

Dans ce chapitre :

Échelles de mesure 564

La valeur nominale 564

La valeur ordinale 564

L'intervalle 566

La proportion 566

Types de statistiques 566

Les statistiques paramétriques 566

Les statistiques non paramétriques 566

Les statistiques descriptives 567

Les statistiques déductives 567

Plan de recherche et test d'hypothèse 567

Les variables indépendantes et dépendantes 568

Le test d'hypothèse 568

Les procédures d'échantillonnage 569

Organisation et présentation des données 570

La distribution de fréquences 570

Les graphiques 571

Les diagrammes à colonnes 573

Les diagrammes circulaires 574

Mesures de tendance centrale 574

La moyenne 574

La médiane 575

Le mode 575

Rang-centile 575

Corrélation 576

Les corrélations positives et négatives 577

Les diagrammes de corrélation 579

Sommaire 579



Explorons les statistiques...

CHAPITRE 23



Introduction aux concepts statistiques

Après avoir terminé ce chapitre, vous devriez pouvoir :

- définir une variété de méthodes et de concepts statistiques ;
- discuter de la valeur et de l'utilité de la statistique ;
- interpréter des résultats statistiques fondamentaux ;
- organiser et présenter diverses informations en un format utilisable.

Les connotations négatives associées à la statistique font de cette dernière un sujet d'anxiété considérable pour plusieurs étudiants. Toutefois, la statistique, comme la recherche d'où elle émerge, a une influence sur toutes les facettes de la vie. La **statistique**, l'étude de la synthèse, de l'analyse et de l'évaluation de données, est probablement l'outil le plus important de la recherche. Son rôle principal est d'extraire des valeurs numériques d'une grande quantité de données révélant une caractéristique des éléments en observation. Une minutieuse analyse statistique de l'information fournit des résultats précis. Pour cette raison, elle est étroitement liée à la méthode de recherche scientifique.

Nous devrions démystifier d'emblée une perception éronnée. L'analyse statistique ne requiert pas de connaissances approfondies en ingénierie ou en mathématiques avancées. Au contraire, la plupart des statistiques peuvent être calculées en utilisant simplement l'arithmétique. Si vous pouvez additionner, soustraire, multiplier, diviser, trouver le carré et la racine carrée sur une calculatrice, vous pouvez inclure la statistique dans vos recherches. Travailler avec des équations peut équivaloir à lire une carte topographique. Une fois que les signes et les symboles sont bien compris, il est aisé de suivre le bon chemin vers la destination désirée ou le résultat final. En outre, des logiciels existent maintenant pour rendre le calcul des statistiques plus aisé que jamais.

Échelles de mesure

Le nombre, à lui seul, nous donne très peu d'informations. Le numéro 99 réfère-t-il à l'uniforme de Wayne Gretzky ? À votre classement dans un marathon ? À la lecture de la température ? Les échelles de mesure nous renseignent sur la complexité des nombres que nous obtenons par la recherche. Elles nous donnent les outils pour décrire les données. Ceci constitue un fait important lorsque vient le temps de choisir les procédures statistiques précises à utiliser (tableau 23.1).

La valeur nominale

Les **données nominales** sont moins complexes parce qu'elles fournissent uniquement de l'information à propos d'une différence, comme les numéros sur les uniformes sportifs : le numéro 23 est différent du numéro 99, lequel est différent du numéro 66. Cette information fondamentale indique une distinction, rien de plus. Les données nominales sont également utilisées pour la catégorisation des données ; lors d'un sondage dans lequel les répondants sont appelés à nommer la chaîne de télévision – « 3 », « 15 » ou « 30 » – qu'ils regardent le plus souvent, par exemple. Ces valeurs numériques servent principalement à classer les réponses en catégories et n'exigent aucun calcul arithmétique tel que l'addition, la soustraction, la multiplication ou la division.

La valeur ordinale

Les **données ordinales** illustrent également une différence. Toutefois, elles indiquent en plus le sens de la différence. En d'autres mots, elles illustrent le rang ; la troisième place d'une athlète dans un marathon, par exemple. La valeur ordinale révèle ici que telle athlète a réalisé une performance supérieure à celle qui a terminé en quatrième position, mais pas une performance aussi bonne que celle de la coureuse qui a fini deuxième. Les données de la mesure ordinale dévoilent plus d'informations que les données de la mesure nominale. Néanmoins, la valeur significative de la différence entre les données est encore inconnue puisqu'aucune égalité n'existe entre les unités. Il se peut que l'athlète qui s'est classée quatrième ait terminé à un dixième de seconde de la troisième place, alors que la gagnante peut avoir gagné avec un écart de plus d'une minute.

En outre, les comités des universités se servent d'une échelle ordinale lorsqu'il s'agit de classer les candidats éligibles à une bourse de mérite. Les notes, classées par rang-centile (terme que nous présenterons ultérieurement), sont un autre exemple de mesures ordinales. Elles sont souvent utilisées lorsque les résultats des tests sont divulgués aux étudiants, puisqu'elles fournissent aussi de l'information sur leur rang.



Tableau 23.1 Les échelles de mesure.

Mesures	Procédures statistiques établies	Exemples
<p>Nominale</p> <ul style="list-style-type: none"> les nombres sont utilisés pour nommer, identifier ou classifier 	<ul style="list-style-type: none"> les procédures statistiques s'appuient sur le calcul 	
<p>Ordinale</p> <ul style="list-style-type: none"> les nombres indiquent le rang ou l'ordre 	<ul style="list-style-type: none"> interprétations de « plus grand que » ou « plus petit que » 	
<p>Intervalle</p> <ul style="list-style-type: none"> les intervalles ou les distances entre chaque nombre sont connus mais la distance qui les sépare du point zéro est inconnue 	<ul style="list-style-type: none"> additions et soustractions les techniques statistiques fondées sur ces opérations arithmétiques sont possibles 	
<p>Proportion</p> <ul style="list-style-type: none"> chaque nombre peut être vu comme une distance mesurée à partir du point zéro 	<ul style="list-style-type: none"> il n'y a aucune limite concernant l'utilité de l'une ou l'autre des techniques arithmétiques ou statistiques 	

L'intervalle

Les données comprises dans un intervalle possèdent les mêmes propriétés que les données ordinales. Elles présentent en plus une caractéristique qui leur est propre : l'égalité des unités. Ceci signifie que nous avons des renseignements au sujet de la différence : le sens de cette différence et la valeur de la différence en unités égales. Une échelle de température en Celsius fournit une image visuelle précise de ce type d'information. Par exemple, une température de 10 degrés est plus chaude que celle de 8 degrés, de 2 degrés de plus, mais elle est plus fraîche qu'une température de 12 degrés, de 2 degrés en moins. Nous pouvons affirmer que la différence entre 8 et 12 degrés est la même qu'entre 10 et 12 degrés.

Les valeurs numériques comprises dans un intervalle n'ont pas un véritable point zéro. Par conséquent, de telles données peuvent être seulement additionnées ou soustraites mais pas divisées. La division présuppose toujours l'existence d'un point zéro exact.

La proportion

Les données dites proportionnelles sont les plus complexes de toutes les échelles de mesure. La proportion présente toutes les propriétés de l'intervalle mais se réfère à un point absolu égal à zéro ; une valeur numérique qui représente l'absence d'une mesure. Le poids, la distance, le temps et la force constituent des exemples significatifs de variables comportant un point zéro absolu.

Le point zéro absolu permet des formulations comme : une mesure de 10 mètres est d'un cinquième moins longue que celle de 50 mètres, un poids de 15 kilogrammes est trois fois plus lourd que celui de 5 kilogrammes ou une durée de 30 secondes est trois fois plus longue qu'une durée de 10 secondes. La mesure de proportion permet, à elle seule, de telles comparaisons précises.

Types de statistiques

La statistique, de nature, peut être classée en

Le zéro absolu

La température ne fait pas partie de la catégorie de la proportion parce qu'à zéro degré Celsius, la température existe toujours. Si vous doutez de ce propos, essayez d'aller dehors en T-shirt ou en culottes courtes alors que le thermomètre indique zéro degré.

Nous ne pouvons pas affirmer, par exemple, qu'une température de 20 degrés est deux fois plus chaude que celle de 10 degrés (bien qu'il fasse plus chaud de 10 degrés), parce que le point zéro sur l'échelle Celsius n'est pas un zéro absolu (le point zéro absolu de la température est 273 degrés Celsius sous zéro, soit -273).

paramétrique ou non paramétrique. Cette dernière peut être également soit descriptive, soit déductive, selon le but à atteindre. Ces concepts sont décrits dans cette section.

Les statistiques paramétriques

Les **statistiques paramétriques** s'avèrent utiles lorsque les données impliquées sont la mesure d'un intervalle et d'une proportion. Ces données sont considérées continues parce qu'elles peuvent être mesurées avec une plus grande précision, selon l'exactitude de l'instrument de mesure. Par exemple, le temps peut être mesuré en millisecondes, en secondes, en minutes, en heures, en semaines, en décennies, etc.

Les statistiques non paramétriques

Les **statistiques non paramétriques** sont utilisées lorsque les données sont de nature nominale ou ordinale. Ces dernières sont considérées discrètes parce que les observations sont limitées à certains nombres, généralement des valeurs entières. Par exemple, lorsqu'on compte les individus participant à un sondage, seules des valeurs entières doivent être utilisées parce qu'on ne peut pas fractionner une personne. La même situation se répète dans le cas du nombre de buts de champ effectués pendant une partie de football,

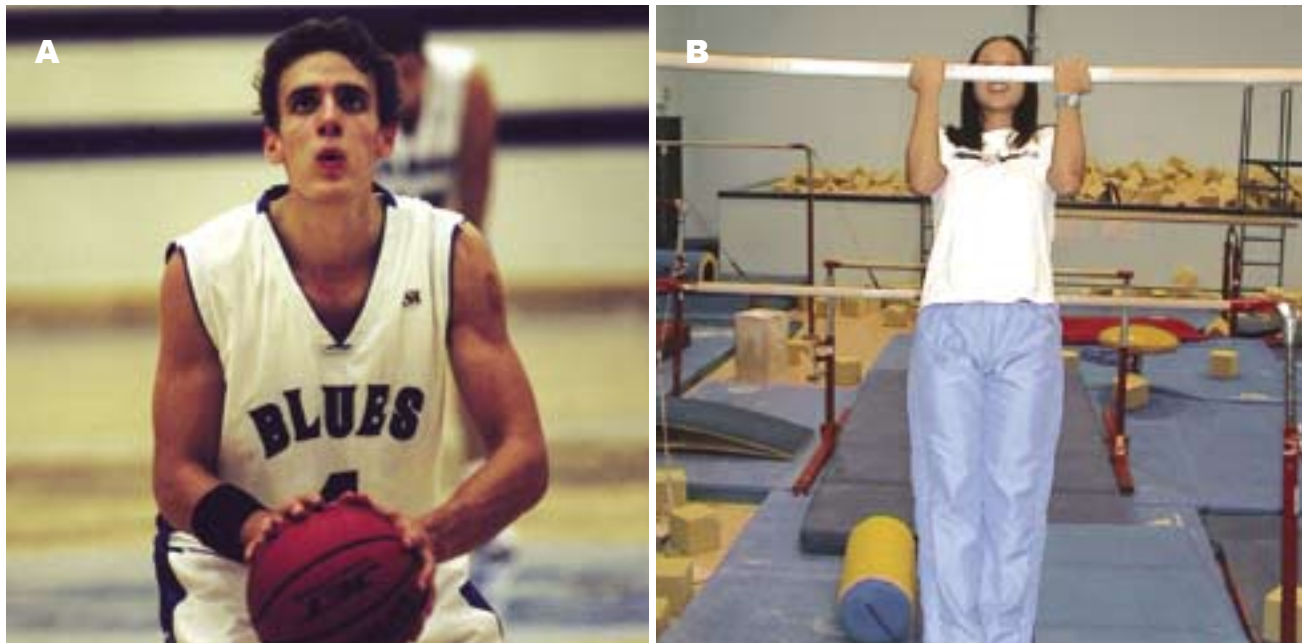


Figure 23.1 A. Le pourcentage des tirs du lancer franc est considéré une statistique descriptive. B. Les résultats des tests de la forme physique d'un petit échantillon sont des exemples de statistiques déductives.

ils ne peuvent être comptés comme trois et demi puisqu'un demi-but de champ n'existe pas. Une variété de statistiques appartiennent à ces deux catégories ; certaines seront présentées ci-dessous.

Les statistiques descriptives

Les **statistiques descriptives** impliquent des mesures qui servent à décrire des événements récents ou certaines conditions sans généraliser les statistiques à d'autres conditions ou groupes. Par exemple, votre pourcentage de lancers francs réussis au basket-ball ou votre note académique cumulative pourrait être considéré comme une statistique descriptive (figure 23.1A).

Les statistiques déductives

Les **statistiques déductives** servent à faire des inférences ou des généralisations d'un petit groupe à un groupe plus grand. Dans la mesure où l'on doit évaluer le niveau de la forme physique des élèves d'une école entière par rapport aux normes nationales, un seul échantillon représentatif de toute la population étudiante est nécessaire pour faire une estimation

de la forme physique relative des étudiants (figure 23.1B). Étant donné que les plus petits groupes sont plus économiques pour la recherche, les statistiques déductives sont alors inestimables.

Plan de recherche et test d'hypothèse

Le principal objectif de la recherche est de répondre aux questions et de résoudre des problèmes. Le **plan expérimental** est le processus de recherche qui implique la manipulation d'événements ou de variables servant à résoudre des problèmes. Dans une expérience, au moins deux groupes sont normalement impliqués : l'un est sujet à de la manipulation de la part du chercheur (**groupe expérimental**) et l'autre ne subit aucune manipulation (**groupe contrôle**).

Dans une étude menée pour déterminer si certaines boissons énergétiques (par exemple, Gatorade) contribuent à l'endurance des athlètes, le groupe expérimental consomme la boisson pendant la période d'entraînement alors que le groupe contrôle boit uniquement de l'eau. Dans cette étude, le chercheur tente de varier les boissons et observe

Tableau 23.2 Exemples de variables indépendantes et dépendantes utilisées lors d'expériences en éducation physique.

Variables indépendantes

- le temps d'entraînement par séance
- le nombre de séances d'entraînement par semaine
- le nombre de répétitions par série d'exercices
- le nombre de séries d'exercices par séance
- l'intensité de l'exercice par unité de temps
- le type de régime
- les méthodes d'entraînement utilisées
- le type de méthode d'entraînement
- l'âge, le sexe ou tout autre variable démographique
- les techniques de motivation
- les médicaments ou les drogues

Variables dépendantes

- la distance atteinte lors d'un lancer ou d'un saut
- le temps de réaction
- le temps mis à parcourir une distance
- le poids levé
- le résultat de la confiance en soi et de la conception de soi
- l'amélioration de la technique du mouvement
- la quantité d'erreurs commises
- l'exactitude de l'accomplissement
- le record académique
- le pourcentage de l'amélioration
- le comportement agressif

les effets de cette variation selon la réaction des participants. Plusieurs recherches sont fondées sur ce genre de plan et nous nous limiterons à cet aspect dans ce chapitre.

Les variables indépendantes et dépendantes

Une étude basée sur la méthode scientifique implique au moins deux variables. D'une part, la **variable indépendante** est contrôlée par le chercheur afin de déterminer l'effet sur certains comportements, d'autre part, la **variable dépendante** est le comportement observé qui est alors mesuré pour déterminer s'il est affecté par la variable indépendante.

Un autre exemple peut être utile pour illustrer davantage cette distinction. Dans une recherche qui observe l'effet de l'entraînement sur le niveau de l'habileté accomplie en gymnastique, le nombre d'heures de pratique (par exemple, deux heures, une heure, une heure et demie) serait la variable indépendante (contrôlée par le chercheur) et le niveau d'habileté accompli représenterait la variable dépendante (le comportement), laquelle est partiellement dépendante du nombre d'heures d'entraînement. Le chercheur tente de découvrir l'effet du nombre de séances d'entraînement sur l'habileté d'une personne à améliorer son niveau d'aptitudes, parmi des individus de même compétence.

D'autres variables indépendantes et dépendantes en éducation physique et dans les sports sont présentées dans le tableau 23.2.

Le test d'hypothèse

Lors de notre présentation sur la rédaction d'un projet de recherche (voir chapitre 22), nous avons introduit l'hypothèse de recherche. Celle-ci prédit les relations entre les variables, fondées sur les recherches précédentes. Dans le plan expérimental, une hypothèse est habituellement formulée dans le sens où elle permet d'analyser, d'un point de vue statistique, l'information pour déterminer si l'hypothèse devrait être acceptée. Néanmoins, le fondement ultime à de tels tests est l'**hypothèse nulle** qui prédit qu'aucune

relation ou qu'aucune différence n'existera entre les groupes expérimental et contrôle impliqués dans l'étude. En d'autres mots, une hypothèse remet en question l'hypothèse de recherche.

Pour l'étude de la relation entre la violence dans les sports à la télévision et les agressions chez les jeunes, l'hypothèse nulle pourrait être : « Il n'y a pas de relation significative entre la violence dans les sports à la télévision et l'agression chez les jeunes. » Si l'hypothèse nulle est statistiquement vraie, cela signifie que l'hypothèse de recherche originale est fautive (les variables ne sont pas significativement reliées). De la même façon, une hypothèse nulle rejetée indique que l'hypothèse de recherche initiale est correcte (les variables sont significativement reliées). En d'autres termes, s'il y a de minimes probabilités que la forme nulle soit vraie, alors nous la rejetons en faveur de l'hypothèse de recherche. Si l'inverse est vrai, l'hypothèse nulle doit être acceptée et une nouvelle recherche expérimentale doit être proposée. Une hypothèse nulle rejetée suggère qu'une différence ou une relation, perçue comme acceptable, existe parmi les variables étudiées.

Les procédures d'échantillonnage

Lorsque les chercheurs s'interrogent, ils orientent

leurs recherches de sorte à estimer des applications à un groupe plus étendu (la population) plutôt que de se limiter aux participants impliqués dans la recherche. Pour effectuer ce genre de procédé, les chercheurs mesurent seulement une portion ou une fraction de la population en observation, appelée **échantillon**. Une **population** se caractérise par des individus, des lieux ou des objets qui partagent au moins une caractéristique commune (tous les garçons âgés de 10 ans dans une école déterminée, par exemple). Un échantillon est un sous-ensemble contenant les caractéristiques essentielles de cette population. Les données recueillies auprès d'un échantillon peuvent alors être généralisées pour estimer les caractéristiques de la population ciblée.

Les populations peuvent être grandes et il s'avère peu réaliste et financièrement impossible de mesurer toutes les variables de chaque membre. L'échantillonnage est donc essentiel. Quoi qu'il en soit, mesurer 2 000 individus, au lieu de 100, d'une population de 20 000 (nombre de spectateurs au court central au Grand Chelem des États-Unis, figure 23.2) refléterait une représentation plus exacte de cette population. Plus grand est l'échantillon, plus la représentation de la population sera exacte et cela diminuera l'erreur de l'estimation. Par conséquent, les chercheurs doivent tenter de maximiser l'étendue de l'échantillon. Cela se fait



Figure 23.2 Comment étudieriez-vous le revenu des spectateurs qui assistent au tournoi de tennis du Grand Chelem des États-Unis ?