Statistiques Appliquées à l'Expérimentation en Sciences Humaines

Christophe Lalanne, Sébastien Georges, Christophe Pallier

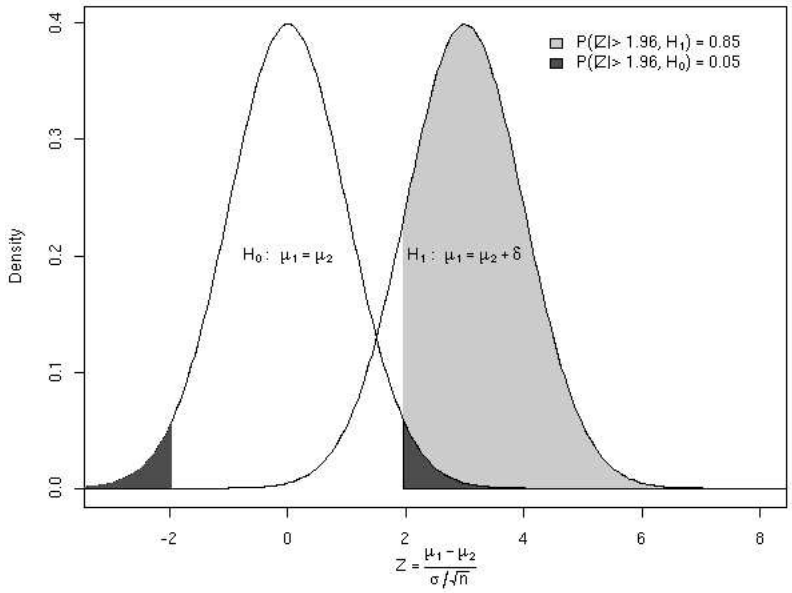

Table des matières

1 Méthodologie expérimentale et recueil des données

1.1 Introduction

L'objet de ce document est de fournir les bases théoriques du traitement statistique des données recueillies lors d'expérimentations en laboratoire sur des sujets humains. Les bases théoriques exposées dans ce document sont illustrées par des études de cas pratiques, afin de fournir un support de réflexion et de travail sur les analyses et interprétations que l'on peut élaborer à partir d'un jeu de données.

Pourquoi le titre *Statistiques Appliquées à l'Expérimentation en Sciences Humaines* ? En fait, la statistique recouvre un vaste domaine d'applications potentielles: psychométrie, agronomie, actuariat, épidémiologie, fiabilité et contrôle, etc. Chacun de ces domaines possède ses propres méthodes d'investigation et surtout d'analyse, et, si les principes de base restent les mêmes, les techniques utilisées varient beaucoup d'un domaine à l'autre. Nous avons donc choisi de nous limiter aux applications en sciences humaines, et au traitement des variables de type numérique mesurées dans le cadre d'un *protocole expérimental*, et en ce sens, ce cours s'apparente beaucoup plus à un cours de biostatistique qu'à un cours complet d'analyse des données (pour de plus amples références sur ce domaine, voir par exemple [5], [4]). Par ailleurs, ce document est structuré dans une optique *applicative*, car nous n'exposons pas les principes de la statistique mathématique (calcul des probabilités, variables et vecteurs aléatoires, lois de distribution et convergence, etc.), qui constitue une discipline en soi ; il est tout à fait possible de comprendre les principes de l'analyse statistique des données et ses applications pratiques sans avoir au préalable suivi un cours approfondi de statistique mathématique. C'est aussi un parti-pris des auteurs que de postuler qu'en commençant par une science appliquée on arrive plus facilement aux théories qui la fondent. Les seules connaissances mathématiques vraiment utiles dans ce document sont les règles de calcul algébrique élémentaire, et quelques notions de probabilités élémentaires.

Ce document ne prétend pas couvrir toutes les analyses possibles dans le domaine de l'expérimentation en sciences humaines, mais expose les bases théoriques des principales techniques d'analyse statistique: analyse descriptive, comparaison de moyennes, analyse de variance, régression linéaire. Des études de cas permettent, dans la plupart des cas, d'illustrer les notions présentées. Néanmoins, le document est organisé de telle sorte que le lecteur trouvera dans chaque chapitre des indications bibliographiques pour approfondir certaines des notions déjà traîtées, ainsi que les nombreuses autres qui ne sont pas couvertes (analyse de variance hiérarchique, régression non-linéaire, tests non-paramétriques, etc.). De même, ce document ne couvre pas les techniques propres à l'analyse des données, c'est-à-dire les méthodes factorielles (A.C.P., A.F.C., A.C.M., Analyse Discriminante, Classification, etc.), qui, de part leur richesse, mériteraient de

figurer dans un grand chapitre à part, voire un autre document.

Ce document présente les techniques de l'analyse descriptive (chapitre 2), qui constitue une étape préalable incontournable avant de poursuivre sur les étapes de l'analyse à visée inférentielle, ainsi que les principes généraux (chapitre 3) et les procédures spécifiques de l'analyse inférentielle (chapitre 4). Ce dernier chapitre est structuré en différentes parties, correspondant aux différents cas de figure que l'on rencontre dans les protocoles expérimentaux, puisque ce sont le type et le statut des variables, ainsi que la structure des données qui déterminent les analyses pertinentes à entreprendre. Cette partie aborde ainsi : la comparaison d'un échantillon à une population de référence, la comparaison de deux échantillons, l'analyse de variance à un ou plusieurs facteurs (ANOVA), l'analyse de l'association et de la liaison entre deux ou plusieurs variables (corrélation et régression linéaire), ainsi que les extensions de ces analyses de base que sont l'analyse de variance multiple (MANOVA) et l'analyse de covariance (ANCOVA). Pour ces deux dernières, seules les principes généraux sont présentés ; le lecteur pourra se reporter aux ouvrages cités en référence pour de plus amples développements. Chaque partie est accompagnée d'un exemple d'application, traité de manière succinte et dans lequel les procédures de test sont effectuées « à la main » afin de sensibiliser le lecteur, d'une part, aux procédures manuelles, et d'autre part, au fait que, si les logiciels le font actuellement beaucoup plus rapidement, il est toujours bon de savoir le faire soi-même puisque cela permet de mieux comprendre les étapes de la procédure et la « logique » de la démarche et des résultats obtenus ; cela peut ainsi permettre de détecter d'éventuelles erreurs ou inconsistences dans les résultats fournis par un logiciel dédié, et qui peuvent résulter, par exemple, d'options de calcul inappropriées.

1.2 Terminologie

La terminologie adoptée ici correspond globalement à celle de [3] (voir également [5]). On désigne la plupart du temps par *individus* ou *observations* les données quantitatives recueillies lors d'un protocole expérimental, par exemple un temps de réaction pour un sujet dans une condition expérimentale. On peut différencier les deux termes en considérant que les observations sont des informations d'une quelconque nature, tandis que les individus sont le support sur lesquelles celles-ci ont été recueillies. On trouve également le terme *unités statistiques* pour désigner les individus. Dans le cadre de ce document orienté sur l'analyse de données qauntitatives recueillies lors d'expérimentation, on parlera plus volontiers d'observations.

Un *échantillon* est un ensemble ou une collection d'individus ou d'observations tirés d'une population plus vaste — *la population parente* (ou de référence) — qui n'est généralement pas observée.

On désigne par *variable*, *facteur*, ou *caractère*, l'objet d'étude (invoqué artificiellement ou naturellement) qui est manipulé par l'expérimentateur.

Un *traitement*, une *condition*, le *niveau* d'un facteur, ou la *modalité* d'une variable représente les différentes « valeurs » prises par la variable d'étude. On désignera par la suite *facteur* la

variable d'étude, qui comportera des niveaux ou des modalités selon que cette variable est numérique ou qualitative[1] (cf. § suivant).

1.3 Typologie des variables

La typologie adoptée ici, illustrée dans la figure 1.1, correspond globalement à celle introduite par Stevens ([8] ; voir également [3] et [5]), bien que celle-ci ait fait l'objet de nombreuses discussions, notamment en ce qui concerne les analyses qui peuvent être menées en fonction du type de variable.

On adoptera également la définition suivante (e.g. [5]) : une variable est constituée d'un ensemble de modalités mutuellement exclusives et constituant le domaine de variation de la variable. Les modalités de la variable peuvent être des valeurs ou des niveaux (i.e. des valeurs ordonnées, dans le cas où la variable est un facteur).

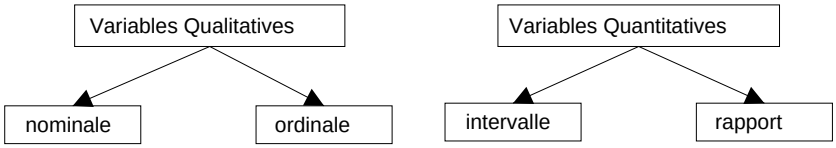

Fig. 1.1: Les deux grandes classes de variables.

1.3.1 Variables qualitatives

Appelée également variable catégorisée, une variable qualitative est une variable dont les modalités ne peuvent pas être « mesurées » sur une échelle spécifique. C'est le cas par exemple de la couleur des cheveux ou du degré d'appréciation d'un certain objet dans le cadre d'un questionnaire avec jugement de préférence. On distinguera les variables *nominales* des variables *ordinales*, qui peuvent être « ordonnées » ou recodées sur une échelle arbitraire. C'est le cas par exemple d'une variable du type « niveau d'expertise » avec les modalités, ou niveaux, « faible », « intermédiaire » et « avancé ».

On incluera dans ce type de variables les variables ouvertes, c'est-à-dire les variables dont on ne peut pas prédire la valeur ou qui ne possèdent pas de domaine de définition : c'est le cas des questions libres posées dans les questionnaires, et pour lesquelles les réponses consistent en des phrases, ou des expressions regroupées. Seront également considérées comme qualitatives les variables binaires ou dichotomisées (e.g. oui/non, présent/absent).

Ce sont dans tous les cas des variables *discontinues*.

[1] On parle en fait de niveaux lorsque les modalités du facteur sont ordonnées, et cela inclut les variables qualitatives de type ordinale.

1.3.2 Variables quantitatives

Les variables quantitatives, ou numériques, possèdent quant à elles une « métrique », c'est-à-dire qu'elles peuvent être représentées sur une échelle spécifique de mesure. On distinguera les variables d'*intervalle*, qui supportent des transformations linéaires (de type $y = ax$), des variables dites *de rapport*, supportant les transformations affines (de type $y = ax + b$). Dans ce dernier cas, il existe une origine, ou un zéro, qui a un sens. Des exemples de telles variables sont : la température en $^{\circ}$C (intervalle), la taille ou un temps de présentation (rapport), etc.

Elles peuvent être en outre *continues* (à valeurs dans l'ensemble des réels), comme par exemple lorsqu'on mesure le temps de réaction à un évènement, ou *discrètes* (à valeurs dans l'ensemble des entiers naturels), comme c'est le cas dans une procédure de comptage de réponses ou d'items.

1.3.3 Variables dépendantes et indépendantes

La variable mesurée lors de l'expérimentation se nomme variable *dépendante*. Il peut bien entendu y en avoir plusieurs. Toutefois, par extension, ce sont également toutes les variables qui seront utilisées en tant qu'observations dans l'analyse (on peut « dériver » de nouvelles variables à partir des variables initiales).

Les différentes conditions de passation de l'expérience constituent la ou les variable(s) *indépendante(s)*. On peut également raffiner la définition (cf. [5]) en distinguant les variables indépendantes *provoquées*, i.e. explicitement déterminées par l'expérimentateur (e.g. intervalle inter-essai, type de consigne, etc.), des variables indépendantes *invoquées*, qui relèvent plutôt des caractéristiques intrinsèques des individus (e.g. âge, sexe, niveau de QI, etc.).

1.4 Planification d'une expérience

La conception de plan d'expériences constitue un domaine d'étude à part entière, et nous nous limiterons à décrire brièvement les principaux concepts associés à la mise en œuvre d'une expérience, et plus particulièrement dans le domaine de la psychologie expérimentale et de la biologie (pour de plus amples références, le lecteur pourra consulter [19], [14], [2], [6]).

En toute généralité, lorsque l'on souhaite mesurer les capacités du sujet humain sur une certaine dimension, on sélectionne généralement un groupe de sujets qui peut être considéré comme représentatif de la population[2]. Ce groupe de sujets est placé dans une « situation expérimentale » particulière, permettant d'observer les performances des sujets en fonction des variations d'un ou plusieurs facteur(s) manipulé(s) par l'expérimentateur et supposé(s) influencer,

[2]Ce groupe d'individus, ou échantillon, n'est pas choisi n'importe comment, mais est « sélectionné » par une méthode d'échantillonnage aléatoire, ce qui constitue la base de *toutes* les procédures inférentielles (cf. chapitre 3).

ou moduler, leurs capacités sur la dimension étudiée. En conséquence, on essaie généralement de minimiser les autres facteurs susceptibles d'influencer les performances : la situation expérimentale est dite *contrôlée*. L'observation de variations des performances en fonction des variations du facteur manipulé (les différentes *conditions* de l'expérience), toutes choses étant égales par ailleurs, peut de la sorte être attribuée à l'effet de ce facteur.

L'allocation des sujets dans les différentes conditions expérimentales est une étape essentielle dans la démarche expérimentale, et influence les traitements statistisques qui seront entrepris lors de l'analyse des données. Les sujets peuvent être répartis de différentes manières (cf. tableau 1.1) : on pourra par exemple constituer deux groupes de sujets, sélectionnés aléatoirement dans une population de référence (par exemple, parmi l'ensemble des étudiants d'une université, de même sexe et de même niveau d'études), et étudier l'effet de deux traitements en administrant de façon exclusive l'un des deux traitements à chacun des deux groupes. Ainsi, les individus du premier groupe seront exposés exclusivement au premier traitement, et ceux de l'autre groupe exclusivement au second traitement. L'un des deux groupes pourrait également ne pas recevoir de traitement, ou servir de « référence » par rapport aux autres traitements, et il serait alors qualifié de groupe *contrôle*. Ce type d'expérience utilise en fait des *groupes indépendants* : les observations dans chacun des deux groupes sont indépendantes les unes des autres puisqu'elles proviennent de sujet différents.

cond. n° 1	cond. n° 2
S_1	S_5
S_2	S_6
S_3	S_7
S_4	S_8
...	...

cond. n° 1	cond. n° 2
S_1	S_1
S_2	S_2
S_3	S_3
S_4	S_4
...	...

	cond. n° 1	cond. n° 2	cond. n° 3
S_1	x_1^1, x_2^1, ...	x_1^2, x_2^2, ...	x_1^3, x_2^3, ...
S_2	x_1^1, x_2^1, ...	x_1^2, x_2^2, ...	x_1^3, x_2^3, ...
S_3	x_1^1, x_2^1, ...	x_1^2, x_2^2, ...	x_1^3, x_2^3, ...
S_4	x_1^1, x_2^1, ...	x_1^2, x_2^2, ...	x_1^3, x_2^3, ...
...	...	...	...

Tab. 1.1: Répartition des sujets entre deux conditions avec des protocoles de groupes indépendants (gauche) ou appariés (milieu), et des protocoles de mesures répétées (droite). Les sujets sont représentés par la lettre S indicée par le n° du sujet. Pour les mesures répétées, la variable mesurée est indiquée par la lettre x indicée par le n° d'essai, l'exposant représentant le n° de la condition.

A l'inverse, on pourrait s'intéresser aux effets d'apprentissage d'une tâche, soit en manipulant directement les conditions d'exécution de la tâche, soit en regardant l'évolution des performances à deux instants temporels donnés. Dans ce cas, on utilisera préférentiellement les mêmes sujets dans les deux conditions expérimentales, et on parle alors de *groupes appariés* : les paires d'observations dans chacun des deux groupes ne sont plus indépendantes les unes des autres puisqu'elles sont issues d'un même sujet, et sont donc liées entre elles (elles demeurent cependant indépendantes entre les sujets à l'intérieur d'un même groupe, ou entre sujets différents dans les deux groupes).

Enfin, une autre technique couramment utilisée consiste à effectuer plusieurs mesures sur le même sujet, dans chacune des conditions expérimentales : on parle dans ce cas de protocole de *mesures répétées*. La répartition des sujets et des répétitions est schématisée dans le tableau 1.1 (tableau de droite).

Bien que ces deux derniers types de protocole soient les plus intéressants, en raison de leur « puissance » intrinsèque (se référer aux ouvrages sur les plans d'expériences déjà cités) et du nombre de sujets requis, il n'est parfois pas possible de les utiliser. Par exemple, lorsque l'on étudie l'effet d'une consigne dans la réussite à un test, il semble difficile de réutiliser les mêmes sujets puisque ceux-ci seraient « biaisés » par leur première expérience avec le matériel expérimental. Comme on le voit, le choix d'un protocole expérimental dépend dans une grande mesure du type de tâche que l'on souhaite administrer lors de l'expérience.

D'autre part, l'expérimentateur s'intéresse souvent à l'étude conjointe de plusieurs facteurs, d'où la nécessité d'adapter ces schémas élémentaires au cas de plusieurs variables, comme nous allons le voir dans le paragraphe suivant.

1.5 Formalisation des plans d'expériences

Dans la conception et l'analyse de plan d'expériences en psychologie expérimentale, on utilise quelque fois un formalisme particulier introduit par [17]. Les variables sont reprises sous une lettre unique, la lettre S étant réservée au facteur sujet. Le nombre de modalités ou de niveaux est indiqué en indice de la lettre désignant le facteur, et on décrit la structuration du protocole expérimental à l'aide de relations k-aires de *croisement* et d'*emboîtement*. Les relations peuvent être de type binaire (lorsqu'il n'y a que 2 variables mises en relation), ternaire, etc. On dit que les variables sont croisées entre elles lorsque toutes les modalités (ou niveaux) des variables sont présentées à chaque sujet, tandis qu'il y a une relation d'emboîtement lorsqu'on introduit un facteur de groupe, c'est-à-dire que seule une partie des sujets est soumis à un sous-ensemble des modalités de la variable. On essaie le plus souvent d'avoir des plans dits *équilibrés*, c'est-à-dire dans lesquels il y a autant d'observations, ou d'individus, dans chacun des groupes introduits par le facteur d'emboîtement[3].

Sans entrer plus loin dans les détails de ce formalisme, on parle de *plans quasi-complets* ([17]) lorsque (i) tous les facteurs croisés deux à deux sont croisés ou emboîtés, et que (ii) les facteurs croisés deux à deux sont croisés dans leur ensemble. Enfin, lorsque le facteur sujet est croisé avec un ou plusieurs facteurs, il s'agit d'un plan de mesures répétées.

Par exemple, un plan classique de groupes indépendants avec k conditions (i.e. on a bien k groupes d'observations indépendantes) sera modélisé par la relation d'emboîtement : $S_{n/k} < G_k >$, le facteur S étant le facteur sujet, et le facteur G un facteur de groupe, à 2 modalités (G est appelé le facteur *emboîtant*, S le facteur *emboîté*). Il y a $\frac{n}{k}$ sujets dans chaque groupe, n désignant le nombre total de sujets participant à l'expérience. A l'inverse, un protocole dans lequel tous les sujets sont confrontés à toutes les modalités i de la ou les variable(s) indépendante(s) implique une relation de croisement : $S_n * T_i$; c'est le cas dans les protocoles utilisant des groupes appariés ou des mesures répétées. Dans le cas précis de mesures répétées avec plusieurs essais par traitement ou condition, on indique généralement le nombre d'essais à l'aide d'un facteur

[3] C'est une condition suffisante, mais pas nécessaire.

supplémentaire, par exemple le facteur « essai » E_j à j répétitions. On peut combiner ces deux relations, comme dans le cas où l'on a à la fois 2 facteurs de groupement dont les différentes modalités des deux variables sont présentées exclusivement à une partie des sujets. La structure des données devient alors $S_{n/(k\times k')} < G_k * G'_{k'} >$. Enfin, on peut également avoir un seul facteur de groupe, et une autre variable dont les modalités sont présentées à l'ensemble des sujets de chaque groupe. Le plan devient $S_{n/k} < G_k > * T_i$ (cf. tableau 1.2).

	A_2	
B_3	s1, s2, s3	s4, s5, s6
	s7, s8, s9	s10, s11, s12
	s13, s14, s15	s16, s17, s18

	A_2	
B_3	s1, s2, s3	s1, s2, s3
	s4, s5, s6	s4, s5, s6
	s7, s8, s9	s7, s8, s9

Tab. 1.2: Exemples d'allocation des sujets dans des plans d'expérience avec deux facteurs A et B emboîtants (gauche) — $S_3 < A_2 * B_3 >$ —, et un facteur d'emboîtement B associé avec un facteur de croisement A (droite) — $S_6 < B_3 > * A_2$ —. Les facteurs A et B ont 2 et 3 modalités (ou niveaux), respectivement.

L'intérêt de cette planification est de permettre d'une part d'optimiser le nombre de sujets nécessaires pour l'étude visée, et, d'autre part, de s'assurer que le partitionnement des *sources de variation*, i.e. l'effet des différents facteurs expérimentaux, pourra être correctement analysé. En particulier, la « dérivation » du plan d'expérience permettra de déterminer les sources de variation de la variable dépendante étudiée. Par exemple, dans une expérience où l'on présente à des participants des deux sexes différentes séries d'images de contenu émotionnel variable (3 modalités : neutre, aversif, agréable), on souhaite étudier l'effet du contenu émotionnel des images sur les participants en fonction de leur sexe. Le plan d'expérience se formalisera sous la forme $S < A > * B$, A désignant le facteur sexe et B le facteur type d'image. Les *facteurs élémentaires* seront ainsi A, B, $S(A)$ (s'il y a une relation d'emboîtement, le facteur élémentaire est accompagné du ou des facteurs emboîtant(s)). Les termes d'interaction d'ordre 1 seront dérivés à partir des relations de croisement, soit ici AB, $BS(A)$; les termes d'interaction d'ordre 2 seront dérivés du croisement des facteurs élémentaires et des termes d'interaction d'ordre 1 (ici il n'y en a pas), et ainsi de suite pour les interactions d'ordre supérieur. Ainsi, dans l'exemple proposé, on retrouve 5 sources de variation potentielles.

De nombreux plans expérimentaux sont possibles, et chacun possède ses spécificités techniques, ainsi que ses avantages et ses inconvénients sur le plan de l'analyse. On peut citer l'exemple classique du plan en *carré latin* qui permet de ventiler plusieurs facteurs avec un nombre limité de sujets. Dans ce type de plan, on peut croiser trois facteurs deux à deux, sans jamais les croiser dans leur ensemble, comme l'illustre le tableau 1.3 : chacun des facteurs a 3 modalités ; on croise 2 des facteurs ensemble, puis on ventile les modalités du dernier facteur par permutation circulaire pour chacun des croisements des deux premiers facteurs.

Un dernier exemple de plan, qui figure parmi les plus intéressants dans le domaine de l'expérimentation, et qui sera développé dans la partie relative à l'analyse de variance (cf. 4.3, p. 70), est le plan *factoriel*. Ce plan, comme le précédent, utilise un croisement entre plusieurs facteurs,

	B_3		
A_3	a1b1c1	a1b2c2	a1b3c3
	a2b1c2	a2b2c3	a2b3c1
	a3b1c3	a3b2c1	a3b3c2

Tab. 1.3: Exemple de plan en carré latin.

ce qui permet de diminuer la quantité de sujets requis dans une expérience.

1.6 Quel logiciel utiliser ?

De nombreuses solutions logicielles sont proposées à l'heure actuelle sur le marché. Pour en citer quelques-unes parmi les plus connues: SAS, SPSS, STATISTICA, SYSTAT, MINITAB, SPAD, STATVIEW, R... (cf. C, p. 123). Chacun de ces logiciels possède ses spécificités et ses domaines de prédilection pour l'analyse des données : SPAD est par exemple spécialisé dans l'analyse factorielle, MINITAB dans l'analyse de variance, etc. Certains logiciels sont également dérivés des autres (e.g. SYSTAT ressemble fortement à SPSS). D'autres ont des possibilités beaucoup plus étendues et constituent de véritables plateformes d'analyse statistique, comme c'est le cas pour STATISTICA ou SPSS ; de surcroît, certains logiciels sont très modulables dans la mesure où des extensions peuvent être ajoutées à la « base » du logiciel: c'est le cas de SAS ou R. Enfin, un dernier critère est la disponibilité ou le tarif de ces plateformes : s'il est possible d'acquérir certains de ces logiciels à des tarifs relativement accessibles (STATISTICA, SPAD), d'autres nécessitent un véritable investissement (SAS par exemple), et ne sont généralement pas destinés à un usage étudiant. A l'inverse, certains logiciels sont proposés en version d'évaluation entièrement fonctionnelle (cas de SPSS par exemple), ou sont disponibles gratuitement : c'est le cas de R. Cette dernière solution est sans doute la plus intéressante et la plus avantageuse. R est sans conteste l'un des logiciels les plus prometteurs à l'heure actuelle dans la mesure où son développement bénéficie d'une communauté grandissante de contributeurs issus à la fois du domaine de la recherche, mais également de la communauté des statisticiens. Par ailleurs, il s'avère très utile tant sur le plan de l'analyse, avec ses algorithmes très performants, que du point de vue de sa souplesse et de son élégance pour les sorties graphiques[4].

[4] L'ensemble des graphiques figurant dans ce document ont été réalisés sous R.

2 Analyse descriptive

2.1 Objet de l'analyse descriptive

L'objectif de l'analyse descriptive est de résumer la masse d'informations numériques accumulées dans le corpus de données en un ensemble synthétique d'indicateurs descriptifs, et d'offrir une représentation graphique des résultats qui permette de voir rapidement leurs principales caractéristiques. Les indicateurs descriptifs les plus utilisés sont :

- les indicateurs de *tendance centrale*, ou de position (cf. 2.4.1),
- les indicateurs de *dispersion*, ou de variabilité (cf. 2.4.2),
- les indicateurs portant sur la *forme de la distribution* des observations (cf. 2.4.3).

Certaines étapes de codage ou de recodage des données peuvent être entreprises au préalable afin de faciliter le traitement des données et produire les informations les plus pertinentes.

L'ensemble de ces étapes doit permettre à terme de situer un individu ou un échantillon parmi le groupe ou la population de référence dans le cas des données univariées, ou de comparer des distributions d'effectifs ou des observations dans le cas de données bivariées. Un principe de base à garder présent à l'esprit lors de l'analyse des données est de toujours procéder à l'analyse univariée avant l'analyse bivariée ou multivariée.

2.2 Codage et recodage des données

2.2.1 Recueil et codage des données

Le recueil de données consiste en la récupération des données, sous forme informatique le plus souvent, au cours ou à la suite de l'expérimentation. Selon le type de protocole (groupes indépendants ou non, mesures répétées) et le nombre de variables étudiées, la quantité de données peut être relativement importante. Pour mener à bien les analyses, il est souvent pratique de coder les variables et les individus à l'aide de symboles univoques (e.g. SEXE pour une variable qui serait le sexe du sujet, DUREE pour un facteur qui serait la durée de présentation d'un stimulus à l'écran, S1 ou I1 pour le sujet/individu n° 1, etc.) : l'idée est de trouver des noms suffisamment explicites pour que l'on sache de quelle variable il s'agit, et surtout d'éviter d'avoir des noms de variables trop longs. D'autre part, il s'avère également commode d'utiliser soit des *variables indicatrices*, c'est-à-dire des variables booléennes valant 1 lorsque la modalité est présente et 0 sinon, soit des variables prenant pour modalités des valeurs entières pour les niveaux d'un facteur, plutôt que de conserver le nom des modalités. Par exemple, si l'on a un facteur « temps de présentation » (du stimulus visuel) à 3 niveaux — 150, 250, 500 ms —, on pourra choisir de le coder comme : facteur TEMPS à 3 niveaux (1: 150, 2: 250 et 3: 500 ms).

Ce type de codage des variables et des individus sera très utile lors de la production de tableaux de résumés numériques ou de graphiques illustratifs pendant la phase initiale d'analyse des données, mais surtout lors du traitement des données dans un logiciel statistique comme R ou STATISTICA. En effet, la plupart des logiciels statistiques utilisent également des variables nommées, pour distinguer les observations issues d'une condition de celles liées à une autre condition.

Il ne faut pas oublier également le cas des données manquantes, c'est-à-dire les observations qui n'ont pu être recueillies, soit pour des raisons matérielles soit pour des contraintes expérimentales (sujet absent, problème d'enregistrement informatique, etc.). Il peut s'avérer judicieux de coder ces valeurs manquantes à l'aide de symboles spécifiques qui permettront d'éviter de les traiter comme les autres. On peut par exemple les coder par la valeur -99, ou par toute autre valeur qui ne soit pas liée à la variable dépendante. On évitera ainsi de coder par 0 les valeurs manquantes dans le cas où la variable mesurée est par exemple un nombre d'items résolus : cette valeur appartient au domaine de définition de la variable et pourrait influencer les calculs subséquents (médiane, moyenne, etc.).

2.2.2 Recodage par discrétisation et dérivation

D'autre part, les données sont souvent « recodées » pour les besoins de l'analyse. C'est le cas par exemple des réponses issues de questionnaires, ou plus généralement des variables qualitatives ordinales. Dans le cas par exemple où l'on a différentes modalités d'une variable ordinale, qui sont ordonnées comme suit : « n'aime pas », « aime peu », « aime moyennement », « aime un peu », « aime beaucoup », on peut préférer les recoder en valeurs numériques discrètes de type : 1, 2, 3, 4, 5. De même, on peut effectuer l'inverse et passer de variables de type numérique à des variables qualitatives (ordinales) en effectuant un regroupement par *classes*. C'est par exemple le cas, dans les analyses d'enquêtes et de sondages, des variables de type âge, qui sont des variables numériques de rapport, et que l'on préfère parfois traiter en classes (e.g. la classe des 20-25 ans, celle des 26-30 ans, etc.), car ce type de partition apporte plus d'informations : on n'attend, par exemple, pas de différences notables entre les 20-21 ans ; en revanche, on a de bonnes raisons de penser qu'il y a des différences entre les personnes âgées de 20 ans et celles âgées de 28 ans.

Les données peuvent également être « recodées » au cours de l'analyse, en *dérivant* les valeurs observées : par exemple, à partir de deux séries de scores obtenus lors de tests, on pourra établir une nouvelle variable numérique qui est la différence (signée) des deux scores. Ce type de recodage n'est pas limité à la dérivation par soustraction : on peut également dériver le protocole par moyennage (somme des scores divisée par le nombre de groupes), ou par différence absolue (valeur absolue de la différence des scores de chaque groupe). Enfin, on peut effectuer des dérivations par restriction, en se limitant à l'étude d'un seul facteur lorsqu'il y en a plusieurs.

2.2.3 Recodage par transformation

Enfin, il est souvent utile de normaliser (on dit également « standardiser ») les données, à l'aide d'une transformation centrée-réduite[1]. Celle-ci consiste simplement à centrer les données x_i par rapport à leur moyenne $\bar{x}$, et à les réduire par rapport à l'écart-type σ_x :

$$x_i \rightarrow z_i = \frac{x_i - \bar{x}}{\sigma_x}$$

Les scores ainsi obtenus s'expriment alors en points d'écart-type par rapport à la moyenne, i.e. on dira d'un sujet ayant un score $z = +1,61$ qu'il est situé à 1,61 points d'écart-type au-dessus de la moyenne. Ceci permet bien évidemment de comparer des sujets issus de différents groupes hétérogènes n'ayant pas forcément les mêmes moyennes de groupe.

D'autres transformations sont également fréquemment utilisées : la transformation en logarithme ($x \rightarrow x' = log(x)$) ou en racine carrée ($x \rightarrow x' = \sqrt{x}$). Ce type de transformation a généralement pour but d'atténuer les variations de la variance en fonction des valeurs de la moyenne. Comme on le verra dans la partie relative à l'analyse inférentielle (cf. chapitres 3 et 4), les postulats de base des procédures de test concernent la distribution normale des observations (ou des résidus), ainsi que l'homogénéité des variances (homoscedasticité) entre les différents niveaux d'un facteur.

2.3 Données numériques et représentations graphiques

2.3.1 Données numériques et type de variable

Selon le type de variable considérée, on n'utilisera pas forcément les mêmes « formats » numériques pour la présentation des données. Par exemple, lorsque l'on travaille sur des variables qualitatives, il est fréquent d'utiliser une distribution d'effectifs, qui pourra s'exprimer soit en unités d'individus, en fréquence ou en pourcentage. Pour les variables quantitatives, on pourra préférer représenter les valeurs observées (i.e. distribution des observations numériques).

En guise d'illustration, on peut considérer les deux exemples suivants :

1. Dans une enquête effectuée au sein d'une promotion d'étudiants, on a relevé les réponses à un questionnaire à choix multiples comportant 10 questions portant sur leurs études antérieures, la situation professionnelle de leurs parents, leur intérêt général dans les discplines enseignées, etc.
2. Dans une expérience de perception visuelle portant sur 20 sujets, on a relevé leurs temps de réaction à l'apparition d'un stimulus à l'écran selon différentes conditions de présentation de la cible.

[1] On trouve également l'appellation scores Z dans la littérature anglo-saxonne, du nom de la loi de distribution centrée-réduite qui est une loi normale de moyenne nulle et d'écart-type 1 (cf. 3.2.1).

Dans le premier cas, les questions du questionnaire sont assimilées à des variables qualitatives (il s'agit d'un Q.C.M., on ne demande pas de réponse numérique), et les réponses proposées constituent les différentes modalités de la variable considére. Dans ce cas, on pourra représenter la fréquence relative des réponses en fonction des modalités de la variable : il s'agit alors d'une représentation des effectifs ou individus.

Dans le second cas, ce sont a priori les valeurs que prend la variable dépendante en fonction des différents niveaux du facteur « présentation » qui sont les plus intéressantes, et celle-ci est une variable numérique de rapport, continue. On aura donc une distribution des observations (i.e. des valeurs observées), qui pourra être résumée à l'aide des moyennes et écarts-type par condition.

2.3.2 Représentations numériques

Variable quantitative discrète

L'exemple classique de variable quantitative discrète est l'âge lorsque celui-ci est arrondi à l'année. Si l'on dispose d'un ensemble d'observations de ce type, on les organise généralement sous la forme d'un tableau dans lequel chaque ligne i correspond à une valeur observée, et on utilise différentes représentations des données en colonnes[2] :

- les *effectifs* n_i (i.e. le nombre de fois où la valeur x_i a été observée) ;
- les *effectifs cumulés* N_i (on rajoute à chaque ligne les effectifs précédents) ;
- les *fréquences* f_i (i.e. les effectifs rapportés à l'effectif total) ;
- les *fréquences cumulées* F_i.

L'usage des fréquences peut s'avérer utile lorsque l'effectif total est important.

Variable quantitative continue

Un exemple typique de données numériques continues est un temps de réaction exprimé en millisecondes. La présentation des données suit le même principe que dans le cas précédent, mais l'on peut utiliser des *classes* pour regrouper les valeurs. Une classe sera définie par son *amplitude* délimitée par deux bornes (inférieure et supérieure), et son *centre* (milieu de l'intervalle défini par les bornes). Lorsque l'amplitude des classes n'est pas systématiquement la même, il est préférable d'utiliser la *densité* plutôt que les effectifs. La densité est simplement l'effectif de la classe rapporté à l'amplitude de celle-ci. De la sorte, les effectifs par classes sont « normalisés » et comparables entre eux.

[2] On peut bien entendu intervertir l'ordre de présentation des données lignes ↔ colonnes

2.3.3 Représentations graphiques

Une ou deux variables numériques

Lorsque l'on souhaite représenter les effectifs observés sur une seule variable numérique, on utilisera les *diagrammes en bâtonnets*, pour une variable à valeurs discrètes (cf. figure 2.1, en bas à gauche), ou les *histogrammes*, pour les variables continues dont les valeurs ont été regroupées en classes (cf. figure 2.1, en haut à gauche). Dans ce dernier cas, on prendra garde au fait que si l'intervalle (ou la largeur) de classe n'est pas constant, il est plus judicieux d'utiliser la densité que l'effectif en ordonnées. Cependant, la qualité de l'estimation d'une distribution au moyen d'un histogramme est fortement dépendante du découpage en classe (amplitude et nombre de classes) : si l'on utilise des classes d'amplitude trop importante, on ne verra pas certains détails caractéristiques de la distribution des effectifs ou observations, alors que si l'on utilise une partition trop fine on risque de voir apparaître trop de distorsions. Il y a donc un compromis à trouver, et d'autres solutions ont été proposées, comme l'utilisation des *méthodes d'estimation fonctionnelle* qui utilisent un paramètre de lissage avec des fonctions construites point par point (noyaux) ou des bases de fonction splines.

Les graphes les plus souvent associés à la relation entre deux variables numériques sont les *graphes bivariés*, les points figurant dans le graphique ayant pour coordonnées les valeurs respectives des deux variables.

Remarque. Lorsque le nombre de variables quantitatives devient plus élevé (e.g. 5-10), des représentations spécifiques liées aux méthodes d'analyse factorielle sont couramment employées. Il s'agit ici, dans le cas de variables numériques, du *graphe des corrélations* et des *graphes factoriels* de l'Analyse en Composantes Principales (A.C.P.).

Une ou deux variables qualitatives

Dans le cas d'une ou de deux variable(s) qualitative(s), on pourra utiliser un ou plusieurs diagrammes en barres, en colonnes ou circulaires (i.e. « camemberts », cf. figure 2.1, en bas à droite).

Remarque. Dans le cas de plusieurs variables qualitatives ou de deux variables qualitatives possédant de nombreuses modalités (e.g. 4), on utilise généralement les méthodes factorielles, comme l'Analyse Factorielle des Correspondances (A.F.C.), qui permettent de représenter le degré de liaison entre les deux variables au moyen d'un graphe plan possédant une métrique, comme pour le graphe factoriel de l'A.C.P. : la distance entre les points traduit généralement l'intensité de l'association (plus les individus sont éloignés, moins ils ont tendance à être associés).

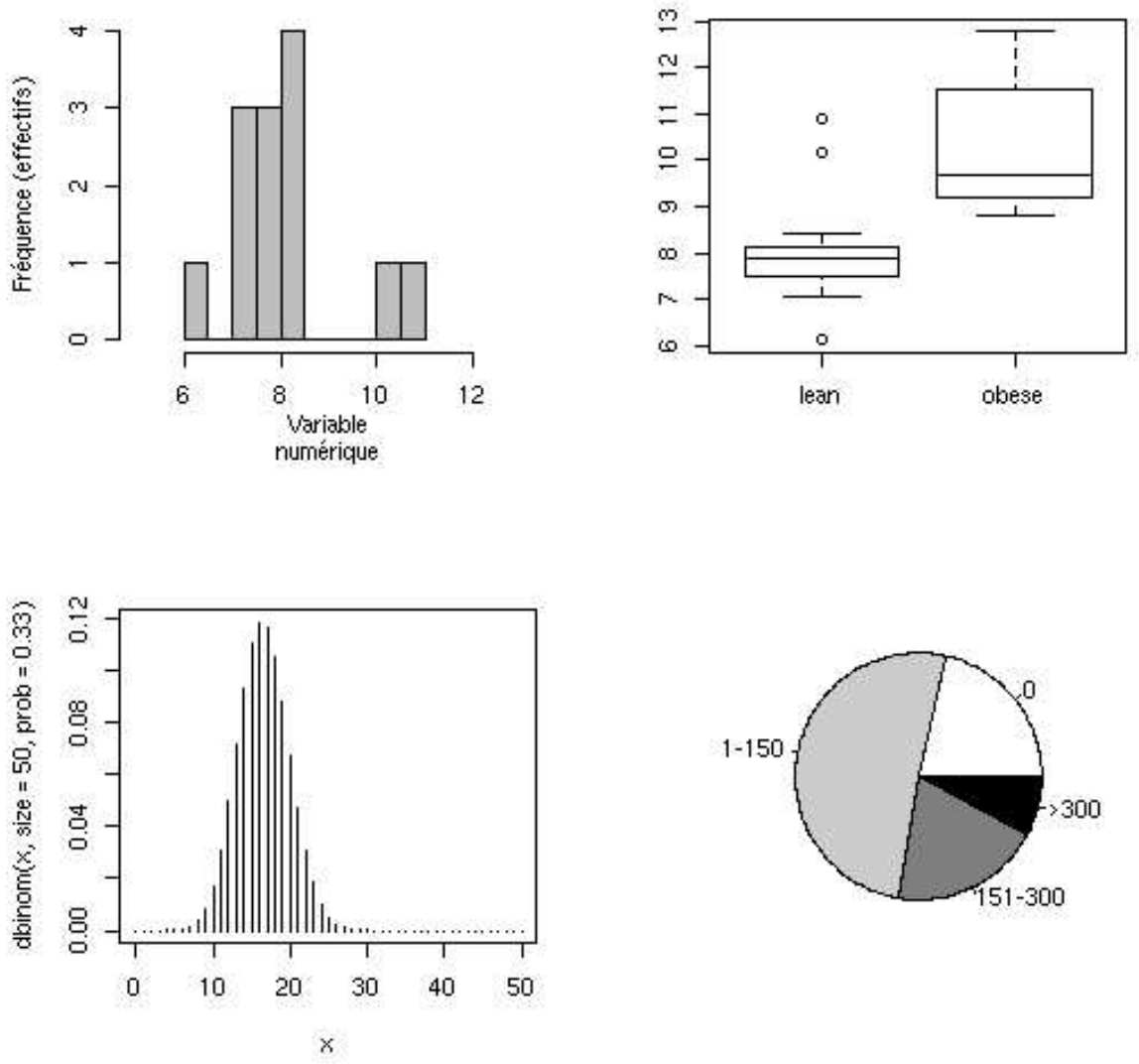

Fig. 2.1: Exemples de représentations graphiques. En haut : à gauche, distribution d'effectifs en fréquence sous forme d'histogramme ; à droite, boîte à moustaches pour les deux modalités d'une variable qualitative. En bas : à gauche, diagramme en barre de la distribution binomiale de paramètre p=0.33 ; à droite, diagramme circulaire (camembert) pour une variable qualitative ordinale à 4 modalités. (Données tirées de [24])

Une variable quantitative et une variable qualitative

Suivant le nombre de modalités de la variable qualitative, on pourra utiliser des représentations sous forme de diagrammes en barres ou circulaires. Il y en aura évidemment autant qu'il

y a de modalités. Un autre graphe très courant est la boîte à moustaches (cf. figure 2.1, en haut à droite), dans laquelle sont résumées les informations de tendance centrale (médiane et/ou moyenne) et de dispersion (intervalle inter-quartile ou écart-type, et valeurs extrêmes). Cette dernière représentation présente l'avantage de ne nécessiter qu'un seul graphe (les modalités de la variable qualitative étant représentées en abscisses), et de résumer l'ensemble des indicateurs descriptifs servant à caractériser la distribution.

2.4 Indicateurs descriptifs

Lorsqu'il s'agit de résumer l'information contenue dans les données recueillies, les principaux indicateurs numériques utilisés sont les indicateurs de tendance centrale, ou de position, et les indicateurs de dispersion, ou de variabilité. Selon la nature de la variable considérée, certains d'entre eux sont plus appropriés que d'autres, comme nous allons le voir. Ces deux indicateurs permettent d'indiquer comment se comportent les observations, en « moyenne », et comment celles-ci se distribuent par rapport à cette valeur centrale. On y adjoint souvent une étude plus fine de la distribution des observations, qui permet de préciser la répartition des observations ou effectifs aux extrémités du domaine de variation (valeurs faibles et élevées, valeurs extrêmes), ainsi que l'homogénéité de la répartition par rapport à la valeur moyenne (valeurs centrales).

2.4.1 Tendance centrale

Selon le type de données recueillies, différents indicateurs de tendance centrale peuvent être utilisés. Il est clair qu'utiliser une moyenne arithmétique sur des variables qualitatives n'aurait aucun sens (e.g. une variable de type couleur des yeux).

Mode

Dans le cas de variables qualitatives nominales, par exemple la couleur des cheveux avec 3 modalités — blond, brun, roux —, le mode est l'indice le plus utile : il indique la modalité la plus observée, i.e. la valeur ou la classe (on parle alors de classe modale) pour laquelle le plus grand nombre d'observations a été observé. Il est utilisable également dans le cas de variables numériques.

Médiane et quantilage

Utile à la fois pour les données qualitatives ordonnées (i.e. variables ordinales) et pour les données quantitatives discrètes, la médiane est la valeur qui partage l'effectif en deux sous-effectifs égaux, i.e. 50 % des observations se trouvent de chaque côté de la valeur médiane. L'avantage

de la médiane est qu'elle est relativement peu sensible aux valeurs extrêmes, et demeure donc l'indicateur de position le plus fidèle de la distribution des observations. Le mode de calcul approché de la médiane se fait de manière assez rapide en cherchant le rang médian ($n/2$ dans le cas d'un nombre pair d'observations, $n/2 + 1/2$ dans le cas d'un nombre impair d'observations) au niveau des effectifs cumulés. Dans le cas des données continues, par interpolation linéaire, on peut également calculer une valeur plus précise de la médiane, à l'aide de la formule suivante :

$$x_{med} = x^{'}_{med} + \left[\frac{(\frac{n}{2} - N_{med-1})}{n_{med}} \times h \right]$$

avec $x^{'}_{med}$ désignant la limite inférieure de la classe médiane, N_{med-1} l'effectif cumulé de la classe précédant la classe médiane, n_{med} l'effectif de la classe médiane, et h l'intervalle de classe.

Moyennes

Dans le cas de variables numériques continues, on utilisera préférentiellement la moyenne, qui indique le centre de gravité de l'ensemble des valeurs observées pour la variable. Contrairement aux deux indicateurs de position précédents, il s'agit bien là d'une valeur de quantification, exprimée dans la même unité de mesure que les variables observées.

Dans le cas d'une moyenne équipondérée (i.e. tous les individus ont le même poids), la moyenne arithmétique est simplement la somme des valeurs observées divisée par l'effectif total n :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Lorsque des poids p_i (avec $p_i = n_i/n > 0$) ont été attribués aux individus, il suffit juste de pondérer la moyenne :

$$\bar{x} = \sum_{i=1}^{n} p_i x_i$$

avec, en général, $\sum_i p_i = 1$.

Contrairement à la médiane, la moyenne arithmétique est influencée par les valeurs extrêmes, notamment dans le cas de petits échantillons et lorsque les données sont équipondérées ou que les valeurs extrêmes possèdent des poids plus importants que les valeurs centrales.

On notera que d'autres types de moyenne peuvent être calculés, comme la moyenne géométrique (pour des données sur des rendements, ou pourcentages) ou la moyenne harmonique (pour des données de variations temporelles, de type vitesse), mais la moyenne arithmétique demeure l'indicateur le plus adéquat pour la majorité des données numériques issues d'un protocole expérimental.

2.4.2 Dispersion

Étendue

L'étendue est simplement l'écart séparant la plus petite valeur observée de la la plus grande, i.e. $x_{max} - x_{min}$. Cette valeur est utile pour caractériser la variable mesurée, mais demeure naturellement sensible aux valeurs extrêmes, et ne reflète par conséquent pas la dispersion moyenne des valeurs observées autour d'une valeur centrale.

Intervalle inter-quantiles

Dans le cas des variables qualitatives, la moyenne n'a aucun sens, pas plus que ses indicateurs de dispersion associés — variance, écart-type — (cf. infra). Pour associer une mesure de dispersion à la valeur centrale indiquée par la médiane, on utilise l'intervalle inter-quantile (IQ) qui représente l'intervalle incluant la médiane et dans lequel se situent 50 % des observations. Lorsqu'on utilise des quartiles, c'est simplement la différence entre le troisième et le deuxième quartile, i.e. $IQ = Q_3 - Q_2$ (pour des centiles, on aurait $IQ = C_{75} - C_{25}$). Rappellons que la valeur du premier quartile est telle que 25 % des observations sont situées avant elle, et de manière générale, on a la relation : $P(x < x_0) = q$, x_0 désignant la valeur du fractile recherché et q la proportion correspondante (dans le cas précédent, $q = 0.25$ et x_0 est la valeur de Q_1). L'examen de la position de la médiane par rapport aux bornes de l'IQ (Q_1 et Q_3, par exemple) permet d'identifier les éventuelles asymétries dans la distribution des effectifs (cf. figure 2.3).

Écarts à la moyenne

Il est possible de situer les observations par rapport à leur valeur moyenne à l'aide des écarts signés $(x_i - \bar{x})$ et absolus $|x_i - \bar{x}|$. Les premiers permettent d'indiquer en outre la position de la variable observée par rapport à la moyenne (cf. figure 2.2), tandis que les seconds n'indiquent que l'ampleur de l'écart à la moyenne. Mais on peut utiliser ces derniers pour évaluer la « distance moyenne à la moyenne », c'est-à-dire l'écart absolu moyen (EAM) :

$$EAM = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|$$

Variance et écart-type

Si la moyenne reflète bien la valeur centrale d'une distribution (sous réserve qu'il n'y ait pas trop de valeurs extrêmes, cf. supra), elle ne renseigne pas sur la distribution des observations autour de cette valeur centrale. Un exemple classique est le cas des salaires : soit 10 individus

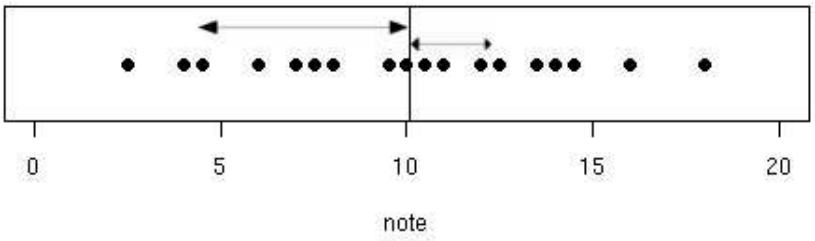

Fig. 2.2: Représentation schématique des écarts à la moyenne pour un ensemble d'observations x_i. La moyenne $\bar{x}$ est figurée par le trait vertical.

d'une société, dont les salaires respectifs (en euros) sont 3 100, 2 500, 2 800, 3 200, 4 000, 2 500, 3 000, 2 700, 3 000, 2 900, et 10 autres individus d'une autre société, dont les salaires sont 1 800, 2 000, 1 900, 4 500, 6 000, 5 000, 1 600, 2 400, 2 500, 2 000 (tiré de [4], p. 12). Si le salaire moyen de ces deux groupes d'individus est identique (2 970 euros), il est clair que les salaires dans la deuxième société sont beaucoup plus « variables », ou « dispersés » autour de leur valeur moyenne, que ceux de la première société.

Variance. La variance mesure la dispersion des valeurs observées autour de la moyenne. Plus précisemment, la variance est la moyenne quadratique des écarts à la moyenne, et s'exprime sous la forme :

$$V(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

encore une fois en supposant l'équipondération des mesures (sinon adapter la formule avec des poids p_i). Contrairement à la plupart des indices descriptifs, la variance ne s'exprime pas dans l'unité de mesure, et est plus difficile à interpréter par rapport aux données observées. En revanche, elle présente certaines propriétés intéressantes en analyse de données. Elle permet de quantifier la « distance carrée moyenne à la moyenne » en prenant les écarts à la moyenne élevés au carré : les écarts sont ainsi sans signe (comme l'EAM), et élevés à la puissance deux, qui donne plus d'importance aux grands écarts (en vertu des propriétés de la fonction $x \mapsto x^2$).

On pourra également utiliser la formule de calcul un peu plus pratique : $V(x) = \frac{1}{n}\sum_i (x_i)^2 - \bar{x}^2$.

Variance corrigée. La variance corrigée, notée s_x^2 et utilisée dans les procédures inférentielles, est calculée de la même manière mais le dénominateur est alors $n-1$. Ceci se justifie par le fait que l'on peut montrer que la variance standard (variance de l'échantillon de taille n), à la différence de la moyenne[3], est un mauvais estimateur de la variance de population, et l'on enlève

[3] sous réserve, comme on l'a vu qu'il n'y aît pas trop de valeurs extrêmes

par conséquent un degré de liberté au dénominateur[4].

Comme on le verra dans la partie relative à l'analyse inférentielle (en particulier dans les procédures d'analyse de variance, cf. 4.2, 4.3), on utilisera alternativement cette notation pour la variance, et les écarts quadratiques à la moyenne (i.e. le numérateur) associés aux degrés de liberté (i.e. le dénominateur). Il est en effet souvent plus commode de travailler avec la somme des écarts quadratiques à la moyenne, que l'on retrouve souvent sous le terme SC ou SCE (somme des carrés des écarts)[5], lorsque l'on s'intéresse à la décomposition des sources de variabilité (cf. 4.2). Les degrés de libertés (dl), liés au nombre d'observations considéré, sont pris en compte ensuite dans les expressions utilisant les SC, et permettent de former les CM, ou CME (carré moyen des écarts[6]), qui reviennent à l'expression initiale de la variance corrigée.

Ecart-type. L'écart-type $\sigma(x)$[7] est simplement la racine carrée de la variance et s'exprime dans la même unité que la variable mesurée. De même que la moyenne, l'écart-type est lui aussi dans une certaine mesure sensible aux valeurs extrêmes, contrairement à l'intervalle inter-quartile.

Ecart-type corrigé. A l'image de la variance corrigée, dans les procédures à visée inférentielle, on utilisera de préférence l'écart-type corrigé s_x, qui est calculé avec un dénominateur égal à $n-1$ et qui est un estimateur non biaisé de l'écart-type de population.

Coefficient de variation

On utilise également un autre indicateur de dispersion relative, sans dimension, semblable à l'écart-type mais pondéré par la moyenne : le coefficient de variation. Il indique en fait le pourcentage de la valeur moyenne que représente l'écart-type. Il se définit somme suit :

$$cv_x = \frac{\sigma_x}{\bar{x}} \times 100$$

Il est notamment utile dans le cas où l'on souhaite comparer deux groupes d'observations,

[4] On corrige en fait la tendance à sous-estimer la variance de population à partir de petits échantillons. En fait, on peut montrer que la variance corrigée est un estimateur sans biais de la variance corrigée parente, dans le cas d'une population de taille infinie. Il est à noter que la plupart des calculatrices calculent par défaut une variance non-corrigée (variance de population), tandis que les logiciels calculent une variance corrigée (variance d'échantillon), comme par exemple Excel, Statistica, ou R.

[5] le terme correspondant dans la littérature anglo-saxone est SS

[6] le terme correspondant dans la littérature anglo-saxone est MS

[7] L'utilisation des notations en caractère grec est généralement réservée aux paramètres de population, et l'on peut trouver des notations comme « ety » pour l'écart-type. Nous utliserons néanmoins la notation $\sigma(x)$ pour désigner l'écart-type (non corrigé) des observations individuelles. En fait, on peut considérer que, dans le cadre de l'analyse descriptive, on se contente de décrire la distribution des observations que l'on assimile à une population, puisqu'il n'y a pas de visée inférentielle ; au contraire, dans le cadre des procédures inférentielles, on tente d'estimer les vrais paramètres de population à l'aide d'un échantillon, et l'on corrige en conséquence les indicateurs de dispersion utilisés pour l'estimation.

dont les moyennes sont sensiblement différentes ou dont l'unité de mesure n'est pas équivalente (e.g. comparer des mesures thermiques effectuées en France, en ˚C, et dans les pays anglosaxons, en ˚F, sans utiliser la conversion $^{\circ}C \leftrightarrow {}^{\circ}F$, ou comparer les performances globales de deux groupes de sujets dichotomisés selon leur niveau de QI).

Exemple d'application I

> Les notes obtenues par 36 élèves dans trois classes différentes ont été relevées afin de comparer l'homogénéité des systèmes de notation (Source : données fictives).

La figure 2.3 illustre l'influence de la distribution des valeurs observées (ici, des effectifs) sur les différents types d'indicateurs de position et de dispersion. On a représenté les notes obtenues par 36 élèves dans 3 classes (chaque note correspond à un point sur le graphique, les élèves ayant obtenu une note identique étant « superposés » verticalement). Les trois distributions possèdent une étendue identique de 15 points (min=4/20, max=19/20) et des moyennes comparables (de haut en bas, 11.03, 10.94 et 11.06, figurées en trait continu de couleur rouge). Les médianes (figurées en trait continu de couleur bleue) sont également relativement alignées sur la même valeur (10.0, 10.5 et 10.5 respectivement, cf. tableau 2.4.2). Par conséquent, d'une part, ces trois distributions ne peuvent être différenciées à partir des seuls indicateurs de tendance centrale, et, d'autre part, ces indicateurs de position ne renseignent pas sur la distribution des valeurs observées par rapport à leurs valeurs respectives.

	x_{min}	Q_1	Q_2	Q_3	IQ	x_{max}	$\bar{x}$	σ_x
classe 1	4	8.75	10.00	14.00	5.25	19	11.03	3.77
classe 2	4	9.00	10.50	13.00	4	19	10.94	2.88
classe 3	4	6.75	10.50	15.25	8.5	19	11.06	4.75

Tab. 2.1: Résumés numériques des notes de 3 classes de 36 élèves (Q_2 = médiane, IQ = intervalle inter-quartile).

Si l'on regarde les indicateurs de dispersion — intervalle inter-quartile (figuré par des tirets bleu sur la figure 2.3) et écart-type (figuré par des tirets rouge) — (cf. également tableau 2.4.2), on constate en revanche que ceux-ci diffèrent entre les 3 classes. L'écart-type est le plus important pour la classe 3 ($\sigma_3 = 4.75$, figure 2.3, en bas), et le plus faible pour la classe 2 ($\sigma_3 = 2.88$, figure 2.3, milieu). Il en va de même de l'intervalle inter-quartile : celui-ci est plus important pour la classe 3, puis pour les classes 1 et 2 ($IQ_1 = 5.25$, $IQ_2 = 4$, $IQ_3 = 8.5$). L'étendue de la moitié des notes (i.e. 18) autour de la valeur centrale est par conséquent beaucoup plus importante pour la classe 3. Ces deux indicateurs de dispersion — écart-type et intervalle inter-quartile — vont dans le même sens et soulignent l'hétérogénéité plus importante de la classe 3 ; néanmoins, ce sont des indicateurs « globaux » de dispersion, et ils ne renseignent

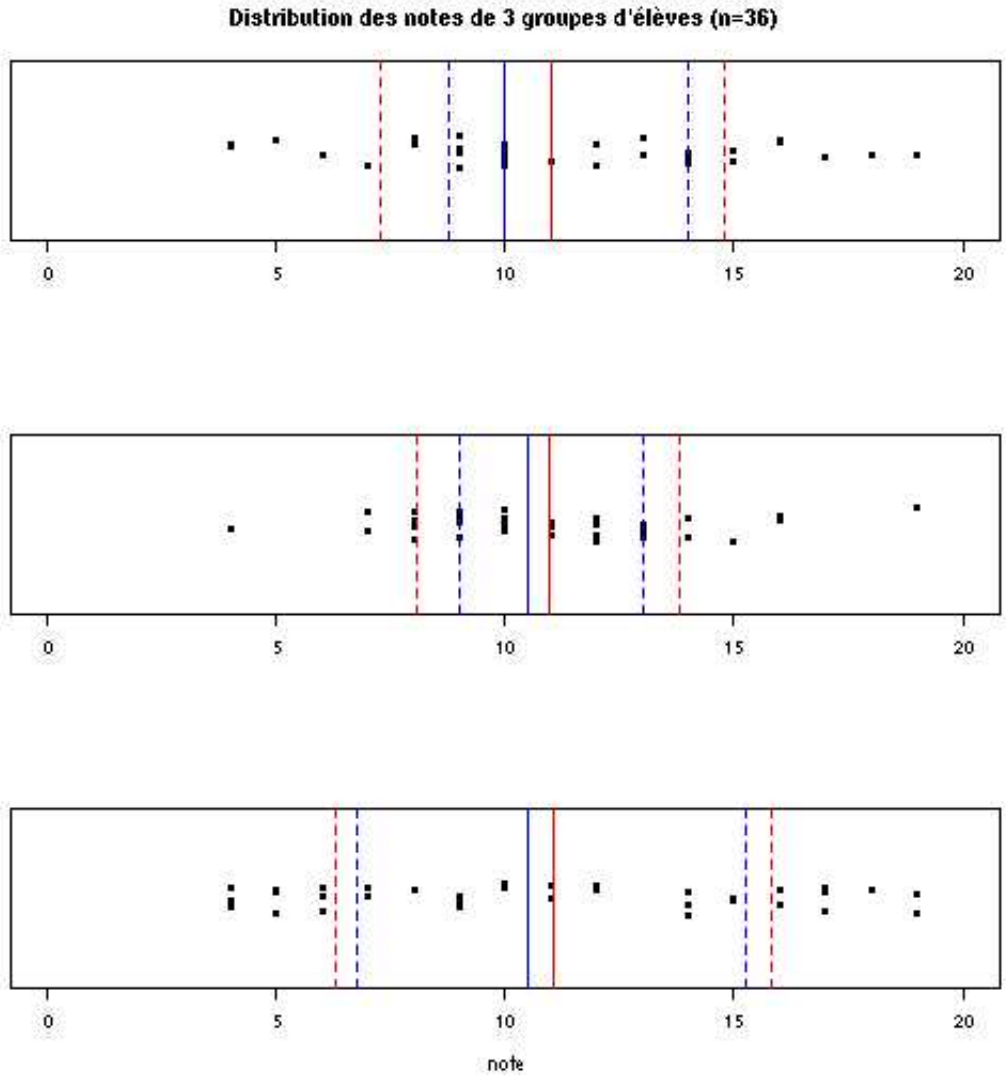

Fig. 2.3: Distributions univariées d'effectifs (notes de 36 élèves pour 3 classes). En rouge sont représentées la moyenne (trait continu) ± son écart-type (tirets), et en bleu la médiane (trait continu), ainsi que le premier et troisième quartile (tirets).

pas sur le degré de symétrie de la distribution des notes autour de la valeur centrale (moyenne ou médiane puisqu'ici elles sont assez proches) qui pourrait expliquer cette variabilité. On pourrait pour cela s'intéresser aux écarts à la moyenne, mais on peut également utiliser le positionnement des premier (Q_1) et troisième (Q_3) quartiles pour savoir comment se répartissent les observations. On constate ainsi que pour la classe 2, Q_1 est plus proche de la médiane que Q_3 : 25 % des notes sont dans la tranche 9-10.5, et 25 % dans la tranche 10.5-13, ce qui traduit une relative asymétrie

dans la distribution des notes. Il en va de même pour la classe 1 qui présente elle aussi un profil asymétrique ($|Q_2 - Q_1| < |Q_3 - Q_2|$). On retrouve cette asymétrie dans la figure 2.3, contrairement à la classe 3 pour laquelle les notes se répartissent plus uniformément de part et d'autre des valeurs centrales. Ainsi, bien que la distribution des notes de la classe 2 soit la plus « ressérée » autour des valeurs centrales, c'est également une classe qui présente une légère asymétrie.

En conclusion, si les notations moyennes, ainsi que les notes les plus basses et les plus hautes, semblent comparables pour ces trois classes, on note cependant la présence d'une variabilité plus importante dans la classe 3, et une légère asymétrie négative des notes dans les classes 1 et 2.

2.4.3 Forme de la distribution des observations

Comparaison des indicateurs de tendance centrale

Lorsque les trois indicateurs de position — mode, médiane et moyenne — sont à peu près « alignés » sur la même valeur ou la même classe, cela indique généralement une distribution relativement symétrique des observations (cf. figure 2.4). Le cas échéant, cela signe une certaine asymétrie de la distribution (cf. figure 2.3). Dans ce cas, la médiane étant toujours située entre le mode et la moyenne (lorsque la distribution est unimodale uniquement), la position de la moyenne par rapport au mode permet de déterminer le sens de l'asymétrie : lorsque l'on a mode $<$ médiane $<$ moyenne, la distribution est asymétrique vers la droite, tandis que lorsque l'on a moyenne $<$ médiane $<$ mode, l'asymétrie est à gauche. L'examen du positionnement des fractiles (e.g. les quartiles, ou pour plus de précision les déciles) permet également de vérifier l'aspect de la distribution, et ses éventuelles asymétries (locales ou globale).

D'autre part, la présence de deux modes indique que la distribution est *bimodale* (ou multimodale s'il y a plus de deux modes relatifs[8]). La distribution peut demeurer « symétrique » lorsque les deux modes sont égaux, mais la présence d'un mode relatif traduit généralement une distribution relativement asymétrique.

Coefficient d'asymétrie

Le coefficient d'asymétrie (appelé également moment centré d'ordre 3) mesure la symétrie relative de la distribution d'effectifs par rapport à la moyenne. Il est donné par la formule suivante :

$$g = \frac{n \sum_{i=1}^{n} (x_i - \bar{x})^3}{(n-1)(n-2)}$$

C'est en fait la moyenne des cubes des valeurs centrées-réduites des observations. Lorsque la

[8] On définit le mode relatif comme un mode ayant une valeur inférieure à celle du mode absolu (mode principal).

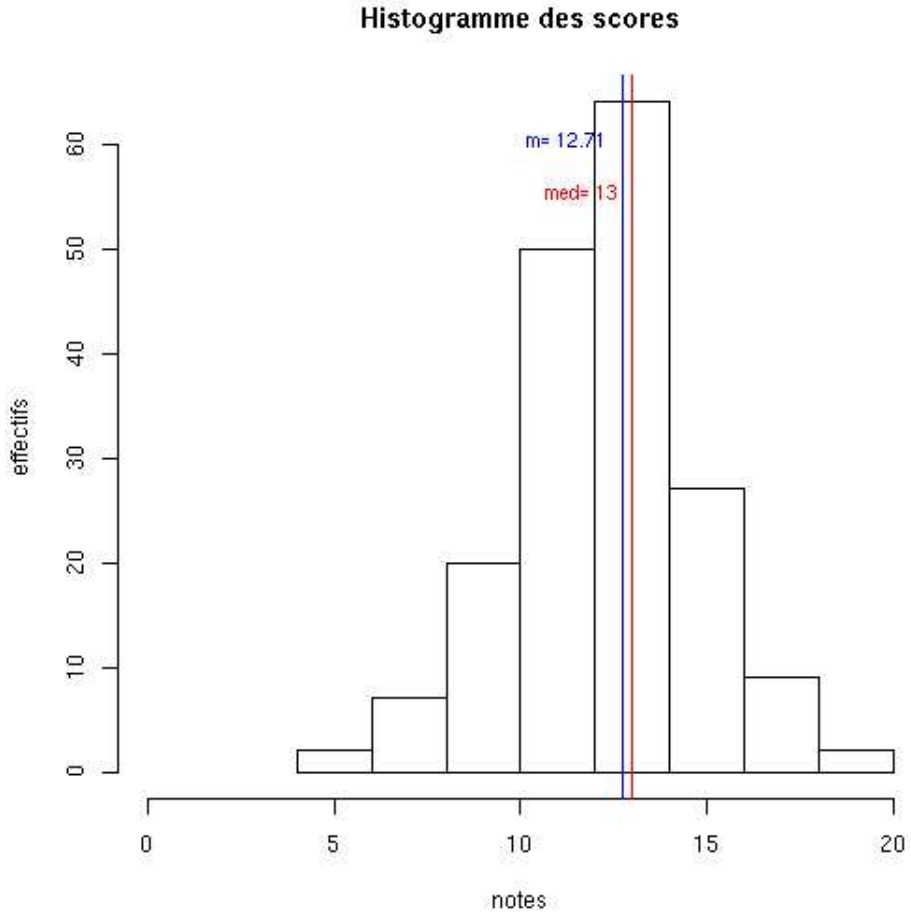

Fig. 2.4: Exemple d'une distribution d'effectifs relativement symétrique ($n = 200$)

distribution est asymétrique à droite, c'est-à-dire que la majorité des effectifs sont concentrés à droite de la moyenne, le coefficient d'asymétrie est positif. A l'inverse, lorsque la distribution est asymétrique à gauche, le coefficient d'asymétrie est négatif et traduit une concentration des valeurs observées en-dessous de la moyenne. Ce coefficient est nul lorsque la distribution est normale.

Comme cela a été dit précédemment, lorsque la distribution est symétrique, la moyenne et la médiane sont (à peu près) confondues. Par conséquent, il est possible de mesurer l'asymétrie en utilisant la formule simplifiée : $g = (\bar{x} - x_{med})/s_x$.

Enfin, cet indicateur n'est utilisable (i.e. a un sens) que dans le cas d'une distribution unimodale, de forme relativement « normale ». En effet, ce coefficient (ainsi que le coefficient d'aplatissement, voir § suivant) a été construit sur la base de la distribution normale, et n'est par conséquent pas utilisable pour d'autres types de distribution, ou en cas de trop fortes asymé-

tries.

Remarques. Deux autres coefficients d'asymétrie peuvent être utilisés :
- le coefficient d'asymétrie de Pearson, $\frac{g}{s^3}$, est l'indice le plus souvent utilisé car il est sans dimension, et il sert d'estimé du paramètre de population dans le cadre d'un test inférentiel de significativité ;
- le coefficient d'asymétrie de Fisher est la valeur obtenue par le calcul suivant : $(\frac{g}{s^3})^2$

Coefficient d'aplatissement

Le coefficient d'aplatissement (appelé également moment centré d'ordre 4) quantifie le degré d'aplatissement de la courbe ; il s'apparente à une mesure de dispersion exprimée en fonction de la valeur de la moyenne ± son écart-type. Il est donné par la formule :

$$k = \frac{n(n+1)(\sum_{i=1}^{n}(x_i - \bar{x})^4)/(n-1) - 3(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}{(n-2)(n-3)}$$

C'est la moyenne des puissances quatrièmes des observations centrées-réduites. Lorsque celui-ci est négatif, la distribution est « aplatie », tandis que lorsqu'il est positif la distribution est dite « aigüe ». On a une distribution normale dans le cas où le coefficient d'aplatissement est nul. De même, cet indicateur n'est utilisable (i.e. a un sens) que dans le cas d'une distribution unimodale, de forme à peu près normale.

Remarques. Deux autres coefficients d'asymétrie peuvent être utilisés :
- le coefficient d'aplatissement le plus utilisé est celui de Pearson, qui est obtenu par le calcul suivant : $\frac{k}{s^4}$
- le coefficient d'aplatissement de Fisher est la valeur obtenue par le calcul suivant : $\frac{k}{s^4} - 3$

2.5 Analyse descriptive des différences et liaisons entre variables

L'ensemble des indicateurs descriptifs décrits dans les paragraphes précédents permettent de procéder au résumé numérique d'une distribution univariée d'observations, ce qui permet à la fois de caractériser un échantillon dans sa globalité, mais également de situer une observation (ou un individu) parmi l'ensemble des observations (resp. individus) qui constituent l'échantillon. Lorsque l'on a deux ou plusieurs échantillons, on souhaite également pouvoir les *comparer* entre eux : c'est le domaine de l'« analyse multivariée » (ou bivariée dans le cas précis de deux échantillons).

Lorsque ces échantillons résultent d'une classification à l'aide d'un facteur, on cherche en fait à les *comparer* entre eux, c'est-à-dire à caractériser l'effet de ce facteur sur la variable mesurée. En

revanche, lorsqu'il s'agit de deux variables numériques ou qualitatives, on peut vouloir chercher le degré d'association entre ces deux variables.

On caractérisera dans tous les cas le *sens*, l'*ampleur* et l'*importance* de l'effet considéré.

2.5.1 Différences quantitatives entre indicateurs descriptifs

Lorsque l'on est en présence de deux groupes d'observations, et que l'on souhaite les comparer, on peut utiliser leurs indicateurs descriptifs de tendance centrale et de dispersion. La différence observée entre deux groupes, nommée d_{obs}, est simplement la différence entre les moyennes respectives de chacun des deux groupes : $d_{obs} = \bar{x}_1 - \bar{x}_2$ (cette différence est signée). Cette différence quantifie l'*effet moyen* du facteur considéré. L'un des groupes peut bien entendu être un groupe de référence (i.e. non observé, mais dont on connaît les caractéristiques). Par exemple, on peut comparer les scores de QI relevés dans un groupe de 20 personnes à ceux de la population parente dont on connaît les caractéristiques générales (moyenne 100, écart-type 15). On peut également comparer les performances entre deux groupes de sujets se distinguant par le sexe dans une tâche de catégorisation : la différence entre le nombre moyen d'items résolus par les garçons et celui résolu par les filles constituerait ainsi l'effet moyen du facteur « sexe ».

L'importance de l'effet moyen est généralement interprétée en fonction de *critères sémantiques*, liés à la connaissance du domaine ou à des études antérieures. Par exemple, pour des temps de réaction simples, on sait que des différences de 20 à 50 ms peuvent être considérées comme importantes, tandis que pour des temps de réaction de choix, on attend plutôt des différences marquées à partir de 50 ms par exemple.

On peut également pondérer cette distance moyenne par l'écart-type, ce que l'on appelle l'*effet calibré* et qui est obtenu comme $EC = d_{obs}/s_{intra}$ ([5])[9]. Cela permet de relativiser l'importance de l'écart en fonction de la variabilité observée à l'intérieur du ou des groupe(s). A la différence de l'effet moyen et de son interprétation au regard d'un critère sémantique, l'effet calibré peut être évalué à l'aide de *critères psychométriques* (cf. [5]). Les critères de décision retenus sont les suivants :

- $EC < 1/3$, l'effet est considéré comme faible ;
- $1/3 \leq EC \leq 2/3$, l'effet est considéré comme intermédiaire ;
- $EC > 2/3$, l'effet est considéré comme important.

[9] Dans le cas où l'on compare un groupe d'observations à un groupe de référence, l'écart-type intra considéré est celui lié au groupe de référence : on parle d'*effet calibré associé à la moyenne*. Dans le cas où deux groupes d'observations sont à comparer, l'écart type considéré est la moyenne des écarts-type intra des deux échantillons : on parle alors d'*effet calibré associé aux moyennes des groupes*.

2.5.2 Liaison entre variables

Liaison entre 2 variables quantitatives

La liaison entre deux variables numériques peut être étudiée grâce au *coefficient de corrélation*. Néanmoins, il faut bien garder présent à l'esprit que le coefficient de corrélation de Bravais-Pearson ne mesure que des relations linéaires, et sa valeur n'est en rien le reflet de l'existence d'un lien de causalité entre les deux variables.

Le coefficient de corrélation de Bravais-Pearson mesure en fait l'intensité de l'association entre les deux variables. Il est donné par la formule suivante :

$$r = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \qquad |r| \leq 1$$

(la fonction $cov(x,y)$ désigne la covariance entre les deux séries d'observations).

Cette valeur est comprise entre -1 et $+1$, les valeurs négatives signifiant une corrélation négative (lorsque les valeurs d'une variable augmentent, les valeurs de l'autre variale ont tendance à diminuer) et les valeurs positives traduisant une corrélation positive (les deux séries d'observations co-varient dans le même sens). Les deux cas extrêmes $r = -1$ et $r = +1$ traduisent des corrélations parfaites, respectivement négative et positive. Enfin, le cas $r = 0$ traduit une absence de corrélation.

Si l'interprétation de la corrélation est facilitée par l'élaboration d'une représentation graphique bivariée, il convient de bien faire attention aux unités utilisées, avant de juger l'inclinaison du nuage de points ou la pente de la droite de régression qui est souvent associée au coefficient de corrélation[10] (cf. figure 2.5) : en effet, la pente de la droite ajustant le nuage de points est dépendante de l'échelle utilisée, tandis que le coefficient de corrélation est indépendant de l'unité de mesure. Par ailleurs, un coefficient de corrélation élevé ne traduit pas toujours une relation linéaire avérée, comme l'illustre la figure 2.6 dans laquelle différentes séries de données, possédant toutes le même coefficient de corrélation ($r = 0.82$), sont représentées : on voit clairement que certaines relations ne sont strictement pas linéaires. Ainsi également, un faible coefficient de corrélation ne signifie pas forcément l'*indépendance* des deux variables considérées, puisque celles-ci peuvent être liées par des relations non-linéaires (e.g. polynomiale, logarithmique, parabolique, etc.).

Enfin, on prendra garde au fait que la corrélation entre deux variables quantitatives peut être liée à une troisième variable agissant de manière indépendante sur celle-ci. Ce type de problématique multivariée peut être adressé à l'aide des corrélations partielles (cf. 4.7.2, p. 98) ; encore faut-il identifier le type de variable susceptible d'influencer les variables d'intérêt... Par exemple, on mentionne généralement l'existence d'une relation entre la pratique d'un sport et la consommation d'alcool. En fait, ce lien disparaît si l'on introduit la variable « âge » (i.e.

[10] La pente de la droite de régression est liée au coefficeint de corrélation par la relation : $\alpha = r\frac{\sigma_x}{\sigma_y}$.

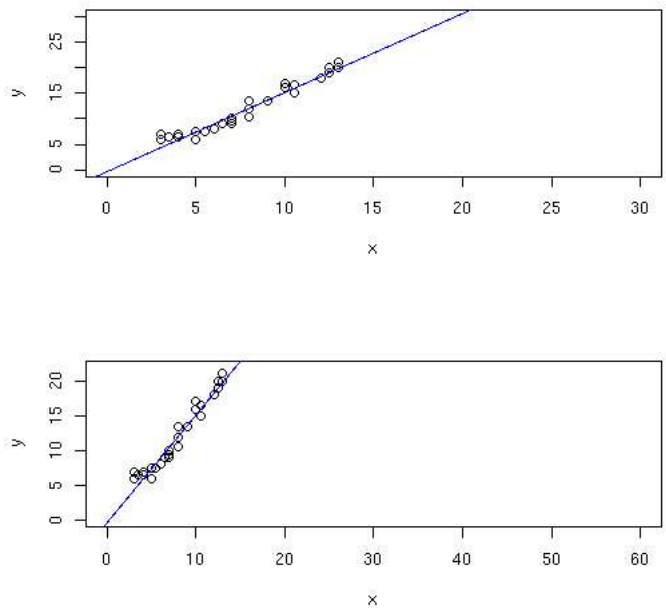

Fig. 2.5: Graphes bivariés. Les mêmes données sont utilisées (corrélation : r=0.97 ; droite de régression : y=1.54x-0.15), seules les limites des axes des abscisses et des ordonnées varient. (Explication dans le texte)

en ajustant les données sur l'âge), tout simplement parce que la consommation d'alcool est fortement liée à l'âge[11].

Le *coefficient de détermination*, R^2, qui est simplement le coefficient de corrélation élevé au carré, indique quant à lui la part de variance prise en compte par l'association linéaire, c'est-à-dire la part de variance expliquée en considérant l'existence d'une relation linéaire entre les deux variables. La part de variance non expliquée par la corrélation, i.e. la variance résiduelle, est quantifiée par le *coefficient d'indétermination* $1 - R^2$.

Critères de décision. Le coefficient de détermination s'interprète selon les critères suivants:

- $0 \leq R^2 < 0.04$: R^2 faible, la part de variance expliquée par l'association entre les deux variables peut être considérée comme négligeable ;
- $0.04 \leq R^2 \leq 0.16$: R^2 intermédiaire, la part de variance expliquée par l'association entre les deux variables peut être considérée comme intermédiaire ;
- $R^2 > 0.16$: R^2 important, la part de variance expliquée par l'association entre les deux variables peut être considérée comme importante (> 16 %).

[11] On appelle *variable de confusion* ce type de variable qui a tendance à annuler ou accentuer la relation statistique entre deux variables.

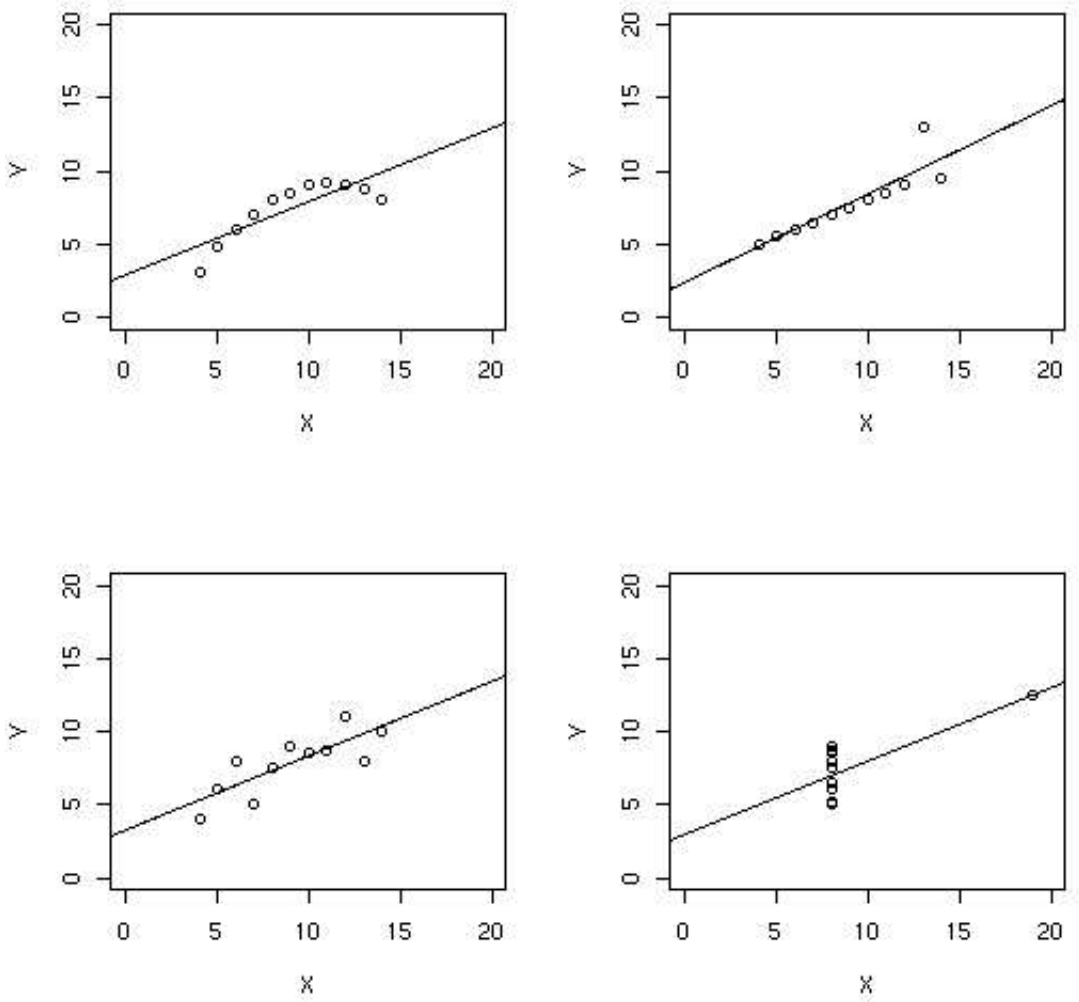

Fig. 2.6: Exemples de corrélation. (Adapté de [9], p. 160)

Liaison entre 1 variable qualitative et 1 variable quantitative

En présence d'une variable numérique et d'une variable qualitative (e.g. variable de groupement), on peut estimer la part de variance expliquée par la prise en compte de la variable qualitative dans la variabilité observée des scores. L'indice η^2 (« eta carré ») est calculé comme le simple rapport entre la variance entre les groupes (variance inter) et la variance à l'intérieur des groupes (variance intra) :

$$\eta^2 = \frac{V_{inter}}{V_{intra}}$$

La variance inter correspond en fait à la variance des moyennes de groupe. La variance intra correspond à la moyenne des variances à l'intérieur des groupes. Si on prend l'exemple de scores observés chez des enfants classés selon 3 groupes d'âge, on peut calculer la part de variance entre les scores qui est expliquée par la prise en compte du facteur de groupe.

Critères de décision. Les critères de décision sont les mêmes que pour le coefficient de détermination R^2 :

- $0 \leq \eta^2 < 0.04$: η^2 faible, la part de variance expliquée par la variable peut être considérée comme négligeable ;
- $0.04 \leq \eta^2 \leq 0.16$: η^2 intermédiaire, la part de variance expliquée par la variable peut être considérée comme intermédiaire ;
- $\eta^2 > 0.16$: η^2 important, la part de variance expliquée par la variable peut être considérée comme importante (> 16 %).

3 Analyse inférentielle

3.1 De la description à l'inférence

3.1.1 Schéma général de la démarche inférentielle

La description des données recueillies lors d'une expérimentation (ou issues d'une démarche d'observation) a été décrite dans le chapitre précédent. Une question que l'on peut se poser est : peut-on *généraliser* les résultats observés sur l'échantillon étudié à la *population parente* dont il est issu[1], et que forcément nous n'avons pas observée ? En d'autres termes, on cherche à *inférer* au niveau de la population parente des résultats observés, e.g. l'effet d'un facteur (ou une caractéristique particulière de la distribution des observations), sur un échantillon représentatif de cette population.

Par exemple, si nous avons observé qu'un groupe de 20 personnes présente des scores, en moyenne, plus élevés (e.g. 2 points sur 20) qu'un autre groupe de même taille, peut-on dire qu'il en ira de même, en moyenne, pour tous les individus présentant les mêmes caractéristiques, c'est-à-dire que, au niveau des populations parentes (i.e. l'ensemble des individus), les individus similaires à ceux du premier groupe réussissent en moyenne mieux que les individus similaires à ceux du deuxième groupe. Les caractéristiques distinguant les deux groupes d'individus peuvent être variées (âge, sexe, niveau d'aptitude dans telle activité, traitement associé, etc.), mais constituent un facteur de classification : si l'on s'intéresse à l'effet du sexe de l'individu sur les performances mesurées, on aura a priori deux populations qui se distinguent par le facteur « sexe ». Dans la démarche du chercheur, l'observation au niveau de l'échantillon d'une différence en termes de performances (les 20 sujets de sexe féminin ont mieux réussi que les 20 autres sujets de sexe masculin), c'est-à-dire d'un *effet* du facteur « sexe » quantifié par les indicateurs appropriés (cf. 2.5.1, p. 30), sera ainsi souvent accompagnée de la question de l'existence d'une différence significative de performances entre les femmes et les hommes au niveau de la population parente (on parlera d'*effet parent*) à laquelle on répondra par un test d'inférence statistique approprié, et une estimation probabiliste du domaine de variation de cet effet parent au moyen d'une procédure de quantification adaptée.

3.1.2 Formalisation

On cherche en fait à décrire les *paramètres* de la *population parente* à l'aide d'un sous-ensemble de cette population, l'*échantillon*. Ces paramètres de population sont *estimés* à partir

[1] On rappelle que l'échantillon a été sélectionné grâce à une méthode d'échantillonnage aléatoire dans la population parente (ou la population de référence).

des indicateurs descriptifs, qui constituent ce que l'on appelle des *statistiques*[2].

Comme on l'a vu dans le chapitre précédent (cf. 2.4.1, p. 21, et 2.4.2, p. 22), les indicateurs de position comme la moyenne et les indicateurs de dispersion — variance et écart-type — sont utilisés pour décrire les observations d'un échantillon tiré d'une population plus vaste, non observée (la population parente). Dans la démarche inférentielle, on utilise donc de préférence ce type d'indicateurs d'échantillon (et les propriétés caractéristiques de leurs distributions), à l'instar de la démarche descriptive où l'on travaille plutôt sur les observations. Mais, il est bien entendu possible de travailler avec d'autres statistiques, comme les proportions, les médianes, les coefficients de corrélation, etc.

Lors de l'analyse descriptive, on situait volontiers un individu parmi la distribution (connue) des effectifs, ou une observation parmi l'ensemble des valeurs observées de la variable quantitative mesurée (cf. également 3.2.2, p. 40). Ici, on s'intéresse toujours aux distributions, mais cette fois-ci on travaillera sur les distributions parentes (la plupart du temps inconnues) : ce n'est plus une observation que l'on va tenter de situer dans la distribution des observations (échantillon), mais une moyenne d'échantillon, i.e. une statistique, dans la distribution parente estimée. On ne raisonnera alors plus en termes d'effectifs ou de fréquences observées, mais en termes probabilistes : en d'autres termes, on cherchera à évaluer la probabilité que la statistique calculée se situe dans un certain intervalle de valeurs de la variable mesurée, ou de manière équivalente au-delà d'une certaine valeur de référence.

3.2 Distribution de probabilités

3.2.1 Distribution d'échantillonnage de la moyenne et loi normale

La loi normale, ou loi de Laplace-Gauss, est une des lois les plus utilisées en statistique, en raison des propriétés de convergence en loi des différentes distributions de probabilités et surtout du théorème d'échantillonnage de la moyenne, que nous présentons « rapidement » dans le pragraphe suivant. En fait, il y a très peu de variables en biologie ou en psychologie qui se distribuent normalement, mais certains développements mathématiques ont permis de montrer que la distribution des moyennes tend vers la normalité, et l'approximation par la loi normale devient acceptable lorsque l'on travaille avec des données moyennées. De même, les autres statistiques potentiellement utilisables (variances, proportions, coefficient de corrélation, ...) possèdent des propriétés connues du point de vue de leurs distributions.

[2] Les caractéristiques principales des statistiques sont l'*exactitude*, la *consistance* et la *précision* (se référer aux ouvrages spécialisés pour de plus amples développements sur le sujet).

Distribution d'échantillonnage de la moyenne

C'est en ce sens qu'il est préférable de travailler avec les indicateurs d'échantillons plutôt que les observations, car ceux-ci ont généralement des distributions connues : si l'on estime à plusieurs reprises la moyenne d'un échantillon de taille suffisamment grande (e.g. $n \geq 1000$), et que l'on construit l'histogramme de ces moyennes, la forme de cet histogramme sera la même que celle de la loi normale, et ce quelque soit la forme des distributions des observations initiales de l'échantillon (cf. pour illustration la figure 3.1) : c'est sur ce principe de *distribution d'échantillonnage*, ici de la moyenne, que reposent tous les tests statistiques. La moyenne de la distribution d'échantillonnage de la statistique considérée est ainsi la moyenne des moyennes de l'ensemble des échantillons que l'on peut former[3], et on peut montrer qu'elle est égale à la moyenne de la population parente, d'où sont tirés ces échantillons. On montre également que la variance de la distribution d'échantillonnage est égale à $\frac{\sigma^2}{n} \times \frac{N-n}{N-1}$, σ^2 représentant la variance de la population parente. Lorsque l'effectif N de la population est très supérieur à celui de l'échantillon n considéré, le rapport $\frac{N-n}{N-1}$ tend vers 1, et la variance de la distribution d'échantillonnage vaut $\frac{\sigma^2}{n}$[4].

On peut donc estimer les paramètres d'une population parente, au travers de la distribution d'échantillonnage, à l'aide des indicateurs de l'échantillon (moyenne $\bar{x}$ et variance corrigée s^2, qui permet d'estimer la variance de population), sans avoir à l'échantillonner de manière exhaustive. La distribution d'échantillonnage représente alors la *distribution de probabilité* de la statistique considérée[5].

Ceci vaut également pour de nombreuses autres statistiques dont la distribution caractéristique est connue (e.g. distribution du χ^2 pour les proportions, distribution du $\mathcal{F}$ de Fisher-Snedecor pour les variances) ou tend vers la loi normale, sous certaines conditions.

[3] A partir d'une population d'effectif N, il est possible de constituer C_N^n échantillons de taille n (C_n^k désignant le nombre binomial).

[4] On remarquera alors que la variance de la distribution d'échantillonnage de la moyenne est inférieure à celle de la population parente : la distribution d'échantillonnage est plus « resserrée » autour de la moyenne.

[5] Intuitivement, on peut « toucher du doigt » cette notion en considérant un grand nombre de lancers successifs d'un dé à 6 faces bien équilibré. La distribution des fréquences observées est approximativement rectangulaire, puisque $P(X = x) = 1/6$, avec $x = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$: on a pratiquement autant de chances de tomber sur l'une des 6 faces. Si maintenant on effectue un grand nombre de lancers successifs de 4 dés, la distribution d'échantillonnage de la moyenne des points obtenus constitue un histogramme symétrique et de forme pyramidale, centré autour de la valeur moyenne 3 : cela s'explique simplement par le fait que la probabilité d'obtenir des valeurs proches de 3 est beaucoup plus grande que les probabilités d'obtenir des valeurs faibles ou élevées (i.e. extrêmes), en raison des combinaisons possibles à partir de 4 dés qui réalisent cet évènement. Or, plus on augmentera le nombre de dés à lancer (e.g. 16 dés, 50 dés) un grand nombre de fois, plus la distribution observée se rapprochera d'une distribution symétrique, en forme de cloche, centré autour de la valeur moyenne : on peut montrer mathématiquement que cette distribution est celle d'une variable aléatoire se distribuant selon une loi normale de moyenne μ et de variance σ^2/n.

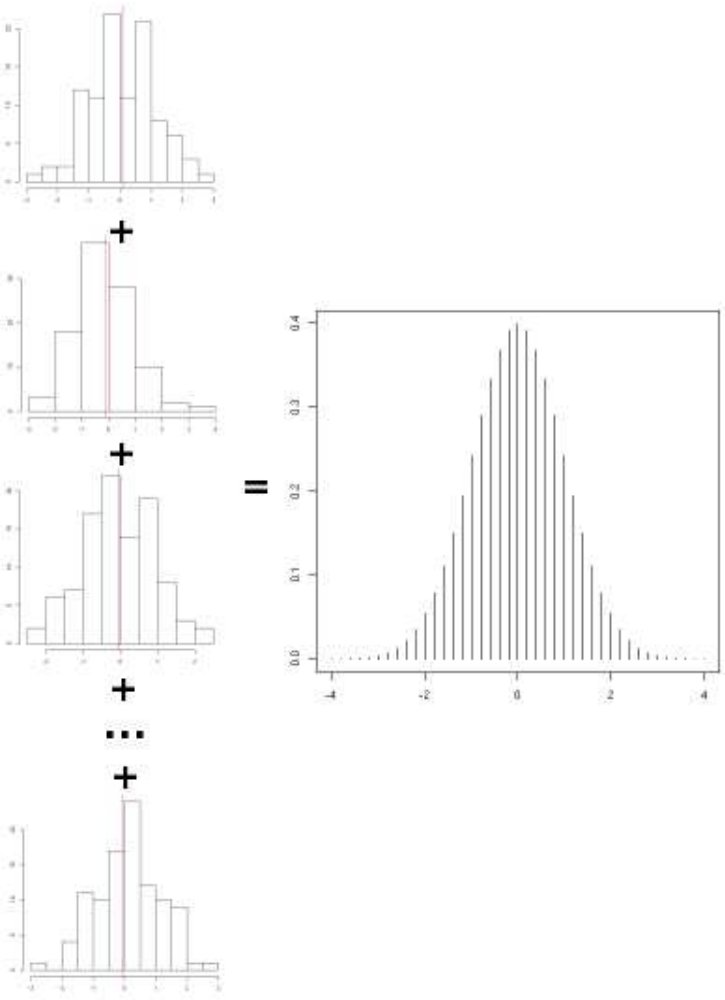

Fig. 3.1: Principe de la distribution d'échantillonnage de la moyenne. La distribution des moyennes de l'ensemble des échantillons tirés d'une population (à gauche) suit une distribution normale (à droite).

Loi normale

La distribution normale est une fonction continue à deux paramètres : la moyenne μ et la variance σ^2 de la distribution, et est définie comme suit :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \qquad (x \in I\!R)$$

Son aspect est celui d'une courbe en cloche, symétrique. Elle est centrée sur la moyenne, et sa largeur à mi-hauteur est égale à deux écarts-type. La figure 3.2 présente deux exemples de

distribution suivant une loi normale de même moyenne (4), mais dont la variance passe de 1 (à gauche) à 3.06 (à droite). Ce triplement de la variance se traduit par un « évasement » de la distribution autour de sa valeur moyenne et un « épaississement » des queues de la distribution.

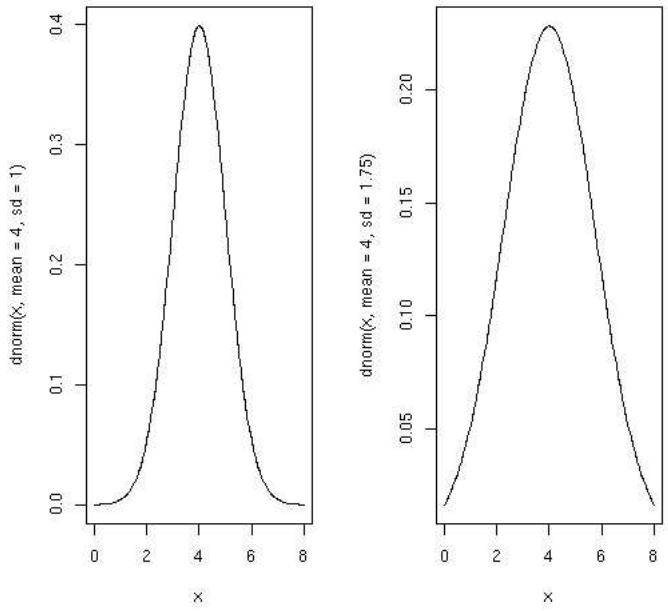

Fig. 3.2: Exemples de distributions normales. A gauche, distribution normale de paramètres $\mu = 4$ et $\sigma^2 = 1^2$; à droite, distribution normale de paramètres $\mu = 4$ et $\sigma^2 = 1.75^2$

Etant donné qu'il existe une infinité de distributions normales, puisque les deux paramètres μ et σ^2 peuvent prendre toutes les valeurs possibles sur $\mathbb{R}$, il s'avère peu commode de travailler directement sur cette distribution (à moins de posséder un logiciel donnant les valeurs de $f(x)$ en fonction de ses paramètres, pour un x donné), qui n'est d'ailleurs pas tabulée. On utilise alors la distribution normale centrée-réduite $z(x)$, qui est une distribution normale de moyenne 0 et de variance 1 (cf. figure 3.3), et pour laquelle on dispose d'une table de certaines valeurs de la fonction de répartition[6]. Cela permet de connaître la valeur de la fonction de répartition pour un z_0 donné, i.e. $P(z < z_0)$ (par définition de la fonction de répartition), c'est-à-dire la proportion d'observations située sous la courbe jusqu'au point d'abscisse z_0.

Rappelons que la transformation centrée-réduite des données est calculée comme suit :

$$z_i = \frac{x_i - \bar{x}}{\sigma}$$

En fait, on centre les données par rapport à la moyenne, et on les réduit par rapport à l'écart-type. De la sorte, toute valeur centrée-réduite z_i s'exprime en point d'écart-type par rapport à

[6]La fonction de répartition peut être « grossièrement » vue comme l'analogue de la distribution des effectifs cumulés construite dans le cas d'un échantillon

la moyenne. Il est alors évident qu'une valeur z_i positive indique que la valeur est supérieure à la moyenne, et, réciproquement, une valeur z_i négative indique que la valeur est inférieure à la moyenne.

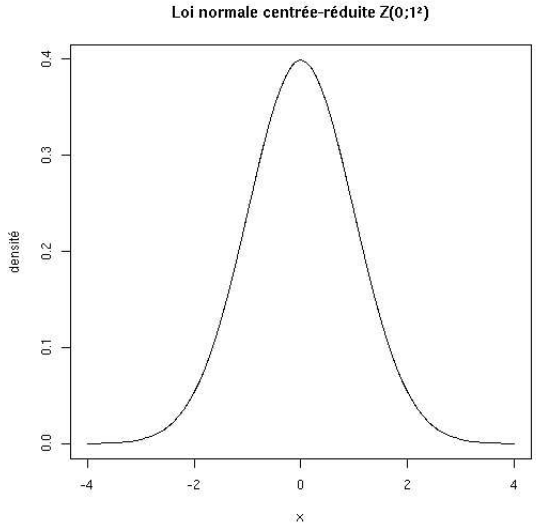

Fig. 3.3: Loi normale centrée-réduite $\mathcal{N}(0; 1^2)$

3.2.2 Calcul élémentaire de probabilités

Lorsque l'on connaît la distribution de la population parente (i.e. ses paramètres, moyenne et variance), ce qui est rarement le cas en pratique, on peut calculer la proportion d'individus ou d'observations qui sont situés par rapport à une valeur de référence, ou dans un intervalle donné. Si l'on prend comme exemple la distibution théorique de la taille des individus (de sexe masculin, français, et dans la tranche d'âge 20-35 ans), celle-ci suit une loi normale de moyenne 170 et d'écart-type 10. On peut donc évaluer la probabilité qu'un individu choisi au hasard parmi la population entière mesure moins de 185 cm, ou plus de 198 cm, ou ait une taille comprise entre 174 et 186 cm.

Puisqu'en principe on ne dispose pas des tables de la distribution correspondante $\mathcal{N}(170; 10^2)$, on utilise la loi normale centrée-réduite $\mathcal{N}(0; 1^2)$, dont la table est disponible dans n'importe quel manuel (cf. Annexe B, p. 113). Ainsi, la probabilité qu'un individu choisi au hasard parmi la population entière mesure moins de 185 cm est de 0.933 ($P(X < 185) = P(Z < \frac{185-170}{10})$, la probabilité qu'un individu mesure plus de 198 cm est de 0.003 ($1 - P(X < 198) = 1 - P(Z < \frac{198-170}{10})$), et la probabilité que sa taille soit comprise entre 174 et 186 est 0.290 ($P(X < 186) - P(X < 174) = P(Z < \frac{186-170}{10}) - P(Z < \frac{174-170}{10})$). Comme on le voit, la probabilité qu'un individu choisi aléatoirement dans une population de moyenne 170 ± 10 mesure plus de 198 cm est très faible (cf. figure 3.4).

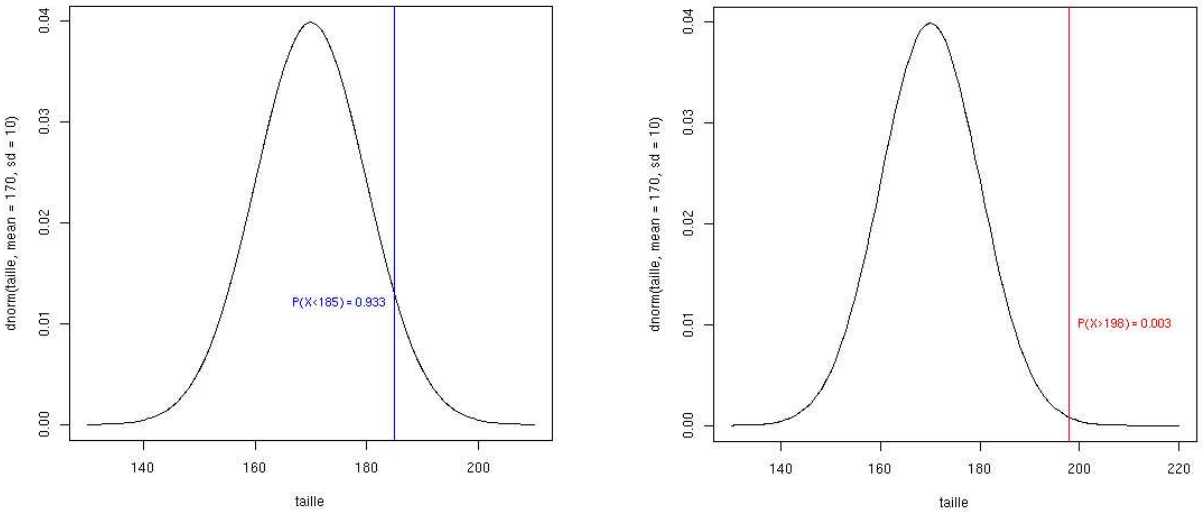

Fig. 3.4: Illustration des probabilités associées à des évènements

Dans le cas de figure exposé ci-dessus, la moyenne et la variance de la population parente sont connues. On peut ainsi calculer les limites d'un intervalle qui contient une proportion donnée de la population, à l'aide de la formule :

$$\mu \pm Z\sigma$$

qui n'est en fait qu'une réexpression de la formule classique $Z = \frac{X-\mu}{\sigma}$. Cet intervalle, appelé *intervalle de confiance* (dans ce cas, c'est un intervalle de confiance pour observations), permet de déterminer si une observation fait vraisemblablement partie de la même population ou non. On peut ainsi calculer l'intervalle contenant 95 % des valeurs : les bornes de cet intervalles seront alors $[\mu - 1.96\sigma\,;\,\mu + 1.96\sigma]$. C'est en fait sur la base de ce calcul de probabilités que repose le *test de typicalité*[7] : un individu, ou un groupe d'individus, sera déclaré atypique ou non représentatif de la population parente dont il est issu, lorsqu'il a une position au moins aussi extrême qu'une certaine position de référence[8]. Cette position de référence est généralement fixée à 5 %. Pour reprendre l'exemple précédent, si on tire au hasard dans la population un groupe de individus de sexe masculin, français, et dont l'âge est compris dans la tranche 20-35 ans, et qu'on constate que la taille moyenne observée est de 198 cm ($P(X > 198) = 0.003 < 0.05$), alors on peut raisonnablement penser — et affirmer — que ce groupe n'est pas représentatif, ou est atypique, de la population dont il est issu (l'ensemble des personnes de sexe masculin, français, et dont l'âge est compris dans la tranche 20-35 ans).

Ce type de démarche constitue la base même de la plupart des tests inférentiels : situer un échantillon observé, caractérisé par des statistiques appropriées, parmi l'ensemble des échantillons que l'on aurait pu constituer à partir de la population parente dont il a été extrait par

[7] appelé aussi *test Z* dans la mesure où il repose sur la loi Z, qui est la loi normale centrée-réduite

[8] il s'agit alors d'un test *inclusif*, mais on peut restreindre la procédure à un test *exclusif*, en excluant la position extrême considérée.

une procédure d'échantillonnage aléatoire. Si sa position est jugée extrême, selon des critères ou normes pré-établis — ici, une valeur de 5 % —, alors on concluera à l'« extrêmalité » de cet échantillon ; en d'autres termes, la probabilité que celle-ci soit dûe à un échantillonnage aléatoire « assez malchanceux » sera jugée comme faible et suffisante pour conclure à la non-typicalité, ou la non-représentativité, de l'échantillon par rapport à la population parente dont il est issu.

3.3 Principes des tests d'inférence et de l'estimation statistique

3.3.1 Principe du test d'hypothèse

Démarche du test inférentiel

Comme nous l'avons mentionné en guise d'introduction à la démarche inférentielle (cf. 3.1.1), l'expérimentateur est généralement amené à se demander si les résultats observés au niveau d'un échantillon sont caractéristiques de la population dont est issu l'échantillon, ou en d'autres termes si l'estimation du paramètre d'une population à partir d'un échantillon se conforme à un modèle précis, explicité sous la forme d'une hypothèse de travail qui sera étudiée au travers d'un schéma d'inférence en tout point conforme à la logique hypothético-déductive.

Les procédures inférentielles, ou *tests statistiques*, visent à tester une hypothèse particulière, dénommée hypothèse nulle, ou H_0, portant sur les paramètres de la distribution parente, estimés à l'aide des statistiques mentionnées (moyenne, variance corrigée, etc.). On peut par exemple postuler que la moyenne parente est égale à telle valeur, c'est-à-dire que l'échantillon observé provient d'une population caractérisée par cette moyenne parente, ou bien que les moyennes parentes pour deux échantillons sont égales, i.e. en d'autres termes que les deux échantillons proviennent de la même population ou de populations de mêmes caractéristiques (notamment du point de vue de la moyenne). On formule généralement une hypothèse alternative, dénommée H_1 ou H_A, qui est la négation logique de l'hypothèse nulle, et qui constitue l'hypothèse de travail (et le fondement de l'étude). C'est cette hypothèse qui sera retenue si le test statistique approprié indique que l'on peut « rejetter » l'hypothèse nulle.

Les tests statistiques, ou *tests d'hypothèse*, visent à trancher entre ces deux hypothèses. Ils indiquent en fait la probabilité p_{obs} d'observer une statistique aussi extrême que la valeur prédite par l'hypothèse nulle lorsque celle-ci est vraie *et* que les conditions d'application du test sont vérifiées. Si cette probabilité est inférieure à une certaine valeur de référence, appelée seuil de décision α et généralement fixée à 5 % (dans le cas d'un test unilatéral, cf. infra), alors on rejettera H_0 à ce seuil. Ce seuil correspond d'une certaine manière au *risque* que l'on accepte de prendre en rejettant l'hypothèse nulle sur la base des paramètres de population estimés à partir d'un échantillon particulier. L'interprétation de p_{obs} se fait de la manière suivante : si $p_{obs} < 0.05$, et si la population parente possède la propriété formulée sous H_0, e.g. $\mu_1 = \mu_2$, alors il est peu probable d'observer un tel échantillon, du seul fait des fluctuations d'échantillonnage. Le test sera jugé significatif au seuil α, et l'hypothèse nulle sera rejetée car elle apparaît incompatible avec

les données observées. Les autres seuils classiquement admis sont les seuils-repères (bilatéraux) 0.01 et 0.001. Si p_{obs} est inférieur au premier seuil repère, i.e. $p_{obs} < 0.05$, on la comparera au deuxième seuil, puis éventuellement au troisième seuil. En revanche, on ne retient que le premier seuil, 0.05, pour déclarer un test non-significatif.

Il est également possible d'utiliser directement les valeurs de la statistique ou valeur de test, e.g. valeur de t_{obs} dans un test t de Student ou valeur de F_{obs} dans un test F de Fisher-Snedecor, et de les comparer aux valeurs critiques, ou théoriques, de la distribution correspondante, au seuil α désiré. Ces valeurs critiques peuvent être lues dans les tables de la distribution concernée ; elles s'interprètent comme les valeurs d'une fonction de répartition classique (un certain nombre de tables pour les lois usuelles sont disponibles en annexe, cf. Annexe B, p. 112). Un test sera jugé significatif si la valeur de test considérée est supérieure à la valeur critique de la distribution correspondante ; par exemple, on concluera à un test t de Student significatif si $t_{obs} > t_{\alpha,\nu}$ (cas d'un test bilatéral). C'est cette procédure qui sera privilégiée lorsque l'on effectue les calculs manuellement, tandis que les logiciels statistiques fournissent les deux valeurs (e.g. t_{obs} et p_{obs} correspondant).

Les tests mis en œuvre peuvent être uni- ou bilatéral, selon les hypothèses postulées. En effet, si l'hypothèse nulle postule toujours l'égalité entre deux paramètres (e.g. $\mu_1 = \mu_2$), ou la nullité d'un paramètre ou d'un écart entre paramètres (e.g. $\mu = \mu_0$, $\mu_1 - \mu_2 = 0$) puisque dans ce dernier cas on se ramène au cas le plus simple d'un seul échantillon, l'hypothèse alternative peut, quant à elle, être orientée ou non : on peut spécifier une hypothèse alternative non-orientée, c'est-à-dire dans laquelle on ne précise pas le sens de l'effet (e.g. $\mu \neq \mu_0$, $\mu_1 \neq \mu_2$ ou $\delta \neq 0$ avec $\delta = \mu_1 - \mu_2$[9]), *ou bien* cette hypothèse H_1 peut être orientée, auquel cas on spécifie le sens de l'effet (e.g. $\mu > \mu_0$, $\mu_1 < \mu_2$ ou $\delta < 0$). Dans les *tests bilatéraux*, H_0 sera donc acceptée si la valeur observée est proche de la valeur théorique, et elle sera rejetée dans le cas contraire, que la valeur observée soit supérieure ou inférieure à la valeur théorique : on établira alors une conclusion en faveur de H_1. Dans les *tests unilatéraux*, l'hypothèse alternative spécifie une relation d'ordre (le paramètre, e.g. $\mu_1 - \mu_2$, est plus petit *ou* plus grand qu'une valeur de référence), et H_0 ne sera rejetée qu'en accord avec cette relation d'ordre. Dans le premier cas (hypothèse non-orientée), le test mis en œuvre sera bilatéral, et p_{obs} sera comparé au seuil de décision 0.05, tandis que dans le second cas (hypothèse orientée), p_{obs} sera comparé à une valeur-repère unilatérale, i.e. 0.05/2. (cf. figure 3.5)). Ces valeurs permettent de spécifier la *zone d'acceptation* et la ou les *zone(s) de rejet*[10], comme étant les zones dans lesquelles se situe le paramètre estimé (e.g. μ ou δ) et permettant, respectivement, d'accepter ou rejeter H_0.

D'autre part, on distingue deux types d'erreur qui peuvent résulter de ce schéma d'inférence : les *erreurs de type I*, ou *risque de première espèce*, et les *erreurs de type II*, ou *risque de deuxième espèce* (cf. tableau 3.1). La première se réfère au rejet d'une hypothèse nulle qui est en réalité vraie, et elle est directement déterminée par le seuil de signification α, appelé encore *seuil de décision*, fixé lors du test ; la probabilité de rejeter H_0 lorsque celle-ci est fausse est ainsi $1 - \alpha$.

[9] l'effet moyen du facteur, quantifié par d_{obs} au niveau de l'échantillon, sera entendu ici comme l'effet parent δ au niveau de la population.

[10] appelée(s) également *région(s) critique(s)*

Le second type d'erreur concerne l'acceptation d'une hypothèse nulle qui est en réalité fausse, et celle-ci n'est généralement pas connue, mais elle demeure inversement liée à la probabilité de commettre une erreur de type I, pour un échantillon d'effectif donné. Plus petites seront les probabilités de commettre une erreur de type I et plus grandes seront les probabilités de commettre une erreur de type II. La seule manière de réduire simultanément ces deux types d'erreur est d'augmenter la taille de l'échantillon. Bien que rarement utilisée, l'erreur de type II présente un intérêt avéré dans certains domaines de recherche (cas des tests cliniques par exemple).

	H_0 vraie	H_0 fausse
Rejet de H_0	α	$1-\alpha$
Acceptation de H_0	$1-\beta$	β

Tab. 3.1: Synthèse des erreurs lors d'un test d'hypothèse (α: erreur de type I ; β: erreur de type II)

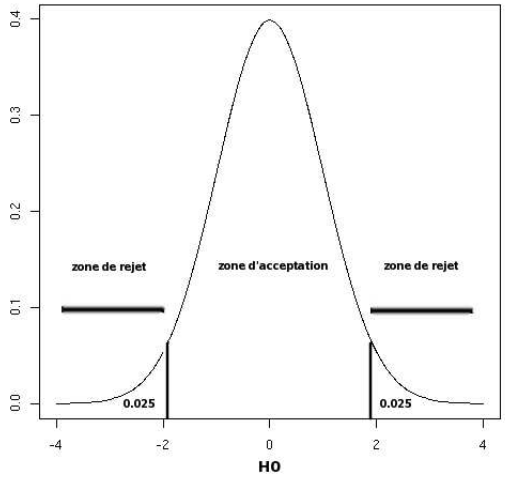

Fig. 3.5: Test bilatéral : zones d'acceptation et de rejet pour une distribution normale et un seuil de décision $\alpha = 0.05$ sous l'hypothèse nulle H_0.

Enfin, la *puissance* du test est liée au risque de deuxième espèce et s'exprime comme la valeur $1-\beta$ (elle est dépendante de la valeur α fixée). Elle indique la probabilité de rejeter correctement H_0 lorsque H_1 est vraie. Celle-ci peut être évaluée manuellement dans les cas simples, et l'utilisation d'un logiciel approprié s'avère indispensable dans les situations un peu plus complexes. Il existe des courbes de puissance des tests statistiques, appelées également *courbes caractéristiques d'efficacité*. Dans certaines situations, en fonction des conditions, différents tests peuvent être mis en œuvre pour éprouver une certaine hypothèse : on choisira alors le test qui est associé à l'erreur β la plus petite (pour un α donné), car c'est le test le plus puissant, qui sera à même de détecter les plus petites différences entre les populations sans pour autant augmenter l'erreur

de première espèce. Nous ne détaillerons pas les procédures de calcul de puissance des tests, ni de taille d'effectif ou de plus petite différence significative, lorsque nous décrirons les tests statistiques dans la partie relative aux procédures inférentielles (cf. chapitre 4), en revanche ils sont détaillés dans de nombreux ouvrages (e.g. [26]).

Un autre critère lié au choix des tests statistiques est celui de *robustesse* : la majorité des tests reposant sur un certain nombre d'hypothèses implicites, ou conditions d'application (e.g. normalité de la distribution des observations, homoscédasticité, etc.), la robustesse d'un test traduit sa tolérance à l'égard de la déviation par rapport à ces conditions d'application. La plupart du temps, les tests statistiques sont construits à partir d'hypothèses sur les distributions des variables étudiées (e.g. distribution selon la loi normale) : on parle alors de *tests paramétriques*. Lorsque la distribution de la population parente n'est pas connue, ou ne peut pas être estimée, on optera pour des *tests non-paramétriques*. Les tests paramétriques sont toujours plus puissants que les tests non-paramétriques, et doivent être employés préférentiellement lorsque les conditions d'application sont vérifiées ; cependant, lorsque celles-ci ne sont pas vérifiées (ou ne peuvent pas être vérifiées, en raison d'un effectif insuffisant par exemple), les tests non-paramétriques doivent être utilisés.

Remarques

On prendra garde au fait que ne pas pouvoir rejeter H_0 ne signifie pas que celle-ci soit vraie : elle est acceptée provisoirement, au travers d'un constat d'ignorance, car les données peuvent être compatibles avec plusieurs hypothèses différentes. Néanmoins, lorsque l'on connaît la valeur de β, celle-ci donne une idée de la plausibilité de cette hypothèse nulle.

D'autre part, on veillera à ne pas confondre l'importance de l'effet que l'on souhaite mettre en évidence (qui est plutôt caractérisée par la différence observée entre les moyennes des échantillons ou l'effet calibré, interprété en fonction de critères psychométriques, cf. 2.5.1, p. 30, et par l'intervalle de confiance associé à la moyenne, cf. § suivant) et la probabilité p_{obs} correspondant à la probabilité d'observer des échantillons aussi extrêmes que l'échantillon observé. Ce n'est pas parce que le résultat d'un test inférentiel est significatif au seuil 0.0001 que l'*effet* est plus important ! Il s'agit d'un test *d'hypothèse*, et on ne se prononce que sur l'*existence* de l'effet.

Enfin, il faut toujours interpréter le résultat d'un test et l'importance de l'effet au regard de la taille de l'échantillon. En effet, plus l'échantillon est grand, plus les estimateurs des paramètres de la population à étudier sont précis et plus le test d'hypothèse fondé sur ces estimateurs devient discriminatoire puisqu'il devient alors très improbable qu'une différence observée entre l'estimateur et la valeur hypothétique soit uniquement attribuable au hasard de l'échantillonnage. On peut, au contraire penser qu'il existe une différence réelle et donc rejeter l'hypothèse de départ, c'est-à-dire favoriser l'existence d'un effet au niveau de la population parente quand bien même l'importance de cet effet serait négligeable. Inversement, travailler avec de petits échantillons induit bien souvent des tests non significatifs, ne permettant pas de rejetter l'hypothèse nulle d'absence d'effet, même si l'effet est important au niveau de la population parente.

3.3.2 Intervalles de confiance

Comme on l'a vu précedemment (cf. 3.2.2), il est possible de construire des intervalles de confiance à x % pour observations, qui indiquent la proportion théorique des observations (x %) que l'on devrait retrouver si on échantillonnait cette population d eparamètres connus (moyenne et variance). Comme on ne travaille plus avec les observations mais avec les indicateurs d'échantillon (moyenne, variance corrigée), qui varient moins que les observations individuelles, on va également construire des intervalles de confiance à x % pour ces statistiques. L'intervalle de confiance le plus utilisé est l'intervalle de confiance à $100(1-\alpha)$, i.e. 95 %, pour la moyenne. Il permet d'indiquer l'intervalle centré sur la statistique considérée (ici, la moyenne empirique) dans lequel se situe avec une probabilité de 95 % le paramètre de la population considéré (i.e. la moyenne théorique).

De même que pour l'intervalle de confiance pour observations, il est nécessaire de connaître la variabilité de la moyenne de population (dans le cas d'observations individuelles, elle est estimée à l'aide de la variance, ou plutôt de l'écart-type) : l'écart-type de la moyenne, appelé *erreur-type*[11], traduit l'incertitude de l'estimé de la moyenne de la population. En d'autres termes, l'erreur-type peut être vue comme l'erreur que l'on commet en estimant la moyenne de la population à partir d'un échantillon unique : elle indique l'erreur d'échantillonnage d'une estimation. L'erreur-type est obtenue en divisant l'écart-type des observations par la racine carrée de l'effectif : $\frac{s}{\sqrt{n}}$. L'intervalle de confiance est obtenu par le même calcul que celui effectué pour les observations, mais en utilisant l'erreur-type à la place de l'écart-type, et la distribution du t de Student à la place de la distribution centrée-réduite[12] :

$$IC_{100(1-\alpha)\%} = \bar{x} \pm t_{\alpha/2,\nu}\frac{s}{\sqrt{n}}$$

où $\bar{x}$ désigne la moyenne de l'échantillon, s l'écart-type corrigé (cf. 2.4.2, p. 22), $\nu = n-1$ les degrés de liberté associés à la valeur critique de la distribution du t de Student au seuil unilatéral $\alpha/2$.

On pourra remarquer d'ailleurs que lorsqu'on construit un intervalle de confiance pour la moyenne, la réduction du seuil α augmente la taille de celui-ci. Dans le cas extrême $\alpha = 0$, $IC_{100(1-\alpha)\%} = \infty$, et $\beta = 1$: toutes les valeurs possibles étant incluses dans l'intervalle de confiance, il n'y aucun risque de conclure de manière erronée qu'une observation fait partie de la même population.

Par ailleurs, on notera que la taille de l'échantillon (reprise au numérateur, et dans les degrés de liberté de la valeur du t) affecte la taille des intervalles de confiance : lorsque l'effectif augmente, la moyenne et l'écart-type se rapprochent des vraies valeurs (i.e. l'estimation est meilleure), et la valeur de t tend de plus en plus vers les valeurs de Z, i.e. la distribution du t de Student tend vers celle de la loi normale centrée-réduite (qui fournit toujours des intervalles de

[11] *standard error* dans la littérature anglo-saxonne

[12] Lorsque la taille de l'échantillon est grande ($n \geq 30$), la distribution du t de Student peut être approximée par celle de la loi normale centrée-réduite

confiance plus étroits dans la mesure où les queues de distribution de la loi Z sont moins épaisses que celle du t de Student).

L'intervalle de confiance et le risque α constituent ainsi une approche complémentaire dans la démarche inférentielle. Dans le cas de l'intervalle de confiance, on se situe plus volontiers dans une approche d'*estimation*, i.e. on cherche à déterminer l'étendue à l'intérieur de laquelle devrait se situer la valeur d'un paramètre (ici la moyenne, mais cela peut concerner n'importe quelle statistique : variance, proportion, etc.). Dans le cas du test d'hypothèse, on travaille plutôt au niveau d'un *modèle* explicatif, conforme à la logique hypothético-déductive, i.e. on soumet la validité d'un modèle à une prise de décision binaire pour vérifier une assertion logique à propos de la valeur du même paramètre.

3.3.3 Conditions générale d'analyse

Les conditions d'application des procédures inférentielles dépendent du type de statistique utilisée, mais en toute généralité on peut mentionner les principales hypothèses implicites liées aux tests d'inférence classiques portant sur la comparaison de moyennes ou l'analyse de variance. Ce sont en particulier :

- l'échantillonnage aléatoire : l'échantillon considéré est constitué d'observations indépendantes et sélectionnées de manière aléatoire dans la population ;
- la normalité de la distribution : les écarts des observations à la moyenne de l'échantillon, appelés encore résidus, se distribuent selon une loi normale $\mathcal{N}(0; s_\varepsilon^2)$;
- l'homogénéité des variances (appelé également homoscédasticité) : lorsqu'il y a plusieurs échantillons, leurs variances respectives sont approximativement égales.

Les conditions d'application spécifiquement liées à certaines procédures (analyse de régression, de covariance, etc.) seront mentionnées lorsque celles-ci seront étudiées dans le chapitre suivant.

3.4 Approche intuitive de l'analyse inférentielle des protocoles expérimentaux

Dans le cadre de l'analyse de données expérimentales, on a souligné l'importance de l'analyse descriptive, et bien évidemment, celle-ci s'accompagne souvent (mais pas nécessairement !) d'une analyse à visée inférentielle. Lorsque l'on cherche à évaluer l'effet d'une variable indépendante sur les performances des sujets, mesurées au travers d'une variable dépendante, on se trouve typiquement en présence de différentes séries d'observations caractérisées par une moyenne et une variance (ou un écart-type). Le modèle implicite est généralement un modèle du type : réponse = effet fixe (ou systématique) + effet aléatoire. Les performances observées peuvent varier d'une condition à l'autre (les différentes modalités de la variable indépendante), et on cherche à évaluer si cette variation entre les conditions peut être considérée comme significative,

au regard de la variabilité qui n'est pas liée à la manipulation de la variable indépendante et qui résulte des fluctuations individuelles à l'intérieur d'une même condition. Ce schéma grossier, que l'on retrouve typiquement dans le cadre d'une étude sur l'effet de deux traitements avec des groupes indépendants, peut être généralisé à de nombreuses autres situations expérimentales et aux techniques qui leur sont dédiées, comme l'ANOVA ou la régression. Dans la démarche inférentielle, on essaie toujours de caractériser les sources de variation de la variable dépendante, en les décomposant selon le protocole expérimental considéré, comme l'illustre la figure 3.6.

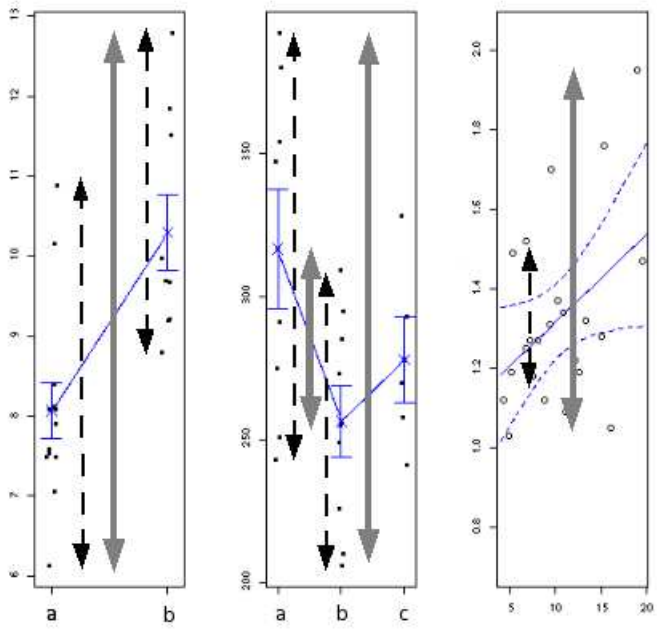

Fig. 3.6: Décomposition de la variabilité pour trois types d'analyses : comparaison de moyennes pour deux échantillons indépendants (à gauche), analyse de variance pour trois échantillons (au milieu), régression linéaire pour deux variables continues (à droite). Les sources de variabilité « intra », dûes aux fluctuations par rapport aux moyennes, sont représentées par les flèches de couleur noire, tandis que les sources de variabilité « totale » sont représentées par des flèches de couleur grise. Dans la figure du milieu, les sources de variabilité « inter » entre deux conditions ont été indiquées par une flèche de couleur grise également.

Le chapitre suivant présente un ensemble de procédures inférentielles adaptées aux différentes situations expérimentales portant sur des variables dépendantes quantitatives, que l'on retrouve classiquement en sciences humaines et en biologie, et qui sont résumées dans la figure **??**. Certaines des techniques, comme la comparaison de moyennes pour deux échantillons, l'analyse de variance à plusieurs facteurs, la corrélation et la régression linéaire, sont traîtées en détails, tandis que d'autres techniques sont simplement présentées dans leurs grandes lignes (régression multiple, MANOVA, ANCOVA), en raison de l'importance des développements qu'elles suppposent et qui dépassent le cadre de ce document. L'accent sera mis principalement sur les tests

d'hypothèse, le calcul des intervalles de confiance correspondant aux différents cas de figure, étant indiqué à la suite des calculs des valeurs de test. Les exemples d'applications proposés ne sont que pour illustrer la démarche de l'analyse, et ne seront pas toujours traités avec le niveau de détails qu'ils mériteraient.

4 Comparaisons, analyses de variance et de liaison

4.1 Comparaison des indicateurs descriptifs pour un ou deux échantillons

4.1.1 Comparaison de moyennes

Principe général

Cette section traite des analyses portant sur la comparaison entre un échantillon et une population de référence (test de conformité), ainsi qu'entre deux échantillons indépendants ou appariés. Cela concerne les situations dans lesquelles on souhaite étudier la conformité d'un ensemble d'observations à une distribution théorique, ou l'effet d'un facteur (variable qualitative) sur une variable dépendante quantitative. Par exemple, on peut vouloir comparer le score moyen observé chez un groupe de sujets lors d'une tâche d'évaluation dont on connaît le taux de réussite moyen car celle-ci a été administrée sur une grande population (cas de la comparaison à une moyenne théorique). On peut également s'intéresser à l'effet de deux traitements pharmacologiques (e.g. facteur P à 2 modalités) sur un indicateur physiologique particulier (e.g. fréquence cardiaque en Hz) en les testant chez deux groupes de sujets différents (cas de la comparaison de deux échantillons indépendants, avec un plan de type $S_{n/2} < P_2 >$). Finalement, on peut étudier les aptitudes spatiales (nombre d'items résolus) d'un groupe de n sujets à l'aide deux tâches administrées successivement (e.g. facteur T à 2 modalités) et se distinguant par la nature des items à reconnaître (cas de la comparaison de deux échantillons appariés, avec un plan de type $S_n * T_2$).

Conditions d'application

Les hypothèses implicites relatives à ce type d'analyse sont les suivantes :
- le ou les échantillons testé(s) ont été tirés au hasard ;
- le ou les échantillons suive(nt) une distribution normale ;
- les échantillons possède(nt) des variances homogènes (homoscédasticité). (*) dans le cas de deux échantillons

Les deux dernières conditions d'application — normalité et homoscédasticité — peuvent être testées à l'aide, respectivement, d'un test d'ajustement à une distribution théorique (test de Kolmogorov-Smirnov, test de Wilks-Shapiro ou test de Lilliefors, cf. Annexe A, p. 110) et d'un test d'homogénéité des variances (test de Fisher, test de Bartlett ou test de Levene, cf. 4.1.4, p.

59).

Hypothèse

Un échantillon. Dans le cas où il n'y a qu'un seul échantillon, et qu'on souhaite comparer sa moyenne à une moyenne théorique de référence μ_0, l'hypothèse nulle formulée est qu'au niveau de la population parente, la moyenne de la population dont est issu l'échantillon n'est pas différente de la moyenne théorique de référence[1] :

$$H_0 : \mu = \mu_0$$

L'hypothèse alternative non orientée sera $H_1 : \mu \neq \mu_0$; dans le cas où celle-ci est orientée, $H_1 : \mu > \mu_0$ (ou $H_1 : \mu < \mu_0$).

Le plus souvent, on prend $\mu_0 = 0$, mais on peut fixer une valeur autre que celle-ci pour la moyenne de référence.

Deux échantillons. Dans le cas d'un test orienté, i.e. unilatéral, sur deux échantillons, on teste l'hypothèse nulle qu'au niveau de la population parente, la moyenne de la population dont est issu l'échantillon 1 n'est pas différente de celle dont est issu l'échantillon 2, c'est-à-dire que les deux échantillons observés sont issus de la même population parente, ou de deux populations ayant les mêmes caractéristiques. L'hypothèse nulle H_0 à tester est ainsi : $H_0 : \mu_1 = \mu_2$, ou en d'autres termes $H_0 : \mu_1 - \mu_2 = 0$, que l'on peut également exprimer comme $H_0 : \delta = 0$, avec δ représentant la différence entre les moyennes au niveau de la population parente, ou *effet parent* (i.e. l'équivalent de d_{obs} au niveau des échantillons). L'hypothèse alternative est que la population parente dont est issu l'échantillon 1 est supérieure (resp. inférieure) à celle dont est issu l'échantillon 2 : $H_1 : \mu_1 > \mu_2$ (resp. $H_1 : \mu_1 < \mu_2$)[2].

Dans le cadre d'un test non orienté, i.e. bilatéral, on fera l'hypothèse que les moyennes parentes ne diffèrent pas (sans préciser le sens de cette différence), ce qui se traduit par l'hypothèse nulle

$$H_0 : \mu_1 = \mu_2$$

et l'hypothèse alternative $H_1 : \mu_1 \neq \mu_2$. De manière générale, on peut également reformuler l'hypothèse nulle comme $H_0 : \mu_1 - \mu_2 = 0$ ou $H_0 : \delta = 0$.

[1] De manière équivalente, cela revient à tester $H_0 : \mu - \mu_0 = 0$, ou $H_0 : \mu = 0$ dans le cas où $\mu_0 = 0$.

[2] Avec la notation $\delta = \mu_1 - \mu_2$ pour l'effet parent, on formule bien évidemment $\delta > 0$ ou $\delta < 0$.

Test d'hypothèse

Pour deux échantillons *indépendants*[3] d'effectifs n_1 et n_2, de moyennes $\bar{X}_1$ et $\bar{X}_2$ et de variances s_1^2 et s_2^2, on utilisera le *test t de Student* reposant sur le calcul suivant :

$$t_{obs} = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}$$

où $s_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{[(n_1-1)s_1^2+(n_2-1)s_2^2(n_1+n_2)]}{(n_1+n_2-2)n_1n_2}}$ représente l'erreur-type de la différence entre les deux moyennes parentes. Cette erreur-type prend en compte les écarts-type des deux échantillons pondérés par leurs effectifs respectifs ; elle peut également être calculée à l'aide de la *variance combinée*, ou commune, $s_c^2 = \frac{SC_1+SC_2}{\nu_1+\nu_2}$, comme $s_{\bar{X}_1-\bar{X}_2} = \frac{s_c^2}{n_1} + \frac{s_c^2}{n_2}$.

Dans le cas de la comparaison d'une moyenne $\bar{X}_1$ à une moyenne de référence, on choisira $\bar{X}_2 = \mu_0$, et l'estimé de la variance sera simplement $s_{\bar{X}_1}$.

Lorsque les échantillons possèdent le même effectif n, et en utilisant $d_{obs} = \bar{X}_1 - \bar{X}_2$, la formule se réduit à :

$$t_{obs} = \frac{d_{obs}}{\sqrt{\frac{s_1^2+s_2^2}{n}}}$$

Cette valeur est à comparer à la valeur critique de la distribution du t de Student avec $\nu = n_1+n_2-2$ degrés de liberté (cf. Annexe B, p. 114), au seuil $\alpha = 0.05$ (cas d'un test bilatéral). S'il s'avère que la valeur de t_{obs} est supérieure à la valeur critique $t_{0.05,\nu}$, on comparera t_{obs} à la valeur critique au seuil suivant 0.01, puis éventuellement au seuil suivant 0.001. Si le test est orienté ($H_1 : \mu_1 > \mu_2$ ou $H_1 : \mu_1 < \mu_2$), le seuil retenu sera unilatéral.

Lorsque le test est significatif, selon que l'hypothèse alternative était ou non orientée, on concluera au seuil unilatéral ou bilatéral respectivement. Dans le cas où le test est non significatif (i.e. $t_{obs} < t_{\alpha,\nu}$), on concluera dans les deux cas au seuil bilatéral[4].

Test approximatif de Welch Lorsque les variances ne sont pas homogènes, un test t corrigé — le *test approximatif de Welch* — peut être effectué en calculant la valeur de test suivante :

$$t_{obs_W} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

[3] L'indépendance signifie qu'il n'y a pas de corrélation entre les deux échantillons du point de vue de la variable mesurée.

[4] Il est à noter que les tests bilatéraux sont beaucoup plus robustes aux violations des conditions d'application du test (normalité et égalité des variances), notamment pour de grands échantillons et des plans équilibrés.

et en la comparant aux mêmes valeurs critiques de la distribution du t de Student, avec $\nu = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$ degrés de libertés.

Lorsque le nombre de degrés de liberté ν n'est pas une valeur entière, on utlisera sa valeur entière approchée par valeur inférieure. On notera que dans ce cas, l'estimé des variances parentes n'est pas un « mélange » des variances intra des deux échantillons comme dans le cas précédent (variance commune), mais les considèrent isolément.

Echantillons appariés

Lorsque les échantillons sont appariés (effectif n), la variabilité se décompose en un terme de variabilité liée au traitement, et un terme de variabilité liée aux individus, mais cette fois-ci les paires d'observations ne sont pas indépendantes. La valeur de test pour l'hypothèse nulle sera donc dérivée à partir de la différence entre chaque paire individu/traitement, et on prendra au numérateur la moyenne des différences, qui est équivalente à la différence des moyennes dans ce cas de figure[5], soit :

$$t_{obs} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_d^2}{n}}}$$

La variance estimée, s_d^2, est la variance corrigée de l'échantillon dérivé par différence, et l'erreur-type de la différence des moyennes est simplement la racine carrée du rapport de cette variance et de la taille de l'échantillon (les deux groupes étant supposés de mêmes effectifs). On se ramène en fait à la situation de comparaison d'une moyenne empirique (moyenne des différences) à une moyenne théorique nulle.

Dans le cas de groupes appariés, on apporte une information supplémentaire par rapport au cas des groupes indépendants. L'appariement sous-entend que les données ne sont pas indépendantes deux à deux, et sont donc corrélées dans une certaine mesure, ce qui permet de ne conserver qu'une source de variabilité (celle du protocole dérivé par différence), contrairement au cas des groupes indépendants. On se réduit en fait à une situation à un seul échantillon. En l'absence de cette corrélation, les deux tests sont strictement identiques. En effet, dans le cas de groupes indépendants, l'erreur-type de la différence estimée des moyennes de groupes est $\sqrt{\frac{s_1^2+s_2^2}{n}}$, lorsque les effectifs des deux groupes sont égaux ($n_1 = n_2 = n$). Dans le cas de groupes appariés, l'erreur-type de la différence estimée des moyennes de groupes repose sur la variance s_d^2 du protocole dérivé par différence, et on peut montrer que celle-ci s'exprime sous la forme $s_d^2 = s_1^2 + s_2^2 - 2\rho s_1 s_2$ (en supposant l'égalité des effectifs), ρ désignant la corrélation parente entre les deux groupes. L'erreur-type pour les groupes appariés, $\sqrt{\frac{s_1^2+s_2^2-2\rho s_1 s_2}{n}}$, sera donc inférieure à

[5] Le lecteur pourra vérifier que la moyenne des différences $\frac{\sum_i (x_i^1 - x_i^2)}{n}$ est en fait égale à la différence des moyennes $\bar{x}_1 - \bar{x}_2$. On trouve souvent la notation d_j pour représenter les différences entre valeurs appariées, et la valeur reprise au numérateur de la statitsique de test est alors $\bar{d}_j$.

celle pour les groupes indépendants, ce qui permet d'augmenter la valeur du t_{obs} dans le premier cas (on prend en compte la connaissance de la relation entre les deux groupes)[6].

4.1.2 Intervalles de confiance

Dans le cas de la comparaison d'un échantillon à une population de référence, l'intervalle de confiance à 95 % pour la moyenne μ peut être calculé à l'aide de la formule :

$$\bar{X} \pm t_{\alpha,\nu} s_{\bar{X}}$$

En présence deux échantillons indépendants de variances homogènes, l'intervalle de confiance pour μ_1 ou μ_2 sera calculé à l'aide de la variance combinée s_c^2, comme suit (ici μ_1) :

$$\bar{X}_1 \pm t_{\alpha,\nu} \sqrt{\frac{s_c^2}{n_1}}$$

Dans ce dernier cas, si on ne peut pas rejetter H_0, on conclue que les deux échantillons proviennent de la même population ou de populations de mêmes caractéristiques, ayant pour estimé de la moyenne parente la moyenne combinée $\bar{X}_c = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 = n_2}$; on peut alors calculer un intervalle de confiance pour cette moyenne parente comme suit :

$$\bar{X}_c \pm t_{\alpha,\nu} \sqrt{\frac{s_c^2}{n_1 + n_2}}$$

Enfin, pour des échantillons appariés, l'intervalle de confiance à 95 % pour la moyenne parente des différences μ_d peut être calculé à l'aide de la formule :

$$\bar{X} \pm t_{\alpha,\nu} s_d$$

Exemple d'application II

> Lors d'une expérimentation médicale, on a relevé le temps de sommeil de 10 patients sous l'effet de deux médicaments. Chaque sujet a pris successivement l'un et l'autre des deux médicaments. Ces données ont été recueillies pour tester l'hypothèse que le médicament m2 est plus efficace que le médicament m1 (Source : [25]).

Dans un premier temps, on isole les variables indépendantes (ou facteurs) de l'étude : la variable sujet S (n=10), et la variable médicament M, à 2 modalités

[6] On vérifie aisément que dans le cas $\rho = 0$, l'erreur-type est identique pour les deux types de groupes.

(m1 : premier médicament, m2 : second médicament). La variable dépendante est le temps de sommeil (en heure). Le facteur M_2 est le facteur systématique principal[7]. Chaque sujet prenant successivement les deux médicaments, on est donc en présence d'un protocole de groupes appariés. Il n'y a donc pas indépendance entre les mesures pour les deux caractères pour un individu, mais il y a bien indépendance entre les individus.

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10
m1	5.7	3.4	4.8	3.8	4.9	8.4	8.7	5.8	5	7
m2	6.9	5.8	6.1	5.1	4.9	9.4	10.5	6.6	9.6	8.4

Tab. 4.1: Temps de sommeil observés pour deux groupes appariés (n=10, individus notés i1 à i10)

L'examen des données (cf. tableau 4.1.2) indique que les valeurs observées pour les deux groupes se distribuent de manière homogène autour de leur moyenne, il ne semble pas y avoir de valeurs atypiques, et la médiane est relativement proche de la moyenne, dans les deux cas. Les étendues des deux groupes m1 et m2 sont comparables (5.3 et 5.6 respectivement). Les moyennes apparaissent donc être des indicateurs fidèles de la tendance centrale.

D'autre part, le groupe ayant reçu le premier médicament (m1) dort en moyenne moins que le groupe ayant reçu le second médicament (m2) : $\bar{x}^1 = 5.75 < \bar{x}^2 = 7.33$, soit une différence observée $d_{obs} = 1.58\,h$. La variance du groupe m2 (3.61) est légèrement supérieure à celle du groupe m1 (2.88) (cf. la distribution bivariée représentée dans la figure 4.1).

Il reste alors à généraliser à la population parente (non observée), dont sont issus ces sujets, ces conclusions descriptives. On se demande ainsi si ces deux groupes sont issus d'une même population parente ou de populations de mêmes caractéristiques ($H_0 : \mu_2 = \mu_1$), ou si ces deux groupes sont issus de deux populations possédant des caractéristiques différentes (H_1). Plus spécifiquement, on précisera l'hypothèse alternative en postulant que le second médicament (m2) est plus efficace que le premier (m1), soit $H_1 : \mu_2 > \mu_1$ (ou $H_1 : \delta < 0$ avec $\delta = \mu_1 - \mu_2$). L'hypothèse alternative est orientée, ce qui signifie que le test mis en œuvre sera également orienté, i.e. unilatéral[8].

La procédure inférentielle appropriée est le test t de Student pour des échantillons appariés, qui revient à tester la moyenne des différences entre les deux groupes contre une moyenne théorique nulle : on se réduit ainsi à une situation à un seul échantillon. On calcule donc le rapport entre la moyenne des différences observées (d_{obs})

[7] Cette variable est *provoquée* puisque c'est l'expérimentateur qui détermine les modalités, i.e. le choix des deux médicaments.

[8] On pourrait également, en plus du sens de la différence, quantifier cette différence théorique, par exemple $\mu_2 - \mu_1 = 1.5$, mais nous nous limitons volontairement à une hypothèse alternative stochastique, conformément aux objectifs de l'étude. Notons cependant que postuler une valeur pour la différence entre les moyennes théoriques influence la puissance du test.

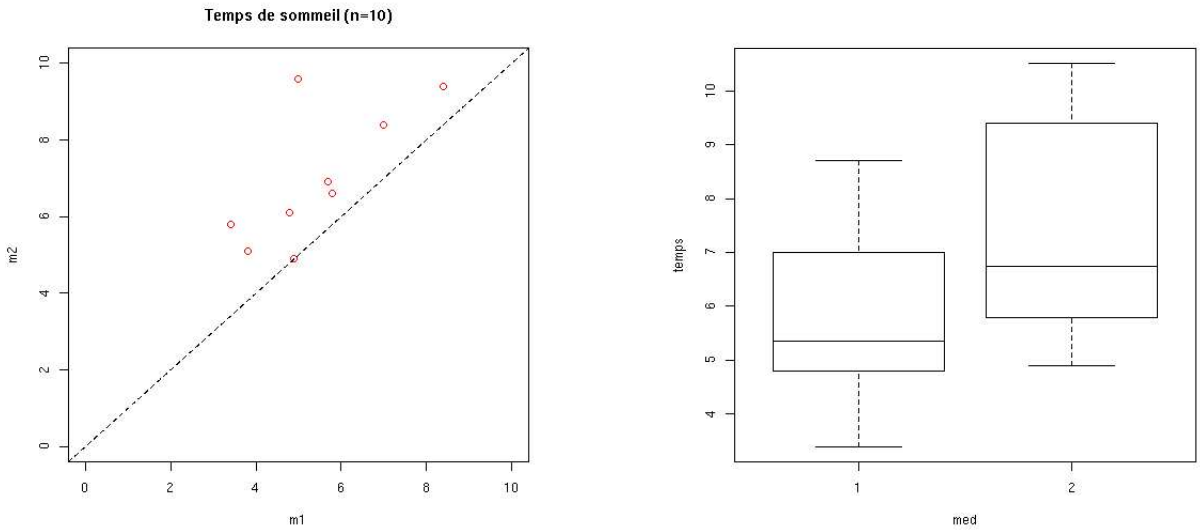

Fig. 4.1: Deux représentations graphiques « complémentaires » de la distribution des observations appariées. A droite, boite à moustaches indiquant le positionnement des 3 quartiles en fonction du médicament ; à gauche, graphe bivarié des données individuelles.

et l'estimé de la variance $\frac{s}{\sqrt{n}}$, et on le compare aux valeurs théoriques $t_{\alpha/2,\nu}$, avec $\nu = n - 1 = 9$ dl et $\alpha/2 = 0.025$.

On a

$$t_{obs} = \frac{d_{obs}}{\sqrt{\frac{s^2}{n}}} = \frac{1.58}{\sqrt{\frac{1.51}{10}}} = 4.06 > t_{0.005,9} = 3.250$$

Le test est donc significatif au seuil unilatéral 0.005, et l'on rejette l'hypothèse nulle H_0 d'une absence de différence entre les moyennes des populations parentes.

On peut donc dire, en toute généralité, que le type de médicament utilisé influence le temps de sommeil des individus de mêmes caractéristiques que les individus de l'échantillon : les individus dorment en moyenne plus longtemps avec le deuxième médicament.

Remarque. On pourrait modifier légèrement les données, en considérant que les groupes sont indépendants, c'est-à-dire que ce ne sont pas les mêmes sujets à qui l'on administre successivement les deux médicaments. On serait amené dans ce cas, pour les mêmes hypothèses H_0 et H_1, à effectuer un test t sur des échantillons indépendants. Le principe est strictement le même, excepté l'estimé de la variance qui mélange les deux variances des deux échantillons, puisque cette fois-ci on doit prendre en compte le fait que les observations sont toutes indépendantes.

La formule de calcul du t est alors le rapport entre la différence des moyennes (d_{obs}) et l'estimé de la variance parente, qui est calculé comme $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, soit $\sqrt{\frac{s_1^2+s_2^2}{n}}$ puisque les effectifs des deux groupes sont égaux. Cette valeur de t_{obs} sera à comparer

aux valeurs critiques de la distribution du t de Student au seuil 0.05/2 pour $\nu = n_1 + n_2 - 2 = 18$ dl.

On a ainsi

$$t_{obs} = \frac{d_{obs}}{\sqrt{\frac{s_1^2+s_2^2}{n}}} = \frac{1.58}{\sqrt{\frac{3.20+4.01}{10}}} = 1.86 < t_{0.025,18} = 2.101$$

Le test est donc non significatif au seuil bilatéral 0.05, et l'on ne peut pas rejetter l'hypothèse nulle H_0 d'une absence de différence entre les moyennes des populations parentes. On ne pourrait pas conclure que les populations parentes dont sont issus les deux échantillons sont différentes du point de vue de leurs caractéristiques.

On voit que le passage d'une situation appariée à une situation indépendante diminue la valeur du t, c'est-à-dire l'importance de l'effet du traitement. En effet, dans le dernier cas (échantillons indépendants), on n'a pas utilisé le fait que les mesures étaient appariées, et de ce fait, on perd de l'information (cf. 4.1.1, p. 53). De manière plus générale, cela illustre également le fait que les protocoles de mesures appariées sont, en règle générale (voir discussion pp. 61-63 dans [9]), plus puissants que les protocoles utilisant des groupes indépendants, et qu'il est préférable, lorsque cela est possible, de construire des expériences utilisant des groupes appariés.

4.1.3 Alternatives non-paramétriques

Nous avons exposé les conditions d'application du test t de Student pour deux échantillons, la plus importante d'entre elles étant l'homoscédasticité. Lorsque cette hypothèse ne peut être acceptée, il convient d'employer des procédures de test ne nécessitant aucune estimation des paramètres de variance de la population parente : ce sont les tests non-paramétriques.

Dans les situations à un ou deux *échantillons indépendants*, on utilisera le *test de Mann-Whitney*, qui est en fait un test reposant sur les rangs plutôt que sur les valeurs prises par la variable. La procédure de calcul consiste à remplacer les données par leurs rangs respectifs, sans se préoccuper du facteur de classement (facteur de groupe). Lorsque l'on travaille avec des données transformées en rangs, un cas épineux se pose en cas d'ex-æquo : on remplace alors le rang correspondant par la valeur moyenne du rang concerné et du rang suivant. Une valeur de test peut être calculée à partir de ceux-ci, et la procédure inférentielle est ensuite comparable à celles décrites précédemment : la valeur observée de la statistique de test est à comparer aux valeurs critiques de la distribution du $\mathcal{U}$ de Mann-Whitney (cf. Annexe B, p. 118).

La valeur de test est calculée comme suit :

$$U = n_1 n_2 + \frac{n_1(n1+1)}{2} - R_1$$

où n_1 et n_2 désignent les effectifs des deux échantillons, avec $n_1 < n_2$[9], et R_1 la somme des rangs

[9] Dans le cas où $n_1 > n_2$, on utilise la statistique $U' = n_1 n_2 + \frac{n_2(n2+1)}{2} - R_2$, où R_2 désigne la somme des rangs de l'échantillon 2. On notera que les deux statistiques sont liées par la relation $U' = n_1 n_2 - U$.

de l'échantillon 1. Cette valeur est à comparer aux valeurs critiques de la distribution du U avec n_1 et n_2 degrés de liberté, au seuil $\alpha = 0.05$.

Pour des *échantillons appariés*, lorsque les observations dérivées par différence ne se distribuent pas normalement, on utilisera le *test de Wilcoxon*, qui repose également sur les données transformées en rangs. La procédure de calcul est relativement simple : on range par ordre croissant les valeurs absolues des observations dérivées par différence, puis on assigne le signe de chacune des différences au rang correspondant. Le traitement des ex-æquo se fait de la même manière que pour le test de Mann-Whitney. On somme ensuite chacun des rangs des deux signes et on forme les deux statistiques de test T_+ et T_-[10]. Pour un test bilatéral, on rejetera H_0 si T_+ ou T_- est inférieure à la valeur critique $T_{\alpha,n}$ dont les tables sont disponibles dans les ouvrages spécialisés (cf. également Annexe B, p. 119, ou [26], [7]).

Exemple d'application III

> On dispose de deux échantillons d'étudiants de sexe masculin et féminin, dont on a relevé la taille. On se demande si les tailles observées peuvent être considérées comme différentes entre les deux groupes (Source : [26]).

Il apparaît que les effectifs des deux groupes sont faibles et inégaux (cf. tableau 4.1.3). On ne met pas en œuvre de procédure de test paramétrique, et on préfère utiliser une statistique de rangs, qui ne présume rien sur les distributions parentes (si ce n'est qu'elles ont à peu près la même forme et des indicateurs de dispersion à peu près comparables). L'hypothèse nulle sera $H_0 : \mu_1 = \mu_2$ (pas de différence entre les deux échantillons du point de vue de la taille), et l'hypothèse alternative, non orientée, $H_1 : \mu_1 \neq \mu_2$.

g1	193	188	185	183	180	178	170
g2	175	173	168	165	163		

Tab. 4.2: Tailles observées (en cm) pour deux groupes d'étudiants de sexe masculin (g1, $n_1 = 7$) et féminin (g2, $n_2 = 5$).

La valeur du U de Mann-Whitney est donnée par la formule :

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = (7)(5) + \frac{(7)(8)}{2} - 30 = 33 > U_{0.05,(5,7)} = 30$$

Le test est significatif au seuil bilatéral 0.05, donc on rejette H_0.

[10] Ces deux valeurs sont liées par la relation $T_+ = \frac{n(n+1)}{2} - T_-$, donc il suffit de calculer une seule des deux sommes.

4.1.4 Autres comparaisons

Comparaison de médianes

Les médianes de deux échantillons peuvent être comparées à partir de la position des valeurs des échantillons par rapport à la médiane de toutes les observations (test des médianes). Lorsque la taille de l'échantillon est grande, on pourra utiliser un test du χ^2 à $\nu = n-1$ degrés de liberté ; le cas échéant, on utilisera le test exact de Fisher. Le lecteur intéressé pourra se reporter à [26] pour le détail des procédures de test.

Comparaison de variances

Le test le plus simple pour comparer deux variances pour deux échantillons distribués *normalement*, d'effectifs n_1 et n_2 et de variances s_1^2 et s_2^2, est le *test de Hartley* qui est comparable à un *test F de Fisher-Snedecor* en prenant comme valeur de test le plus grand des deux rapports des variances. L'hypothèse nulle est l'égalité des variances parentes ($H_0 : \sigma_1^2 = \sigma_2^2$). La valeur de test, dans le cas où les effectifs n_1 et n_2 sont égaux, est donc la suivante :

$$F = max(\frac{s_1^2}{s_2^2}, \frac{s_2^2}{s_1^2})$$

et elle est à comparer aux valeurs critiques de la distribution du F avec $(\nu_1, \nu_2) = (n_1 - 1, n_2 - 1)$ degrés de liberté. Il est important de bien respecter l'ordre des degrés de liberté — ν_1 au numérateur et ν_2 au dénominateur —, en fonction de la statistique retenue, avant de consulter la valeur critique dans les tables du $\mathcal{F}$ de Fisher-Snedecor. On remarquera, après consultation de la table, que l'hypothèse d'égalité des variances ne sera rejetée que dans des cas assez « extrêmes » : par exemple, pour 2 groupes de 20 sujets, la valeur de test doit être supérieure à 2.12 (au seuil $\alpha = 0.05$) !

Lorsque le test ne permet pas de rejeter H_0, on peut alors former la variance combinée, ou variance commune, $s_c^2 = \frac{SC_1+SC_2}{\nu_1+\nu_2}$ qui est le meilleur estimateur des variances parentes (cf. § sur les échantillons indépendants, p. 52).

Il est également possible d'utiliser le *test de Bartlett*, que l'on retrouve en analyse de variance (cf. partie suivante, p. 61) puisqu'il peut être utilisé pour tester l'égalité de plus de deux variances. La valeur de test pour k échantillons est calculée comme suit :

$$B = ln(s_c^2)\left(\sum_{i=1}^{k}\nu_i\right) - \sum_{i=1}^{k}\nu_i ln(s_i^2)$$

où $\nu_i = n_i - 1$ avec n_i l'effectif du i-ème échantillon. La variance combinée s_c^2 est celle présentée dans le paragraphe précédent. Lorsqu'il y a plus de deux groupes d'observations (e.g. cas de l'ANOVA), elle s'exprime sous la forme générale : $s_c^2 = \sum_{i=1}^{k}(n_i - 1)s_i^2/(n-k)$, ou en utilisant

les sommes des écarts quadratiques, $s_c^2 = \sum_{i=1}^{k} SC_i / \sum_{i=1}^{k} \nu_i$. En normalisant cette statistique par le facteur de correction C calculé comme suit :

$$C = 1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^{k} \frac{1}{\nu_i} - \frac{1}{\sum_{i=1}^{k} \nu_i} \right)$$

la statistique $B_c = \frac{B}{C}$ peut être comparée aux valeurs critiques de la distribution du χ^2 avec $k-1$ dl.

Lorsque la condition de normalité n'est pas vérifiée, il est déconseillé d'utiliser le test F, et le test de Bartlett ne constitue pas non plus une alternative très fiable étant donné leurs sensibilités respectives aux déviations à la normalité. En particulier, si les données ne suivent pas une distribution normale mais que les variances sont égales, le test de Bartlett tend à indiquer incorrectement la présence d'hétéroscédasticité.

En dernier lieu, le *test de Levene* d'homogénéité des variances peut également être utilisé lorsque la condition de normalité n'est pas vérifiable. Ce test est légèrement moins sensible que le test de Bartlett aux déviations par rapport à la normalité[11]. Il s'apparente à un test t calculé sur les différences en valeurs absolues entre les observations dans chaque échantillon et les moyennes respectives de chacun des échantillons. La valeur de test pour k échantillons (effectif total n) est calculée à l'aide de la formule :

$$W = \frac{(n-k) \sum_{i=1}^{k} n_i (\bar{z}_i - \bar{z})^2}{(k-1) \sum_{i=1}^{k} \sum_{j=1}^{n_i} n_i (z_{ij} - \bar{z}_i)^2}$$

où les $z_{ij} = |x_{ij} - \bar{x}_i|$ représentent les écarts absolus des observations à la moyenne du groupe i, $\bar{z}_i$ indiquent les moyennes des groupes et $\bar{z}$ la moyenne générale. La valeur calculée est à comparer à la distribution du $\mathcal{F}$ de Fisher-Snedecor avec $(\nu_1, \nu_2) = (k-1, n-k)$ degrés de liberté, au seuil $\alpha = 0.05$

Comparaison sur les indicateurs de distribution et autres indicateurs descriptifs

Il est également possible de comparer les indicateurs liés à la forme de la distribution (coefficients de symétrie et d'aplatissement), ainsi que les proportions, les coefficients de variation, etc. Le lecteur intéressé pourra consulter [26].

[11] En fait, on peut montrer que ce test est à la fois plus robuste et plus puissant dans la mesure où (i) il ne détectera pas incorrectement l'hétéroscédasticité lorsque les observations s'écartent de la normalité et que les variances sont en réalité égales (robustesse), et (ii) il détectera l'hétéroscédasticité lorsque les variances sont en réalité inégales (puissance).

4.2 Analyse de variance d'ordre 1

4.2.1 Principe général

Cette partie est consacrée à l'étude des situations expérimentales dans lesquelles l'effet d'un facteur (variable qualitative) quantifié par une même variable dépendante est étudié au travers de k échantillons ($k \geq 2$) indépendants. Le plan correspondant est $S_{n/k} < G_k >$, où G désigne le facteur de classification à k modalités. On appelle ce type d'analyse l'*analyse de la variance*, ou analyse de variance (ANOVA). En cela, elle constitue une extension à la section précédente dans laquelle on s'intéressait à la comparaison de deux échantillons indépendants (cf. section 4.1, p. 50). Dans le cas particulier où $k = 2$ (2 groupes d'observations), les deux techniques d'analyse fournissent le même résultat[12]. Ce type d'analyse sera mis en œuvre lorsque, par exemple, on étudie l'effet du temps de présentation d'une cible à l'écran (facteur T à 3 niveaux, t1=20 ms, t2=80 ms, t3=150 ms) sur le temps de réaction simple pour détecter cette cible chez 15 sujets répartis en 3 groupes indépendants (le plan serait ainsi $S_5 < T_3 >$).

4.2.2 Types d'ANOVA

On distingue trois types d'ANOVA : les types I, II et III. Le type I est qualifié de *modèle à effets fixes*, les niveaux de chacun des facteurs étant déterminés délibéremment (i.e. fixés) par l'expérimentateur. C'est le cas dans la plupart des protocoles expérimentaux, et c'est celui que nous développerons dans ce chapitre. Au cours de l'analyse, on cherchera principalement à déterminer si les moyennes diffèrent entre elles dans leur globalité, et, lorsque c'est le cas, quelles sont les paires de moyennes qui sont significativement différentes. Le type II est appelé *modèle à effets aléatoires*, et dans ce type de modèle les niveaux du facteur d'étude sont déterminés de manière aléatoire. On s'intéresse alors principalement à la variabililté entre les échantillons par rapport à la variabilité à l'intérieur d'un échantillon. Enfin, le type III est un *modèle à effets fixes et aléatoires*, que l'on rencontre uniquement dans les ANOVA à plusieurs critères de classification ou facteurs (cf 4.3, p. 70).

Le choix du plan expérimental revêt ici toute son importance puisque ce sont le statut des variables et les relations qu'elles entretiennent entre elles qui guident la démarche d'analyse. Pour les analyses de variance à un seul facteur, le facteur d'étude est le plus souvent déterminé par l'expérimentateur, et on se trouve dans une situation de type I. C'est le cas lorsque l'expérimentateur choisit délibéremment les niveaux du facteur d'intérêt (par exemple, *fixer* les niveaux d'un facteur « durée de stimulation dermique » aux valeurs 50, 100 et 250 ms). A l'inverse, si l'expérimentateur choisit aléatoirement les niveaux du facteur, le modèle d'ANOVA sous-jacent est de type II.

[12] En fait, la valeur du F de Fisher-Snedecor, qui sert de valeur de test dans l'ANOVA à un seul facteur, correspond à la valeur du t de Student, utilisé dans la comparaison de deux moyennes, élevée au carré. Le seuil de significativité exact p_{obs} est strictement identique dans les deux cas.

4.2.3 Modèle général

Pour une ANOVA de type I (modèle à effets fixes), les variations de la variable dépendante Y en fonction des k modalités ou niveaux de la variable indépendante peuvent se modéliser sous la forme :

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

où y_{ij} représente la valeur de l'observation j dans la condition i, μ désigne la moyenne générale, α_i représente l'effet des k conditions, i.e. la différence entre la moyenne des valeurs de Y dans la condition i et la moyenne générale[13], et les ε_{ij} représentent les valeurs résiduelles, qui se distribuent selon une loi normale $\mathcal{N}(0; s^2_\varepsilon)$.

Dans ce modèle, on fait explicitement l'hypothèse que les différences entre les échantillons, i.e. les différences inter-groupes, sont dûes au facteur manipulé par l'expérimentateur, et que la variabilité totale autour de la moyenne générale (tous groupes confondus) se décompose en cette variabilité inter-groupes (variance inter) et la variabilité résiduelle (variance intra), liées aux fluctuations inter-individuelles dans chaque groupe et qui n'est pas expliquée par le facteur. Le modèle général postulé au niveau de la population peut ainsi se réécrire, pour l'échantillon considéré, sous la forme :

$$(y_{ij}) - \bar{y} = (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i)$$

où le membre de gauche représente la variance totale (l'écart de chacune des observations à la moyenne générale), et le membre de droite se décompose en un terme de variabilité dûe au facteur, $(\bar{y}_i - \bar{y})$ (l'écart entre les k moyennes des conditions et la moyenne générale), et un terme de variabilité résiduelle, $(y_{ij} - \bar{y}_i)$ (l'écart entre les observations intra-condition et leurs moyennes respectives).

4.2.4 Conditions d'application

Les hypothèses implicites de l'analyse de variance pour k échantillons, ou groupes, tirés au hasard dans la population sont les suivantes :

- les écarts à la moyenne dans chaque groupe, ou *résidus*, sont indépendants et se distribuent selon une loi normale ;
- la variance des résidus est homogène entre les conditions (homoscédasticité).

Les tests d'ANOVA sont relativement robustes aux déviations par rapport à la normalité, comme dans le cas précédent (test t de Student pour échantillons indépendants ou appariés), mais en revanche sont sensibles à l'hétérogénéité des variances, notamment en cas d'effectifs inégaux. Cette dernière hypothèse implicite peut être vérifiée, selon la règle d'usage, en examinant la distribution de l'ensemble des observations par condition. En cas de fortes inhomogénéité entre les groupes, l'hétéroscédasticité peut être testée à l'aide des tests d'homogénéité de variances présentés dans la partie précédente (cf. 4.1.4, p. 59), mais il faut garder présent à l'esprit que ces tests sont peu robustes lorsque les résidus s'écartent fortement de la normalité, le test le plus robuste demeurant le test de Levene.

[13] Pour les ANOVA de type II, on remplace généralement l'effet fixe α_i par l'effet aléatoire A_i.

4.2.5 Hypothèses

On teste l'hypothèse d'une absence de différence entre les k moyennes au niveau de la population parente, i.e. que les k échantillons proviennent de la même population ou de populations ayant des caractéristiques comparables. L'hypothèse nulle H_0 à tester est ainsi : $H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$, l'hypothèse alternative (toujours non orientée) étant que les échantillons sont issus de populations différentes, i.e. qu'*au moins* deux des moyennes parentes diffèrent entre elles : $H_1 : \mu_1 \neq \mu_2 \, ou \, \mu_1 \neq \mu_3 \, ou \, \mu_2 \neq \mu_3 \ldots$

4.2.6 Décomposition de la variance et test d'hypothèse

Tableau de décomposition de la variance

Comme on l'a exposé dans le chapitre consacré aux bases de l'analyse inférentielle (cf. chap. 3), lorsque l'on s'intéresse à l'effet d'un facteur à plusieurs niveaux ou modalités, on étudie en fait les sources potentielles responsables de la variabilité observée. Dans un modèle d'ANOVA à un seul facteur, la variabilité totale est ainsi décomposée en deux sources : une source liée aux différences entre les conditions (variance inter), et une autre source liée aux fluctuations aléatoires propres à chaque condition (variance intra), i.e. la part de variance qui n'est pas expliquée par le facteur, soit en résumé : $V_{totale} = V_{inter} + V_{intra}$. La variance inter est apportée par les éventuelles différences entre les moyennes des groupes d'observations (c'est la variance des moyennes de groupe), tandis que la variance intra est la moyenne des variances de chaque groupe d'observations (c'est la moyenne des variances corrigées classiques pour des échantillons).

On présente généralement cette décomposition de la variance en un tableau résumant, pour les différentes sources de variation (groupe et erreur), les sommes des carrés des écarts à la moyenne (SC) ainsi que les degrés de liberté associés à chaque SC. Les estimés des variances inter et intra, ou carré moyen (CM) associés aux groupes (ou facteur) et à l'erreur, se retrouvent en faisant le rapport des SC sur leurs degrés de liberté respectifs (cf. tableau 4.4). Dans la mesure où les variances des populations sont supposées égales (hypothèse d'homoscédasticité), on utilise ainsi le carré moyen résiduel comme estimé de la variance parente.

Du point de vue des notations adoptées, on utilisera les suivantes : X_{ij} pour la j-ème observation dans la condition i, $\bar{X}$ pour la moyenne générale (tous groupes confondus), et $\bar{X}_i$ pour la moyenne de la condition (ou groupe) i.

Test d'hypothèse

Pour k échantillons d'effectifs égaux n (condition suffisante pour un plan équilibré), on utilisera le *test F de Fisher-Snedecor*, qui est simplement le rapport entre le carré moyen des groupes

g_1	g_2	$\dots$	g_i	$\dots$	g_k	
X_{11}	X_{21}	$\dots$	X_{i1}	$\dots$	X_{k1}	
X_{12}	X_{22}	$\dots$	X_{i2}	$\dots$	X_{k2}	
X_{13}	X_{23}	$\dots$	X_{i3}	$\dots$	X_{k3}	
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	
X_{1j}	X_{2j}	$\dots$	X_{ij}	$\dots$	X_{kj}	
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	
X_{1n_1}	X_{2n_2}	$\dots$	X_{in_i}	$\dots$	X_{kn_i}	
X_1	X_2	$\dots$	X_i	$\dots$	X_k	X

Tab. 4.3: Distribution des observations dans un protocole à un seul facteur.

Variance	SC	dl	CM	F
Totale	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X})^2$	n-1	CMt=SCt/dl	
Groupes	$\sum_{i=1}^{k} n_i(\bar{X}_i-\bar{X})^2$	k-1	CMg=SCg/dl	CMg/CMe
Erreur	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2$	n-k	CMe=SCe/dl	

Tab. 4.4: Tableau de décomposition de la variance en fonction des sources de variabilités (SC représente la somme des carrés, dl le nombre de degrés de libertés, CM le carré moyen et F la valeur du F de Fisher-Snedecor)

(CMg) et le carré moyen de l'erreur (CMe) :

$$F_{obs} = \frac{CMg}{CMe}$$

qui est à comparer aux valeurs critiques de la distribution du $\mathcal{F}$ de Fisher-Snedecor au seuil $\alpha = 0.05$ avec $\nu_1 = k-1$ et $\nu_2 = n-k$ degrés de libertés (cf. Annexe B, pp. 116-117). Le test étant non-orienté, le résultat du test est établi sur la base d'un seuil bilatéral.

4.2.7 Comparaisons multiples

Lorsque le résultat de l'ANOVA indique que l'hypothèse nulle peut être rejetée, c'est-à-dire que les moyennes des groupes ne sont pas *toutes* les mêmes, il est nécessaire de savoir lesquelles diffèrent entre elles. Il serait tentant d'effectuer de simples tests t de Student sur les moyennes prises deux à deux, mais cela s'avère une approche invalide. En effet, bien que chaque comparaison puisse être effectuée au seuil α désiré, la probabilité totale de commettre une erreur de type I (rejeter l'hypothèse nulle lorsque celle-ci est en réalité vraie) parmi l'ensemble des comparaisons est proportionnelle au nombre de comparaisons effectuées. On peut montrer que cette probabililté est de $1-(1-\alpha)^m$, où $m = \frac{k(k-1)}{2}$ désigne le nombre de comparaisons à effectuer, en considérant k moyennes de groupe. A titre illustratif, lorsque l'on souhaite comparer 5 groupes d'observations entre eux au seuil 0.05, la probabilité de commettre une erreur de type I sur l'ensemble des comparaisons de moyenne est de $1-(1-0.05)^5 = 1-0.95^5 = 0.40$, ce

qui constitue une probabilité guère acceptable. Cela signifie que si les 5 moyennes comparées sont égales (hypothèse nulle), on détectera au moins une paire de moyennes significativement différente dans 40 % des cas !

Une approche plus appropriée consiste à utiliser les différentes techniques de *comparaisons multiples*, chacune ayant leurs avantages et leurs inconvénients, notamment en ce qui concerne le contrôle des erreurs de type I. On regroupe sous cette appellation les comparaisons dites *planifiées* et les comparaisons *non-planifiées*, ou *post-hoc*. Les premières sont réservées à l'analyse des protocoles dont les conditions expérimentales ont été spécifiquement désignées pour tester des différences entre conditions (en référence à une théorie ou des observations préliminaires). Mais la plupart du temps, les comparaisons interviennent à la suite de l'observation d'une différence entre les conditions, telle que l'indique une ANOVA. Ce sont des comparaisons *a posteriori*, puisqu'elles dépendent des résultats de l'ANOVA préalable. Nous nous intéresserons à ce deuxième type de comparaisons, dans la mesure où ce sont celles qui sont le plus souvent mises en œuvre.

Comme on l'a dit précedemment, la probabilité de commettre une erreur de type I augmente avec le nombre de comparaisons à effectuer. Par conséquent, il est nécessaire soit de réduire le seuil de décision α, soit d'utiliser une version modifiée de la statistique du t de Student. Parmi l'ensemble des tests de comparaisons multiples existant, nous ne présenterons que trois méthodes, par souci de clarté. Pour de plus amples informations sur les comparaisons multiples (planifiées ou non), le lecteur intéressé peut se référer à [**?**] ou [26] par exemple.

Enfin, il est inutile de mettre en œuvre ce type de tests lorsque le résultat de l'ANOVA indique qu'on ne peut pas rejetter H_0, car cette dernière est la plus puissante des méthodes de détection de différences de moyenne. Un test libéral (e.g. test de Student-Newman-Keuls) pourrait en effet détecter une paire de moyennes significativement différentes alors que l'ANOVA indique l'égalité des variances, mais ce serait commettre une erreur grossière que d'outrepasser le modèle initial stipulant l'hypothèse nulle.

Test de Bonferroni

La *méthode de Bonferroni* consiste à ajuster le seuil α pour chaque comparaison de paires de moyennes de sorte que le critère de décision concernant l'égalité de toutes les moyennes soit égal au seuil α désiré (typiquement $\alpha = 0.05$). S'il y a k comparaisons à effectuer à l'aide d'un test t, on choisira ainsi un seuil $\alpha' = \frac{\alpha}{k}$. L'avantage de cette méthode est d'être « adaptable » au nombre de comparaisons total à effectuer : si parmi 6 comparaisons (cas de 4 moyennes de groupe) il y en a 2 qui n'intéressent pas l'expérimentateur, le seuil α' peut être ajusté en conséquence (i.e. $\frac{\alpha}{4}$ au lieu de $\frac{\alpha}{6}$).

Il est à noter cependant que cette méthode est *conservatrice* (les valeurs de p obtenus lors du test t sont légèrement trop importantes), et donc ce test est moins puissant et peut ne pas détecter de petites différences significatives.

Test de Scheffé

La *méthode de Scheffé* utilise une version modifiée du t de Student : la valeur de test est calculée comme celle du t, mais en prenant la valeur absolue de la différence des moyennes (au numérateur), et en comparant le résultat à une valeur critique pondérée de la distribution du $\mathcal{F}$ de Fisher-Snedecor $\sqrt{(k-1)F_{\alpha,(k-1,n-k)}}$ ($k-1$ et $n-k$ désignent les degrés de liberté associés au carré moyen des groupes et à l'erreur résiduelle).

L'avantage de cette méthode est d'être relativement *consistante* avec l'ANOVA, et ce test ne détectera pas de différences significatives si le résultat de l'ANOVA est non significatif.

Test de Tukey-HSD

La *méthode de Tukey-HSD* reprend une statistique similaire à celle du t, mais celle-ci est comparée à une table modifiée des valeurs critiques (« studentized range distribution »), qui est dépendante du nombre de degrés de liberté associés à l'erreur résiduelle ainsi que du nombre de comparaisons à effectuer.

4.2.8 Intervalles de confiance

Les intervalles de confiance associés aux moyennes des k groupes sont calculés comme de simples intervalles de confiance associés à une moyenne. Lorsqu'une moyenne a été déclarée comme significativement différente des autres, la formule suivante donne l'intervalle de confiance à 95 % associé à cette moyenne :

$$\bar{x_i} \pm t_{\alpha/2,n_i-1}\sqrt{\frac{CM_e}{n_i}}$$

où $CM_e = s^2$ représente la variance résiduelle, et n_i l'effectif du i-ème groupe.

Exemple d'application IV

> Lors d'une expérimentation pédagogique, on désire comparer l'efficacité de quatre méthodes d'enseignement. On dispose des notes (sur 25 points) obtenues à un examen par quatre groupes d'élèves ayant chacun reçu un des 4 types d'enseignement a, b, c ou d (Source : données fictives).

Analyse de l'effet principal. L'examen des données (plan considéré : $S < P_4 >$), résumées dans le tableau 4.5, révèle que les groupes c et d présentent des moyennes

a	b	c	d
10	14	18	16
17	17	16	20
17	11	14	22
13	14	14	14
17	13	19	18
14	13	16	18
13	13	16	15
17	15	17	16
...	...	...	...
14.88	**14**	**16.54**	**16.28**
2.09	*1.78*	*2.15*	*2.74*

Tab. 4.5: Notes des élèves selon le type de pédagogie (tableau partiel). Les moyennes et écarts-type associés sont indiqués dans les deux dernières lignes du tableau (en gras et en italique, respectivement).

supérieures aux 2 premiers groupes, bien que celles-ci soient également associées aux écarts-type les plus élevés (2.15 et 2.74). On a donc bien des variations entre les groupes, mais il y a également des fluctuations inter-individuelles à l'intérieur de chacun des groupes. On décompose les différentes sources de variabilité dans ce protocole de la manière suivante : (i) la variabilité inter-groupe, liée à l'effet du facteur systématique (type de pédagogie), et (ii) la variabilité résiduelle, liée à la variance intra-groupe (variance inter-individuelle à l'intérieur de chaque groupe). La variance résiduelle est en fait la variance qui n'est pas expliquée par le facteur. S'il y a un effet global du facteur « pédagogie » sur la variable dépendante (notes), alors la variabilité entre les groupes devrait être supérieure à la variabilité moyenne à l'intérieur des groupes.

Source de Variabilité	SC	dl	CM	F
Pédagogie	95.47	3	31.82	6.64***
Erreur	417.22	87	4.80	

Tab. 4.6: Tableau d'ANOVA : variance liée au facteur « pédagogie » et variance résiduelle. [Le seuil p de significativité est indiqué suivant la convention : (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.]

Les valeurs calculées sont résumées dans le tableau d'ANOVA (cf. tableau 4.6), et la valeur du F s'avère significative. On rejette donc l'hypothèse nulle d'absence de différences entre les moyennes, et on peut conclure, en toute généralité, que les moyennes parentes diffèrent (i.e. les échantillons observés ne proviennent pas tous de la même population).

Lors de l'inspection de la figure 4.2, on pourrait penser que les groupes a et d,

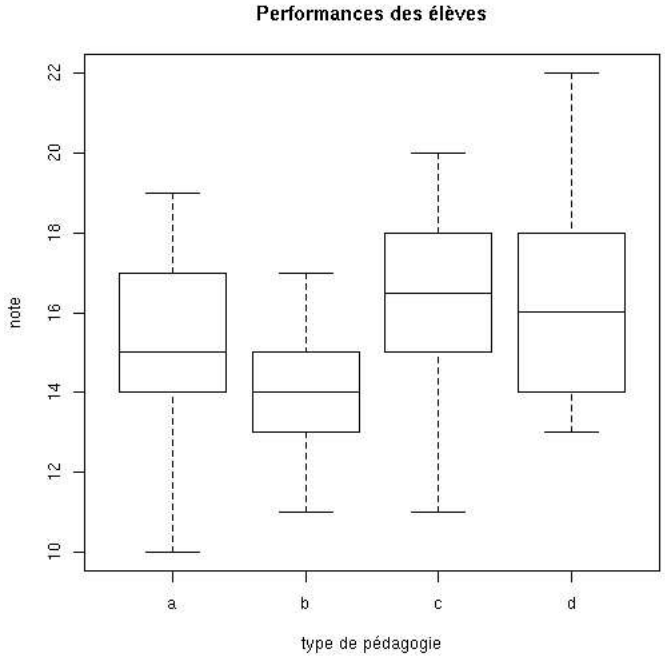

Fig. 4.2: Distribution des notes en fonction du type de pédagogie (a, b, c, d). Les boîtes à moustaches indiquent le positionnement des trois quartiles.

ainsi que les groupes c et d ne diffèrent pas entre eux étant donné leurs relatives homogénéités ainsi que leurs moyennes et variances respectives (cf. tableau 4.5); en revanche, on peut se demander si les groupes b et c diffèrent entre eux du point de vue de leurs moyennes. Le test d'ANOVA ne permet pas d'indiquer quelles sont précisemment les moyennes qui diffèrent significativement, et pour cela on utilisera un test de comparaisons multiples (non planifiées).

Comparaisons multiples. Pour comparer les paires de moyennes, nous utiliserons le test de Tukey-HSD, qui utilise une version modifiée du t de Student (cf. 4.2.7). Il consiste en fait à tester chaque paire de moyennes, comme dans le cas d'échantillons indépendants de variances égales. Suivant les objectifs de la recherche (modèle/estimation), on peut indiquer uniquement les seuils p associés à chaque test d'hypothèse (e.g. $H_0 : \mu_a = \mu_b$, ou en d'autres termes $H_0 : \mu_a - \mu_b = 0$, c'est-à-dire $H_0 : \delta = 0$), ou bien les intervalles de confiance à 95 %[14] (cf. figure 4.3).

Les résultats indiquent que seules les paires c-a, c-b et d-b sont significativement différentes du point de vue de leur moyenne, au seuil $\alpha = 0.05$ (les intervalles de confiance respectifs ne comprennent pas la valeur $\delta = 0$). Cela permet d'affiner la conclusion inférentielle générale énoncée à l'issue du test d'ANOVA, et préciser pour quel type de pédagogie on observe des différences significatives au niveau des notes

[14] Le calcul des IC pour les comparaisons multiples peut être trouvé dans [26].

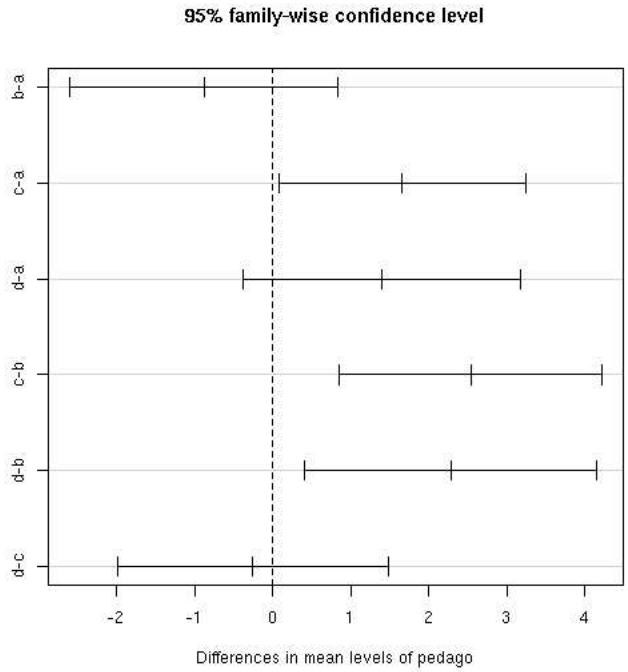

Fig. 4.3: Intervalles de confiance à 95 % des différences de moyenne pour chaque paire de groupe (Méthode de Tukey-HSD).

moyennes.

4.2.9 Alternatives non-paramétriques

Lorsque les conditions d'application de l'ANOVA ne sont pas vérifiées, on peut utiliser le *test de Kruskall-Wallis*. C'est en fait une ANOVA basée sur les rangs des observations, dont la valeur de test H est calculée comme suit :

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i}{n_i} - 3(N+1)$$

où N est le nombre total d'observations, k le nombre de groupes, n_i l'effectif partiel du groupe i, et R_i la somme des rangs pour les observations dans le groupe i.

Comme dans toute statistique reposant sur les observations transformées en rangs, la présence d'ex-æquo devient problématique, surtout lorsqu'ils sont en grand nombre (ce qui est évidemment plus fréquent lorsque l'échantillon est grand). On ajustera dans ce cas la statistique de test H en la divisant par :

$$C = 1 - \frac{\sum_{i=1}^{m} t_i^3 - t_i}{N^3 - N}$$

où m est le nombre de groupes comprenant des ex-æquo et t_i le nombre d'ex-æquo dans le groupe i.

Des tables spécifiques sont disponibles pour cette statistique de test (cf. B, p. 121, ou [26], [7]), mais lorsque les effectifs sont grands ou s'il y a plus de 6 conditions, la statistique de test H tend vers la loi du χ^2 à $k-1$ degrés de liberté.

4.3 Analyse de variance d'ordre n

4.3.1 Principe général

Cette section est consacrée à l'étude des situations expérimentales dans lesquelles l'effet de plusieurs facteurs (variables qualitatives) est étudié simultanément, c'est-à-dire dans le même protocole expérimental. En cela, elle constitue une extension à la situation précédente dans laquelle on n'étudiait qu'un seul facteur à la fois (cf. 4.2). Si l'on s'intéresse, par exemple, à l'effet du degré d'abstraction des items de test et du type de consigne sur les performances des sujets dans une tâche de raisonnement, on pourrait réaliser deux expériences dans lesquelles on manipulerait chacun des deux facteurs, et analyser les résultats à l'aide d'ANOVA d'ordre 1 s'il y a plus de 2 modalités ou niveaux pour chaque facteur : on saurait ainsi si le degré d'abstraction affecte sensiblement les performances mesurées, et également si le type de consigne affecte les performances. Mais, on ne saurait pas si l'effet du type de consigne est le même quelque soit le degré d'abstraction des items de test ; en d'autres termes, on perd l'information concernant l'interaction entre ces deux facteurs.

Les modèles d'ANOVA d'ordre n sont donc comparables sur le fond au modèle précédent de l'ANOVA d'ordre 1, mais incluent en plus de l'étude des effets principaux, celle du ou des effet(s) d'interaction.

Les hypothèses implicites de l'ANOVA sont les mêmes (cf 4.2.5, p. 63).

4.3.2 Plans factoriel et hiérarchique

Lorsque l'on manipule différents facteurs au cours d'un même protocole expérimental, le choix d'un plan d'expérience s'avère déterminant, comme nous l'avons déjà dit dans le premier chapitre (cf. 1.5, p. 11). Les observations peuvent être arrangées selon des relations de croisement ou d'emboîtement, mais il convient de bien distinguer la manière dont les niveaux d'un facteur varient en fonction des niveaux des autres facteurs. Dans le cas de deux facteurs, lorsque les

niveaux du premier facteur sont les mêmes pour tous les niveaux du second facteur, on parle de *plan factoriel*. Lorsque ce n'est pas le cas, il s'agit d'un *plan hiérarchique* dans lequel on a introduit un sous-groupe défini par un facteur de groupement : dans ce cas, il y a toujours un facteur aléatoire (la source de variabilité liée au sous-groupe), et l'ANOVA est forcément de type II (effets aléatoires) ou III (effets mixtes).

D'autre part, il peut y avoir plusieurs observations pour chaque combinaison de facteurs (*plan avec réplication*[15]), ou il peut n'y en avoir qu'une seule (*plan sans réplication*). Les observations pour chaque combinaison de facteurs peuvent également être effectuées sur le même individu, auquel cas on parle d'un *plan à mesures répétées*.

Enfin, le plan peut être *équilibré* (e.g. même nombre d'observations pour chaque combinaison de facteurs[16]) ou *déséquilibré*. Lorsque le plan n'est pas équilibré, l'ANOVA se complique sensiblement car les différentes SC ne sont plus indépendantes les unes des autres, et l'effet des facteurs n'est par conséquent plus additif[17]. Les procédures de calcul manuel deviennent « monstrueuses » et l'emploi d'un logiciel statistique est vivement recommandé (cf. à ce sujet l'article de [15]).

A chaque situation correspond un type d'analyse de variance spécifique. Nous ne traiterons pas ici des analyses sur les plans hiérarchiques, et nous nous limiterons aux analyses menées sur des plans factoriels à 2 facteurs, équilibrés, avec ou sans réplication, ainsi que sur des plans à mesures répétées. Les plans correspondants déclinés dans cette partie sont ainsi $S_{n/2k} < A_k * B_k >$ et $S_n * A_k$, respectivement.

4.3.3 Effets principaux et interaction d'ordre n

A la différence de l'ANOVA d'ordre 1 dans laquelle on analyse l'effet du seul facteur présent dans le protocole (et les différences entre les niveaux à l'aide des comparaisons multiples), on considère ici l'effet de chaque facteur, qualifié d'*effet principal*, et l'*effet d'interaction* entre les facteurs. Il y a donc autant d'effets principaux qu'il y a de facteurs, et les effets d'interaction, ainsi que leur ordre, sont fonction du nombre de facteurs présents dans l'analyse. Lorsqu'on travaille avec deux facteurs A et B, on n'a qu'un seul terme d'interaction AxB : il s'agit d'une interaction de premier ordre. Lorsqu'on a 3 facteurs A, B et C, il y a trois interactions de premier ordre (AxB, AxC, BxC) et une interaction d'ordre 2 (AxBxC). L'effet d'interaction traduit simplement le fait que l'effet d'un facteur dépend de l'autre facteur, c'est-à-dire que l'effet d'un facteur n'est pas le même selon les niveaux de l'autre facteur (dans le cas où il y a deux facteurs).

[15] On utilise également le terme de « répétition », mais nous utiliserons le terme « réplication » pour éviter la confusion avec les mesures répétées

[16] Comme nous l'avons déjà mentionné, la notion de plan équilibré n'est pas simplement lié au fait qu'il y aît le même nombre d'observations pour chaque combinaison de facteurs ; un plan équilibré peut être obtenu lorsque les effectifs marginaux sont proportionnels selon les niveaux des facteurs considérés, mais nous ne traiterons pas de cette alternative dans ce document. Le lecteur intéressé peut consulter [26], pp. 245-246, et [9], pp. 136-137.

[17] ce qui signifie également que l'ordre d'étude des facteurs influence la valeur du F calculée

Le nombre d'hypothèses testées étant directement liées au nombre de facteurs susceptibles de moduler les performances observées, il y aura ainsi autant d'hypothèses nulles à formuler qu'il y a d'effets principaux et d'effets d'interaction. Pour deux facteurs A et B à 2 niveaux chacun (a1 et a2, b1 et b2, respectivement), inclus dans un plan de type $S_{n/4} < A_2 * B_2 >$, on aura donc une hypothèse nulle concernant :

1. l'effet du facteur A ($H_0 : \mu_{a1} = \mu_{a2}$; $H_1 : \mu_{a1} \neq \mu_{a2}$)
2. l'effet du facteur B ($H_0 : \mu_{b1} = \mu_{b2}$; $H_0 : \mu_{b1} \neq \mu_{b2}$)
3. l'effet d'interaction AxB ($H_0 : pas\ d'interaction$[18] ; $H_1 : interaction$)

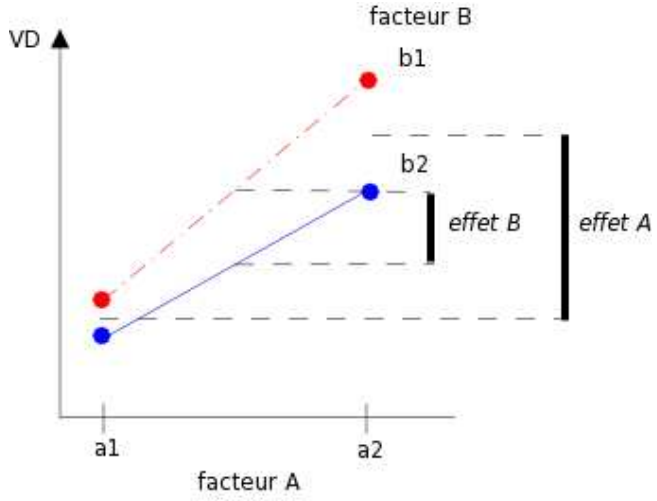

Fig. 4.4: *Graphique d'interaction.* Données moyennées dans un plan factoriel classique pour deux facteurs A et B à deux niveaux chacun ($S_{n/4} < A_2 * B_2 >$), illustrant un faible effet de B, un effet plus marqué de A et une faible interaction entre les deux facteurs.

		A_2	
		a1	*a2*
B_2	*b1*	a1b1	a2b1
	b2	a1b2	a2b2

Tab. 4.7: Données moyennées dans un plan factoriel classique pour deux facteurs A et B à deux niveaux chacun (a1 et a2, b1 et b2).

Dans ce cas précis (2 facteurs), l'effet principal du facteur A est la différence moyenne entre les deux niveaux a1 et a2, indépendemment du facteur B. L'effet principal du facteur B s'exprime de la même manière (cf. figure 4.4). En schématisant les données moyennes comme dans le tableau 4.7, cela se traduit par :

- effet A = $\frac{a2b1+a2b2}{2} - \frac{a1b1+a1b2}{2}$
- effet B = $\frac{a1b2+a2b2}{2} - \frac{a1b1+a2b1}{2}$

[18] on peut la formaliser sous la forme : $\mu_{a1b1} - \mu_{a2b1} = \mu_{a1b2} - \mu_{a2b2}$

L'absence d'effet d'interaction se traduit par le fait que les différences entre les niveaux a1 et a2 ne sont pas significativement différentes entre les niveaux b1 et b2. Si l'on représente graphiquement les observations moyennées sur un graphique (la variable dépendante en ordonnées et un des facteurs en abscisse), l'absence d'interaction se traduit par des profils de variation parallèles lorsqu'ils sont considérés par rapport au second facteur (cf. figure 4.5). Le cas échéant, l'interaction peut être (cf. figure 4.5) :

- *croisée*, l'effet du facteur A s'inverse en fonction des niveaux du facteur B, e.g. a1b1 < a2b1, et a1b2 > a2b2 ;
- *ordonnée*, l'effet du facteur A est consistant quelque soit le niveau du facteur B, e.g. a1 > a2 pour les niveaux b1 et b2.

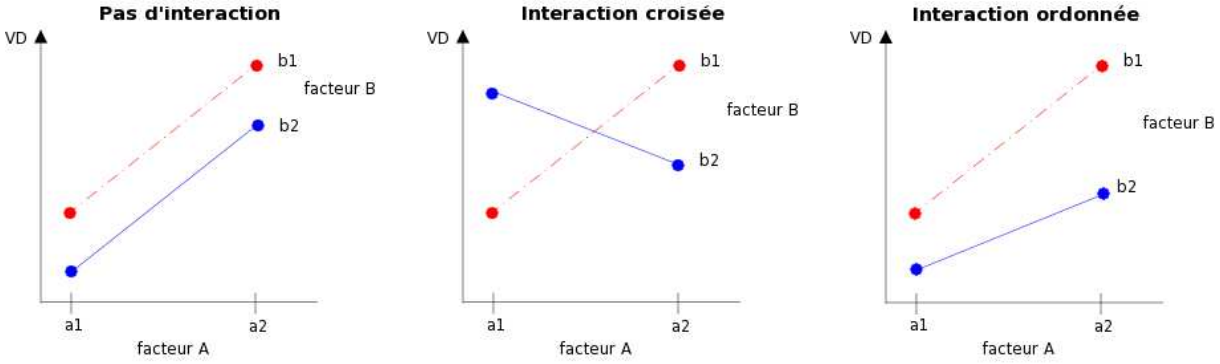

Fig. 4.5: Différents types d'effet d'interaction : à gauche, absence d'interaction ; au milieu, interaction croisée ; à droite, interaction ordonnée. (Explication dans le texte)

Il convient de prendre garde à l'interprétation des interactions d'ordre 3 ou plus, puisque celles-ci peuvent s'avérer parfois très complexes à interpréter, voire arte-factuelles (cf. pour une discussion, [9], pp. 134-136).

4.3.4 Plan factoriel avec réplication

Décomposition de la variance

Lorsque le protocole expérimental comporte des réplications, c'est-à-dire qu'il y a plusieurs observations pour chaque combinaison de chaque niveau des différents facteurs (on parle de *traitement* pour désigner une condition résultant du croisement de plusieurs facteurs), on décompose la variabilité totale en ses différentes sources (SC) liées à l'effet de chacun des facteurs, ainsi que le ou les effets d'interaction et l'erreur résiduelle liée aux fluctuations aléatoires (cf. tableau 4.9).

La *SC totale* est calculée comme dans le cas de l'ANOVA d'ordre 1 : c'est la somme des écarts quadratiques de chacune des observations par rapport à la moyenne générale. La *SC liée à l'erreur résiduelle* (équivalente à la variance intra dans l'ANOVA d'ordre 1) est la somme des écarts de chacune des observations par rapport à leur moyenne respective (e.g. toutes les

observations individuelles dans la condition a1b1 par rapport à la moyenne a1b1). La *SC globale liée aux facteurs* (équivalente à la variance inter dans l'ANOVA d'ordre 1) est calculée comme la somme des écarts quadratiques de chaque combinaison de niveaux des facteurs par rapport à la moyenne générale. Elle se décompose en différentes sources, liées aux effets principaux des facteurs ainsi qu'à l'effet d'interaction. Les *SC liées aux facteurs A et B* sont calculées en les considérant isolément l'un de l'autre, comme on le fait en ANOVA d'ordre 1[19]. Quant à la *SC liée à l'interaction*, c'est simplement ce qu'il reste de variation non expliquée par les deux facteurs : il suffit donc de retrancher les SC des deux facteurs à la SC globale liée aux facteurs.

a_1			a_2			...	a_i				...
b_1	b_2	...	b_1	b_2	...		b_1	b_2	...	b_j	...
X_{111}	X_{121}	...	X_{211}	X_{221}	...		...	...		X_{ij1}	...
X_{112}	X_{122}	...	X_{212}	X_{222}	...		...	...		X_{ij2}	...
X_{113}	X_{123}	...	X_{213}	X_{223}	...		...	...		X_{ij3}	...
...	...	...	...	...	...		...	...		...	...
X_{11k}	X_{12k}	...	X_{21k}	X_{22k}	...		...	...		X_{ijk}	...
...	...	...	...	...	...		...	...		...	...
X_{11n}	X_{12n}	...	X_{21n}	X_{22n}	...		...	...		X_{ijn}	...

Tab. 4.8: Distribution schématique des observations dans un plan factoriel avec réplication $S < A * B >$.

Variance	SC	dl	CM	F
Totale	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(X_{ijk}-\bar{X})^2$	$nab-1$	CMt=SCt/dl	
Facteur A	$nb\sum_{i=1}^{a}(X_i-\bar{X})^2$	$a-1$	CMa=SCa/dl	CMa/CMe
Facteur B	$na\sum_{j=1}^{b}(\bar{X}_j-\bar{X})^2$	$b-1$	CMb=SCb/dl	CMb/CMe
Inter. AxB	$n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{X}_{ij}-\bar{X}_i-\bar{X}_j+\bar{X})^2$	$(a-1)(b-1)$	CMi=SCi/dl	I
Erreur	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(X_{ijk}-\bar{X}_{ij})^2$	$ab(n-1)$	CMe=SC/dl	

Tab. 4.9: Tableau de décomposition de la variance en fonction des sources de variabilité, pour un plan factoriel à 2 facteurs avec réplication (a et b représentent le nombre de modalités ou niveaux des facteurs A et B, n l'effectif pour chaque combinaison des facteurs (i.e. nombre de réplications pour un plan équilibré), SC représente la somme des carrés, dl le nombre de degrés de libertés, CM le carré moyen et F la valeur du F de Fisher-Snedecor).

Le partitionnement des degrés de libertés associés à ces SC pour le cas de 2 facteurs est indiqué dans le tableau 4.9. Les notations adoptées sont similaires à celles déjà présentées dans le cas de l'ANOVA à un seul facteur ; les indices i, j et k désignent, respectivement, les niveaux des 2 facteurs, et le numéro de réplication. Dans le cas où les 2 facteurs possèdent a et b niveaux, il y a $a-1$ et $b-1$ dl associés à chacune des SC de ces facteurs. Les dl associés à l'interaction valent $(a-1)(b-1)$. Les dl associés à la SC totale valent $N-1$, où N désignent l'effectif total, tandis que les dl associés à l'erreur résiduelle sont ceux qui restent lorsque l'on a ôté aux dl de la SC totale les dl liés aux facteurs et à l'interaction.

[19] En fait, on décompose la variance inter en ses différentes composantes, puisqu'ici il y a plusieurs facteurs qui entrent en jeu.

Les carrés moyens (CM) sont formés à partir des SC de chacun des facteurs, y compris l'interaction, et de l'erreur résiduelle.

Hypothèses

Les hypothèses implicites de l'ANOVA d'ordre n sont les mêmes que celles de l'ANOVA d'ordre 1.

Test d'hypothèse

Pour les tests d'inférence concernant les différentes hypothèses nulles formulées, la valeur de test sera calculée, comme dans l'ANOVA d'ordre 1, à partir du rapport des CM liés aux facteurs et à l'interaction sur le CM de l'erreur. Ainsi, dans le cas de deux facteurs, on aura les trois statistiques de test suivantes :

- $H_0 : \mu_{a1} = \mu_{a2}$, $F_{obs(A)} = \frac{CM_A}{CMe}$
- $H_0 : \mu_{b1} = \mu_{b2}$, $F_{obs(B)} = \frac{CM_B}{CMe}$
- $H_0 : absence\ d'interaction$, $F_{obs(AxB)} = \frac{CM_{A\times B}}{CMe}$

Les valeurs obtenues sont à comparer aux valeurs critique de la distribution du $\mathcal{F}$ de Fisher-Snedecor, avec les degrés de liberté correspondant et au seuil α désiré (test bilatéral).

Remarque. Ces tests d'hypothèses concernant l'ensemble des sources de variabilité ne sont valables que dans le cas d'un modèle de type I, dans lequel tous les effets sont fixes. Dans le cas où l'un des facteurs est aléatoire (modèle de type III) ou lorsque tous les facteurs sont aléatoires (modèle de type II), le calcul est un peu plus complexe. Il faut dans un premier temps tester si l'interaction est significative comme dans le cas du modèle I (comparaison du carré moyen de l'interaction sur le carré moyen résiduel). Si l'interaction est significative, on testera l'effet de chacun des facteurs, en comparant le carré moyen associé à chaque facteur au carré moyen de l'interaction. Le cas échéant, lorsque l'interaction n'est pas significative, il devient difficile d'obtenir un bon estimé de la variance résiduelle, et différentes « règles » ont été proposées, mais nous laissons le soin au lecteur de consulter les ouvrages appropriés.

Comparaisons multiples et intervalles de confiance

Lorsque les tests sur les effets principaux et sur l'effet d'interaction s'avèrent significatifs, on procèdera à la comparaison paire par paire des moyennes, avec les mêmes techniques de comparaisons multiples que celles décrites précédemment dans le cadre de l'ANOVA d'ordre 1 (cf. 4.2.7, p. 64). En d'autres termes, lorsque l'interaction est significative, on comparera

les moyennes entre les niveaux d'un facteur pour chaque niveau de l'autre facteur[20]. Il est à noter qu'en cas d'interaction significative, la simple mention dans une conclusion de l'effet des facteurs principaux n'est pas suffisante dans la mesure où l'effet d'un facteur est alors dépendant des niveaux de l'autre.

De la même manière, les intervalles de confiance de la moyenne peuvent être calculés à l'aide des formules déjà présentées dans le cadre de l'ANOVA d'ordre 1. Dans les deux cas — comparaisons multiples et calcul des intervalles de confiance —, il suffit de remplacer l'erreur-type de la moyenne par le CM de l'erreur résiduelle.

Lorsque l'interaction n'est pas significative, on procède comme dans l'ANOVA d'ordre 1 et on compare les moyennes de chaque niveau du premier facteur en regroupant les données de chacun des niveaux du second facteur.

Exemple d'application V

> Les données sont issues d'une expérience dans laquelle la concentration de calcium dans le plasma a été mesurée chez 20 sujets des deux sexes ayant subi ou non l'administration d'un traitement hormonal (Source : [26]).

Les données individuelles en fonction des deux facteurs à deux niveaux (sexe et hormone) sont résumées dans le tableau 4.6 et dans la figure 4.6. Le plan d'analyse correspondant est $S_5 < H_2 * I_2 >$ (I pour individus, afin d'éviter l'usage de la lettre S utilisée pour les sujets).

Pour tester l'effet des deux facteurs, on procède à une analyse de variance d'ordre 2 ; les deux facteurs étant fixes, il s'agit d'une ANOVA de type I. Par ailleurs, il y a plusieurs observations pour chacune des combinaisons de facteurs (h1i1, h1i2, h2i1, h2i2) : il s'agit d'un plan factoriel avec réplication.

Les hypothèses à tester sont les suivantes :

1. $H_0 : \mu_{h1} = \mu_{h2}$ ($H_1 : \mu_{h1} \neq \mu_{h2}$), pour l'effet du facteur hormone ;
2. $H_0 : \mu_{i1} = \mu_{i2}$ ($H_0 : \mu_{i1} \neq \mu_{i2}$), pour l'effet du facteur sexe ;
3. $H_0 : pas\, d'interaction\, H \times I$ ($H_1 : interaction\, H \times I$), pour l'effet d'interaction entre les deux facteurs.

Le tableau 4.10 résume la décomposition de la variance pour le jeu de données considéré.

Les tests sur les effets principaux et l'effet d'interaction pour répondre aux hypothèses initiales sont ainsi :

1. $H_0 : \mu_{h1} = \mu_{h2}$, $F_{obs(H)} = \frac{CM_H}{CMe} = \frac{1386.11}{22.90} = 60.50 > F_{0.001,(1,16)} = 16.01$, on rejette H_0

[20] On remarquera qu'en présence de 2 facteurs A et B à 2 et 3 niveaux respectivement, et d'une interaction AxB significative, cela mène à $\frac{ab(ab-1)}{2}$, i.e. 15 comparaisons de moyennes à effectuer au total.

h1		h2	
s1	s2	s1	s2
16.5	14.5	39.1	32.0
18.4	11.0	26.2	23.8
12.7	10.8	21.3	28.8
14.0	14.3	35.8	25.0
12.8	10.0	40.2	29.3

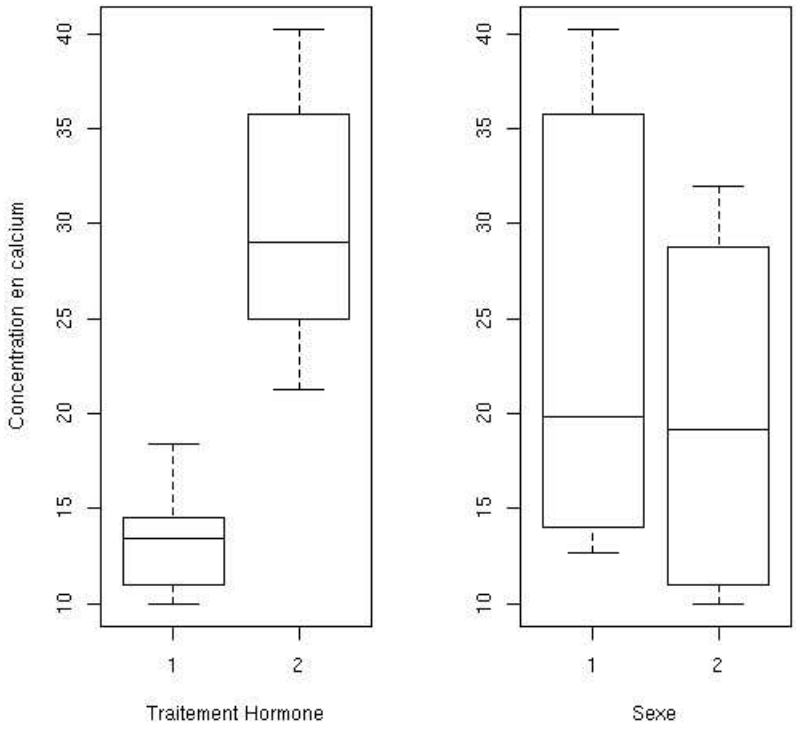

Fig. 4.6: Concentration en calcium (en mg/100 ml) des individus en fonction de leur sexe et de l'administration préalable d'hormone.

Source	SC	dl	CM
Totale	1827.70	19	
Hormone	1386.11	1	1386.11
Sexe	70.31	1	70.31
Hormone x Sexe	4.90	1	4.90
Erreur	366.37	16	22.90

Tab. 4.10: Analyse de variance pour les données de l'exemple.

2. $H_0 : \mu_{i1} = \mu_{i2}$, $F_{obs(S)} = \frac{CM_S}{CMe} = \frac{70.31}{22.90} = 3.07 < F_{0.05,(1,16)} = 4.49$, on ne peut pas rejetter H_0
3. H_0 : *absence d'interaction* $H \times I$, $F_{obs(HxI)} = \frac{CM_{HxI}}{CMe} = \frac{4.90}{22.90} = 0.21 <$

$F_{0.05,(1,16)} = 4.49$, on ne peut pas rejetter H_0

Les résultats des tests indiquent un effet significatif du facteur H_2 (hormone) au seuil bilatéral 0.001, tandis que l'effet du facteur I_2 (sexe) n'est pas significatif. Il n'y a par ailleurs pas d'effet d'interaction entre les deux facteurs : l'effet du facteur H_2 est le même quelque soit le sexe des individus. Cette absence d'interaction se retrouve sur le graphique d'interaction 4.7 reprenant l'ensemble des résultats, appelé également graphique d'interaction : les segments reliant les valeurs moyennes en fonction des conditions de l'expérience sont pratiquement parallèles.

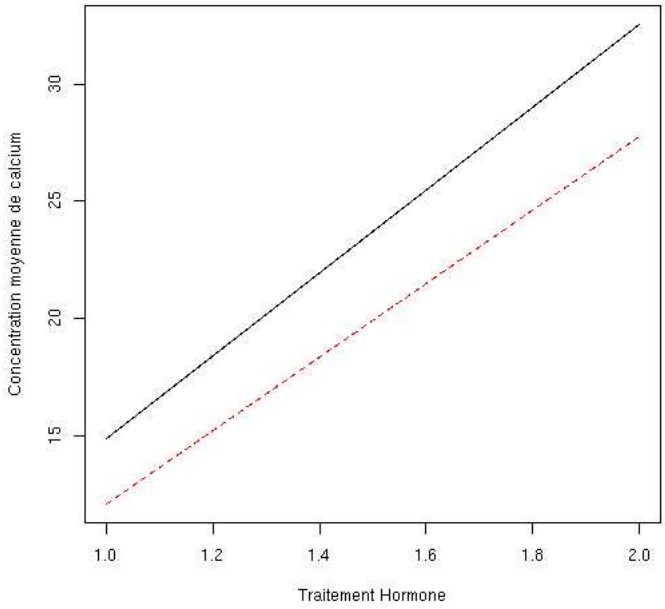

Fig. 4.7: Graphique d'interaction entre les deux facteurs : « sans hormone » (1) et « avec hormone » (2) en abscisses, « femme » en rouge, « homme »$\bar{1}$ en noir.

4.3.5 Plan factoriel sans réplication

Décomposition de la variance

Lorsque le protocole expérimental ne comporte pas de réplication, c'est-à-dire qu'on ne dispose que d'une observation par combinaison de facteurs, il est toujours possible d'effectuer une ANOVA, mais l'analyse est limitée à la seule étude des effets principaux, avec comme hypothèse implicite supplémentaire l'absence d'interaction entre les facteurs. En effet, puisqu'il y n'y a qu'une seule observation pour chaque combinaison de chaque niveau des différents facteurs, il n'est plus possible d'estimer la variabilité intra pour cette combinaison particulière et l'on ne peut plus estimer la variabilité résiduelle à partir de ces variabilités intra. Celle-ci doit donc être

estimée à partir du CM de l'interaction (en fait, les composantes d'interaction et résiduelles sont confondues).

La décomposition de la variance suit le même principe que dans le cas des plans avec réplication (exception faite de l'indice de réplication) et est illustrée dans le tableau 4.11.

Variance	SC	dl	CM	F
Totale	$\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{X}_{ij}-\bar{X})^2$	$ab-1$	CMt=SC/dl	
Facteur A	$b\sum_{i=1}^{a}(X_i-X)^2$	$a-1$	CMa=SC/dl	CMa/CMe
Facteur B	$a\sum_{j=1}^{b}(\bar{X}_j-\bar{X})^2$	$b-1$	CMb=SC/dl	CMb/CMe
Erreur	$\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{ij}-X_i-X_j+Y)^2$	$(a-1)(b-1)$	CMe=SC/dl	

Tab. 4.11: Tableau de décomposition de la variance en fonction des sources de variabilité, pour un plan factoriel à 2 facteurs sans réplication (SC représente la somme des carrés, dl le nombre de degrés de libertés, CM le carré moyen et F la valeur du F de Fischer-Snedecor).

4.3.6 Plan factoriel à mesures répétées

Décomposition de la variance

Dans un plan factoriel à mesures répétées, plusieurs observations sont réalisées sur un même sujet dans différentes conditions, contrairement aux autres cas (ANOVA à un ou plusieurs facteurs) dans lesquels chacune des observations était issue d'un individu spécifique. Dans ce cas de figure, la procédure de décomposition de la variance totale est similaire aux précédentes, excepté que l'on ajoute un terme de variance liée au facteur sujet. Les facteurs de croisement sont qualifiés de *facteurs intra* (l'effet est également appelé effet intra), et les facteurs emboîtants, ou de groupement, des *facteurs inter*. Si l'on prend comme exemple un protocole dans lequel on étudie l'effet d'un facteur sur un ensemble de sujets, la variance totale se décompose en un terme de variance inter-sujets, lié au facteur sujet, et un terme de variance intra-sujet, lié au facteur d'étude. Les dl associés à la SC du facteur « sujet » valent $n-1$, où n désigne le nombre de sujets participant au protocole. Quant aux dl associés à la SC intra-sujets, ils sont partionnés en deux sources : les dl associés à l'effet du facteur, valant $k-1$ (où k désigne le nombre de niveaux du facteur), et les dl associés à l'erreur résiduelle qui valent $(k-1)(n-1)$.

On formera les CM liés au facteur d'étude[21] et à l'erreur résiduelle, comme précédemment, à l'aide des SC et des dl adéquats.

Bien que les nous ayons présenté succintement le principe général des plan factoriels à mesures répétées, d'autres protocoles expérimentaux sont couramment employés, en psychologie expérimentale par exemple, et incluent deux ou plusieurs facteurs croisés (plan de type $S*A*B$)

[21] le facteur sujet n'intervient que pour le calcul de l'erreur résiduelle, on n'attend pas d'effet ayant un sens de ce type de facteur.

— il s'agit toujours de facteurs intra —, ou des mélanges de facteurs inter et intra (plan de type $S < A > * B$). Ces différents cas de figure donneront naturellement lieu à des procédures d'analyse un peu plus élaborées, puisqu'il faut prendra en compte les différents facteurs ainsi que les relations qu'ils entretiennent. D'autre part, il est courant d'effectuer de multiples observations, i.e. d'administrer plusieurs essais, pour un même traitement[22], et l'on ne conserve généralement que la moyenne des observations, notamment pour les résumés numériques et les représentations graphiques. Cette procédure expérimentale nécessite cependant de prendre en compte le facteur de répétition des essais dans l'analyse initiale des données. Si l'effet de celui-ci s'avère non-significatif, alors ce facteur pourra être exclu des analyses subséquentes et les séries de valeurs par traitement pourront être remplacées par leurs moyennes respectives, afin de reprendre une ANOVA à mesures répétées classique à plusieurs facteurs.

Une contrainte supplémentaire par rapport aux hypothèses implicites de l'ANOVA (homogénéité des variances parentes) et liée spécifiquement à ce type de plan expérimental est que l'on suppose que les corrélations entre les conditions expérimentales (i.e. les niveaux d'un facteur, ou les combinaisons de niveaux de plusieurs facteurs) prises deux à deux sont les mêmes. En d'autres termes, on s'attend à ce qu'un sujet qui réussit mieux dans la première condition qu'un autre sujet, réussisse également mieux dans la deuxième condition, etc. L'ensemble de ces hypothèses est connu également sous le terme de *circularité* (ou *sphéricité* dans la littérature anglo-saxonne, cf. [16], et discussion dans [26], p. 259). Une manière de s'affranchir de ces contraintes est d'utiliser des facteurs de correction, ou d'ajustement, tel que l'*ajustement de Greenhouse-Geisser*, ou de recourir à une analyse de variance multivariée (MANOVA, cf. 4.4, p. 82), mais ces procédures débordent du cadre fixé dans ce document.

Test d'hypothèse

Les hypothèses seront formulées sur la base des facteurs autres que le facteur « sujet ». La valeur de test portant sur les hypothèses nulles liées à l'effet des facteurs est similaire à celle développée dans les cas précédents, c'est-à-dire qu'elle est calculée à partir du rapport des CM liés au(x) facteur(s) sur le CM lié à l'erreur résiduelle.

Exemple d'application VI

> Les données sont issues d'une expérience dans laquelle la concentration de cholestérol dans le sang a été mesurée chez 7 sujets après administration répétée de 3 traitements différents (Source : [26]).

Les données individuelles en fonction des deux facteurs à deux niveaux (sexe et hormone) sont résumées dans le tableau 4.12 et dans la figure 4.8.

[22] On parle souvent de *traitement* pour désigner le croisement des niveaux ou modalités de plusieurs facteurs, par exemple pour 2 facteurs A et B à 2 modalités, a1b1 représente un des 4 traitements.

Sujet	I	II	III
s1	164	152	178
s2	202	181	222
s3	143	136	132
s4	210	194	216
s5	228	219	245
s6	173	159	182
s7	161	157	165

Tab. 4.12: Concentration de cholestérol (en mg/dl) pour 7 sujets en fonction des 3 traitements reçus successivement (I, II, III).

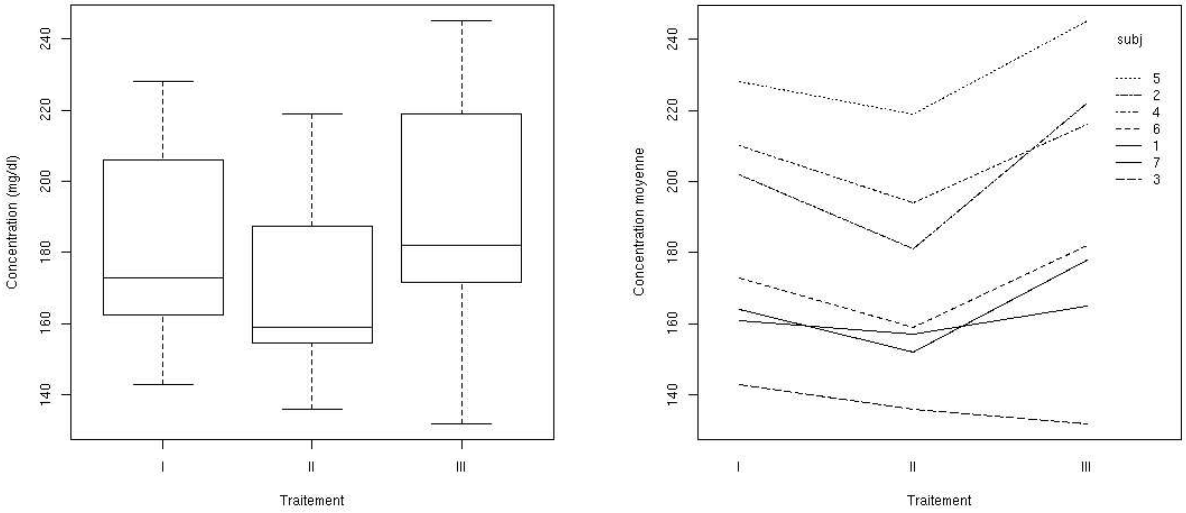

Fig. 4.8: Concentration de cholestérol en fonction du type de traitement (à gauche), et graphique des effets individuels (à droite).

Dans ce contexte, on isole les deux facteurs « sujet » et « traitement », et le plan d'analyse considéré sera $S_7 * T_3$. On fixe l'hypothèse nulle $H_0 : \mu_1 = \mu_2 = \mu_3$ pour l'effet du facteur traitement, l'hypothèse alternative étant que les moyennes parentes ne sont pas toutes les mêmes. La variabilité observée sera décomposée en 2 sources principales de variation : la variance dûe au facteur sujet, et la variance intra-sujet qui se décompose en un terme de variance liée aux différents niveaux du facteur « drogue » et un terme de variance résiduelle, liées aux fluctuations aléatoires (cf. tableau 4.13).

On effectue le test d'inférence comme dans l'exemple précédent : $F_{obs(H)} = \frac{CM_T}{CMe} = \frac{727.00}{57.94} = 12.6 > F_{0.005,(2,12)} = 8.51$, donc le test est significatif au seuil bilatéral $\alpha = 0.005$ et on rejette H_0.

Source	SC	dl	CM
Totale	20880.57	20	
Sujets	18731.24	6	
Traitement	1454.00	2	727.00
Erreur	695.33	12	57.94

Tab. 4.13: Analyse de variance pour les données de l'exemple.

4.3.7 Alternatives non-paramétriques

Les alternatives non-paramétriques pour les ANOVA à plusieurs facteurs sont (i) le *test de Kruskal-Wallis*, présenté dans le cas de l'ANOVA à un seul facteur (cf. 4.2.9, p. 69) pour les plans factoriels avec réplication ; (ii) le *test de Friedman* pour les plans à mesures répétées. Pour de plus amples informations sur la mise en œuvre de ces tests, le lecteur pourra consulter [26] et [7].

4.4 Analyse de variance multidimensionnelle

4.4.1 Principe général

L'analyse de variance multidimensionnelle, ou MANOVA, est en fait une généralisation de l'ANOVA à un ou plusieurs facteurs (variables qualitatives) dans laquelle 2 ou plusieurs variables dépendantes continues sont mesurées simultanément[23]. La MANOVA permet d'examiner s'il y a des effets principaux et d'interaction des facteurs sur ces variables *considérées dans leur ensemble*. Cette méthode est ainsi utilisée lorsque l'on possède plusieurs mesures distinctes, ou indices, caractérisant une même variable dépendante « globale ». Par exemple, on pourrait s'intéresser au degré de « distanciation » (variable dépendante « globale », non mesurable directement) dans une étude sur le comportement non-verbal d'un ensemble de personnes, en mesurant différents indicateurs comme le nombre de fois où la personne s'est croisée les bras, la distance corporelle et l'inclinaison du torse (ensemble des variables dépendantes mesurées). Ces variables dépendantes, bien que toutes quantitatives, sont le plus souvent différentes (par exemple, du point de vue des échelles de mesure), mais elles sont susceptibles d'être fortement corrélées entre elles. Il est dès lors envisageable de traiter ces variables dépendantes ensemble lors de l'analyse de l'effet des facteurs, c'est-à-dire de prendre en compte cette corrélation intrinsèque dans l'analyse de l'effet des variables indépendantes. Si on observe un effet d'une variable indépendante, cela signifiera que l'ensemble des variables dépendantes est globalement affecté par cette celle-ci sans que l'on sache nécessairement quelles sont précisement les variables dépendantes qui sont affectées.

[23] On remarquera également que l'ANOVA à mesures répétées apparaît comme un cas particulier de la MANOVA, sa particularité résidant dans le fait qu'on s'intéresse à l'effet d'une variable intra-sujets répétée.

Nous ne présentons dans cette partie que les grandes lignes de la démarche d'analyse en MANOVA, et laissons le soin au lecteur intéressé par les techniques avancées d'analyse multivariée de consulter les (nombreux) ouvrages spécialisés.

4.4.2 Conditions d'application

Les hypothèses implicites de la MANOVA sont les suivantes :
- les observations de chaque individu sont indépendantes et ont été sélectionnées de manière aléatoire ;
- les observations de chaque individu se distribuent suivant une loi normale multivariée ;
- les variances des k groupes sont homogènes ;
- les covariances entre chacune des VD prises deux à deux pour les k groupes sont homogènes.

4.4.3 Hypothèse

A la différence de l'ANOVA, dans une MANOVA avec 2 variables X_1 et X_2 et k groupes d'observations, on formulera dans une même hypothèse nulle l'égalité *conjointe* des moyennes parentes pour les deux variables, c'est-à-dire

$$H_0 : \mu_{11} = \mu_{12} = \ldots = \mu_{1k} \; et \; \mu_{21} = \mu_{22} = \ldots = \mu_{2k}$$

l'hypothèse alternative étant que les k populations n'ont pas la même moyenne pour la variable 1 ni la même moyenne pour la variable 2. H_0 sera donc rejetée dès que l'une des égalités de moyennes ne sera pas vérifiée.

4.4.4 Test d'hypothèse

Plusieurs statistiques de test ont été proposées dans le cadre de la MANOVA, et la plupart du temps elles sont approximées par la valeur de test vue dans le cas des ANOVA, le F de Fisher-Snedecor. Leur interprétation en est ainsi grandement facilitée, puisque l'on retrouve des repères comparables à ceux de l'ANOVA.

Les quatre principales statistiques utilisées sont :
- le *lambda de Wilks* (Λ), est la valeur de test la plus utilisée, et elle est souvent transformée pour être interprétée comme un test F avec sa probabilité p associée. Λ est un indice variant de 0 à 1 indiquant la part de variance non expliquée par la prise ne considération du facteur[24]. De manière analogue au coefficient $\eta^2 = \frac{V_{inter}}{V_{intra}}$, la part de variance expliquée par les niveaux du facteur est ici : $\eta^2 = 1 - \Lambda$.

[24] C'est en quelque sorte l'équivalent du coefficient d'indétermination $1 - R^2$ en corrélation et régression.

– la *trace de Pillai* et la *trace de Hotelling*, reposent elles-même sur une transformation en une valeur de test de type F de Fisher-Snedecor ; la première est conseillée par de nombreux auteurs dans le cas général, en raison notamment de sa robustesse à l'égard des violations des hypothèses implicites d'homogénéité des variances et des covariances parentes.
– la *racine de Roy*, est également approximée par un F.

4.5 Corrélation

4.5.1 Principe général

Lorsque l'on est en présence de deux variables quantitatives — une variable dépendante Y et une variable indépendante X —, on peut s'intéresser au degré d'*association linéaire* existant entre les deux, i.e. à leur covariation dans leurs domaines de définition respectifs. Pour cela, on utilise le coefficient de corrélation de Bravais-Pearson, r_{XY} , qui est une mesure de cette association linéaire (cf. 2.5.2, p. 31), sans postuler aucune relation de causalité. L'hypothèse d'une absence de corrélation au niveau de la population parente peut ensuite être testée à l'aide d'une statistique appropriée.

4.5.2 Conditions d'application

Les hypothèses implicites pour l'analyse de la corrélation entre deux variables sont les suivantes :
– les valeurs des variables dépendante et indépendante se distribuent *toutes les deux* normalement (critère de binormalité) ;
– leurs variances respectives sont indépendantes (non nécessairement égales) ;
– la relation liant les deux variables est linéaire.

4.5.3 Hypothèse

On fera généralement l'hypothèse que la corrélation au niveau de la population parente est nulle : $H_0 : \rho = 0$, l'hypothèse alternative étant que la corrélation parente entre les deux variables n'est pas nulle ($H_1 : \rho \neq 0$). Cette hypothèse alternative peut également être orientée ($H_1 : \rho \neq 0$ ou $H_1 : \rho \neq 0$), auquel cas le test sera orienté (i.e. des seuils unilatéraux seront retenus pour la significativité).

Dans le cas où l'on souhaite tester une autre hypothèse nulle, i.e. une corrélation parente égale à une certaine valeur théorique ρ_0, alors il sera nécessaire de transformer les données au préalable.

4.5.4 Test d'hypothèse

Afin de tester l'hypothèse nulle d'une corrélation parente nulle, un *test t de Student* (pour la corrélation) peut être calculé comme suit :

$$t = \frac{r_{XY}}{\sqrt{\frac{1-r^2}{n-2}}}$$

où $\sqrt{\frac{1-r^2}{n-2}}$ représente l'erreur-type associée au coefficient de corrélation. La valeur de t obtenue doit être comparée aux valeurs théoriques de la distribution du t de Student à $\nu = n-2$ degrés de liberté, au seuil de décision α (cas d'un test bilatéral).

Comparaison de plusieurs échantillons. Dans le cas où l'on souhaite comparer les coefficients de corrélation r_1 et r_2 observés sur des échantillons différents, en postulant comme hypothèse nulle que les deux échantillons proviennent de la même population parente (i.e. $H_0 : \rho_1 = \rho_2$), il est nécessaire de transformer les coefficients de corrélation en coefficients réduits, à l'aide de la formule :

$$z = \frac{1}{2} ln \left(\frac{1+r}{1-r} \right)$$

On utilisera ces coefficients réduits pour calculer la valeur de test suivante :

$$Z = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1-3}} + \sqrt{\frac{1}{n_2-3}}}$$

qui doit être comparée aux valeurs critiques de la distribution normale centrée-réduite $\mathcal{N}(0; 1^2)$, au seuil α (test bilatéral ou unilatéral selon H_1).

Il est possible également de comparer plus de 2 coefficients de corrélation, et d'effectuer des comparaisons multiples (e.g. [26], section 19.7, pp. 390-394), comme pour les techniques d'ANOVA à un seul facteur (cf. 4.2, p. 61).

Exemple d'application VII

> L'exemple suivant est tiré de [26] : il s'agit de données concernant la longueur des ailes et de la queue d'un échantillon de moineaux. On s'intéresse à la corrélation linéaire entre ces deux caractéristiques chez les moineaux[25]. Le jeu de données est indiqué dans le tableau 4.14.

La distributrion univariée des observations pour chacune des deux variables ne présente pas d'asymétrie particulière, exceptée une plus grande concentration des effectifs pour les valeurs les plus basses (cf. figure 4.9). L'association linéaire entre les

longueur ailes	10.4	10.8	11.1	10.2	10.3	10.2	10.7	10.5	10.8	11.2	10.6	11.4
longueur queue	7.4	7.6	7.9	7.2	7.4	7.1	7.4	7.2	7.8	7.7	7.8	8.3

Tab. 4.14: Longueur des ailes et de la queue (en cm) d'un groupe de 12 moineaux.

deux caractéristiques (longueur des ailes et longueur de la queue) de cet ensemble de 12 individus est mesurée à l'aide du coefficient de corrélation de Bravais-Pearson :

$$r = \frac{\sum_{i=1}^{12}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{12}(x_i - \bar{x})^2 \sum_{i=1}^{12}(y_i - \bar{y})^2}} = \frac{1.32}{\sqrt{1.72 \times 1.35}} = 0.87$$

La corrélation linéaire entre les deux variables apparaît relativement satisfaisante, étant donné la valeur élevée (0.87) du coefficient de corrélation. Les deux variables mesurées ont tendance à *covarier* ensemble : pour l'échantillon de moineaux considéré, lorsque la longueur des ailes augmente, la longueur de la queue augmente également (cf. figure 4.9).

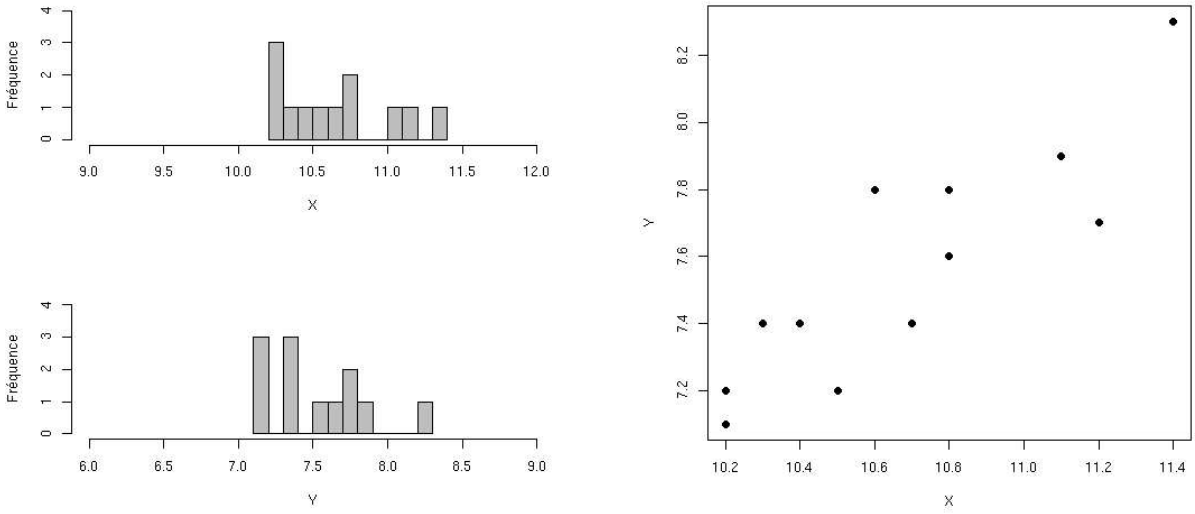

Fig. 4.9: Distributions uni- (à gauche) et bivariées (à droite) de la longueur des ailes (X) et de la longueur de la queue (Y) pour un ensemble de 12 moineaux.

Le coefficient de détermination vaut quant à lui : $R^2 = 0.87^2 = 0.76$. La prise en compte de la corrélation entre les deux variables permet d'expliquer 76 % de la variabilité observée sur la variable Y (longueur de la queue)[26].

Pour inférer l'existence d'une corrélation au niveau de la population parente (i.e. la population dont sont issus les moineaux observés), on teste l'hypothèse nulle d'absence de corrélation parente, $H_0 : \rho = 0$ ($H_1 : \rho > 0$):

[26] Contrairement à la régression linéaire (cf. 4.6), les indices utilisés en corrélation sont symétriques, et s'appliquent aussi bien dans un sens que dans l'autre : la variabilité expliquée peut ainsi porter sur la variable X car $r_{XY} = r_{YX}$ et le coefficient de détermination est le même dans les deux cas.

– à l'aide d'un test t : $t_{obs} = \frac{t}{s_r}$ avec $s_r = \sqrt{\frac{1-R^2}{n-2}} = 0.156$, soit $t_{obs} = \frac{0.870}{0.156} = 5.58 > t_{0.0005,10} = 4.587$; le test est significatif au seuil unilatéral $\alpha = 0.0005$ et donc on rejette H_0.
– à l'aide d'un test F : $F_{obs} = \frac{1+|r|}{1-|r|} = \frac{1.870}{0.130} = 14.4 > F_{0.001,(10,10)} = 8.75$; le test est significatif au seuil bilatéral $\alpha = 0.001$ et donc on rejette H_0.

D'une manière générale, on peut conclure à l'existence d'une corrélation positive entre la longueur des ailes et la longueur de la queue des moineaux.

4.5.5 Alternative non-paramétrique

Lorsque les données ne satisfont pas au critère de binormalité, il est préférable d'utiliser un coefficient de corrélation basé sur les rangs, comme le *coefficient de rang de Spearman* r_s. La statistique utilisée est calculée à partir de la somme du carré d_i de la différence entre les rangs de la variable X et de la variable Y :

$$r_s = 1 - \frac{6 \sum_{i=1} n(R_{x_i} - R_{y_i})^2}{n^3 - n}$$

Il existe des distributions tabulées pour ce type de coefficient, auxquelles il faut se référer pour éprouver la significativité du test (cf. Annexe B, p. B.10, ou [26], [7]). On peut également utiliser une autre statistique de test — le *tau de Kendall* —, qui mène généralement aux mêmes conclusions que le coefficient de Spearman, excepté lorsque la taille de l'échantillon est grande auquel cas le test reposant reposant sur ce dernier est plus puissant.

Exemple d'application VIII

> On s'intéresse à l'association linéaire entre deux séries de notes observées lors de deux examens différents (Source : données adaptées de [26]). Le jeu de données figure dans le tableau 4.15.

Pour mesurer la corrélation entre les deux séries de notes, on calcule le coefficient de corrélation de Spearman :

$$r_{s_{obs}} = 1 - \frac{6 \sum_{i=1}^{10} d_i^2}{n^3 - n} = 1 - \frac{6 \times 72}{10^3 - 10} = 0.564$$

En stipulant pour hypothèse nulle : $H_0 : \rho_s = 0$ ($H_1 : \rho_s \neq 0$, on ne spécifie pas le sens de la liaison), on compare cette valeur de la valeur de test aux valeurs critiques de la distribution appropriée. Or $r_{s_{obs}} < rs_{0.05,10} = 0.648$, donc on ne peut pas rejeter l'hypothèse nulle d'absence de corrélation entre les deux séries de notes.

sujet	note X	rang X	note Y	rang Y	d_i^2
1	11.4	3	16.6	7	16
2	9	1	7.4	1	0
3	14.4	7	8.2	2	25
4	15.6	8	16.8	8	0
5	10.6	2	11.2	3	1
6	12.6	5	17	9	16
7	17.2	9	15.4	6	9
8	19.6	10	17.4	10	0
9	11.8	4	14	5	1
10	14.2	6	11.8	4	4

Tab. 4.15: Notes obtenues à deux examens (X et Y) par 10 élèves d'une classe.

4.6 Régression linéaire simple

4.6.1 Principe général

La régression permet de détecter et de quantifier l'effet d'une variable indépendante X sur une variable dépendante Y. Dans le cadre d'une analyse de régression, on fait l'hypothèse implicite que la variable indépendante, appelée également variable *prédictrice* ou explicative, ou régresseur, est responsable d'une partie de la variation de la variable dépendante, mais qu'en revanche la variable dépendante n'affecte pas la variable indépendante. Le *modèle* postulé dans le cadre d'une analyse de régression permet d'*expliquer* et de *mesurer* l'influence d'une variable quantitative sur une autre variable quantitative.

La technique de régression linéaire est très utile et couramment employée en sciences expérimentales dans la mesure où elle permet non seulement d'effectuer des tests d'hypothèses quant à l'effet d'une variable sur une autre, mais également de prédire les valeurs de la variable dépendante, sous certaines conditions et avec une certaine tolérance, ce qui revient d'une certaine manière à quantifier l'effet de la variable indépendante.

La corrélation et la régression sont deux techniques qui sont souvent associées, mais qui ne doivent pas être confondues. En ce qui concerne la régression, une différence majeure est qu'elle est dépendante du statut attribué aux variables : l'une des variables est considérée comme variable prédictrice, tandis que la corrélation est un indice « symétrique »[27] qui n'implique pas de relation de causalité. D'autre part, tandis que la corrélation mesure le *degré d'association* entre deux variables, la régression mesure l'*intensité de l'effet* (implicite) d'une variable sur l'autre. Ces liens sont souvent sources de confusion dans les techniques utilisées : dans le cadre d'une analyse de régression, on utilisera bien entendu les indicateurs de corrélation pour qualifier la qualité de l'ajustement linéaire, mais s'il n'y a aucune raison de postuler l'existence d'une relation de causalité entre les deux variables, alors l'emploi de la régression s'avère inapproprié. La taille

[27] La corrélation (x,y) est la même que la corrélation (y,x). Le statut, ou le rôle, des deux variables est parfaitement interchangeable.

d'un enfant est, en règle générale, fortement corrélée de manière positive avec son poids, et cette relation biométrique est bien de nature causale et est clairement orientée (influence de la taille sur le poids) : plus le corps est grand, plus la masse corporelle augmente afin de compenser les demandes énergétiques, etc. En revanche, dans les premiers mois de la vie du nourrisson, la longueur des pieds est également corrélée positivement avec le temps de fixation des objets présents dans le champ visuel de l'enfant, mais il paraît délicat de postuler une relation de causalité !

En guise remarque, la figure 4.10 illustre les différentes techniques applicables dans le cadre de l'étude des covariations entre variables selon le type de variables et les relations postulées entre ces variables.

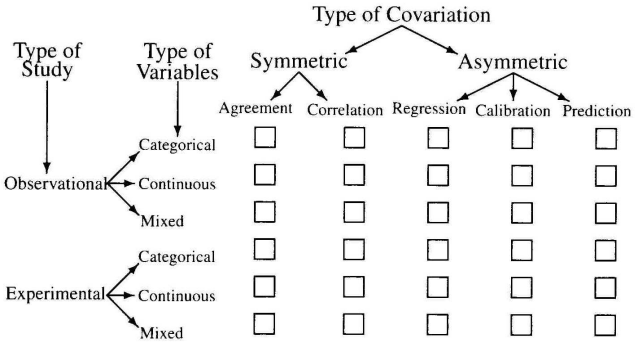

Fig. 3.1 Taxonomy of measures of covariation.

Fig. 4.10: Diversité des techniques d'étude de la covariation entre les variables, selon le type de variable et leurs relations.

4.6.2 Modèle de la régression linéaire

Le modèle général de la régression linéaire postule que les variations observées de la variable dépendante Y sont attribuables à l'effet de la variable indépendante X (effet de causalité) et à des variations non-liées à la variable indépendante, ou résiduelles (effet aléatoire). Il est la plupart du temps formalisé selon l'équation :

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

où α et β représentent respectivement l'ordonnée à l'origine et la pente, ou *coefficient de régression*, les ε_i représentant les résidus, supposés indépendants et se distribuant selon une loi normale de moyenne nulle et de variance fixe $\mathcal{N}(0; \sigma_\varepsilon^2)$.

Les données de l'échantillon seront ainsi exprimées sous la forme d'un modèle de régression de la forme $y = ax + b$, a désignant le *coefficient de régression*, i.e. la pente de la droite de régression, et b l'ordonnée à l'origine. a indique de combien d'unités varie Y lorsque X varie

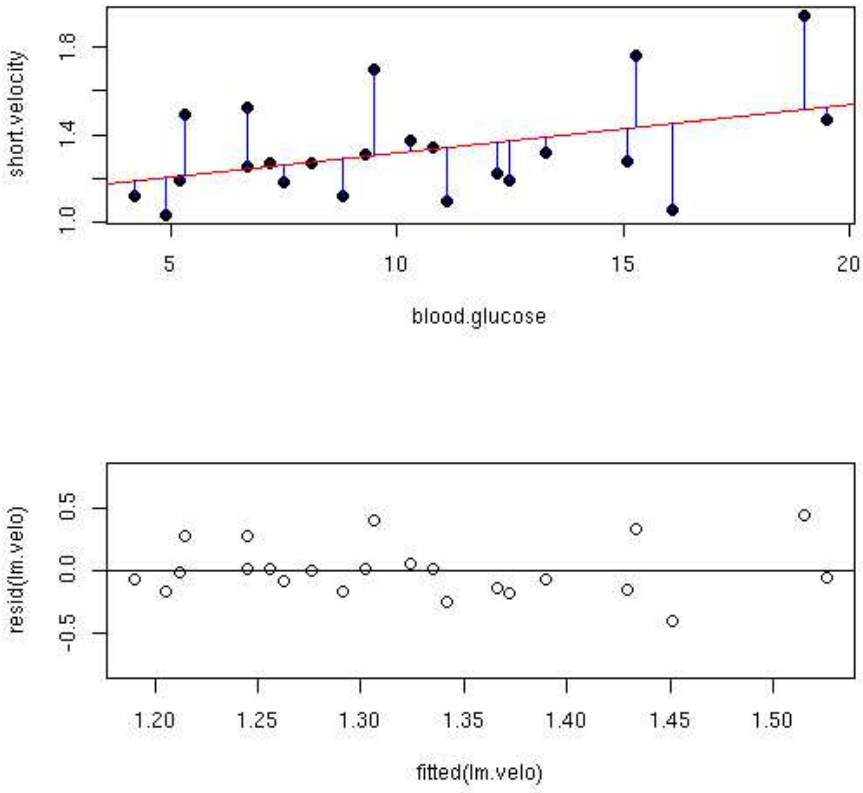

Fig. 4.11: Exemples de régression et représentation des résidus. (Tiré de [24])

d'une unité ; b indique la valeur de Y lorsque $X = 0$. L'estimation des paramètres du modèle au niveau de la population parente, α et β, peut être effectuée à partir de cette droite de régression pour l'échantillon observé, et on notera $\hat{\alpha}$ (i.e. a) et $\hat{\beta}$ (i.e. b) les estimés de ces paramètres.

La figure 4.11 schématise ce modèle, en présentant à la fois la droite de régression (en haut), où les écarts à la droite de régression sont représentés par des segments de couleur bleu, ainsi que la distribution des résidus (en bas). La droite de régression, $\hat{y} = \hat{\alpha} + \hat{\beta}x$, modélise la relation linéaire[28] prédite, ou théorique, entre la variable dépendante Y et la variable indépendante explicative, ou prédictrice, X. Les paramètres de la droite de régression sont estimés par la méthode des moindres carrés, qui minimise la distance des points projetés parallèlement à l'axe des ordonnées (i.e. verticalement) sur la droite de régression[29]. On peut montrer que les paramètres

[28] La *droite* de régression est en fait *affine* puisqu'on introduit une ordonnée à l'origine, non forcément nulle, mais la *relation* est considérée comme linéaire en raison de l'additivité des termes du modèle.

[29] Ce principe de minimisation par la méthode des moindres carrés, $min \sum_i (y_i - \hat{y_i})^2$ se retrouve dans de nombreux tests d'ajustement à des modèles (e.g. modèles linéaires, loi de distribution, etc.), mais peut également être effectué de manière différente, par exemple $min \sum_i (x_i - \hat{x_i})^2$ lorsque l'on s'intéresse à de la prédiction

ainsi estimés sont :

$$\hat{\beta} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2/(n-1)}$$

et

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

la droite de régression passant toujours par le centre de gravité du nuage de points, de coordonnées $(\bar{x}, \bar{y})$.

Les résidus, i.e. les écarts à la droite de régression $y_i - \hat{y}_i$ indiquent l'écart entre les valeurs réelles, *observées*, et les valeurs attendues, *prédites* (par le modèle linéaire). Comme on peut le voir sur la figure 4.12, leur distribution suit à peu près une distribution normale $\mathcal{N}(0, 0.04)$.

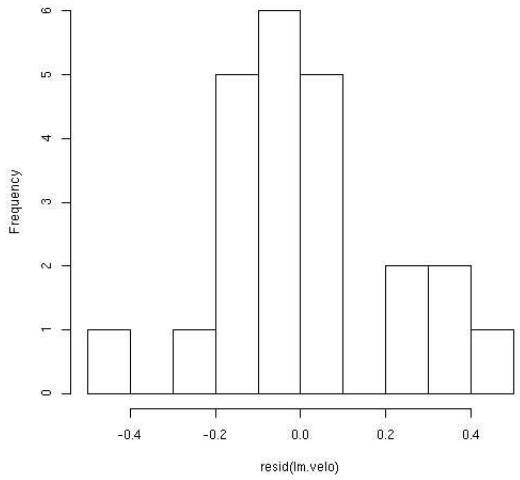

Fig. 4.12: Distribution des résidus ($\overline{(y_i - \hat{y})} \approx 0$, $\sigma^2_{(y_i - \hat{y})} = 0.04$)

La variabilité totale est ainsi la somme d'une variabilité liée à la prédictrice et d'une variabilité résiduelle, liée aux fluctuations autour de la droite du modèle de régression[30]. La qualité d'ajustement du modèle peut être évaluée à l'aide du coefficient de détermination, R^2, qui mesure la part de variance des y_i expliquée les x_i (variable prédictrice). Celui-ci peut-être calculé comme dans le cas de la corrélation, mais également comme le rapport entre la variances des valeurs de Y prédites et celle des valeurs de Y observées.

inversée dans le cadre de la régression, ou $min \sum_i ((x_i, y_i) - (\hat{x_i}, \hat{y_i}))^2$ lorsqu'on s'intéresse à l'ajustement des données observées à des courbes bidimensionnelles (e.g. cercle ou ellipse).

[30] On retrouve là une décomposition de la variance comparable à celle qui a été présentée lors de l'analyse de variance — $V_{totale} = V_{inter} + V_{intra}$ —, entre la variance inter, liée à l'effet de la variable indépendante, et la variance intra, liée à la variabilité intra-groupe. Dans le cas du modèle de régression, la décomposition se formule dans les mêmes termes, la variabilité inter-groupe étant la variance liée à la prédictrice, et la variabilité intra-groupe devenant la variabilité liée aux résidus, soit $V_{totale} = V_{predictrice} + V_{residus}$

4.6.3 Conditions d'application

Les conditions d'application en régression linéaire sont les suivantes :
- les valeurs résiduelles suivent une distribution normale ;
- il y a homogénéité des variances des résidus pour l'ensemble des valeurs de la variable dépendante[31] ;
- la relation entre la variable prédictrice et la variable dépendante est linéaire ;
- l'erreur de mesure sur la variable indépendante est faible.

Il n'y a guère de contrôle possible en ce qui concerne la dernière hypothèse implicite, et les autres conditions d'application peuvent être vérifiées à l'aide des tests usuels décrits dans les parties précédentes.

4.6.4 Hypothèse

Sous les conditions énoncées précédemment, on postule généralement une hypothèse nulle d'absence de liaison linéaire, c'est-à-dire un coefficient de régression β nul : $H_0 : \beta = 0$, l'hypothèse alternative étant qu'il existe une liaison linéaire, $H_1 : \beta \neq 0$, si on ne précise pas le sens de la liaison — positive ou négative, c'est-à-dire $H_1 : \beta > 0$ ou $H_1 : \beta < 0$, respectivement —.

4.6.5 Test d'hypothèse

On distingue classiquement deux types de test pouvant être mis en œuvre pour éprouver l'hypothèse nulle H_0 : (i) un test t de Student sur l'estimé de la pente, et (ii) un test F de Fisher-Snedecor sur la variance expliquée par le modèle.

Test sur le coefficient de régression

Pour le test sur l'estimé de la pente, ou coefficient de régression, l'erreur-type associée à l'estimé de la pente $\hat{\beta}$ correspond en fait à la variance résiduelle (i.e. celle qui n'est pas expliquée par la variable indépendante) rapportée à la somme des carrés des écarts (SC) de la variable indépendante. Elle est donc calculée comme suit :

$$s_b = \sqrt{\frac{s^2_{Y \cdot X}}{\sum_i x_i^2}}$$

La valeur de test consiste alors à calculer

$$t_{obs} = \frac{\hat{\beta}}{s_b}$$

[31] cette condition est utile dans le cas de la régression avec mesures répétées sur X

et à comparer cette valeur aux valeurs repères de la distribution du t de Student, avec $\nu = n - 2$ degrés de liberté, au seuil $\alpha = 0.05$ (seuil bilatéral dans le cas d'une hypothèse non orientée).

Remarque. On remarquera que plus l'étendue des valeurs de la variable indépendante (X) sera grande, plus l'erreur-type de la pente sera petite. La capacité de détecter une pente qui diffère significativement de 0 est ainsi inversement proportionnelle à l'étendue de la variable dépendante mesurée. Il peut donc s'avérer intéressant de mesurer la variable dépendante à des valeurs extrêmes de la variable indépendante, puisque cela influence la puissance du test (e.g. [9]).

Analyse de variance

La même hypothèse $H_0 : \beta = 0$ peut être éprouvée à l'aide de la technique d'analyse de variance ; en revanche, l'hypothèse alternative n'est pas orientée. La variabilité totale est décomposée en deux sources de variation : la variance liée à la relation linéaire entre les deux variables (régression de Y sur X), et la variance résiduelle. La SC liée à la régression linéaire est simplement la somme des écarts quadratiques entre les valeurs prédites et la valeur moyenne pour la variable dépendante. La SC résiduelle quant à elle est la somme des écarts quadratiques entre les valeurs observées et les valeurs prédites. Il n'y a qu'un seul degré de liberté associé au modèle, puisqu'il y a 2 paramètres à estimer ($\hat{\alpha}$ et $\hat{\beta}$).

On peut résumer cela, comme en ANOVA, dans un tableau de décomposition de la variance, comme illustré dans le tableau 4.16.

Source	SC	dl	CM
Totale	$\sum_i (y_i - \bar{y})^2$	n-1	
Modèle	$\frac{(\sum_i (x_i - \bar{x})(y_i - \bar{y}))^2}{\sum_i (x_i - \bar{x})^2}$	1	SCm/dl
Résidus	$\sum_i (y_i - \hat{y}_i)^2$	n-2	SCr/dl

Tab. 4.16: Tableau de décomposition de la variance en régression linéaire.

La valeur de test F_{obs} est simplement le rapport entre la variance liée au modèle et la variance résiduelle, c'est-à-dire entre le CM de régression (SCm/dl) et le CM résiduel (SCr/dl). Elle est à comparer aux valeurs critiques de la distribution du $\mathcal{F}$ de Fisher-Snedecor à $(\nu_1, \nu_2) = (1, n - 1)$ degrés de liberté, au seuil $\alpha = 0.05$ (bilatéral).

4.6.6 Estimation et prédiction : calcul des intervalles de confiance

Intervalles de confiance pour la pente et la moyenne

L'intervalle de confiance pour le coefficient de régression peut être calculé à partir de l'erreur-type du paramètre estimé (la pente) comme suit :

$$IC_{95\%} = [b - t_{\alpha,n-2}s_b\,;\, b + t_{\alpha,n-2}s_b]$$

L'interprétation de l'intervalle de confiance se fait tout naturellement comme dans les autres cas d'estimation vus précédemment. On dira ici, avec une certitude de 95 %, que l'estimé de la pente est compris dans l'intervalle $b \pm t_{\alpha,n-2}s_b$ (il y a 5 % de chances que cela ne soit pas vrai et que nous commettions une erreur en affirmant cela). La droite de régression théorique peut ainsi posséder une pente variant autour de ces bornes limites, c'est-à-dire que la droite de régression peut « tourner » autour du point moyen $(\bar{x}, \bar{y})$ dans les limites imposées à la pente.

L'intervalle de confiance pour la valeur moyenne de y à une valeur donnée x s'obtient comme suit :

$$\hat{y} \pm t_{\alpha,n-2}\sqrt{s^2_{XY}\left[\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_i (x_i - \bar{x})^2}\right]}$$

On représente généralement les bornes de l'IC pour la moyenne à l'aide d'une *bande de confiance* composée de deux « courbe » situées de part et d'autre de la droite de régression, comme illustré dans la figure 4.13. Ces bandes sont généralement incurvées (ce sont en fait des arcs d'hyperbole), plus particulièrement lorsque l'on s'éloigne du point moyen $(\bar{x}, \bar{y})$[32].

Intervalles de confiance pour les valeurs prédites

L'intervalle de confiance pour les valeurs prédites peut être calculé comme suit :

$$\hat{y}_i \pm t_{\alpha,n-2}\sqrt{s^2_{XY}\left[1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_i (x_i - \bar{x})^2}\right]}$$

A la différence de l'intervalle de confiance pour la moyenne, il s'agit ici d'un intervalle de confiance pour des données individuelles (observations prédites), et cet intervalle est plus large que le précédent. On le représente également à l'aide d'une bande de confiance autour de la droite de régression (cf. figure 4.13).

[32] La droite de régression est en fait mieux estimée autour du point moyen $(\bar{x}, \bar{y})$, puisqu'elle passe toujours par celui-ci, et l'erreur augmente avec la distance à ce point moyen.

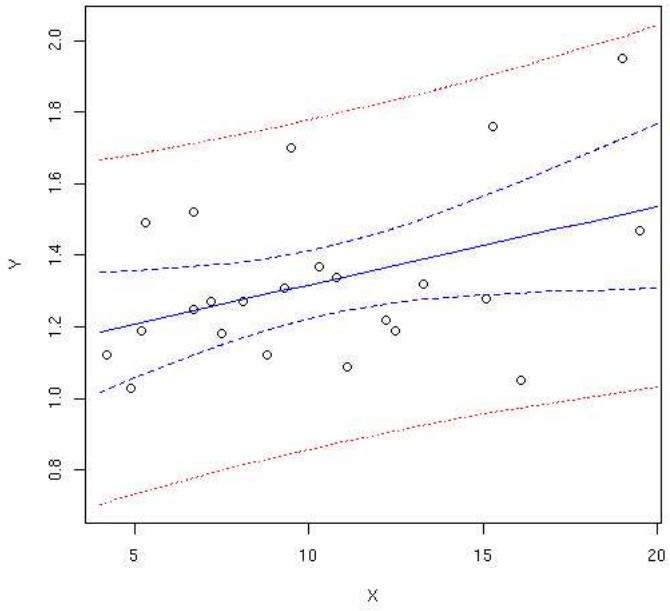

Fig. 4.13: Intervalles de confiance pour la pente (en bleu) et pour les valeurs prédites (en rouge).

Exemple d'application IX

> L'exemple suivant est tiré de [26] : il s'agit de données concernant la longueur des ailes d'un échantillon de moineaux d'âges variables (cf. tableau 4.17). On s'intéresse à la prédiction de la longueur des ailes en fonction de l'âge.

âge	3.0	4.0	5.0	6.0	8.0	9.0	10.0	11.0	12.0	14.0	15.0	16.0	17.0
longueur	1.4	1.5	2.2	2.4	3.1	3.2	3.2	3.9	4.1	4.7	4.5	5.2	5.0

Tab. 4.17: Age (en jours) et longueur des ailes (en cm) d'un groupe de 13 moineaux.

On laissera au lecteur le soin d'effectuer les analyses descriptives univariées et bivariées appropriées, et on va tester la significativité de la relation fonctionnelle entre la longueur des ailes et l'âge des individus, à l'aide des deux méthodes exposées précédemment. On remarquera que contrairement à l'exemple cité dans le cas de la corrélation (cf. 4.5.4, p. 85), ici il y a bien lieu de supposer une relation de causalité entre les deux variables (l'augmentation de l'âge *entraîne* généralement l'augmentation de la longueur des ailes).

Droite de régression. Le calcul de la droite de régression $y = ax + b$ se fait en deux étapes :

1. on calcule dans un premier temps la pente de régression, b, comme le rapport de la covariance entre les deux variables X et Y et de la variance de X, soit $b = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{Var(x)} = \frac{5.9}{21.83} = 0.270\ cm/jour$;
2. puis, on calcule l'ordonnée à l'origine à l'aide de la relation $a = \bar{y} - b\bar{x}$, soit $a = 3.415 - 0.270 \times 10 = 0.715\ cm$ (approx.).

Ces résultats indiquent que lorsque X varie d'une unité (jour), Y varie de 0.270 unité (cm), et qu'à la naissance ($X = 0$) la longueur des ailes vaudrait 0.715 cm (cf. figure 4.14). On peut de la sorte prédire les valeurs de Y pour certaines valeurs de X. Par exemple, la longueur prédite des ailes pour un moineau âgé de 13 jours serait :

$$\hat{y} = a + b \times x_i = 0.715 + 0.270 \times 13 = 4.225\ cm$$

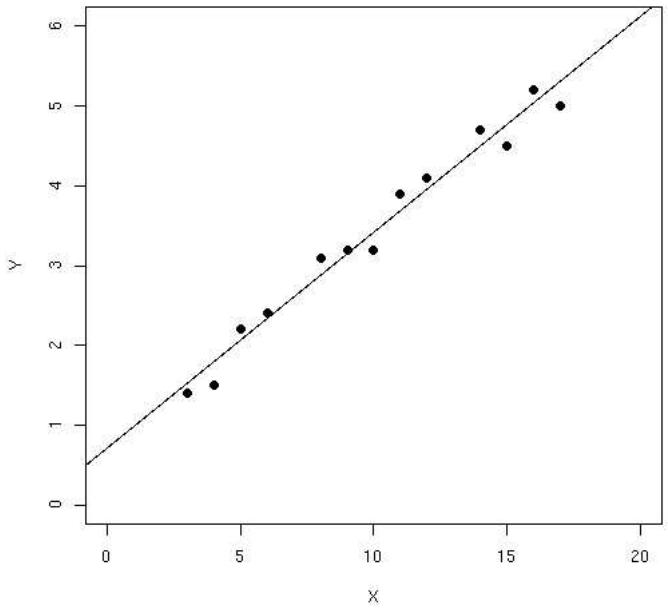

Fig. 4.14: Evolution de la longueur des ailes (Y) en fonction de l'âge des moineaux (X). La droite de régression est représentée en trait continu.

Test t sur le coefficient de régression. On pose l'hypothèse nulle $H_0 : \beta = 0$, i.e. la pente de régression est nulle au niveau de la population parente. L'hypothèse alternative est ici orientée ($H_O : \beta > 0$). L'erreur-type, s_b, associée à l'estimé de la pente est calculée comme :

$$s_b = \sqrt{\frac{s^2_{Y \cdot X}}{\sum_{i=1}^{13} (x_i - \bar{x})^2/(n-1)}} = \sqrt{\frac{0.048}{262.000}} = 0.0135(cm/jour)$$

La valeur de test est ainsi : $t_{obs} = \frac{b}{s_b} = \frac{0.270}{0.0135} = 20.00 > t_{0.0005,11} = 4.437$. Le résultat du test étant significatif, on est amené à rejetter H_0 au seuil unilatéral $\alpha = 0.0005$.

Analyse de variance. En posant l'hypothèse nulle $H_0 : \beta = 0$ contre l'hypothèse alternative $H_1 : \beta \neq 0$, on peut décomposer la variabilité totale, comme dans l'approche classique de l'ANOVA, en deux sources de variation : (1) la variance liée à la relation linéaire entre les deux variables et (2) la variance résiduelle liées aux fluctuations aléatoires, que l'on peut résumer dans un tableau de décomposition de la variance (cf. tableau 4.18).

Source	SC	dl	CM
Totale	19.657	12	
Modèle	19.132	1	19.132
Résidus	0.525	11	0.048

Tab. 4.18: Analyse de variance pour les données de l'exemple.

On a ainsi :

$$F_{obs} = \frac{19.132}{0.048} = 401.1 > F_{0.001,(1,11)} = 19.7$$

On rejette donc H_0 au seuil $\alpha = 0.001$ (bilatéral).

Le coefficient de détermination R^2, qui mesure d'une certaine manière l'intensité de la relation entre X et Y ou la qualité de l'ajustement linéaire, est calculé comme le rapport entre la SC du modèle et la SC totale, et on a $R^2 = \frac{19.132}{19.657} = 0.97$: 97 % de la variance est expliquée lorsque l'on prend en compte le modèle linéaire.

Les deux tests permettent donc de rejetter H_0 (en pratique, on n'en utilise qu'un seul des deux pour la conclusion !) : d'une manière générale, il existe une relation linéaire positive et forte entre la longueur des ailes et l'âge des individus considérés.

Estimation. L'intervalle de confiance pour le coefficient de régression calculé vaut : $IC_{95\%} = b \pm t_{0.05,11} s_b = 0.270 \pm (2.201 \times 0.0135) = 0.270 \pm 0.030$ cm/jour, soit

$$IC_{95\%} = [0.240\,;\, 0.300]$$

Nous pouvons ainsi affirmer avec un seuil de confiance de 95 % que l'estimé de la pente se situe entre ces deux limites 0.240 et 0.300 (il n'y a pas plus de 5 % de chances que nous commettions une erreur dans cette affirmation).

4.7 Régression linéaire multiple

4.7.1 Principe général

La régression linéaire multiple étend l'analyse de régression linéaire simple présentée dans la partie précédente (cf. 4.6, p. 88) au cas de la prédiction des valeurs d'une variable dépendante Y en fonction de plusieurs variables indépendantes quantitatives (continues) X_j. Cette technique

est tout à fait adaptée dans le cadre d'une étude visant à étudier, par exemple, la vitesse d'exécution de mouvements précis de pointage en environnements virtuels en fonction de la taille, du poids, de l'âge, et du temps de réaction moyen dans des tâches de détection de cibles visuelles des participants.

Avant de décrire le modèle général de la régression linéaire multiple, nous présenterons rapidement les outils d'analyse de corrélation adaptés au cas de plusieurs variables.

4.7.2 Corrélation partielle

Lorsque l'on est en présence non plus de deux variables quantitatives, mais d'un ensemble de k variables quantitatives, il est toujours possible d'étudier le degré d'association entre ces variables mais en utilisant cette fois-ci la corrélation multiple, et les coefficients de corrélation partielle.

Ce serait le cas dans une étude où on s'intéresse à l'efficacité de méthodes d'entraînement (e.g. score moyen obtenu lors d'une épreuve) dans une discipline sportive particulière et où on disposerait à la fois des données de taille, de poids, d'âge, de capacité respiratoire (e.g. $V0_2max$), de durée d'entraînement hebdomadaire, du nombre de cigarettes fumées par jour, etc. pour un ensemble de sujets. Dans un cas comme celui-ci, on pourrait ne s'intéresser qu'aux relations bivariées entre les variables à l'aide du coefficient de corrélation de Bravais-Pearson (cf. 4.5, p. 84), mais ce serait négliger les liaisons existant entre les variables : la capacité respiratoire est certainement corrélée avec l'âge du sujet, mais également avec la durée d'entraînement hebdomadaire, ou le nombre de cigarettes fumées par jour, alors que cette dernière présente sans doute très peu de liaison avec la taille de l'individu.

La *matrice de corrélation* de l'ensemble des variables permettrait de résumer de manière synthétique les liaisons bivariées, mais indépendemment des autres variables. Néanmoins, lorsque l'on souhaite prédire les valeurs de la variable dépendante (dans cet exemple, le score moyen obtenu lors d'une épreuve) en fonction d'une autre variable (e.g. la capacité respiratoire), on peut se demander si cette variable est une bonne prédictrice, c'est-à-dire si elle apporte une information spécifique par rapport aux autres variables avec laquelle elle est corrélée. On utilisera pour cela les corrélations partielles, qui permettent d'étudier le degré de liaison entre deux variables en maintenant constante la ou les autre(s) variable(s), c'est-à-dire en éliminant les variations dûes à ces variables additionnelles. La corrélation partielle entre deux variables X et Y, parmi 3 variables X, Y et Z, est généralement notée $r_{XY.Z}$. Elle est calculée comme suit :

$$r_{XY.Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}$$

et elle s'interprète comme une corrélation simple entre deux variables. Lorsqu'on l'élève au carré, il s'agit d'un coefficient de détermination classique qui représente la part de variance de Y expliquée par X, lorsque Z est maintenu constant. Dans le cas où il y a plus de 3 variables, le

calcul manuel des coefficients de corrélation partielle est relativement fastidieux, et l'usage d'un ordinateur s'avère bien souvent salvateur !

L'usage des corrélations partielles et du coefficient de détermination est essentiel lorsqu'on est en présence de plusieurs variables (prédictrices) et que l'on souhaite expliquer les variations de la variable dépendante, puisque ceux-ci permettent de séparer les sources de variabilité afin d'étudier l'effet d'une variable *conditionnellement* aux autres variables, qui partagent plus ou moins de covariation ensemble.

4.7.3 Modèle général de la régression multiple

Le modèle général pour la régression multiple avec k variables indépendantes s'exprime sous la forme :

$$y_i = \alpha + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i$$

où α désigne l'ordonnée à l'origine du modèle, β_j le coefficient de régression partielle de la variable dépendante sur la variable indépendante j, x_j la variable indépendante j, et ε_i le résidu pour l'observation i[33]. En fait, dans ce modèle, le coefficient de régression partielle est égal à la pente de la régression linéaire de Y sur la variable indépendante X_j lorsque *toutes* les autres variables indépendantes sont maintenues constantes.

Afin de comparer l'*effet relatif*, ou *pouvoir prédicteur*, de chacune des variables indépendantes sur la variable dépendante, il est nécessaire de normaliser l'équation de régression multiple, à l'aide d'une transformation centrée-réduite des y_i et des x_{ij}, qui permet d'éliminer l'effet des différentes échelles (de mesure)[34]. Les variables d'étude deviennent ainsi :

$$y_i^* = \frac{y_i - \bar{y}}{s_y} \qquad x_{ij}^* = \frac{x_{ij} - \bar{x}}{s_{x_j}}$$

A partir de ces expressions normalisées des données, on obtient les coefficients normalisés de régression partielle β_j' qui sont également les coefficients de régression partielle pondérés par le rapport des écarts-type de la variable indépendante X_j sur la variable dépendante Y :

$$\beta_j' = \beta_j \frac{s_{x_j}}{s_y}$$

Ces coefficients normalisés indiquent le taux de changement de la variable dépendante Y en unités d'écart-type par unité d'écart-type de la variable indépendante X_j lorsque *toutes* les

[33] A la différence du cas bivarié (cf. régression linéaire simple, 4.6, p. 88), ce modèle linéaire ne définit plus une simple droite de régression, mais un hyperplan à k dimensions : une valeur y est ainsi une combinaison linéaire des x_j.

[34] Si la matrice initiale des y_i et des x_{ij} était sous forme centrée-réduite, il est bien entendu inutile d'effectuer cette transformation.

autres variables sont maintenues constantes, et quantifient de la sorte l'*effet relatif* de chacune des variables indépendantes sur la variable dépendante.

La qualité de l'ajustement du modèle, c'est-à-dire la part de variance des y_i expliquée par l'ensemble des k variables prédictrices (les x_{ij}), peut être évaluée à l'aide du coefficient de corrélation multiple r, qui mesure la corrélation entre les y_i observés et les $\hat{y}_i$ estimés, ou prédits. Comme dans le cas de la régression linéaire simple, on utilisera pour l'interprétation le coefficient de détermination R^2, et préférentiellement le coefficient de détermination ajusté, R^{*2}, ce dernier prenant en compte le nombre de variables prédictrices. Il sera notamment utile dans le cas où l'on souhaite comparer différents modèles comprenant un nombre différent de variables explicatives (cf. infra).

4.7.4 Conditions d'application

Conditions générales d'application

Dans ce type d'analyse, on postule que :
- les résidus sont indépendants, et se distribuent selon une loi normale ;
- la variance des résidus est homogène ;
- les relations entre la variable dépendante et les variables indépendantes sont linéaires ;
- l'effet de chaque variable indépendante est additif, c'est-à-dire qu'il n'y a pas d'interactions entre ces variables).

Ces hypothèses implicites peuvent être éprouvées comme dans le cas de la régression linéaire simple.

Problème de multicolinéarité

Le problème fréquemment rencontré dans la régression linéaire multiple est celui de la corrélation entre les variables indépendantes, ou *multicolinéarité*. Dans le cas idéal où les variables indépendantes sont deux à deux orthogonales (i.e. non corrélées), les estimés des coefficients de régression partielle, e.g. la pente d'une des droites de régression, peuvent être obtenus à l'aide de régressions simples. Cela n'est plus possible dès lors que les variables indépendantes sont corrélées, puisque l'on ne peut plus décomposer le modèle de régression linéaire multiple en un mélange additif des modèles simples de régression.

Le prix à payer est une augmentation des estimés des erreurs-type des coefficients, et par conséquent une diminution de la puissance des tests. La détection de la multicolinéarité peut être effectuée, entre autres, en examinant la matrice de corrélation, ou à l'aide d'un indicateur statistique spécifique comme le *coefficient d'inflation de la variance*.

4.7.5 Démarche de l'analyse et test d'hypothèse

La signification statistique du modèle complet peut être éprouvée par un test F comparant le rapport entre le carré moyen de la régression et le carré moyen résiduel aux valeurs critiques de la distribution du $\mathcal{F}$ de Fisher-Snedecor, au seuil $\alpha = 0.05$.

L'erreur-type de l'estimation correspond en fait à l'écart-type corrigé des valeurs résiduelles $y_i - \hat{y}_i$. Cette valeur de test permet de tester l'hypothèse nulle $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$, l'hypothèse alternative étant qu'au moins un des coefficients de régression partielle n'est pas nul.

Des tests t de Student peuvent ensuite être mis en oeuvre afin de tester la significativité des coefficients de régression partielle individuels, i.e. pour chacune des k variables indépendantes (prédictrices).

Différentes méthodes ont été proposées pour la sélection des prédictrices et la recherche du modèle décrivant le mieux les relations entre la variable dépendante et celles-ci, c'est-à-dire celui qui explique le mieux les variations de la variable dépendante en prenant en compte les prédictrices les plus pertinentes. Le critère de sélection le plus souvent retenu repose sur le coefficient de détermination (R^2 ou R^{*2}), et, lorsque le nombre de variables n'est pas trop élevé[35], on peut générer tous les modèles de régression, incluant 1, puis 2, puis 3, ... prédictrices, et sélectionner le modèle qui maximise le critère retenu. Bien que cette méthode *exhaustive* soit coûteuse en temps de calcul, elle assure de trouver le « meilleur » modèle. Nous n'aborderons pas ici la problématique liée à la définition du meilleur modèle ; cela dépend des critères que l'on se donne, du statut des variables et des hypothèses sous-jacentes à l'étude.

Lorsque la recherche exhaustive du modèle adéquat n'est pas envisageable, il est nécessaire d'utiliser des *méthodes de sélection de modèle*. Les principales méthodes de recherche du modèle par étapes, lorsque le nombre de variables est important (≥ 6), sont :

- l'approche par *sélection progressive* : on sélectionne une première variable, et on ajoute successivement les autres variables jusqu'à ce que le coefficeint de détermination n'augmente plus, c'est-à-dire jusqu'à ce que la qualité de l'ajustement ne soit plus améliorée par l'ajout de nouvelles prédictrices ;
- l'approche par *élimination rétrograde* : on débute avec un modèle incluant l'ensemble des prédictrices, et on élimine progressivement une à une celles qui ne contribuent pas significativement à réduire la variance résiduelle ;
- l'approche par *régression pas-à-pas* : elle est basée sur les deux méthodes précédentes, mais permet une évaluation séquentielle et non plus seulement hiérarchisée de chaque variable pour l'inclusion (ou l'exclusion) dans le modèle.

[35] Pour k variables indépendantes, le nombre de régressions à évaluer est de $2^k - 1$, ce qui donne par exemple 127 modèles à évaluer dans le cas de 7 variables.

Approche par sélection progressive

On évalue toutes les prédictrices pour sélectionner la variable la plus significative à inclure dans le modèle initial (modèle le plus simple à une seule variable). Ensuite, à chaque étape, on évalue l'ensemble des variables qui ne figurent pas dans le modèle et on inclue celle qui permet d'expliquer une part additionnelle significative de variabilité (indiqué par un test F). La condition d'arrêt de ce processus itératif est l'absence de variables additionnelles significatives.

Le principal désavantage de cette méthode est qu'une fois que les variables ont été incluses dans le modèle elles y restent, même si l'inclusion de nouvelles variables significatives les rend non-significatives.

Approche par élimination rétrograde

On débute avec le modèle complet (k prédictrices), et à chaque étape on enlève du modèle la variable qui est la moins significative (tel qu'indiqué par un test F), jusqu'à ce que les variables restantes soient toutes significatives.

Le même problème que dans l'approche par sélection progressive se pose : la significativité des variables peut dépendre de la présence/absence des autres variables.

Approche par régression pas-à-pas

On procède comme dans les cas précédents, mais à chaque nouvelle inclusion (ou exclusion dans le cas de l'élimination rétrograde) la significativité des variables présentes dans le modèle est évaluée : si l'inclusion d'une nouvelle variable rend une variable présente dans le modèle non-significative, alors cette dernière est retirée du modèle.

Exemple d'application X

> L'exemple suivant est tiré de [24] : il s'agit de données concernant les caractéristiques vitales et les capacités pulmonaires de sujets atteints d'une maladie infectieuse (Source : [24], chap. 9, pp. 149-158).

Les calculs manuels étant longs et fastidieux, ce sont les sorties texte du logiciel de statistique R qui sont repris pour illustrer très rapidement le principe de la régression multiple. La figure 4.15 résume le protocole multivarié en présentant les graphiques de corrélation pour l'ensemble des 10 variables numériques étudiées (de haut en bas : âge, sexe, taille, poids, masse corporelle, volume expiratoire, volume résiduel, capacité fonctionnelle résiduelle, capacité pulmonaire totale, pression expiratoire maximale).

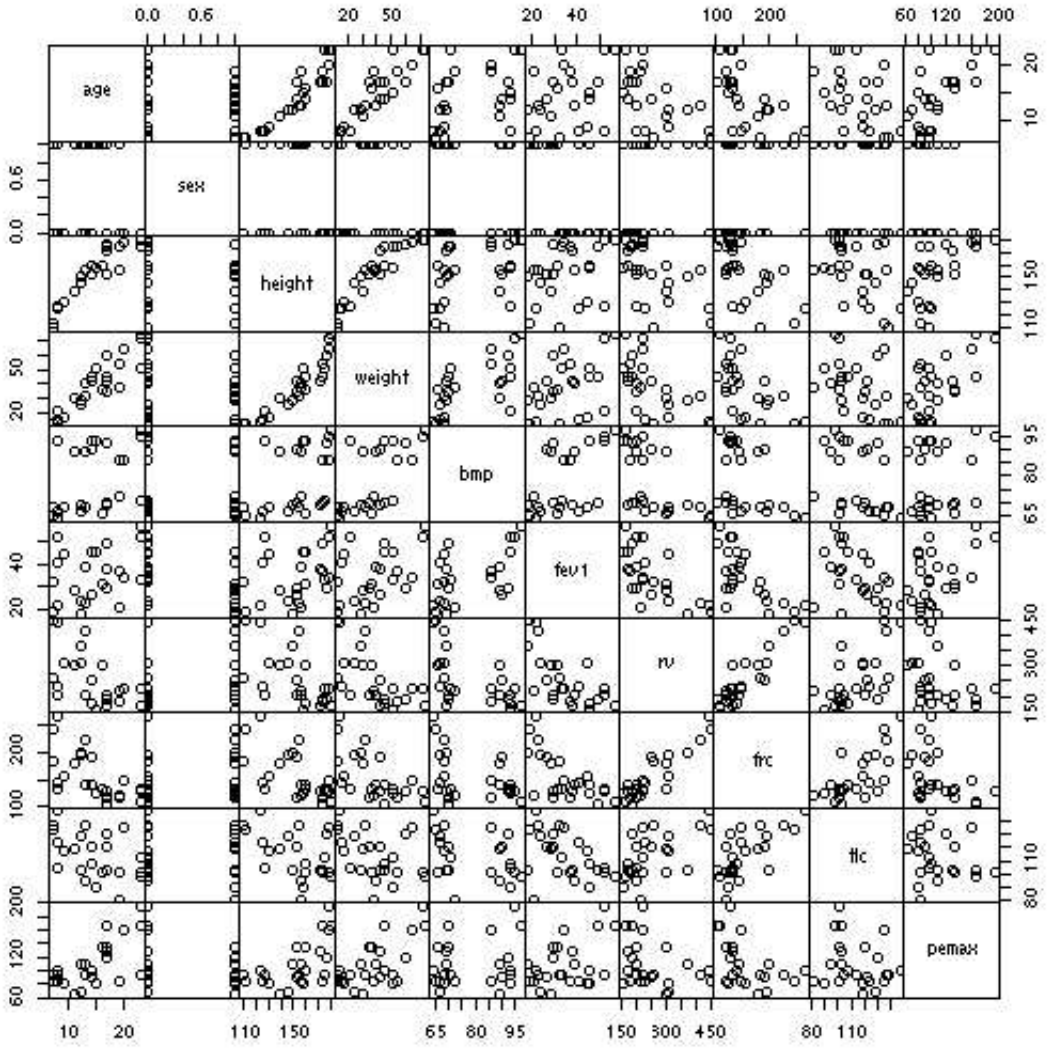

Fig. 4.15: Matrice des distributions bivariées pour les 10 variables numériques considérées.

A l'inspection de la figure, il apparaît certaines associations visibles à l'oeil nu, comme par exemple entre les variables taille (*height*) et poids (*weight*), taille et âge (*age*), age et poids. Ces relations ne sont pas surprenantes, mais on constate également des associations entre le volume résiduel (*rv*) et la capacité fonctionnelle résiduelle (*frc*), par exemple, ainsi que de nombreuses absences de corrélation clairement marquées, comme entre les variables taille et volume résiduel.

La variable d'étude étant la pression expiratoire maximale (*pemax*), on cherche à expliquer la variabilité observée en fonction des diféfrentes variables. Pour cela, on formule un modèle linéaire du type : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_9 x_9 + \varepsilon$, dans lequel

y désigne la pression expiratoire maximale, et les $x_1, \ldots, x_9$ les 9 variables d'étude (âge, sexe, ..., capacité pulmonaire résiduelle). On va chercher en fait le « meilleur » modèle qui permet d'expliquer le maximum de la variabilité observée, en étudiant la contribution de chacun des termes (i.e. des variables explicatives ou prédictrices) aux résultats observés pour la variable dépendante, conditionnellement aux autres variables.

Tests t. En testant chacune des 9 variables isolément à l'aide de tests t de Student, on constate qu'aucun des tests n'est significatif. Néanmoins, le test F sur la part de variabilité expliquée par le modèle liénaire est significatif ($F_{obs} = 2.929$, $p < 0.05$), comme l'indique le listing ci-dessous :

```
Residuals:
    Min      1Q  Median      3Q     Max
-37.338 -11.532   1.081  13.386  33.405

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 176.0582   225.8912   0.779    0.448
age          -2.5420     4.8017  -0.529    0.604
sex          -3.7368    15.4598  -0.242    0.812
height       -0.4463     0.9034  -0.494    0.628
weight        2.9928     2.0080   1.490    0.157
bmp          -1.7449     1.1552  -1.510    0.152
fev1          1.0807     1.0809   1.000    0.333
rv            0.1970     0.1962   1.004    0.331
frc          -0.3084     0.4924  -0.626    0.540
tlc           0.1886     0.4997   0.377    0.711

Residual standard error: 25.47 on 15 degrees of freedom
Multiple R-Squared: 0.6373,     Adjusted R-squared: 0.4197
F-statistic: 2.929 on 9 and 15 DF,  p-value: 0.03195
```

La part de variabilité expliquée par le modèle incluant toutes les variables indépendantes est indiquée par le coefficeint de détermination $R^2 = 0.6373$ (ou $R^{*2} = 0.4197$ lorsque celui-ci est ajusté en fonction du nombre de variables étudiées). Celle-ci peut être considérée comme importante (~ 64 %).

En fait, les tests t calculés n'indiquent pas ce qui se passe lorsque l'on enlève une variable explicative du modèle général, c'est-à-dire dans un modèle réduit. Cela indique simplement dans le cas du modèle général qu'aucun des 9 coefficients de régression partielle n'est statistiquement significatif. En revanche, le test F indique qu'au moins un des coefficients de régression partielle est différent de zéro

Analyse de variance. Le tableau de décomposition de la variance est indiqué dans le listing ci-dessous :

```
Analysis of Variance Table

Response: pemax
```

```
          Df  Sum Sq Mean Sq F value   Pr(>F)
age        1 10098.5 10098.5 15.5661 0.001296 **
sex        1   955.4   955.4  1.4727 0.243680
height     1   155.0   155.0  0.2389 0.632089
weight     1   632.3   632.3  0.9747 0.339170
bmp        1  2862.2  2862.2  4.4119 0.053010 .
fev1       1  1549.1  1549.1  2.3878 0.143120
rv         1   561.9   561.9  0.8662 0.366757
frc        1   194.6   194.6  0.2999 0.592007
tlc        1    92.4    92.4  0.1424 0.711160
Residuals 15  9731.2   648.7
--
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Le tableau d'ANOVA indique qu'une fois que le facteur « âge » est inclus dans le modèle, l'inclusion progressive des autres facteurs n'améliore pas sensiblement le modèle initial. On peut évaluer la part de variance totale expliquée par les 8 autres facteurs et considérer un nouveau modèle dans lequel seraient inclus deux facteurs seulement : les facteurs « âge » et « autres », ce dernier comprenant les 8 facteurs autres que le facteur « âge ».

```
Analysis of Variance Table

Model 1: pemax ~ age + sex + height + weight + bmp + fev1 + rv + frc +
    tlc
Model 2: pemax ~ age
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     15  9731.2
2     23 16734.2 -8   -7002.9 1.3493 0.2936
```

On itèrerait ensuite le processus selon les méthodes de sélection décrites précedemment ...

4.8 Analyse de covariance

4.8.1 Principe général

L'analyse de covariance constitue un compromis entre l'analyse de variance à plusieurs facteurs, dans laquelle on étudie l'effet de plusieurs variables indépendantes qualitatives, sur la variable dépendante, et la régression multiple, dans laquelle on travaille avec des variables de type quantitatives. L'analyse de covariance, ou ANCOVA, permet en effet d'étudier l'effet de variables indépendantes des deux types — qualitative (i.e. discontinue) et quantitative (i.e. continue) — sur la variable dépendante.

L'ANCOVA est typiquement utilisée lorsque l'on s'intéresse à l'effet d'une variable indépendante quantitative (continue) sur la variable dépendante mesurée, en fonction d'une autre

variable indépendante qualitative (souvent dichotomisée). Par exemple, on pourrait s'intéresser à l'influence du taux d'alcool dans le sang (variable indépendante, numérique continue) sur les temps de réaction à des stimuli visuels présentés en vision périphérique (variable dépendante, numérique continue), en fonction du sexe du sujet (variable indépendante, nominale).

4.8.2 Modèle de l'ANCOVA

Le modèle général de l'analyse de covariance, très semblable aux modèles linéaires de l'ANOVA d'ordre n et de la régression, est généralement formulé comme suit :

$$Y_{ij} = \mu + \alpha_i + \beta(X_{ij} - \bar{X}_i) + \varepsilon_{ij}$$

où i désigne l'indice des catégories de la variable qualitative, et j l'indice des observations dans chaque catégorie. Les paramètres intervenant dans ce modèle sont: μ, la moyenne générale, α_i, l'écart entre la moyenne du groupe i et la moyenne générale, β, la pente de la relation entre la variable dépendante et la variable indépendante quantitative (continue), et X_i, la moyenne de la variable indépendante quantitative pour la catégorie i.

4.8.3 Conditions d'application

Les hypothèses implicites de l'ANCOVA recouvrent celles de l'ANOVA et de la régression, et on postule que :

- les résidus sont indépendants, et se distribuent normalement ;
- la variance des résidus est homogène entre les différentes modalités de la variable indépendante (homoscédasticité) ;
- la relation entre la variable indépendante et la variable dépendante est linéaire ;
- la pente de cette relation linéaire est la même pour les différentes modalités de la variable.

4.8.4 Hypothèse

Lorsque le protocole expérimental comporte deux variables indépendantes, dont l'une est de type quantitative continue et l'autre de type qualitative à k modalités, la question qui se pose est de savoir si la régression de la variable dépendante sur la variable indépendante quantitative est la même quelque soit la modalité ou le niveau de la variable indépendante qualitative. En d'autres termes, cela revient à formuler l'hypothèse nulle que les différentes pentes de régression, pour chaque modalité de la variable qualitative, ne diffèrent pas, soit $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k$. L'hypothèse alternative la plus appropriée est généralement la non-égalité d'au moins deux des pentes, que l'on formalise, comme dans le modèle de l'ANOVA : $H_1 : \beta_1 \neq \beta_2 \neq \ldots \neq \beta_k$.

4.8.5 Test d'hypothèse

Il s'agit ici, comme dans la régression, d'un test d'ajustement à un modèle linéaire. L'analyse se fait étape par étape, en ajustant une série de modèles, du plus complexe (i.e. le modèle de base incluant tous les termes) au plus simple, et en les comparant deux à deux de manière hiérarchique (le premier au deuxième, le deuxième au troisième, etc.). A chaque étape, le modèle de base est évalué en testant la significativité des différents termes inclus dans le modèle à l'aide d'une comparaison de la variance résiduelle du modèle à celle du modèle plus simple duquel le terme testé a été exclus. On simplifie ainsi le modèle de base en éliminant les termes qui n'expliquent pas une quantité significative de la variabilité observée.

Le modèle utilisé pour le test d'ajustement linéaire est le même que le modèle général exposé précédemment, mais inclut une pente différente pour chaque modalité de la variable qualitative, c'est-à-dire :

$$Y_{ij} = \mu + \alpha_i + \beta_i(X_{ij} - \bar{X}_i) + \varepsilon_{ij}$$

En bref, on effectue des analyses de régression classiques pour chaque modalité de la variable qualitative, et les procédures de test sont identiques à celles que nous avons décrites dans le chapitre correspondant (cf. section 4.6).

Pour de plus amples détails concernant l'ANCOVA, le lecteur intéressé peut consulter [26], [20], [18] et [21] (pour les développements mathématiques).

Ouvrages de base

[1] P. Dalgaard. *Introductory Statistics with R.* Springer-Verlag, 2002.

[2] J. Goupy. *Introduction aux plans d'expériences.* Dunod, 2001.

[3] H. Rouanet, B. Le Roux, M.-C. Bert. *Statistique en Sciences Humaines : Procédures Naturelles.* Dunod, 1987.

[4] J.-M. Bouroche, G. Saporta. *L'Analyse des Données.* Presses Universitaires de France, 2002.

[5] M. Wolff, D. Corroyer. *L'Analyse Statistique des Données en Psychologie.* A. Colin, 2003.

[6] S. Miller. *Experimental Design and Statistics.* Brunner-Routledge, 1984.

[7] S. Siegel, N.J. Castellan. *Nonparametric Statistics for the Behavioral Sciences.* McGraw-Hill, 1988.

[8] S.S. Stevens. *Psychophysics.* Transaction Books, 2000.

[9] G. van Belle. *Statistical Rules of Thumb.* Wiley & Sons, 2002.

[10] J. H. Zar. *Biostatistical Analysis.* Prentice Hall, 1996.

[11] J.-P. Marques de Sá. *Applied Statistics using SPSS, Statistica and Matlab.* Springer-Verlag, 2003.

[12] Y. Dodge. *Analyse de régression appliquée, manuel et exercices.* Dunod, 2000.

[13] H. Rouanet. *L'inférence statistique dans la démarche du chercheur.* Peter Lang, 1991.

Pour aller plus loin

[14] D. Benoist, S. Tourbier, Y. Tourbier. *Plans d'expériences, construction, analyse.* Tec-Doc, 1994.

[15] D. Corroyer, E. Devouche, J.-M., Bernard, P. Bonnet, Y. Savina. Comparaison de six logiciels pour l'analyse de la variance d'un plan s<a2*b2> déséquilibré. *L'Année Psychologique*, 2003.

[16] H. Rouanet, D. Lépine. Comparison between treatments in a repeated-measurement design: Anova and multivariate methods. *British Journal of Mathematical and Statistical Psychology*, 1970.

[17] H. Rouanet, D. Lépine. Structures linéaires et analyse des comparaisons. *Mathématiques et Sciences Humaines*, 1976.

[18] J.-P. Nakache, J. Confais. *Statistique explicative appliquée.* Technip, 2003.

[19] D. Montgomery. *Design and Analysis of Experiments.* Wiley & Sons, 1997.

[20] N. Draper, H. Smith. *Applied Regression Analysis.* Wiley & Sons, 1998.

[21] R.A. Johnson, D.W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice-Hall, 1992.

[22] G. Saporta. *Probabilités, Analyse des Données, Statistique*. Technip, 1990.

[23] W.N. Venables, B.D. Ripley. *Modern Applied Statistics with S*. Springer-Verlag, 2002.

Sources de données

[24] P. Dalgaard. *Introductory Statistics with R*. Springer-Verlag, 2002.

[25] Student. The probable error of a mean. *Biometrika*, 1908.

[26] J. H. Zar. *Biostatistical Analysis*. Prentice Hall, 1996.

A Tests d'ajustement à des distributions théoriques

Principe général

L'objet des tests d'ajustement à des distributions théoriques est de déterminer si une distribution empirique, c'est-à-dire une distribution d'effectifs observés à l'issue d'une expérimentation, se conforme à une distribution théorique donnée. La technique généralement employée par ce type de test consiste à mesurer l'écart entre les deux distributions, soit à l'aide des fréquences observées et des fréquences théoriques, soit à l'aide de la distribution cumulée des données et de la distribution cumulée théorique.

Ces tests sont utiles lorsque l'on travaille avec de *petits échantillons*, puisque pour les grands échantillons les tests paramétriques demeurent relativement insensibles aux déviations à la normalité.

Nous présentons dans ce chapitre les principaux tests d'ajustement à une distribution normale, puisque ce sont ceux-ci qui sont utilisés lorsque l'on travaille avec des données numériques. Pour les données qualitatives, les tests appropriés sont ceux du χ^2 et du G, et le lecteur intéressé pourra consulter [26] (chap. 22) pour de plus amples informations sur ces tests.

Tests d'ajustement à une distribution normale

Test de Kolmogorov-Smirnov

Ce test, applicable à des données continues, permet de comparer la distribution relative cumulée observée à la distribution théorique. La valeur de test, D_{max}, correspond à la valeur absolue de la différence maximum entre les deux distributions cumulées :

$$d_{max} = max(|f_i - \hat{f}_i|, |f_{i-1} - \hat{f}_i|)$$

où $f_i = \frac{i}{n}$ représentent les fréquences relatives cumulées empiriques ordonnée par ordre croissant, i désignant le rang de l'observation et n l'effectif total (i.e. le nombre total d'observations). Les fréquences cumulées théoriques $\hat{f}_i$ sont calculées à l'aide de la table des proportions de la distribution $Z(0; 1^2$ après transformation centrée-réduite des données empiriques $\frac{x_i - \bar{x}}{s}$. Cette valeur de test doit être comparée aux valeurs critiques correspondant au test de Kolmogorov-Smirnov (cf. Annexe B, p. 120, ou [26], [7]).

Test de Lilliefors

Le test de Lilliefors est une variante du test de Kolmogorov-Smirnov, préférentiellement utilisé lorsque les paramètres de la population parente — moyenne et variance — ne sont pas connus.

Test de Wilks-Shapiro

Ce test repose sur l'utilisation d'un diagramme de probabilité, incluant une échelle de probabilité normale selon laquelle une distribution normale apparaît comme une droite. La valeur de test, W, est le carré du coefficient de corrélation entre les valeurs observées x_i et leurs équivalents z_i basés sur leur fréquence relative cumulée. Cela permet de mesurer le degré d'alignement des valeurs observées sur une droite. De nombreux auteurs s'accordent sur le fait que ce test est le plus puissant dans le cas de petits échantillons.

B Lois de distribution et tables statistiques

Tab. B.1: Loi normale centrée réduite. Fonction de répartition de la loi $\mathcal{N}(0,1)$.

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Tab. B.2: Loi normale centrée réduite. Fractiles $u(\beta)$ d'ordre β de la loi $\mathcal{N}(0,1)$.

β	**0.000**	**0.001**	**0.002**	**0.003**	**0.004**	**0.005**	**0.006**	**0.007**	**0.008**	**0.009**
0.50	0.0000	0.0025	0.0050	0.0075	0.0100	0.0125	0.0150	0.0175	0.0201	0.0226
0.51	0.0251	0.0276	0.0301	0.0326	0.0351	0.0376	0.0401	0.0426	0.0451	0.0476
0.52	0.0502	0.0527	0.0552	0.0577	0.0602	0.0627	0.0652	0.0677	0.0702	0.0728
0.53	0.0753	0.0778	0.0803	0.0828	0.0853	0.0878	0.0904	0.0929	0.0954	0.0979
0.54	0.1004	0.1030	0.1055	0.1080	0.1105	0.1130	0.1156	0.1181	0.1206	0.1231
0.55	0.1257	0.1282	0.1307	0.1332	0.1358	0.1383	0.1408	0.1434	0.1459	0.1484
0.56	0.1510	0.1535	0.1560	0.1586	0.1611	0.1637	0.1662	0.1687	0.1713	0.1738
0.57	0.1764	0.1789	0.1815	0.1840	0.1866	0.1891	0.1917	0.1942	0.1968	0.1993
0.58	0.2019	0.2045	0.2070	0.2096	0.2121	0.2147	0.2173	0.2198	0.2224	0.2250
0.59	0.2275	0.2301	0.2327	0.2353	0.2378	0.2404	0.2430	0.2456	0.2482	0.2508
0.60	0.2533	0.2559	0.2585	0.2611	0.2637	0.2663	0.2689	0.2715	0.2741	0.2767
0.61	0.2793	0.2819	0.2845	0.2871	0.2898	0.2924	0.2950	0.2976	0.3002	0.3029
0.62	0.3055	0.3081	0.3107	0.3134	0.3160	0.3186	0.3213	0.3239	0.3266	0.3292
0.63	0.3319	0.3345	0.3372	0.3398	0.3425	0.3451	0.3478	0.3505	0.3531	0.3558
0.64	0.3585	0.3611	0.3638	0.3665	0.3692	0.3719	0.3745	0.3772	0.3799	0.3826
0.65	0.3853	0.3880	0.3907	0.3934	0.3961	0.3989	0.4016	0.4043	0.4070	0.4097
0.66	0.4125	0.4152	0.4179	0.4207	0.4234	0.4261	0.4289	0.4316	0.4344	0.4372
0.67	0.4399	0.4427	0.4454	0.4482	0.4510	0.4538	0.4565	0.4593	0.4621	0.4649
0.68	0.4677	0.4705	0.4733	0.4761	0.4789	0.4817	0.4845	0.4874	0.4902	0.4930
0.69	0.4959	0.4987	0.5015	0.5044	0.5072	0.5101	0.5129	0.5158	0.5187	0.5215
0.70	0.5244	0.5273	0.5302	0.5330	0.5359	0.5388	0.5417	0.5446	0.5476	0.5505
0.71	0.5534	0.5563	0.5592	0.5622	0.5651	0.5681	0.5710	0.5740	0.5769	0.5799
0.72	0.5828	0.5858	0.5888	0.5918	0.5948	0.5978	0.6008	0.6038	0.6068	0.6098
0.73	0.6128	0.6158	0.6189	0.6219	0.6250	0.6280	0.6311	0.6341	0.6372	0.6403
0.74	0.6433	0.6464	0.6495	0.6526	0.6557	0.6588	0.6620	0.6651	0.6682	0.6713
0.75	0.6745	0.6776	0.6808	0.6840	0.6871	0.6903	0.6935	0.6967	0.6999	0.7031
0.76	0.7063	0.7095	0.7128	0.7160	0.7192	0.7225	0.7257	0.7290	0.7323	0.7356
0.77	0.7388	0.7421	0.7454	0.7488	0.7521	0.7554	0.7588	0.7621	0.7655	0.7688
0.78	0.7722	0.7756	0.7790	0.7824	0.7858	0.7892	0.7926	0.7961	0.7995	0.8030
0.79	0.8064	0.8099	0.8134	0.8169	0.8204	0.8239	0.8274	0.8310	0.8345	0.8381
0.80	0.8416	0.8452	0.8488	0.8524	0.8560	0.8596	0.8633	0.8669	0.8705	0.8742
0.81	0.8779	0.8816	0.8853	0.8890	0.8927	0.8965	0.9002	0.9040	0.9078	0.9116
0.82	0.9154	0.9192	0.9230	0.9269	0.9307	0.9346	0.9385	0.9424	0.9463	0.9502
0.83	0.9542	0.9581	0.9621	0.9661	0.9701	0.9741	0.9782	0.9822	0.9863	0.9904
0.84	0.9945	0.9986	1.0027	1.0069	1.0110	1.0152	1.0194	1.0237	1.0279	1.0322
0.85	1.0364	1.0407	1.0450	1.0494	1.0537	1.0581	1.0625	1.0669	1.0714	1.0758
0.86	1.0803	1.0848	1.0893	1.0939	1.0985	1.1031	1.1077	1.1123	1.1170	1.1217
0.87	1.1264	1.1311	1.1359	1.1407	1.1455	1.1503	1.1552	1.1601	1.1650	1.1700
0.88	1.1750	1.1800	1.1850	1.1901	1.1952	1.2004	1.2055	1.2107	1.2160	1.2212
0.89	1.2265	1.2319	1.2372	1.2426	1.2481	1.2536	1.2591	1.2646	1.2702	1.2759
0.90	1.2816	1.2873	1.2930	1.2988	1.3047	1.3106	1.3165	1.3225	1.3285	1.3346
0.91	1.3408	1.3469	1.3532	1.3595	1.3658	1.3722	1.3787	1.3852	1.3917	1.3984
0.92	1.4051	1.4118	1.4187	1.4255	1.4325	1.4395	1.4466	1.4538	1.4611	1.4684
0.93	1.4758	1.4833	1.4909	1.4985	1.5063	1.5141	1.5220	1.5301	1.5382	1.5464
0.94	1.5548	1.5632	1.5718	1.5805	1.5893	1.5982	1.6072	1.6164	1.6258	1.6352
0.95	1.6449	1.6546	1.6646	1.6747	1.6849	1.6954	1.7060	1.7169	1.7279	1.7392
0.96	1.7507	1.7624	1.7744	1.7866	1.7991	1.8119	1.8250	1.8384	1.8522	1.8663
0.97	1.8808	1.8957	1.9110	1.9268	1.9431	1.9600	1.9774	1.9954	2.0141	2.0335
0.98	2.0537	2.0749	2.0969	2.1201	2.1444	2.1701	2.1973	2.2262	2.2571	2.2904
0.99	2.3263	2.3656	2.4089	2.4573	2.5121	2.5758	2.6521	2.7478	2.8782	3.0902

Tab. B.3: Table des valeurs critiques du ⊔ de Student.

α /2 / α / ddl	.025 / .05	.005 / .01	.0005 / .001
1	12.706	63.657	636.619
2	4.303	9.925	31.599
3	3.182	5.841	12.924
4	2.776	4.604	8.610
5	2.571	4.032	6.869
6	2.447	3.707	5.959
7	2.365	3.499	5.408
8	2.306	3.355	5.041
9	2.262	3.250	4.781
10	2.228	3.169	4.587
11	2.201	3.106	4.437
12	2.179	3.055	4.318
13	2.160	3.012	4.221
14	2.145	2.977	4.140
15	2.131	2.947	4.073
16	2.120	2.921	4.015
17	2.110	2.898	3.965
18	2.101	2.878	3.922
19	2.093	2.861	3.883
20	2.086	2.845	3.850
21	2.080	2.831	3.819
22	2.074	2.819	3.792
23	2.069	2.807	3.768
24	2.064	2.797	3.745
25	2.060	2.787	3.725
26	2.056	2.779	3.707
27	2.052	2.771	3.690
28	2.048	2.763	3.674
29	2.045	2.756	3.659
30	2.042	2.750	3.646
40	2.021	2.704	3.551
60	2.000	2.660	3.460
120	1.980	2.617	3.373
30000	1.960	2.576	3.291

Tab. B.4: Loi de distribution du χ^2. Fractiles $k_m(\beta)$ d'ordre β de la loi du Khi-2 à m degrés de liberté.

$\mathbf{m}\beta$	**0.001**	**0.005**	**0.010**	**0.025**	**0.05**	**0.1000**	**0.5000**	**0.9000**	**0.9500**	**0.9750**	**0.9900**	**0.9950**	**0.9990**
1	0.000	0.000	0.000	0.001	0.004	0.016	0.455	2.706	3.841	5.024	6.635	7.879	10.828
2	0.002	0.010	0.020	0.051	0.103	0.211	1.386	4.605	5.991	7.378	9.210	10.597	13.816
3	0.024	0.072	0.115	0.216	0.352	0.584	2.366	6.251	7.815	9.348	11.345	12.838	16.266
4	0.091	0.207	0.297	0.484	0.711	1.064	3.357	7.779	9.488	11.143	13.277	14.860	18.467
5	0.210	0.412	0.554	0.831	1.145	1.610	4.351	9.236	11.070	12.833	15.086	16.750	20.515
6	0.381	0.676	0.872	1.237	1.635	2.204	5.348	10.645	12.592	14.449	16.812	18.548	22.458
7	0.598	0.989	1.239	1.690	2.167	2.833	6.346	12.017	14.067	16.013	18.475	20.278	24.322
8	0.857	1.344	1.646	2.180	2.733	3.490	7.344	13.362	15.507	17.535	20.090	21.955	26.124
9	1.152	1.735	2.088	2.700	3.325	4.168	8.343	14.684	16.919	19.023	21.666	23.589	27.877
10	1.479	2.156	2.558	3.247	3.940	4.865	9.342	15.987	18.307	20.483	23.209	25.188	29.588
11	1.834	2.603	3.053	3.816	4.575	5.578	10.341	17.275	19.675	21.920	24.725	26.757	31.264
12	2.214	3.074	3.571	4.404	5.226	6.304	11.340	18.549	21.026	23.337	26.217	28.300	32.909
13	2.617	3.565	4.107	5.009	5.892	7.042	12.340	19.812	22.362	24.736	27.688	29.819	34.528
14	3.041	4.075	4.660	5.629	6.571	7.790	13.339	21.064	23.685	26.119	29.141	31.319	36.123
15	3.483	4.601	5.229	6.262	7.261	8.547	14.339	22.307	24.996	27.488	30.578	32.801	37.697
16	3.942	5.142	5.812	6.908	7.962	9.312	15.338	23.542	26.296	28.845	32.000	34.267	39.252
17	4.416	5.697	6.408	7.564	8.672	10.085	16.338	24.769	27.587	30.191	33.409	35.718	40.790
18	4.905	6.265	7.015	8.231	9.390	10.865	17.338	25.989	28.869	31.526	34.805	37.156	42.312
19	5.407	6.844	7.633	8.907	10.117	11.651	18.338	27.204	30.144	32.852	36.191	38.582	43.820
20	5.921	7.434	8.260	9.591	10.851	12.443	19.337	28.412	31.410	34.170	37.566	39.997	45.315
21	6.447	8.034	8.897	10.283	11.591	13.240	20.337	29.615	32.671	35.479	38.932	41.401	46.797
22	6.983	8.643	9.542	10.982	12.338	14.041	21.337	30.813	33.924	36.781	40.289	42.796	48.268
23	7.529	9.260	10.196	11.689	13.091	14.848	22.337	32.007	35.172	38.076	41.638	44.181	49.728
24	8.085	9.886	10.856	12.401	13.848	15.659	23.337	33.196	36.415	39.364	42.980	45.559	51.179
25	8.649	10.520	11.524	13.120	14.611	16.473	24.337	34.382	37.652	40.646	44.314	46.928	52.620
26	9.222	11.160	12.198	13.844	15.379	17.292	25.336	35.563	38.885	41.923	45.642	48.290	54.052
27	9.803	11.808	12.879	14.573	16.151	18.114	26.336	36.741	40.113	43.195	46.963	49.645	55.476
28	10.391	12.461	13.565	15.308	16.928	18.939	27.336	37.916	41.337	44.461	48.278	50.993	56.892
29	10.986	13.121	14.256	16.047	17.708	19.768	28.336	39.087	42.557	45.722	49.588	52.336	58.301
30	11.588	13.787	14.953	16.791	18.493	20.599	29.336	40.256	43.773	46.979	50.892	53.672	59.703
31	12.196	14.458	15.655	17.539	19.281	21.434	30.336	41.422	44.985	48.232	52.191	55.003	61.098
32	12.811	15.134	16.362	18.291	20.072	22.271	31.336	42.585	46.194	49.480	53.486	56.328	62.487
33	13.431	15.815	17.074	19.047	20.867	23.110	32.336	43.745	47.400	50.725	54.776	57.648	63.870
34	14.057	16.501	17.789	19.806	21.664	23.952	33.336	44.903	48.602	51.966	56.061	58.964	65.247
35	14.688	17.192	18.509	20.569	22.465	24.797	34.336	46.059	49.802	53.203	57.342	60.275	66.619
36	15.324	17.887	19.233	21.336	23.269	25.643	35.336	47.212	50.998	54.437	58.619	61.581	67.985
37	15.965	18.586	19.960	22.106	24.075	26.492	36.336	48.363	52.192	55.668	59.893	62.883	69.346
38	16.611	19.289	20.691	22.878	24.884	27.343	37.335	49.513	53.384	56.896	61.162	64.181	70.703
39	17.262	19.996	21.426	23.654	25.695	28.196	38.335	50.660	54.572	58.120	62.428	65.476	72.055
40	17.916	20.707	22.164	24.433	26.509	29.051	39.335	51.805	55.758	59.342	63.691	66.766	73.402
41	18.575	21.421	22.906	25.215	27.326	29.907	40.335	52.949	56.942	60.561	64.950	68.053	74.745
42	19.239	22.138	23.650	25.999	28.144	30.765	41.335	54.090	58.124	61.777	66.206	69.336	76.084
43	19.906	22.859	24.398	26.785	28.965	31.625	42.335	55.230	59.304	62.990	67.459	70.616	77.419
44	20.576	23.584	25.148	27.575	29.787	32.487	43.335	56.369	60.481	64.201	68.710	71.893	78.750
45	21.251	24.311	25.901	28.366	30.612	33.350	44.335	57.505	61.656	65.410	69.957	73.166	80.077
46	21.929	25.041	26.657	29.160	31.439	34.215	45.335	58.641	62.830	66.617	71.201	74.437	81.400
47	22.610	25.775	27.416	29.956	32.268	35.081	46.335	59.774	64.001	67.821	72.443	75.704	82.720
48	23.295	26.511	28.177	30.755	33.098	35.949	47.335	60.907	65.171	69.023	73.683	76.969	84.037
49	23.983	27.249	28.941	31.555	33.930	36.818	48.335	62.038	66.339	70.222	74.919	78.231	85.351
50	24.674	27.991	29.707	32.357	34.764	37.689	49.335	63.167	67.505	71.420	76.154	79.490	86.661

Tab. B.5: Table des valeurs critiques du $\mathcal{F}$ de Fisher-Snedecor.

Ddl1	!	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ddl2	α															
2	.05	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4
	.01	98.5	99.0	99.2	99.3	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
	.001	998	999	999	999	999	999	999	999	999	999	999	999	999	999	999
3	.05	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.72	8.70
	.01	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	27.1	27.0	26.9	26.9
	.001	167	149	141	137	135	133	132	131	130	129	129	128	128	128	127
4	.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86
	.01	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.5	14.4	14.3	14.2	14.2
	.001	74.1	61.2	56.2	53.4	51.7	50.5	49.7	49.0	48.5	48.1	47.7	47.4	47.2	46.9	46.8
5	.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62
	.01	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89	9.83	9.77	9.72
	.001	47.2	37.1	33.2	31.1	29.8	28.8	28.2	27.6	27.2	26.9	26.6	26.4	26.2	26.1	25.9
6	.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94
	.01	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.66	7.61	7.56
	.001	35.5	27.0	23.7	21.9	20.8	20.0	19.5	19.0	18.7	18.4	18.2	18.0	17.8	17.7	17.6
7	.05	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51
	.01	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.41	6.36	6.31
	.001	29.2	21.7	18.8	17.2	16.2	15.5	15.0	14.6	14.3	14.1	13.9	13.7	13.6	13.4	13.3
8	.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22
	.01	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.61	5.56	5.52
	.001	25.4	18.5	15.8	14.4	13.5	12.9	12.4	12.0	11.8	11.5	11.4	11.2	11.1	10.9	10.8
9	.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01
	.01	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	5.05	5.01	4.96
	.001	22.9	16.4	13.9	12.6	11.7	11.1	10.7	10.4	10.1	9.89	9.72	9.57	9.44	9.33	9.24
10	.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.84
	.01	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.65	4.60	4.56
	.001	21.0	14.9	12.6	11.3	10.5	9.93	9.52	9.20	8.96	8.75	8.59	8.45	8.32	8.22	8.13
11	.05	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72
	.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.34	4.29	4.25
	.001	19.7	13.8	11.6	10.3	9.58	9.05	8.66	8.35	8.12	7.92	7.76	7.63	7.51	7.41	7.32
12	.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62
	.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.10	4.05	4.01
	.001	18.6	13.0	10.8	9.63	8.89	8.38	8.00	7.71	7.48	7.29	7.14	7.00	6.89	6.79	6.71
13	.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53
	.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.91	3.86	3.82
	.001	17.8	12.3	10.2	9.07	8.35	7.86	7.49	7.21	6.98	6.80	6.65	6.52	6.41	6.31	6.23
14	.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46
	.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.75	3.70	3.66
	.001	17.1	11.8	9.73	8.62	7.92	7.44	7.08	6.80	6.58	6.40	6.26	6.13	6.02	5.93	5.85
15	.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40
	.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.61	3.56	3.52
	.001	16.6	11.3	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.94	5.81	5.71	5.62	5.54
16	.05	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35
	.01	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.50	3.45	3.41
	.001	16.1	11.0	9.01	7.94	7.27	6.80	6.46	6.19	5.98	5.81	5.67	5.55	5.44	5.35	5.27
17	.05	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31
	.01	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46	3.40	3.35	3.31
	.001	15.7	10.7	8.73	7.68	7.02	6.56	6.22	5.96	5.75	5.58	5.44	5.32	5.22	5.13	5.05

Dd1	!	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Dd2	α															
18	.05	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27
	.01	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.32	3.27	3.23
	.001	15.4	10.4	8.49	7.46	6.81	6.36	6.02	5.76	5.56	5.39	5.25	5.13	5.03	4.94	4.87
19	.05	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23
	.01	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.24	3.19	3.15
	.001	15.1	10.2	8.28	7.27	6.62	6.18	5.85	5.59	5.39	5.22	5.08	4.97	4.87	4.78	4.70
20	.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20
	.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18	3.13	3.09
	.001	14.8	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.94	4.82	4.72	4.64	4.56
21	.05	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20	2.18
	.01	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24	3.17	3.12	3.07	3.03
	.001	14.6	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11	4.95	4.81	4.70	4.60	4.51	4.44
22	.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17	2.15
	.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.07	3.02	2.98
	.001	14.4	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99	4.83	4.70	4.58	4.49	4.40	4.33
23	.05	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.18	2.15	2.13
	.01	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	3.02	2.97	2.93
	.001	14.2	9.47	7.67	6.70	6.08	5.65	5.33	5.09	4.89	4.73	4.60	4.48	4.39	4.30	4.23
24	.05	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.15	2.13	2.11
	.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03	2.98	2.93	2.89
	.001	14.0	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80	4.64	4.51	4.39	4.30	4.21	4.14
25	.05	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11	2.09
	.01	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	3.06	2.99	2.94	2.89	2.85
	.001	13.9	9.22	7.45	6.49	5.88	5.46	5.15	4.91	4.71	4.56	4.42	4.31	4.22	4.13	4.06
30	.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01
	.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.79	2.74	2.70
	.001	13.3	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.11	4.00	3.91	3.82	3.75
35	.05	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.07	2.04	2.01	1.99	1.96
	.01	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.80	2.74	2.69	2.64	2.60
	.001	12.9	8.47	6.79	5.88	5.30	4.89	4.59	4.36	4.18	4.03	3.90	3.79	3.70	3.62	3.55
40	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95	1.92
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61	2.56	2.52
	.001	12.6	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.75	3.64	3.55	3.47	3.40
45	.05	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	2.01	1.97	1.94	1.92	1.89
	.01	7.23	5.11	4.25	3.77	3.45	3.23	3.07	2.94	2.83	2.74	2.67	2.61	2.55	2.51	2.46
	.001	12.4	8.09	6.45	5.56	5.00	4.61	4.32	4.09	3.91	3.76	3.64	3.53	3.44	3.36	3.29
50	.05	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.92	1.89	1.87
	.01	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.63	2.56	2.51	2.46	2.42
	.001	12.2	7.96	6.34	5.46	4.90	4.51	4.22	4.00	3.82	3.67	3.55	3.44	3.35	3.27	3.20
60	.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.89	1.86	1.84
	.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.44	2.39	2.35
	.001	12.0	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69	3.54	3.42	3.32	3.23	3.15	3.08
80	.05	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.91	1.88	1.84	1.82	1.79
	.01	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.48	2.42	2.36	2.31	2.27
	.001	11.7	7.54	5.97	5.12	4.58	4.20	3.92	3.70	3.53	3.39	3.27	3.16	3.07	3.00	2.93
100	.05	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.82	1.79	1.77
	.01	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.43	2.37	2.31	2.27	2.22
	.001	11.5	7.41	5.86	5.02	4.48	4.11	3.83	3.61	3.44	3.30	3.18	3.07	2.99	2.91	2.84
1000	.05	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.80	1.76	1.73	1.70	1.68
	.01	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.27	2.20	2.15	2.10	2.06
	.001	10.9	6.96	5.46	4.65	4.14	3.78	3.51	3.30	3.13	2.99	2.87	2.77	2.69	2.61	2.54

Tab. B.6: Table des valeurs critiques du $\mathcal{U}$ de Mann-Whitney.

N_1 \ N_2	9	10	11	12	13	14	15	16	17	18	19	20
1												
2	0	0	0	1	1	1	1	1	2	2	2	2
3	2	3	3	4	4	5	5	6	6	7	7	8
4	4	5	6	7	8	9	10	11	11	12	13	13
5	7	8	9	11	12	13	14	15	17	18	19	20
6	10	11	13	14	16	17	19	21	22	24	25	27
7	12	14	16	18	20	22	24	26	28	30	32	34
8	15	17	19	22	24	26	29	31	34	36	38	41
9	17	20	23	26	28	31	34	37	39	42	45	48
10	20	23	26	29	33	36	39	42	45	48	52	55
11	23	26	30	33	37	40	44	47	51	55	58	62
12	26	29	33	37	41	45	49	53	57	61	65	69
13	28	33	37	41	45	50	54	59	63	67	72	76
14	31	36	40	45	50	55	59	64	67	74	78	83
15	34	39	44	49	54	59	64	70	75	80	85	90
16	37	42	47	53	59	64	70	75	81	86	92	98
17	39	45	51	57	63	67	75	81	87	93	99	105
18	42	48	55	61	67	74	80	86	93	99	106	112
19	45	52	58	65	72	78	85	92	99	106	113	119
20	48	55	62	69	76	83	90	98	105	112	119	127

Tab. B.7: Table des valeurs critiques pour le test de Wilcoxon.

n	$\alpha(2)$: 0.50 $\alpha(1)$: 0.25	0.20 0.10	0.10 0.05	0.05 0.025	0.02 0.01	0.01 0.005	0.005 0.0025	0.001 0.0005
4	2	0						
5	4	2	0					
6	6	3	2	0				
7	9	5	3	2	0			
8	12	8	5	3	1	0		
9	16	10	8	5	3	1	0	
10	20	14	10	8	5	3	1	
11	24	17	13	10	7	5	3	0
12	29	21	17	13	9	7	5	1
13	35	26	21	17	12	9	7	2
14	40	31	25	21	15	12	9	4
15	47	36	30	25	19	15	12	6
16	54	42	35	29	23	19	15	8
17	61	48	41	34	27	23	19	11
18	69	55	47	40	32	27	23	14
19	77	62	53	46	37	32	27	18
20	86	69	60	52	43	37	32	21
21	95	77	67	58	49	42	37	25
22	104	86	75	65	55	48	42	30
23	114	94	83	73	62	54	48	35
24	125	104	91	81	69	61	54	40
25	136	113	100	89	76	68	60	45
26	148	124	110	98	84	75	67	51
27	160	134	119	107	92	83	74	57
28	172	145	130	116	101	91	82	64
29	185	157	140	126	110	100	90	71
30	198	169	151	137	120	109	98	78
31	212	181	163	147	130	118	107	86
32	226	194	175	159	140	128	116	94
33	241	207	187	170	151	138	126	102
34	257	221	200	182	162	148	136	111
35	272	235	213	195	173	159	146	120
36	289	250	227	208	185	171	157	130
37	305	265	241	221	198	182	168	140
38	323	281	256	235	211	194	180	150
39	340	297	271	249	224	207	192	161
40	358	313	286	264	238	220	204	172
41	377	330	302	279	252	233	217	183
42	396	348	319	294	266	247	230	195
43	416	365	336	310	281	261	244	207
44	436	384	353	327	296	276	258	220
45	459	402	371	343	312	291	272	233
46	477	422	389	361	328	307	287	246
47	499	441	407	378	345	322	302	260
48	521	462	426	396	362	339	318	274
49	543	482	446	415	379	355	334	289
50	566	503	466	434	397	373	350	304
51	590	525	486	453	416	390	367	319
52	613	547	507	473	434	408	384	335
53	638	569	529	494	454	427	402	351
54	668	592	550	514	473	445	420	368
55	688	615	573	536	493	465	438	385
56	714	639	595	557	514	484	457	402
57	740	664	618	579	535	504	477	420
58	767	688	642	602	556	525	497	438
59	794	714	666	625	578	546	517	457
60	822	739	690	648	600	567	537	476

Tab. B.8: Table des valeurs critiques pour le test de Kolmogorov.

n	$P = .80$	$P = .90$	$P = .95$	$P = .98$	$P = .99$
1	.90000	.95000	.97500	.99000	.99500
2	.68377	.77639	.84189	.90000	.92929
3	.56481	.63604	.70760	.78456	.82900
4	.49265	.56522	.62394	.68887	.73424
5	.44698	.50945	.56328	.62718	.66853
6	.41037	.46799	.51926	.57741	.61661
7	.38148	.43607	.48342	.53844	.57581
8	.35831	.40962	.45427	.50654	.54179
9	.33910	.38746	.43001	.47960	.51332
10	.32260	.36866	.40925	.45662	.48893
11	.30829	.35242	.39122	.43670	.46770
12	.29577	.33815	.37543	.41918	.44905
13	.28470	.32549	.36143	.40362	.43247
14	.27481	.31417	.34890	.38970	.41762
15	.26588	.30397	.33760	.37713	.40420
16	.25778	.29472	.32733	.36571	.39201
17	.25039	.28627	.31796	.35528	.38086
18	.24360	.27851	.30936	.34569	.37062
19	.23735	.27136	.30143	.33685	.36117
20	.23156	.26473	.29408	.32866	.35241
21	.22617	.25858	.28724	.32104	.34427
22	.22115	.25283	.28087	.31394	.33666
23	.21645	.24746	.27490	.30728	.32954
24	.21205	.24242	.26931	.30104	.32286
25	.20790	.23768	.26404	.29516	.31657
26	.20399	.23320	.25907	.28962	.31064
27	.20030	.22898	.25438	.28438	.30502
28	.19680	.22497	.24993	.27942	.29971
29	.19348	.22117	.24571	.27471	.29466
30	.19032	.21756	.24170	.27023	.28987
31	.18732	.21412	.23788	.26596	.28530
32	.18445	.21085	.23424	.26189	.28094
33	.18171	.20771	.23076	.25801	.27677
34	.17909	.20472	.22743	.25429	.27279
35	.17659	.20185	.22425	.25073	.26897
36	.17418	.19910	.22119	.24732	.26532
37	.17188	.19646	.21826	.24404	.26180
38	.16966	.19392	.21544	.24089	.25843
39	.16753	.19148	.21273	.23786	.25518
40	.16547	.18913	.21012	.23494	.25205
41	.16349	.18687	.20760	.23213	.24904
42	.16158	.18468	.20517	.22941	.24613
43	.15974	.18257	.20283	.22679	.24332
44	.15796	.18053	.20056	.22426	.24060
45	.15623	.17856	.19837	.22181	.23798
46	.15457	.17665	.19625	.21944	.23544
47	.15295	.17481	.19420	.21715	.23298
48	.15139	.17302	.19221	.21493	.23059
49	.14987	.17128	.19028	.21277	.22828
50	.14840	.16959	.18841	.21068	.22604

Tab. B.9: Table des valeurs critiques du $\mathcal{H}$ de Kruskal-Wallis. Si le nombre de niveaux est supérieur à 3, ou si le nombre de sujets est supérieur à 5, utiliser la distribution du χ^2 avec df=nombre de niveaux-1.

n_1	n_2	n_3	p£0.05	p£0.01
2	1	1	-	-
2	2	1	-	-
2	2	2	-	-
3	1	1	-	-
3	2	1	-	-
3	2	2	4.714	-
3	3	1	5.142	-
3	3	2	5.361	-
3	3	3	5.6	7.2
4	1	1	-	-
4	2	1	-	-
4	2	2	5.333	-
4	3	1	5.208	-
4	3	2	5.444	6.444
4	3	3	5.727	6.746
4	4	1	4.967	6.667
4	4	2	5.455	7.036
4	4	3	5.599	7.144
4	4	4	5.692	7.654
5	1	1	-	-
5	2	1	5	-
5	2	2	5.16	6.533
5	3	1	4.96	-
5	3	2	5.251	6.909
5	3	3	5.649	7.079
5	4	1	4.986	6.955
5	4	2	5.273	7.118
5	4	3	5.631	7.445
5	4	4	5.618	7.76
5	5	1	5.127	7.309
5	5	2	5.339	7.269
5	5	3	5.706	7.543
5	5	4	5.643	7.791
5	5	5	5.78	7.98

Tab. B.10: Table des valeurs critiques pour le coefficient des rangs de Spearman.

α / n	0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
4	0.600	1.000	1.000						
5	0.500	0.800	0.900	1.000	1.000				
6	0.371	0.657	0.829	0.886	0.943	1.000	1.000		
7	0.321	0.571	0.714	0.786	0.893	0.929	0.964	1.000	1.000
8	0.310	0.524	0.643	0.738	0.833	0.881	0.905	0.952	0.976
9	0.267	0.483	0.600	0.700	0.783	0.833	0.867	0.917	0.933
10	0.248	0.455	0.564	0.648	0.745	0.794	0.830	0.879	0.903
11	0.236	0.427	0.536	0.618	0.709	0.755	0.800	0.845	0.873
12	0.224	0.406	0.503	0.587	0.671	0.727	0.776	0.825	0.860
13	0.209	0.385	0.484	0.560	0.648	0.703	0.747	0.802	0.835
14	0.200	0.367	0.464	0.538	0.622	0.675	0.723	0.776	0.811
15	0.189	0.354	0.443	0.521	0.604	0.654	0.700	0.754	0.786
16	0.182	0.341	0.429	0.503	0.582	0.635	0.679	0.732	0.765
17	0.176	0.328	0.414	0.485	0.566	0.615	0.662	0.713	0.748
18	0.170	0.317	0.401	0.472	0.550	0.600	0.643	0.695	0.728
19	0.165	0.309	0.391	0.460	0.535	0.584	0.628	0.677	0.712
20	0.161	0.299	0.380	0.447	0.520	0.570	0.612	0.662	0.696
21	0.156	0.292	0.370	0.435	0.508	0.556	0.599	0.648	0.681
22	0.152	0.284	0.361	0.425	0.496	0.544	0.586	0.634	0.667
23	0.148	0.278	0.353	0.415	0.486	0.532	0.573	0.622	0.654
24	0.144	0.271	0.344	0.406	0.476	0.521	0.562	0.610	0.642
25	0.142	0.265	0.337	0.398	0.466	0.511	0.551	0.598	0.630
26	0.138	0.259	0.331	0.390	0.457	0.501	0.541	0.587	0.619
27	0.136	0.255	0.324	0.382	0.448	0.491	0.531	0.577	0.608
28	0.133	0.250	0.317	0.375	0.440	0.483	0.522	0.567	0.598
29	0.130	0.245	0.312	0.368	0.433	0.475	0.513	0.558	0.589
30	0.128	0.240	0.306	0.362	0.425	0.467	0.504	0.549	0.580
31	0.126	0.236	0.301	0.356	0.418	0.459	0.496	0.541	0.571
32	0.124	0.232	0.296	0.350	0.412	0.452	0.489	0.533	0.563
33	0.121	0.229	0.291	0.345	0.405	0.446	0.482	0.525	0.554
34	0.120	0.225	0.287	0.340	0.399	0.439	0.475	0.517	0.547
35	0.118	0.222	0.283	0.335	0.394	0.433	0.468	0.510	0.539
36	0.116	0.219	0.279	0.330	0.388	0.427	0.462	0.504	0.533
37	0.114	0.216	0.275	0.325	0.383	0.421	0.456	0.497	0.526
38	0.113	0.212	0.271	0.321	0.378	0.415	0.450	0.491	0.519
39	0.111	0.210	0.267	0.317	0.373	0.410	0.444	0.485	0.513
40	0.110	0.207	0.264	0.313	0.368	0.405	0.439	0.479	0.507

C Logiciels statistiques

Voici une liste (non exhaustive) de solutions logicielles utilisées en Analyse de données et en Statistique inférentielle :

- Statistica (http://www.statsoft.com/, version d'évaluation, disposant d'un nombre limité de modules, téléchargeable gratuitement ;
- Spss (http://www.spss.com/), version d'évaluation, entièrement fonctionnelle mais limitée à 30 jours d'utilisation, téléchargeable gratuitement sur le site ;
- R (http://cran.cict.fr/), version de base et modules additionnels, téléchargeables gratuitement

Table des figures

Liste des tableaux

Index

U

V