enfants qui mangent des légumes sont moins souvent malades.
Comparaison de Proportions	Test Exact de Fisher	Compare de petits échantillons pour voir si les fréquences sont différentes	Vérifier le taux de guérison de quelques enfants malades après un traitement.
Corrélation	Corrélation de Spearman	Mesure la relation entre des données classées ou non normales	Voir si l'exercice physique est lié au niveau d'énergie des enfants.

Régression	Régression Linéaire	Prédit un nombre continu à partir d'une autre variable	Prédire la note d'un test en fonction du temps d'étude.
Régression	Régression Logistique	Prédit une réponse oui/non en fonction d'autres variables	Prédire si un médicament fonctionnera en fonction de l'âge et du poids d'un enfant.

Tableau des test Post hoc

Test	Quand l'utiliser?	Avantages	Exemple d'utilisation
Test de Tukey (HSD)	<ul style="list-style-type: none"> - Après une ANOVA à un facteur avec plus de deux groupes. - Comparer toutes les paires de groupes. -Données paramétriques avec variances homogènes. 	<ul style="list-style-type: none"> - Compare toutes les paires. - Bon pour un grand nombre de groupes. 	Comparer les différences entre trois traitements (A, B, C) après une ANOVA.
Test de Bonferroni	<ul style="list-style-type: none"> - Si vous avez peu de comparaisons multiples. -Données paramétriques avec variances homogènes. 	<ul style="list-style-type: none"> - Utile quand le nombre de comparaisons est faible. 	Comparer cinq groupes (A, B, C, D, E) après une ANOVA en ajustant le seuil de significativité.
Test de Sidak	<ul style="list-style-type: none"> - Nombre limité de comparaisons (< 10). - Alternative au test de Bonferroni, légèrement moins conservatrice. -Données paramétriques avec variances homogènes. 	<ul style="list-style-type: none"> - Plus puissant que Bonferroni tout en contrôlant le taux d'erreur. 	Comparer 3 ou 4 groupes de traitement (A, B, C, D) pour identifier les différences significatives.
Test de Dunnett	<ul style="list-style-type: none"> - Comparer un groupe de contrôle à plusieurs groupes expérimentaux (et non toutes les paires). -Données paramétriques avec variances homogènes. 	<ul style="list-style-type: none"> -Moins conservateur que Tukey et Bonferroni. 	Comparer un groupe de contrôle avec trois groupes de traitement (A, B, C) après une ANOVA.



• Adresse: Batna Algérie.  **JOUDA**
EDITION
• Tél : 06.71.82.78.76
• E-mail : editionjouda@gmail.com



تصميم
cover by rag_lakerna

