



Sommaire

	Préface	VII
	Introduction	IX
	Les auteurs	XI
Chapitre 1	• Analyser pour décider	1
Chapitre 2	• Décrire les données	29
Chapitre 3	• Simplifier les données	51
Chapitre 4	• Segmenter	79
Chapitre 5	• L'analyse de variance	107
Chapitre 6	• La régression linéaire	133
Chapitre 7	• L'analyse conjointe	155
Chapitre 8	• Communiquer les résultats	177
	Bibliographie générale	195
	Index	197



Chapitre

Analyser pour décider

1. Études et recherche en marketing.....2
2. Des données aux variables.....7
3. Mesurer à l'aide d'un questionnaire.....16

Exercices

1. Quand Pampers collecte des données.....23
2. L'audience de la super star.....24
3. L'enquête « point de vente ».....25

Une bonne décision consiste à choisir la plus optimale des solutions parmi une série d'alternatives. Le marketing – et en particulier sa dimension études – s'est longtemps cantonné à un rôle purement descriptif. Mais les bonnes décisions n'arrivent pas par hasard : elles doivent être fondées sur des informations fiables et valides. Tour à tour, les outils d'études de marchés et les techniques d'analyse se sont considérablement enrichis. L'avènement d'Internet, la sophistication et l'exhaustivité des données de panel, la montée en puissance des bases de données clients et du data mining ont repoussé les limites des études de marchés traditionnelles, favorisant l'émergence d'une information marketing de grande qualité et d'analyses explicatives, voire prédictives, des comportements.

Ce chapitre présente les grandes familles d'études de marchés et pose les bases de l'analyse de données en marketing en abordant les concepts de données, de variables et de mesure.

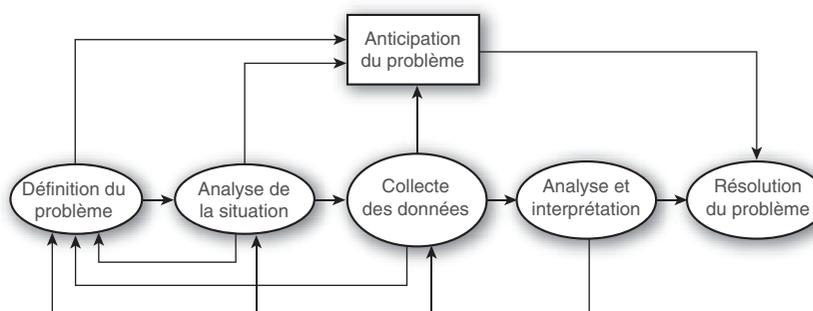
1 Études et recherche en marketing

Les études et recherche marketing ont pour but d'aider le responsable marketing à résoudre un problème spécifique, à contrôler ses performances, à planifier les décisions (Evrard, Pras et Roux, 2003). **Leur objectif est de lier l'entreprise à son environnement en développant des instruments de mesure, en collectant et en analysant des données, et en communiquant les résultats et leur interprétation.** Telle est la définition du processus de recherche en marketing qui nous guidera tout au long des huit chapitres de cet ouvrage.

1.1 LA DÉMARCHE D'ÉTUDE

À partir de la définition précédente, nous pouvons résumer la démarche d'étude à cinq étapes principales, reprises à la figure 1.1 ci-après.

Figure 1.1
Les cinq étapes d'une démarche d'étude.



La première étape de la démarche d'étude est d'identifier le problème managérial : le besoin d'étude est donc déterminé par l'existence d'un problème à résoudre. Plusieurs types de problèmes et plusieurs types de résolutions peuvent être envisagés, comme le montre le tableau 1.1.

Tableau 1.1 : Du problème managérial à la technique d'étude

Problème managérial	Objectifs d'étude	Techniques d'étude
Existe-t-il un marché potentiel pour un nouveau produit ?	<ul style="list-style-type: none"> – Tester les réactions des consommateurs à l'idée – Tester la composition du produit – Estimer le taux d'essai et de réachat 	<ul style="list-style-type: none"> – Test de concept – Test de formule – Marché-test simulé – Marché-témoin
	<ul style="list-style-type: none"> – Connaître les concurrents 	<ul style="list-style-type: none"> – Analyse de la concurrence – Panels

Tableau 1.1 : Du problème managérial à la technique d'étude (suite)

Problème managérial	Objectifs d'étude	Techniques d'étude
	– Connaître les attentes des consommateurs	– Identifier les bénéfices recherchés par les consommateurs – Étude de segmentation
	– Détecter les forces et faiblesses de la marque	– Étude du capital marque – Analyse des images de marque des concurrents
	– Déterminer un prix de vente	– Test de prix psychologiques – Analyse conjointe

Source : adapté de Vernet, 2000.

La formulation correcte d'un problème permet de faire le lien entre un besoin de décision et la mise en œuvre d'une démarche de recherche, de collecte, d'analyse et d'interprétation d'informations. La fonction « étude » doit donc être envisagée autour de ce paradigme informationnel. Son rôle consiste à transformer des informations brutes en données utiles dans la recherche de nouvelles opportunités, à mettre en place des systèmes d'écoute du marché et de veille concurrentielle, et à prescrire les comportements à adopter sur les marchés. Au confluent des flux d'informations de l'entreprise, elle acquiert aujourd'hui une dimension stratégique croissante.

Si la collecte et l'analyse de données sont au cœur du métier d'analyste en marketing, ces derniers font de plus en plus appel à des données secondaires et à des données stockées dans des entrepôts de données (*data warehouses*). Cette tendance est accentuée par le recours à Internet qui, en combinant habilement sites de marque et techniques de marketing direct, s'avère être une source inépuisable d'informations sur les marchés, les consommateurs, les concurrents.

L'existence de cette profusion de données fait évoluer les besoins d'étude dans l'entreprise et modifie par suite le recours aux différentes techniques. Auparavant, la conduite d'une étude de marché était principalement entendue comme la nécessité de procéder à une collecte de données terrain, souvent par le biais du questionnaire, de l'entretien ou de réunions de consommateurs. Dorénavant, l'accès aisé à des données secondaires, à la dissémination interfonctionnelle d'une intelligence marketing dans l'organisation modifie quelque peu la donne.

Cette vaste quantité d'informations disponibles rend nécessaire, pour le chargé d'étude comme pour le chef de produit, une compétence accrue en analyse de données. Elle permettra d'éviter les erreurs d'interprétation et de maîtriser la qualité d'études souvent réalisées par des instituts. Un besoin croissant d'opérationnalité se fait sentir en la matière. Cette opérationnalité passe tout d'abord par le développement de mesures pertinentes et valides supportant des construits psychologiques (décisions d'achat, notoriété, intérêt pour la marque, le produit, etc.), afin de bien mesurer ce qui se rapporte au problème managérial. Elle passe ensuite par la mise en œuvre d'analyses qui permettent d'expliquer et de prédire des comportements, afin de bien comprendre le problème managérial pour le résoudre et agir.

1.2 LES TECHNIQUES D'ÉTUDES

Les techniques d'études sont regroupées en deux catégories principales, selon leurs objectifs et leurs limites. Les études quantitatives dominent largement le marché des études, même si, dernièrement, les départements marketing ont manifesté un intérêt croissant pour les études qualitatives. Le tableau 1.2 montre la répartition des différentes techniques en fonction des méthodes de collecte les plus fréquemment utilisées en marketing.

Tableau 1.2 : Répartition des différentes techniques d'étude

Techniques	2005	2006
Quantitatives		
Études quantitatives <i>via</i> Internet	13 %	20 %
Études par téléphone	29 %	30 %
Tests en salle	11 %	10 %
Études en face-à-face	42 %	37 %
Études postales	5 %	4 %
Qualitatives		
Réunions de groupe	57 %	55 %
Entretiens individuels	22 %	22 %
Études qualitatives <i>via</i> Internet	5 %	17 %
Autres techniques qualitatives	16 %	5 %

Source : adapté de SEMO, 2008 (Syntec Études Marketing et Opinion).

Parmi les techniques les plus largement utilisées, on peut retenir :

- **P'étude *ad hoc*** : étude quantitative ou qualitative réalisée pour le compte d'un seul client ;
- **P'étude omnibus** : étude quantitative réalisée à date régulière. Le questionnaire regroupe l'ensemble des questions de différents souscripteurs ;
- **le baromètre** : étude réalisée à date fixe comme l'omnibus, mais avec le même questionnaire d'une étude à l'autre, pour le compte d'un ou de plusieurs clients ;
- **le panel** : investigation approfondie réalisée périodiquement pour plusieurs clients. Les interviewés sont identiques d'une vague à l'autre. Il s'appuie sur des échantillons importants de 2 000 à 10 000 individus ;

- **le marché-test** : étude quantitative visant à prévoir les ventes et parts de marché d'un nouveau produit; on parle également de marché-test pour des observations de type expérimental en magasin;
- **l'entretien individuel** : étude qualitative dont l'objectif est de recueillir le discours individuel. On distingue l'entretien non directif (libre propos), semi-directif (intervention et thèmes), directif (guide d'entretien strict, questions ouvertes), associatif ou projectif (analogie, associations de mots, compléments de phrases, jeux de rôle, etc.);
- **la réunion de groupe** : étude qualitative libre et non structurée d'un groupe de 8 à 12 participants, conduite par un animateur. La discussion libre repose sur les phénomènes de psychologie collective des groupes restreints, fondée notamment sur les travaux de Kurt Lewin.

Le tableau 1.3 représente les objectifs et les limites des approches qualitatives et quantitatives.

Tableau 1.3 : Objectifs et limites des approches qualitative et quantitative

Approche	Objectifs	Limites
Qualitative	Répertorier Explorer Générer Comprendre	Généralisation des résultats
Quantitative	Dénombrer Hiérarchiser Pondérer Résumer	Biais déclaratifs Mémorisation des répondants

Source : adapté de Vernet, 2000.

Les études qualitatives sont utilisées dans une dimension principalement exploratoire, afin de comprendre en profondeur des comportements de consommateurs par exemple. Si elles ne permettent pas de généraliser les résultats qu'elles produisent, elles n'en sont pas moins utiles pour dépasser les mesures d'attitudes des questionnaires. Elles permettent d'accéder à une étude approfondie des processus liés aux comportements de consommation, grâce notamment aux entretiens et aux réunions de consommateurs, et d'accéder plus profondément à l'explication de ces comportements, en levant le voile sur des facteurs inconscients (le non-verbal, le « non-dit »), en d'autres termes le monde interne des consommateurs et notamment leur rapport aux marques.

Les études qualitatives se distinguent également par la place qu'elles occupent dans la démarche de recherche. Souvent considérées comme un prélude à l'étude quantitative ou limitées à la confirmation des résultats d'une enquête par questionnaire, elles se substituent de plus en plus aux approches traditionnelles, grâce notamment à l'utilisation d'Internet et de ses potentialités multimédias, et à la nécessité croissante pour le marketing d'être connecté au terrain. Garnier, par exemple, a lancé, il y a peu, une vaste opération de type ethnographique baptisée *Consumer Connect*, dont l'objectif était avant tout d'immerger les chefs de produit parmi les consommateurs et d'observer leur utilisation du produit *in situ*. L'avènement d'Internet a contribué à repopulariser cette technique auprès des instituts d'étude : on peut citer l'émergence de la netnographie (voir ci-après) ou encore le *Home Use Blog* (HUB), développé conjointement par Danone et la société Repères.

EXEMPLE

La netnographie

On constate, depuis quelques années, un intérêt grandissant pour l'information collectée à partir de l'observation de communautés virtuelles, nouvelles formes de communautés dont Internet a permis l'émergence. Ainsi, de nombreuses firmes ont réalisé des études sur la base d'informations issues de forums de discussion et n'ont pas tardé à saisir les opportunités offertes par ces nouveaux types d'interactions sociales. Kozinets a développé récemment une approche nouvelle – l'ethnographie sur Internet ou netnographie – qu'il définit comme « une nouvelle méthode de recherche qualitative qui adapte la méthode de l'ethnographie à l'étude des cultures et des communautés qui émergent grâce aux communications informatisées » (Kozinets, 2002, p. 62). En tant que technique de recherche en marketing, la netnographie utilise l'information publique disponible sur les forums en ligne afin d'identifier et de comprendre les besoins et les influences qui pèsent sur les décisions d'achat de groupes de consommateurs présents sur Internet. Pour Laurent Florès, CEO de la société d'étude crmmatrix, spécialiste de l'écoute client, le canal Internet permet aux marques de participer à de véritables conversations et de s'appuyer sur un puissant levier du marketing : le bouche à oreille. Il est désormais possible de quantifier le volume de ces conversations, d'analyser leur contenu et le profil des intervenants, avec un avantage important sur les techniques traditionnelles, puisque cette approche n'altère pas le contexte étudié par l'intervention d'un analyste mais collecte plutôt une information en langage naturel.

Les techniques quantitatives, auxquelles cet ouvrage est essentiellement consacré, constituent la part dominante des études marketing. Leur objectif est avant tout de mesurer, de quantifier et de permettre de généraliser les résultats à partir de l'échantillon de la population concernée. Ce type d'étude repose généralement sur un grand nombre d'observations et sur des informations structurées (valeurs numériques, échelles ou valeurs nominales) par opposition aux informations non structurées (discours, texte libre/questions ouvertes, etc.). Plus précisément, trois types d'études quantitatives peuvent être distingués, en fonction du contexte de découverte de l'information : décrire, expliquer, prédire.

Les **études descriptives** sont fondées sur des mesures dont le but est de collecter des données brutes afin de créer des structures décrivant les caractéristiques d'une population cible ou d'un marché. Elles peuvent être utiles, entre autres, pour faire la photographie d'un marché, de la satisfaction des consommateurs, de la notoriété d'une marque. La dimension descriptive est l'objectif premier traditionnellement assigné aux études marketing. Cette étape importante a pour objet de mesurer la force d'association entre deux variables, par exemple, et permet de poser un cadre d'analyse nécessaire aux études explicatives et prédictives.

Les **études explicatives** ont pour objet de transformer des données brutes en structures expliquant des relations de causalité entre deux ou plusieurs variables. L'approche explicative est utile lorsque l'étude a pour objectif de comprendre les causes directes d'un phénomène. Ce type d'étude peut permettre, par exemple, de modéliser l'impact de la publicité sur les ventes. L'approche explicative est particulièrement utile dans un contexte d'aide à la décision, où le but assigné à l'étude n'est plus simplement de décrire mais aussi de comprendre, de la manière la plus fiable et la plus valide, les déterminants affectant la performance des décisions marketing.

Les **études prédictives**, quant à elles, ont pour objet de transformer les données brutes collectées sur les caractéristiques comportementales des consommateurs ou des entreprises/marchés pour créer des modèles prédictifs à des fins d'optimisation. Ces approches,

Exercices

EXERCICE 1 QUAND PAMPERS COLLECTE DES DONNÉES

Énoncé

À Scwallbach, près de Francfort en Allemagne, plus de 1500 mères de famille fréquentent chaque semaine le centre d'innovation de Procter & Gamble. Elles viennent prendre des couches pour les tester et remplissent, en échange, des questionnaires. Dans l'espace de jeu à disposition, des chercheurs étudient les attitudes et comportements de bébés venus s'y amuser une partie de leur journée. Plus loin, des pièces au sol très mou – pour simuler la marche d'un tout petit –, et aux meubles géants, mettent les salariés du groupe dans la peau de jeunes enfants à différents stades de leur évolution. En France, les salariés en charge du marketing peuvent faire des « séjours d'immersion » dans des familles avec bébés, se levant la nuit avec les parents. Cette approche visant à scruter les usages et leur évolution s'inscrit dans une nouvelle démarche qui commence avec la traditionnelle boîte remise à la maternité. Des mailings prennent ensuite le relais. Les parents d'un premier enfant sont en général avides d'informations : un site internet de la marque Pampers met en avant conseils et données, des jeux en ligne – très appréciés – pour se mettre dans la peau d'un bébé, nourrissant débats, échanges, autant d'informations étudiées de près par les spécialistes de la marque.

1. Quel est le principal intérêt de la démarche de Pampers ? Quelle est la méthode utilisée, et quels en sont les principaux avantages ?
2. Comment, à votre avis, les équipes de Pampers valorisent-elles les données collectées ? Selon vous, à quels outils d'analyse ont-elles recours ?
3. Quel type de méthode, complémentaire, pourraient-elles mettre en place ? Argumentez.

Solution

1. Le principal intérêt de la démarche de Pampers est de mettre le consommateur au centre du processus de collecte de données. La méthode utilisée est à l'évidence qualitative. Elle permet d'étudier en profondeur les processus de consommation en interrogeant les parents, plus particulièrement les mères, et en observant les interactions mère-enfant. De plus, en simulant les attitudes et les comportements des bébés, elle permet surtout d'immerger les salariés du groupe dans la peau de jeunes enfants à différents stades de leur évolution. Nous sommes ici dans une démarche *orientée marché*, où la dissémination de l'information sur les consommateurs vers l'ensemble de l'organisation tient une place importante. Deux méthodes sont donc principalement utilisées : une expérimentation à Scwallbach, où les chercheurs peuvent observer et tester les comportements des bébés ; l'ethnographie en France, où les marketeurs font des séjours d'immersion dans des familles. Des outils quantitatifs d'enquête classiques prennent ensuite le relais à partir des données collectées dans les maternités.

2. Deux dimensions doivent être prises en considération. Les données issues des approches qualitatives font l'objet d'analyses de plusieurs ordres : des analyses de contenu par exemple, afin de faire émerger des thèmes, des discours, mais également un traitement des

données issues de l'expérimentation où il s'agit d'observer l'impact sur certaines variables d'une variable dont on contrôle les effets. Pour traiter des données d'expérimentation, on pourra utiliser l'analyse de variance (voir chapitre 4) ou l'analyse conjointe (voir chapitre 7), par exemple, en fonction des contraintes liées à la nature des variables.

3. Les données issues de la campagne de marketing direct et provenant du site de marque sont d'une grande richesse et peuvent nourrir de nombreuses analyses. On peut étudier les retours de la campagne de marketing direct en mettant en relation les profils sociodémographiques des parents ayant reçu la boîte d'échantillons avec la probabilité d'achat. En ce qui concerne le site internet, les *fichiers logs*, ou fichiers regroupant l'ensemble des événements survenus sur un serveur, peuvent servir de base à des analyses poussées, comme le fait Amazon.com pour customiser sa page d'accueil en fonction des profils de navigation des internautes.

EXERCICE 2 L'AUDIENCE DE LA SUPER STAR

Énoncé

Reprenons l'exemple de la mesure d'audience utilisée pour illustrer l'intervalle de confiance. Un sondeur réalise une étude d'audience par téléphone pour connaître les caractéristiques sociodémographiques et les comportements des téléspectateurs de la *Super Star*, émission de télé-réalité diffusée en prime time sur le câble et le satellite. Il sélectionne 1 000 numéros de téléphone par tirage aléatoire simple dans la base de données des abonnés de la chaîne (qui en compte 120 000 sur le câble et 2 100 000 sur le satellite). On pose l'hypothèse que les 1 000 personnes répondent effectivement aux enquêteurs. On constate que l'émission absorbe 36,8 % de l'audience des personnes interrogées de moins de 35 ans, et que le montant moyen dépensé par cette cible en SMS et appels téléphoniques est de 6,2 €, avec un écart type de 2,2 €.

1. Quel aurait été l'intervalle de confiance si l'étude d'audience avait porté sur 5 000 abonnés de la chaîne?
2. Un annonceur souhaite investir en devenant sponsor de l'émission à condition qu'elle réalise 40 % d'audience sur les moins de 35 ans. Lui recommanderiez-vous l'investissement publicitaire?

Solution

1. Si l'on avait interrogé 5 000 abonnés de la chaîne, on aurait calculé l'intervalle de confiance de la manière suivante :

$$p = 0,368$$

$$q = 1 - p = 0,632$$

$$0,368 - 1,96 \sqrt{\frac{0,368 - 0,632}{5000}} \leq \pi \leq 0,368 + 1,96 \sqrt{\frac{(0,368 - 0,632)}{5000}}$$

Soit : 35,4 % $\leq \pi \leq$ 38,1 %

Le sondage réalisé permet donc d'estimer cette proportion avec une précision absolue de 2,99 % (au degré de confiance 0,95). En augmentant la taille de l'échantillon, on diminue l'amplitude de l'intervalle de confiance.

2. Dans le cadre de la première étude d'audience, l'intervalle de confiance se situait entre 33,8 % et 39,7 % (au degré de confiance 0,95). La borne supérieure restant en deçà de la mesure plancher souhaitée par l'annonceur, il n'est donc pas souhaitable de réaliser l'investissement publicitaire. Il peut être intéressant de refaire le calcul pour un degré de confiance plus faible, à 0,90 ($z = 1,64$), à titre d'illustration. On obtient alors les résultats suivants :

$$0,368 - 1,64 \sqrt{\frac{0,368 - 0,632}{1000}} \leq \pi \leq 0,368 + 1,64 \sqrt{\frac{0,368 - 0,632}{1000}}$$

Soit : $34,3 \% \leq \pi \leq 39,3 \%$

EXERCICE 3 L'ENQUÊTE « POINT DE VENTE »

Énoncé

Reprenons l'exemple sur les points de vente que nous avons utilisé dans la section 2.2 (pointdevente.sav). Si l'on résume l'ensemble des questions de l'enquête dans le tableau suivant, on obtient :

- Fréquentez-vous ce point de vente au moins toutes les deux semaines ?
- Quel montant moyen dépensez-vous par mois dans ce type de point de vente ?
- Seriez-vous prêt à faire vos achats dans ce (nouveau) point de vente ?
- À combien estimez-vous le prix moyen d'une paire de chaussures dans ce point de vente ?
- Vous décririez-vous comme un auditeur régulier de radio ?
- Quel type de programme de radio écoutez-vous le plus souvent ?
- Regardez-vous régulièrement le journal télévisé ?
- Quel journal TV regardez-vous le plus fréquemment ?
- Lisez-vous la presse quotidienne ?
- Quelle rubrique de presse quotidienne lisez-vous le plus souvent ?
- Êtes-vous abonné à un titre de presse magazine ?
- La décoration de la boutique est importante à mes yeux.
- Je préfère un point de vente situé à moins de 30 minutes de chez moi.
- Je préfère être conseillé(e) par des vendeurs(ses).
- J'aime que les collections soient originales.
- J'aime qu'il y ait de nombreuses références dans les collections.
- J'aime qu'il y ait des marques connues dans les collections.
- Je préfère une décoration sobre.
- Je préfère une décoration sophistiquée.
- Je préfère une musique d'ambiance classique.
- Je préfère une musique d'ambiance rock.
- Quelle est votre année de naissance ?
- Quel est votre niveau d'étude ?

- Quel est votre statut marital?
 - En incluant les enfants de moins de 18 ans, quelle est la taille de votre foyer?
 - Quels sont approximativement les revenus de votre foyer?
 - Quel est votre sexe?
 - Possédez-vous une carte de fidélité de l'enseigne?
1. Décrivez le type d'échelle associé à chacune des questions du tableau.
 2. Donnez trois exemples de tests que vous pourriez mettre en œuvre à partir de ces variables.

Solution

1. Vous pouvez reprendre le tableau en y incluant le type de variable.

Fréquentez-vous ce point de vente au moins toutes les deux semaines?	Nominale
Quel montant moyen dépensez-vous par mois dans ce type de point de vente?	Numérique
Seriez-vous prêt à faire vos achats dans ce (nouveau) point de vente?	Échelle métrique
À combien estimez-vous le prix moyen d'une paire de chaussures dans ce point de vente?	Numérique
Vous décririez-vous comme un auditeur régulier de radio?	Nominale
Quel type de programme de radio écoutez-vous le plus souvent?	Nominale (échelle)
Regardez-vous régulièrement le journal télévisé?	Nominale
Quel journal TV regardez-vous le plus fréquemment?	Nominale (échelle)
Lisez-vous la presse quotidienne?	Nominale
Quelle rubrique de presse quotidienne lisez-vous le plus souvent?	Nominale (échelle)
Êtes-vous abonné à un titre de presse magazine?	Nominale
La décoration de la boutique est importante à mes yeux.	Échelle métrique
Je préfère un point de vente à moins de 30 minutes de chez moi.	Échelle métrique
Je préfère être conseillé(e) par des vendeurs(euses).	Échelle métrique
J'aime que les collections soient originales.	Échelle métrique
J'aime qu'il y ait de nombreuses références dans les collections.	Échelle métrique
J'aime qu'il y ait des marques connues dans les collections.	Échelle métrique
Je préfère une décoration sobre.	Échelle métrique

Je préfère une décoration sophistiquée.	Échelle métrique
Je préfère une musique d'ambiance classique.	Échelle métrique
Je préfère une musique d'ambiance rock.	Échelle métrique
Quelle est votre année de naissance?	Numérique
Quel est votre niveau d'étude?	Nominale (échelle)
Quel est votre statut marital?	Nominale (échelle)
En incluant les enfants de moins de 18 ans, quelle est la taille de votre foyer?	Numérique
Quels sont approximativement les revenus de votre foyer?	Nominale (échelle)
Quel est votre sexe?	Nominale
Possédez-vous une carte de fidélité de l'enseigne?	Nominale

2. De nombreux tests sont envisageables :

- a. un tri croisé entre le montant moyen dépensé dans le point de vente et le niveau d'études par exemple, afin de mettre en évidence un impact de la CSP sur les achats;
- b. une analyse typologique afin de classer les individus de l'enquête en fonction de leur profil de réponse;
- c. une analyse de variance multiple (MANOVA) dont l'objet serait d'expliquer le montant moyen dépensé par une série de variables explicatives comme, par exemple, le niveau d'études, le statut marital, etc.

Focus 2.1 Les fractiles

Les fractiles sont les valeurs d'une variable quantitative qui divisent les données triées en classes par centième. Les quartiles (25^e, 50^e et 75^e centiles) divisent les observations en quatre classes de taille égale. On les définit dans SPSS à partir de la boîte de dialogue **Effectifs > Statistiques** (voir figure 1.1), en sélectionnant **Partition en *n* classes égales** (*n* définissant le niveau de partition souhaité). Vous pouvez également spécifier des centiles particuliers (par exemple le 95^e centile), autrement dit les valeurs au-dessus de 95 % des observations.

Mesures de la dispersion

Les **mesures de la dispersion** reposent sur les indicateurs suivants : l'étendue, la variance, l'écart type et le coefficient de variation. L'**étendue** (ou **intervalle**) est la différence entre la plus grande et la plus petite des valeurs observées. La **variance** est la mesure de la dispersion autour de la moyenne, égale à la somme des carrés des écarts par rapport à la moyenne, divisée par le nombre d'observations moins un. Lorsque les données se concentrent autour de la moyenne, la variance est faible. Si les données sont dispersées autour de la moyenne, la variance est élevée. Il s'agit d'une mesure plus fine de la dispersion, au sens où toutes les données sont prises en compte. En revanche, elle est sensible aux valeurs extrêmes. L'**écart type** est la mesure de la dispersion autour de la moyenne, exprimée dans la même unité que la variable. L'écart type est la racine carrée de la variance. On l'écrit de la manière suivante :

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

Le **coefficient de variation** est le rapport de l'écart type à la moyenne ($CV = \frac{s}{\bar{X}}$), exprimé en pourcentage. Son objet est de mesurer le degré de variation de la moyenne d'un échantillon à l'autre, lorsque ceux-ci sont issus de la même distribution.

Mesures de la distribution

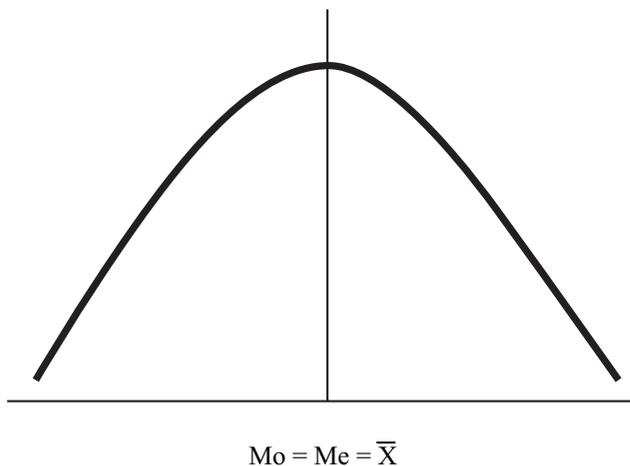
On **mesure la symétrie et la forme de la distribution** par l'asymétrie et l'aplatissement. Ces statistiques sont présentées avec leur erreur standard.

Le **coefficient de symétrie (skewness)** mesure l'asymétrie d'une distribution. Une distribution normale est symétrique (voir figure 2.4), c'est-à-dire que les valeurs sont les mêmes de part et d'autre du centre de la distribution, et possède une valeur de skewness de 0. Une distribution avec un skewness positif significatif est une distribution asymétrique à droite (la distribution prend la forme d'une longue queue à droite) et une distribution avec un skewness négatif significatif est une distribution asymétrique à gauche (la distribution prend la forme d'une longue queue à gauche). Cette asymétrie s'explique par le fait que les écarts sont plus importants dans une direction que dans l'autre.

Le **coefficient d'aplatissement (kurtosis)** permet de mesurer le relief ou la platitude d'une courbe issue d'une distribution de fréquences. En d'autres termes, le coefficient d'aplatissement permet de mesurer le degré de concentration des observations dans les queues de la courbe. Le coefficient de kurtosis est de 0 pour une distribution normale (gaussienne).

Un kurtosis négatif indique donc que les queues comptent un plus grand nombre d'observations que dans une distribution gaussienne. Les coefficients de kurtosis et de skewness peuvent être utilisés pour s'assurer que les variables suivent une distribution normale, condition nécessaire pour de nombreux tests statistiques. On estime que le coefficient de symétrie ou skewness doit être inférieur à 1 et le coefficient d'aplatissement ou kurtosis doit être inférieur à 1,5 pour considérer que la variable suit bien une loi normale.

Figure 2.4
Représentation d'une distribution normale.



SPSS

Reprenons notre exemple avec SPSS (pointsdevente.sav) : rappelez la boîte de dialogue de la procédure précédente (**Effectifs**) en cliquant sur l'icône  dans la barre d'outils. Procédez aux mêmes opérations mais cette fois pour la variable *montant*. Dans la boîte de dialogue **Effectifs** que vous venez de rappeler, cliquez sur l'onglet **Statistiques** et cochez les statistiques de mesure de la tendance centrale, de dispersion et de distribution, puis sélectionnez un graphique (un histogramme avec courbe gaussienne par exemple) pour représenter la distribution.

Les figures 2.5 et 2.6 reprennent les statistiques descriptives de la variable *montant*.

Figure 2.5
Description de la variable *montant*.

Statistiques		
Quel montant moyen dépensez-vous par mois dans ce type de point de vente?		
	Valide	
N	400	
	Manquante	0
Moyenne		153.5100
Erreur std. de la moyenne		4.55739
Médiane		172.0000
Mode		.00
Ecart-type		91.14782
Variance		8307,925
Asymétrie		-,067
Erreur std. d'asymétrie		,122
Aplatissement		-,085
Erreur std. d'aplatissement		,243
Intervalle		444.00
Minimum		.00
Maximum		444.00
Somme		61,404.00

3.1 L'HYPOTHÈSE STATISTIQUE

Une hypothèse statistique est un énoncé quantitatif concernant les caractéristiques d'une population ou, plus précisément, une affirmation portant sur une ou plusieurs variables. Elle se présente traditionnellement sous la double forme d'une première hypothèse, appelée **hypothèse nulle**, et d'une seconde hypothèse, appelée **hypothèse alternative**. Son objectif est de réfuter l'hypothèse nulle, laquelle concerne le plus souvent un *statu quo* ou une absence de différence, au profit de l'hypothèse alternative.

Exemple : on peut poser l'hypothèse nulle H_0 qu'il n'existe pas de différence de ventes entre les points de vente situés en centre-ville et ceux de la périphérie urbaine, et l'hypothèse alternative H_1 qu'elles sont différentes en centre-ville et en périphérie urbaine.

Les tests statistiques étant conçus pour la réfutation d'hypothèses et non pour leur confirmation, l'hypothèse alternative est celle qui sera acceptée si l'hypothèse nulle est rejetée. Accepter une hypothèse revient donc à dire que l'hypothèse est non rejetée plutôt qu'acceptée, c'est-à-dire que les données recueillies au cours d'une expérience particulière sont compatibles avec l'hypothèse alternative proposée.

L'objectif de l'analyse de données est donc de prendre une décision : en l'occurrence, rejeter ou non l'hypothèse nulle H_0 . Les tests étant fondés sur des informations incomplètes issues d'observations portant sur un échantillon de la population, il est nécessaire de définir le **seuil de signification** du test, seuil formulé en pourcentage de chances de rejeter l'hypothèse nulle alors qu'en réalité celle-ci était vraie. Le seuil de signification est habituellement noté α et exprimé en pourcentage. Le choix du seuil est lié au niveau de risque accepté (1 % ou 5 % étant les valeurs usuelles). Son complément $(1 - \alpha)$, appelé **seuil de confiance**, correspond au pourcentage de cas où on acceptera l'hypothèse nulle à juste titre. On appelle **erreur de type I** le fait de rejeter, à la suite des résultats d'un test statistique, une hypothèse qui serait en réalité vraie (condamner un innocent) et **erreur de type II** l'erreur liée au fait d'accepter une hypothèse qui serait en réalité fautive (innocenter un coupable). La probabilité de commettre ce type d'erreur est notée β ; on appelle **puissance du test** son complément $(1 - \beta)$, lequel correspond à la probabilité de rejeter une hypothèse qui serait réellement fautive (voir tableau 2.1).

Tableau 2.1 : Types d'erreurs dans un test statistique

		Situation dans la population	
		Ho vraie	Ho fautive
Décision	Ho acceptée	Décision correcte (seuil de confiance = $1 - \alpha$)	Erreur de type II (β)
	Ho rejetée	Erreur de type I (seuil de signification = α)	Décision correcte (puissance du test = $1 - \beta$)

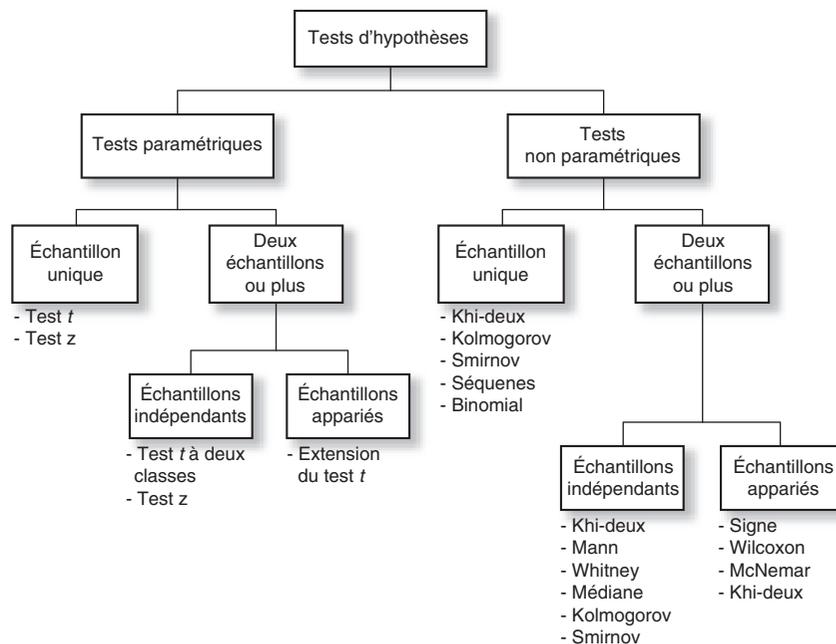
Bien que l' α établisse le niveau de signification du test, c'est la puissance du test $(1 - \beta)$ qui donne une estimation de la probabilité de trouver des différences significatives – si elles existent – dans les données. Pourquoi, dès lors, ne pas prendre en compte l' α et le β en tant que niveaux de confiance? La raison évoquée est que l'erreur de type I et l'erreur de

type II sont inverses : plus l'erreur de type I devient restrictive (proche de 0) et plus la probabilité d'une erreur de type II augmente; de même, réduire l'erreur de type I réduit la puissance du test. L'analyste doit donc trouver le juste équilibre entre le degré de confiance (α) et la puissance du test qui en résulte. La seule manière de faire baisser simultanément α et β est d'augmenter la taille de l'échantillon étudié.

3.2 LES TESTS D'HYPOTHÈSES

Les tests d'hypothèses, ou tests d'inférence, ont pour objectif de mesurer l'effet d'une variable indépendante sur une variable dépendante, en fonction du nombre d'échantillons et en fonction de la nature des variables étudiées. On nomme **tests paramétriques** les approches reposant sur des données métriques (et par suite sur des paramètres connus tels que la moyenne ou l'écart type, par exemple), et **tests non paramétriques** les approches reposant sur des données non métriques (et qui, par suite, peuvent s'affranchir de conditions de distribution particulières). Les tests non paramétriques étant peu sensibles à la taille de l'échantillon et aux données aberrantes, ils sont utilisés en marketing où les échantillons peuvent parfois être de petite taille (moins de 30 individus). Le nombre d'échantillons joue également un rôle important dans le choix du test approprié. En effet, deux situations doivent être distinguées : lorsque l'on étudie deux populations distinctes sur une même variable, on parle de **mesures indépendantes** (comparer les clients et les non-clients); et lorsque les mêmes individus sont mesurés sur une même variable dans deux situations distinctes, on parle de **mesures appariées** (comparer les niveaux de prix à deux périodes distinctes). Ces éléments affectent de manière importante les statistiques de tests (voir figure 2.15).

Figure 2.15
Tests paramétriques et tests non paramétriques (Malhotra et al., 2007).



3.3 TESTS PARAMÉTRIQUES

Les deux principaux tests paramétriques sont le test t et le test Z , qui ont pour objet de tester des différences de moyenne. Ces tests sont souvent mis en œuvre en marketing, car ils permettent, par exemple, de comparer la moyenne d'une variable dépendante métrique en fonction des modalités d'une variable nominale. On formule alors une hypothèse nulle qui sera vérifiée par le test t ou le test Z . Pour plus de simplicité, ces deux tests sont présentés ici pour des échantillons uniques.

Test t

Le test t est directement lié à la statistique t de Student, qui suppose que la variable adopte une distribution normale, que la moyenne soit connue et que la variance, lorsqu'elle est inconnue, soit estimée sur l'échantillon. On le calcule de la manière suivante :

$$t = (\bar{X} - \mu) / s_{\bar{X}}$$

Où : \bar{X} : moyenne de l'échantillon
 μ : moyenne de la variable
 $s_{\bar{X}}$: variance de l'échantillon

Dans SPSS, ce test paramétrique peut être estimé avec la procédure suivante : menu **Analyse > Comparer les moyennes > Test T pour échantillon unique...**, procédure que nous avons utilisée au chapitre 1 pour estimer l'intervalle de confiance. Pour comparer les moyennes de deux échantillons indépendants (comparaison des clients et des non-clients par exemple), on utilisera une analyse de variance (ANOVA) à 1 facteur (voir chapitre 4). Pour comparer les moyennes de deux échantillons appariés (comparaison de relevés de prix à deux périodes distinctes par exemple), on suivra une extension du test t pour échantillons appariés qui est disponible dans la même boîte de dialogue.

Test Z

Le test Z peut être mis en place lorsque la variance de l'échantillon est connue. La valeur de Z s'obtient par la formule suivante :

$$Z = (\bar{X} - \mu) / \sigma_{\bar{X}} \text{ où } : \sigma_{\bar{X}} \text{ est l'écart type de la population}$$

Ce test peut également être étendu pour tester des proportions.

3.4 TESTS NON PARAMÉTRIQUES

Les tests non paramétriques sont souvent mis en œuvre dans la pratique en marketing : ils s'appliquent aux variables qualitatives et s'avèrent relativement performants sur de petits échantillons, même s'ils sont moins puissants que les tests paramétriques. Voici les principaux tests paramétriques présentés ici : un **test d'ajustement** (le test de Kolmogorov-Smirnov), des **tests de comparaison d'échantillons indépendants** (le test U de Mann-Whitney et le test de la médiane), ainsi que des **tests de comparaison d'échantillons appariés** (le test de Wilcoxon, le test du signe et le test de McNemar).

- **Test de Kolmogorov-Smirnov (K-S)**

Le test de Kolmogorov-Smirnov est un test dit d'ajustement, car il permet d'établir si une population donnée suit une distribution particulière (normale, uniforme ou poisson par exemple), condition exigée par de nombreux tests. Le K-S est calculé à partir de la plus grande différence (en valeur absolue) entre les fonctions de distribution théorique et observée cumulées :

$$K = \text{Max } |A_i - O_i|$$

Le K-S pour un échantillon s'obtient dans SPSS à partir du menu **Analyse > Tests non paramétriques > K-S à 1 échantillon...**

- **Test U de Mann-Whitney**

Le test de Mann-Whitney permet de vérifier que deux échantillons (ou groupes) proviennent bien de la même population. On peut l'utiliser, par exemple, pour comparer les réponses dans un département par rapport aux réponses nationales. La statistique du test *U* réunit les deux échantillons et ordonne les observations par ordre croissant de taille. Le test calcule le nombre de fois où un résultat du groupe 1 précède un résultat du groupe 2, ainsi que le nombre de fois où un résultat du groupe 2 précède un résultat du groupe 1. *U* est d'autant plus petit que les groupes sont différents.

Pour calculer le *U* de Mann-Whitney dans SPSS, il faut d'abord définir la variable qui servira à scinder les données en deux échantillons : **Analyse > Test non paramétrique > 2 échantillons indépendants...**, puis sélectionner une variable de regroupement (**Facteur**) et cliquer sur **Définir les niveaux**. Pour définir les groupes, vous devez indiquer les valeurs pour le groupe 1 et celles pour le groupe 2. Sélectionnez ensuite le **test U de Mann-Whitney** dans la boîte de dialogue.

- **Test de la médiane**

Ce test, moins puissant que le *U* de Mann-Whitney, permet de déterminer si deux groupes sont issus de populations ayant la même médiane, en estimant la position de chaque observation par rapport à la médiane globale des deux échantillons.

Pour calculer le test de la médiane dans SPSS, vous devez suivre la procédure suivante : **Analyse > Test non paramétrique > K échantillons indépendants...**, puis sélectionner le **test de la médiane** dans le menu du type de test envisagé.

- **Test de Wilcoxon**

Le test de Wilcoxon est utilisé dans le cas de la comparaison de deux échantillons appariés, c'est-à-dire lorsque l'on souhaite, par exemple, comparer deux types de réponses : avant/après l'exposition à un message publicitaire, attitude par rapport à une marque A et une marque B, etc. La statistique *z* du test de Wilcoxon s'obtient en calculant la différence entre les scores des deux observations par paires d'observations, puis en calculant le rang de toutes les différences, et enfin la somme des rangs positifs et des rangs négatifs. On rejette l'hypothèse nulle (absence de différence entre les deux groupes) s'il y a une différence entre la somme des rangs positifs et la somme des rangs négatifs. Le sens de la statistique indique le sens de la différence de la paire examinée.

Dans SPSS, ouvrez le menu **Analyse > Test non paramétrique > 2 échantillons liés...**, puis sélectionnez le test que vous souhaitez mettre en œuvre (**Wilcoxon, Signe, McNemar**), comme le montre la figure 2.16.

- **Test du signe**

Le test du signe est relativement proche du test de Wilcoxon, mais il est plus limité et par suite moins puissant. Il ne s'attache en effet qu'à une comparaison des signes des différences, sans procéder à un classement comme le fait le test de Wilcoxon.

I Principes de validation d'une échelle de mesure

Nous avons vu dans le chapitre 1 que les concepts étaient mesurés avec plusieurs questions ou items. Par exemple, pour estimer l'attitude du client à l'égard d'un produit, le chargé d'étude pose des questions qui permettent de bien saisir les différentes facettes de ce concept (part affective, cognitive, etc..). Ensuite, il faut vérifier que ces différentes questions ou items mesurent bien ce que l'on cherche à mesurer, afin d'obtenir au final des résultats plus proches de la réalité.

Dans l'article intitulé « Un paradigme pour développer de meilleures mesures des construits marketing », Churchill (1979) propose une procédure pour renforcer la validité et la fiabilité des mesures. Après avoir sélectionné des échelles (jeu d'items pour mesurer un concept), il s'agit, dans un premier temps, de les soumettre à l'analyse factorielle exploratoire puis au test de la fiabilité ¹.

1.1 LA VALIDITÉ D'UNE ÉCHELLE DE MESURE

La validité d'une échelle de mesure désigne sa capacité à appréhender un phénomène (Hair *et al.*, 2006). Les tests de validité ont pour objectif de vérifier si les différents items d'un instrument sont une bonne représentation du phénomène étudié : mesure-t-on ce que l'on cherche à mesurer? (Evrard *et al.*, 2003).

La validité prend plusieurs formes; il existe donc plusieurs techniques pour la vérifier :

- **la validité faciale ou de contenu** : il s'agit de savoir si la mesure capture les différents aspects du phénomène étudié. Elle est fondée sur le jugement du chercheur et de ses pairs. Par exemple, lors du test du questionnaire, des experts du domaine peuvent émettre un avis sur la capacité des items à recouvrir tous les aspects d'un concept;
- **la validité de trait ou de construit** : est-ce que les différents indicateurs offrent une bonne représentation du phénomène étudié ? Il faut vérifier si les indicateurs censés mesurer le même phénomène sont corrélés (validité convergente) et s'ils se distinguent des indicateurs censés mesurer des phénomènes différents (validité discriminante) (Evrard *et al.*, 2003) :
 - **la validité convergente** est établie lorsque les mesures d'un même construit sont corrélées;
 - **la validité discriminante** est destinée à s'assurer que les indicateurs de mesure d'un construit sont faiblement corrélés aux indicateurs de mesure d'autres construits, conceptuellement distincts du premier. L'analyse factorielle exploratoire (AFE) permet de tester ces deux validités;
- **la validité nomologique ou prédictive** résulte de la conformité des relations entre les mesures d'un concept et celles d'autres concepts avec les prédictions de la théorie (Evrard *et al.*, 2003). Cette étape de validation intervient au cours de la phase confirmatoire.

1. Puis, dans une phase de validation, les échelles modifiées après suppressions d'énoncés subissent une deuxième fois ces procédures, on parle d'analyse confirmatoire. Cette seconde étape vise à connaître les qualités psychométriques des instruments de mesure.

1.2 LA FIABILITÉ D'UNE ÉCHELLE DE MESURE

La fiabilité correspond au degré avec lequel les instruments utilisés mesurent de façon constante le construit étudié (Evrard *et al.*, 2003). Par conséquent, une échelle est fidèle si l'on retrouve plusieurs fois les mêmes résultats sur les mêmes sujets. Trois méthodes permettent de tester la fiabilité d'une mesure :

- **la méthode du « test/retest »** : le questionnaire est administré deux fois à la même population et les résultats obtenus sont comparés. Cette technique est particulièrement appropriée pour la mise au point d'instrument de mesure;
- **la méthode du « Split half », ou des deux moitiés** : le questionnaire est administré au même moment à des échantillons différents (l'échantillon est scindé en deux) et les résultats sont comparés. Il existe cependant un risque de sélection; les deux échantillons sont-ils appariés? se ressemblent-ils?
- **la technique des formes alternatives** : il s'agit d'introduire dans le questionnaire plusieurs questions sur le même phénomène mais formulées différemment. Le questionnaire est administré aux mêmes individus. Le coefficient alpha de Cronbach est calculé pour vérifier si les énoncés partagent des notions communes, et s'ils sont en cohérence entre eux.

Focus 3.1 Estimer la fiabilité avec le coefficient alpha de Cronbach

L'**alpha de Cronbach** est un coefficient de fiabilité qui mesure la cohérence interne d'une échelle construite à partir d'un ensemble d'items. La pratique consiste à réduire un grand nombre d'items initiaux dans un processus itératif de conservation/élimination des items en fonction de la valeur du coefficient alpha, qui varie entre 0 et 1. Plus la valeur de l'alpha est proche de 1, plus la cohérence interne de l'échelle (sa fiabilité) est forte. On élimine donc les items qui diminuent le score, et on conserve ceux qui contribuent à augmenter l'alpha. L'examen de l'alpha de Cronbach évite au chargé d'étude de tomber dans un travers fréquent qui consiste à reprendre un questionnaire existant sans se préoccuper de ses problèmes de mesure.

Le seuil d'acceptabilité de l'alpha varie selon l'objectif de la recherche. Pour une étude exploratoire, un coefficient plus faible est acceptable (0,7) alors que dans le cadre d'une recherche fondamentale, il doit être plus élevé (> 0,8) (Nunnally et Bernstein, 1994). Lorsqu'une échelle est utilisée pour comparer des groupes, un alpha de 0,8 est satisfaisant, et il est inutile d'essayer d'obtenir un niveau supérieur. De Vellis (2003) propose une typologie (voir tableau 3.1).

Tableau 3.1 : Les valeurs de l'alpha de Cronbach

< 0,6	Insuffisant
entre 0,6 et 0,65	Faible
entre 0,65 et 0,7	Minimum acceptable
entre 0,7 et 0,8	Bon
entre 0,8 et 0,9	Très bon
> 0,9	Considérer la réduction du nombre d'items

Il existe une relation entre le nombre d'items et la valeur de l'alpha : un nombre restreint d'items (de deux à trois) donne un alpha généralement plus faible (0,6) qu'une mesure de quatre énoncés (0,7). Au-delà de 0,9, l'alpha risque, en revanche, de traduire davantage une redondance inter-items, appauvrissant ainsi le domaine conceptuel étudié (Peterson, 1995). Il est, par conséquent, recommandé de ne pas dépasser le seuil de 0,9.

Le logiciel SPSS fournit les niveaux du coefficient d'alpha de l'échelle lorsque chaque item est supprimé. Les items dont la suppression améliore sensiblement le coefficient ne sont généralement pas retenus si la validité de contenu ne s'en trouve pas amoindrie.

Focus 3.2 Le traitement des items inversés

La conception d'un questionnaire demande des précautions (plusieurs items, non-réponse possible, clarté de la question, ordre des questions, etc.) car la formulation des questions peut influencer la mesure d'un concept. L'inversion d'item est souvent employée pour s'assurer de la validité et de la fiabilité de la mesure (par exemple, la satisfaction à l'égard d'un service est mesurée par un item : « je suis pleinement satisfait par ce service » et un autre, inversé, « ce service ne me satisfait pas pleinement »).

Nous cherchons à évaluer dans quelle mesure les items utilisés sont de bons indicateurs des concepts qu'ils sont censés mesurer. Pour cela, il est généralement conseillé de réaliser une analyse factorielle exploratoire pour vérifier que les items se « regroupent » bien de la manière prévue, et de calculer ensuite le coefficient alpha de Cronbach qui évalue la fiabilité de chaque échelle (Churchill, 1979).

2 L'analyse factorielle

L'**analyse factorielle** est une méthode exploratoire d'analyse des tableaux de contingence développée essentiellement par J.-P. Benzecri durant la période 1970-1990. Elle désigne un ensemble de méthodes statistiques multivariées dont le principal objectif est de définir la structure des corrélations entre un grand nombre de variables (par exemple, les réponses à un questionnaire) en déterminant un ensemble de dimensions communes appelés facteurs.

2.1 LES UTILISATIONS DE L'ANALYSE FACTORIELLE

L'analyse factorielle sert à identifier les **dimensions** de la structure et à déterminer dans quelle mesure chaque variable peut expliquer chaque dimension.

Les deux objectifs de l'analyse factorielle sont :

- **Résumer les données.** L'analyse factorielle fait ressortir les dimensions sous-jacentes qui, une fois interprétées, décrivent les données de manière synthétique.
- **Réduire les données.** Elle calcule des scores pour chaque dimension et les substitue aux variables originelles.

Alors que dans les autres méthodes (régressions, analyse de variance, etc.) les variables sont considérées comme des variables soit dépendantes, soit indépendantes, dans l'analyse factorielle, toutes les variables sont considérées chacune par rapport aux autres. Les **facteurs** sont formés pour maximiser l'explication de l'ensemble des variables et non pour prédire des variables dépendantes. Dès lors, l'analyse factorielle est appropriée dans une optique exploratoire (analyse factorielle exploratoire ou AFE).

EXEMPLE

Les critères importants dans l'évaluation d'un club de sport

Dans une enquête sur les attentes des clients vis-à-vis de leur salle de sport, on interroge les individus sur une vingtaine de critères. L'analyse factorielle sert à regrouper les attentes en trois ou quatre points plus simples. Elle agrège les variables en facteurs ou combinaisons de variables. L'objectif est de rendre l'information plus synthétique et facile à lire sur une carte factorielle (voir tableaux 3.2 et 3.3).

Tableau 3.2 : Exemple d'application de l'analyse factorielle

	Rencontre	Muscles	Esthétisme	Dé foulé ment	Santé	Dynamisme	Prise en charge	Confort	Économie	Lieu agréable
1	4	1	4	2	4	1	1	2	1	2
2	1	2	4	5	4	1	1	1	1	1
3	2	4	2	4	3	1	1	2	4	2
4	3	4	2	4	3	3	3	2	1	2
5	1	4	3	4	4	4	4	3	2	3
6										

À titre d'exemple, le confort, les aspects dé foulé ment, dynamisme et santé représentent peut-être en fait la même chose : être en forme.

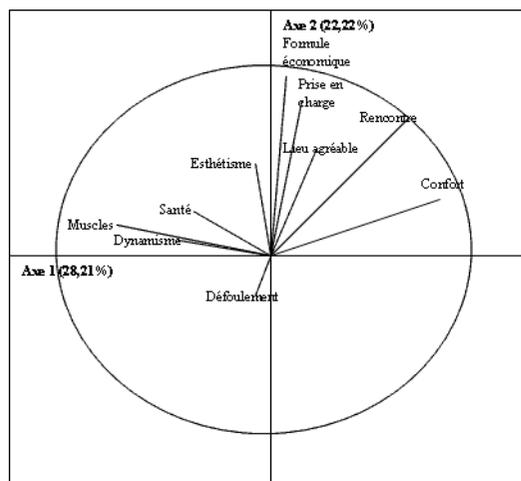
Tableau 3.3 : Exemple d'application de l'analyse factorielle (suite)

	Facteur 1 Forme	Facteur 2 Contact	Facteur 3
1			
2			
3			
4			
5			
6			

La solution de l'analyse factorielle est trouvée par essai/erreur et le jugement s'établit en fonction des concepts (voir figure 3.1). Sur l'axe horizontal de la figure, à gauche les atten-

tes des clients portent sur la forme physique; à droite, sur le confort de la salle. Sur l'axe vertical s'opposent le côté sociable du club de sport et le besoin de s'y défouler. Au total, la variance restituée par ces deux axes (les deux premiers facteurs) est de 50,43.

Figure 3.1
Représentation graphique de l'analyse factorielle.



L'analyse factorielle exploratoire permet d'identifier des groupes d'items qui covarient les uns avec les autres et semblent représenter des variables latentes pertinentes. Autrement dit, l'AFE consiste à explorer la relation entre des variables mesurées, afin de déterminer si ces relations peuvent être résumées par un nombre moins important de construits latents.

L'AFE permet de vérifier le nombre de dimensions ou, plus souvent, l'unidimensionalité d'un concept. En effet, un concept peut comporter une ou plusieurs facettes. Par exemple, l'implication comporte une composante affective, une composante calculée et une composante normative. Lorsque l'on fait appel à des échelles de mesure déjà utilisées, l'AFE permet de vérifier si l'on retrouve, pour l'échantillon étudié, la même structure factorielle. Elle fera alors ressortir autant de facteurs que le construit a de dimensions (un seul facteur si le construit est unidimensionnel). Dans le cadre du développement de nouveaux instruments, l'AFE permet de constater si les items correspondent effectivement aux concepts présentés aux répondants.

2.2 LES CONDITIONS ET OPTIONS DE L'ANALYSE FACTORIELLE

En fonction des caractéristiques de l'échantillon et des données collectées, plusieurs options sont possibles pour la réalisation d'une AFE (analyse factorielle exploratoire). Nous verrons, dans un premier temps, la taille de l'échantillon requise avant de présenter les différentes options et tests permettant de valider les résultats d'une AFE.

La taille de l'échantillon nécessaire

La taille de l'échantillon dépend du nombre d'items soumis à l'AFE. Il faut un minimum de cinq observations par item (un ratio de 10 pour 1 est préférable). Le nombre total d'observations doit être d'au moins 50 et il est souhaitable d'interroger au moins 100 individus.

La méthode d'extraction utilisée

La méthode d'extraction la plus employée est l'analyse en composantes principales (ACP). L'ACP a pour objet de résumer l'ensemble des données quantitatives d'un tableau individus/variables. En effet, l'ACP synthétise les données en construisant un petit nombre de variables nouvelles, les composantes principales. Les éléments critiques de la grille peuvent alors être captés rapidement, à l'aide de représentations graphiques établies à partir des ACP.

Le choix de la matrice des données

Il est possible de travailler sur la matrice de corrélation ou sur la matrice de covariance. Pour simplifier, ce choix s'effectue ainsi :

- **matrice de corrélation** : lorsque les variables sont mesurées avec des échelles différentes;
- **matrice de covariance** : lorsque l'on applique l'analyse factorielle à plusieurs groupes avec des variances différentes pour chaque variable.

L'adéquation des données

Avant de réaliser l'analyse, il est important de s'assurer que les données sont factorisables. Elles doivent former un ensemble cohérent pour pouvoir y chercher des dimensions communes qui aient un sens et qui ne soient pas des artefacts statistiques (Evrard *et al.*, 2003). La matrice des données doit comporter suffisamment de corrélations pour justifier la réalisation d'une AFE. Plusieurs indicateurs peuvent être utilisés :

- **La matrice des corrélations anti-image** représente la valeur négative des corrélations partielles. Des corrélations anti-image importantes indiquent que la matrice des données n'est peut-être pas adaptée à l'AFE.
- **Le test de Sphéricité de Bartlett** examine la matrice des corrélations dans son intégralité et fournit la probabilité de l'hypothèse nulle selon laquelle toutes les corrélations sont de zéro.
- **La « Measure of Sampling Adequacy » (MSA) ou Kaiser-Meyer-Olkin (KMO)** indique dans quelle proportion les variables retenues forment un ensemble cohérent et mesurent de manière adéquate un concept. Elle teste si les corrélations partielles entre les variables sont faibles.

Des valeurs de KMO comprises entre 0,3 et 0,7 représentent des solutions factorielles acceptables. Ce test, d'abord réalisé pour chaque variable, doit ensuite être repris avec l'ensemble des variables (Hair *et al.*, 2006).

L'extraction des facteurs

Il n'existe pas de base quantitative exacte pour déterminer le nombre de facteurs à extraire. Les critères sont souvent choisis sur la part de variance de chaque item qu'un facteur permet d'expliquer :

- **L'« eigenvalue », ou règle des valeurs propres > 1 ou règle de Kaiser-Guttman** : une valeur propre représente la quantité d'informations capturée par un facteur. Un facteur qui aurait une valeur propre inférieure à 1 représenterait moins d'informations qu'un simple item.
- **le « Scree Test », ou test du coude ou de l'éboulis** : ce test se fonde également sur les valeurs propres des facteurs mais dans une perspective relative et non absolue. Étant

donné que chaque facteur est extrait d'une matrice qui est le résidu de l'extraction précédente, la quantité d'informations contenue dans les facteurs successifs décroît. Lorsque, entre deux facteurs, la décroissance en termes d'informations devient faible ou nulle, on peut estimer que le dernier facteur ne contient pas suffisamment d'informations pour être retenu.

- **le critère du pourcentage de variance** : il s'agit d'une approche par laquelle on observe les pourcentages cumulés de la variance extraite par les facteurs successifs. L'objectif est de s'assurer qu'un facteur explique une quantité significative de variance. Il est souvent conseillé d'arrêter l'extraction lorsque 60 % de la variance expliquée est extraite (Hair *et al.*, 2006).

La rotation des facteurs

Afin de pouvoir interpréter les facteurs, il est généralement nécessaire de réaliser une rotation. Celle-ci permet d'identifier des groupes de variables fortement liés les uns aux autres. La rotation fait en sorte que chaque item ne soit fortement lié qu'à un seul facteur. Cette opération est réalisée par une redistribution de la variance des premiers facteurs extraits aux facteurs successifs, afin d'aboutir à une structure factorielle plus simple (Hair *et al.*, 2006). Lorsque les axes sont maintenus à 90 degrés, on parle de rotation orthogonale; lorsque les axes ne sont pas contraints à être indépendants, on parle de rotation oblique.

Il existe plusieurs méthodes de rotation :

- **Varimax** : rotation orthogonale qui minimise le nombre de variables ayant de fortes corrélations sur chaque facteur. Simplifie l'interprétation des facteurs.
- **Oblimin direct** : rotation oblique, c'est-à-dire dans laquelle les axes se positionnent en fonction des items et ne sont donc pas orthogonaux.
- **Quartimax** : méthode qui minimise le nombre de facteurs requis pour expliquer chaque variable. Simplifie l'interprétation des variables observées.
- **Equamax** : méthode de rotation qui minimise à la fois le nombre de variables qui pèsent fortement sur un facteur et le nombre de facteurs requis pour expliquer une variable (combinaison des méthodes Varimax et Quartimax).

Focus 3.3 L'analyse factorielle exploratoire : rotation orthogonale ou oblique ?

Les critères de choix entre la **rotation orthogonale** (Varimax) et la **rotation oblique** sont les suivants :

La rotation orthogonale maintient les axes de l'espace factoriel en angle droit. Ce type de rotation permet de minimiser le nombre d'items ayant des contributions élevées sur un axe et donc de simplifier les facteurs. Elle permet d'obtenir une structure factorielle plus claire.

Si la corrélation entre facteurs est faible, inférieure à 0,15 (De Vellis, 2003) ou à 0,3 (Nunnally et Bernstein, 1994), la rotation orthogonale sera préférée pour sa simplicité. Toutefois, si l'on a des raisons de penser que des items ou facteurs sont corrélés, il est logique de réaliser une rotation oblique. On peut également comparer la solution avec rotation oblique et rotation orthogonale. S'il est possible d'assigner un item au même facteur dans les deux cas, alors la rotation orthogonale sera choisie pour sa simplicité.

Dans la grande majorité des cas, une rotation orthogonale est suffisante pour aboutir à une structure simple. Hair *et al.* (2006) estiment cependant que la rotation oblique est conseillée si l'on souhaite déterminer des facteurs représentant des concepts qui seront analysés postérieurement car la structure factorielle obtenue possède une plus grande stabilité.

2.3 L'ÉPURATION DES DONNÉES

L'AFE pour vérifier le nombre de dimensions d'un concept

L'analyse factorielle est utilisée pour vérifier la **validité de trait ou de construit**. Il s'agit de tester et de purifier les échelles d'un questionnaire. L'AFE permet de s'assurer que l'échelle évalue précisément et exclusivement le construit qu'elle est censée mesurer. Lorsque le construit est unidimensionnel, l'AFE fera ressortir un seul facteur, plusieurs pour **les construits multidimensionnels**. Il est aussi possible de fixer *a priori* le nombre de facteurs sous SPSS.

Nous traitons ici de la question des items et facteurs à retenir ou, au contraire, à supprimer, suite à une AFE. En effet, lorsque les facteurs sont extraits, il est nécessaire d'évaluer la validité convergente et discriminante au niveau de l'item ainsi que la fiabilité des échelles. La validité convergente concerne le fait que les réponses obtenues par différents indicateurs du même construit soient fortement corrélées; la validité discriminante est démontrée lorsque la mesure d'un construit déterminé est faiblement corrélée à une mesure d'un autre construit.

Ces analyses sont réalisées pour chaque échelle. Les items présumés mesurer un même construit doivent donc être fortement corrélés les uns aux autres (validité convergente) et faiblement corrélés aux items censés mesurer d'autres construits (validité discriminante). Le niveau du coefficient structurel de l'AFE (composante) sert à déterminer si l'item satisfait au critère de validité convergente. Le tableau 3.4 présente les niveaux de significativité des contributions factorielles des items selon la taille de l'échantillon étudié.

Tableau 3.4 : Niveau de significativité des coefficients structurels selon la taille de l'échantillon

Niveau des coefficients structurels	Taille de l'échantillon nécessaire
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

Source : adapté de Hair *et al.*, 2006.

L'épuration d'une échelle de mesure se fait en deux temps :

D'une part, pour **les coefficients structurels ou composantes**, un seuil est déterminé en fonction de la taille de l'échantillon. Par exemple, pour un test d'échelle sur un échantillon de 200 individus, un seuil de 0,40 sera retenu. Pour les échelles multidimensionnelles, sont éliminés les items dont les poids factoriels sont supérieurs à 0,30 sur plusieurs facteurs et ceux n'ayant aucune contribution supérieure ou égale à 0,50 sur l'une des composantes principales identifiées. Ces seuils peuvent aussi varier en fonction de la taille de l'échantillon (Hair *et al.*, 2006).

D'autre part, la formation des facteurs repose sur l'importance des variables initiales sur ces facteurs. **Les « communalités »** (part de variance expliquée par l'item) doivent dépasser 0,5 et si possible 0,7. Le niveau de représentation est considéré comme moyen pour un seuil de 0,40, bon pour un seuil de 0,65 et excellent lorsque la communalité dépasse 0,80 (Evrard *et al.*, 2003).

SPSS

Dans cet exemple, nous testons l'échelle destinée à mesurer l'ambition professionnelle. Cette échelle unidimensionnelle de 10 items est issue de la littérature. Les réponses aux questions sont collectées grâce à une échelle de Likert à cinq échelons allant de « Pas du tout d'accord » à « Tout à fait d'accord » (voir tableau 3.5).

Tableau 3.5 : Exemple de l'échelle destinée à mesurer l'ambition

Item 1 - J'aimerais avoir un poste plus important et que les autres m'envient.
Item 2 - J'aime bien discuter avec des gens importants.
Item 3 - Je veux être une personne importante dans la communauté.
Item 4 - J'admire beaucoup les gens qui ont gravi les échelons et sont au sommet.
Item 5r ¹ - Si j'avais suffisamment d'argent, je ne travaillerais plus*.
Item 6 - Même si je gagnais beaucoup d'argent au jeu, je continuerais à exercer mon métier.
Item 7r - Si je pouvais toucher le chômage, je préférerais ne pas travailler*.
Item 8 - J'aime être admiré(e) pour ma réussite.
Item 9r - Je n'aime pas être remarqué(e)*.
Item 10 - J'aime que des employés me demandent conseil.

1. Le r signifie que cet item est inversé.

Les 10 items sont, dans un premier temps, soumis à une analyse factorielle exploratoire (méthode de l'ACP), afin de vérifier la structure du construit mesuré. Cette échelle est testée avec un échantillon de 106 individus.

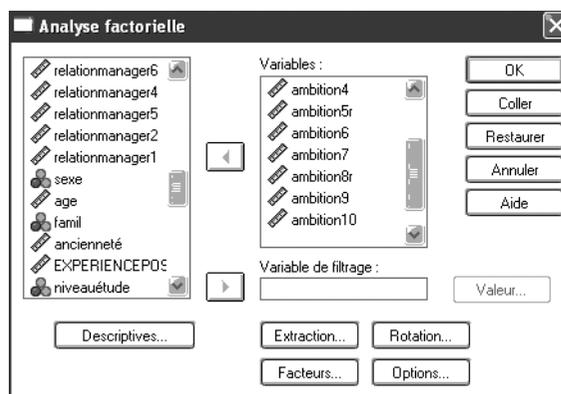
Ouvrez le fichier « challenge »¹. Allez dans le menu **Analyse > Positionnement > Analyse factorielle**. Une boîte de dialogue apparaît (voir figure 3.2).

1. Vous trouverez ce fichier à l'adresse : <http://www.pearsoneducation.fr>.

Transférez les items destinés à mesurer l'ambition en les sélectionnant chacun à leur tour et en cliquant sur la flèche.

Figure 3.2

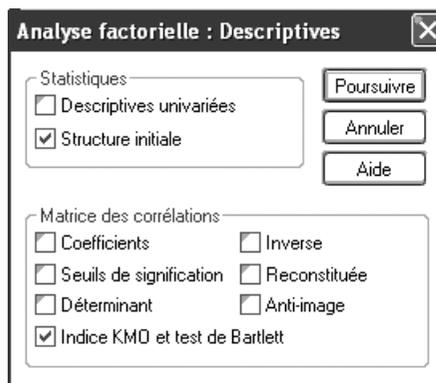
Demande d'analyse factorielle.



Avant de lancer l'AFE, plusieurs commandes sont à effectuer. Afin de vérifier l'adéquation des données, on peut demander l'indice KMO et le test de Bartlett par le bouton **Descriptives** dans la boîte de dialogue précédente. La structure initiale (précochée) donne les communalités, les valeurs propres et la part de variance expliquée initiale (voir figure 3.3).

Figure 3.3

Demande de l'indice KMO et du test de Bartlett.



Cliquez ensuite sur **Poursuivre** pour revenir à la boîte de dialogue initiale.

Pour sélectionner la méthode de l'analyse factorielle, cliquez sur **Extraction** et la boîte de dialogue de la figure 3.4 apparaît.

Nous pouvons choisir **la méthode de l'analyse factorielle** (composantes principales ; facteurs communs, etc.). Nous sélectionnons **Composantes principales**.

Pour obtenir les facteurs, le logiciel présélectionne les valeurs propres supérieures à 1. Mais il est aussi possible de déterminer le nombre de facteurs. Dans une optique exploratoire, nous laissons libre ce nombre de facteurs.

Le choix de la matrice de départ est aussi fixé dans cette boîte de dialogue : la matrice de corrélation est présélectionnée. Nous gardons cette matrice pour l'analyse.

On peut demander un graphique des valeurs propres qui sert à éliminer des facteurs avec le test du coude.

Cliquez ensuite sur **Poursuivre** pour revenir à la boîte de dialogue initiale.

I Les principes de l'analyse conjointe

Les travaux de Green dans les années 1970 marquent le début de la prise en compte de l'analyse conjointe dans la recherche en marketing. La méthode dite de l'« analyse des mesures conjointes », au développement croissant depuis les années 1980, vise à mieux comprendre le comportement des individus et, en particulier, du consommateur.

Le sketch de Coluche illustre la problématique de l'analyse conjointe sur la préférence entre être « *grand, riche, beau et intelligent* » et « *petit, pauvre, moche et bête* ». Si l'on présente les combinaisons suivantes « *grand, riche, moche et bête* » ou « *petit, pauvre, beau et intelligent* », l'individu est alors amené à faire des compromis dans lesquels l'avantage d'une caractéristique compense une autre qu'il doit rejeter. L'analyse conjointe permet de répondre aux questions suivantes : Quelle est l'importance de tel ou tel attribut (prix, dimensions, etc.) du produit pour le consommateur ? Quelle est l'importance de certains niveaux d'attributs (niveau de prix, dimensions en cm, etc.) par rapport à d'autres dans l'esprit du consommateur ? Cette méthode détermine à la fois l'importance relative de chaque attribut et les niveaux des attributs préférés des répondants.

Lorsqu'on dispose d'informations sur les répondants (données démographiques ou autres), l'analyse conjointe permet d'identifier les segments de marché pour lesquels des produits spécifiques seront plus adaptés. Par exemple, une personne appartenant à une CSP+ et un étudiant peuvent avoir des goûts différents auxquels des offres de produits distincts pourront répondre.

L'analyse conjointe repose sur la décomposition de la préférence en utilités partielles. Pour déterminer l'utilité totale d'un produit, on suppose que l'individu additionne les utilités partielles des attributs du produit. On parle de *modèle additif*. Au final, l'individu choisit parmi les produits celui qui lui procure l'utilité totale la plus élevée. L'estimation permet ainsi d'obtenir, pour chaque facteur et ses niveaux, des utilités partielles ainsi que l'importance de chaque attribut. Ce qui compte est donc l'individu tel qu'il réagit dans une situation déterminée.

L'analyse conjointe appartient aux modèles de décomposition (voir focus 7.1 sur le modèle compensatoire) où l'importance des caractéristiques est estimée à partir des préférences déclarées du consommateur et de ses notations des différents produits sur plusieurs caractéristiques. Elle permet d'analyser l'importance des caractéristiques du produit dans la formation des préférences.

Focus 7.1 Le modèle compensatoire

Le modèle d'attitude implicite de l'analyse conjointe est un modèle compensatoire, où l'évaluation se fonde sur le principe du compromis, c'est-à-dire qu'un peu moins d'un attribut peut être compensé par un peu plus d'un autre.

Par exemple, un individu qui cherche un appartement peut avoir plusieurs critères : le montant du loyer, la superficie, le nombre de pièces, la luminosité, la proximité des transports publics, etc. Si l'un de ces critères n'est pas satisfait (superficie insuffisante), il peut être compensé par un montant du loyer plus faible dans un modèle compensatoire (ce n'est pas le cas dans le modèle non compensatoire).

1.1 LES UTILISATIONS DE L'ANALYSE CONJOINTE EN MARKETING

L'analyse conjointe est largement utilisée en marketing pour l'identification d'un nouveau concept, pour divers tests (prix, produits, publicité, distribution, etc.), pour l'analyse concurrentielle ou la segmentation du marché (voir tableau 7.1). Il s'agit, par exemple :

- de déterminer l'importance relative d'attributs dans le processus de choix des consommateurs;
- d'estimer la part de marché des marques qui diffèrent au niveau des attributs;
- de déterminer la composition d'objets les plus appréciés;
- de segmenter le marché à partir des similarités de préférences pour des niveaux d'attributs.

Tableau 7.1 : Les applications de l'analyse conjointe

Pour les biens de consommation courante	
Nouveaux produits	72 %
Prix	61 %
Segmentation	48 %
Publicité	39 %
Distribution	7 %

1.2 LES CONDITIONS D'APPLICATION DE L'ANALYSE CONJOINTE

L'analyse des mesures conjointes nécessite que **les variables explicatives soient qualitatives ou nominales et que les variables à expliquer soient quantitatives**. Ces dernières peuvent être évaluées :

- à l'aide d'une échelle de mesure;
- à l'aide d'un ordre de préférence (classement) entre différentes combinaisons de niveaux de facteurs.

Par exemple, dans le cadre d'une étude sur la préférence des clients à l'égard d'une offre de transport aérien, on pourra demander aux individus d'évaluer différentes offres de compagnies avec une note de 1 à 9, autrement dit de les classer. Ces offres incluront, par exemple, le prix du billet (bas, moyen, élevé), la qualité du service à bord (excellente, moyenne, mauvaise), le nombre d'escales, etc. Il sera ainsi possible d'estimer, parmi ces facteurs et leurs niveaux, celui qui a le plus d'importance pour les clients dans leur choix d'une offre de transport aérien. L'objectif est ensuite d'élaborer une offre optimale pour la clientèle.

1.3 LES ÉTAPES DE L'ANALYSE CONJOINTE

La méthodologie de l'analyse conjointe est jalonnée par trois grandes étapes (Green et Srinivasan, 1990) :

- **collecte des données** : choix du plan factoriel complet ou fractionné, de la forme des questions et de la méthode de recueil;
- **définition de l'échelle de mesure de la variable dépendante** : choix de la mesure (classement, notation des combinaisons, comparaison de paires de combinaisons);
- **estimation** : étape liée à la nature de la mesure de la variable dépendante (ANOVA si la variable dépendante est quantitative, analyse monotone de la variance si elle est ordinale).

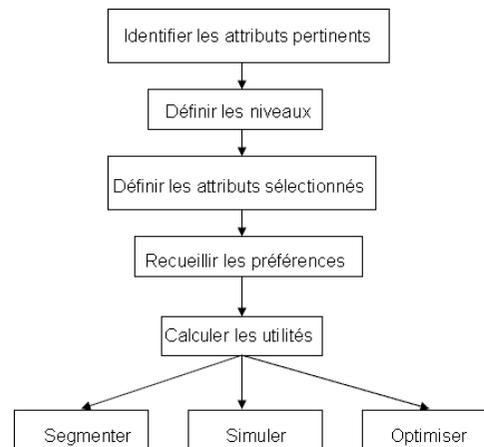
Ces étapes sont développées au cours de la section suivante.

2 La préparation de l'analyse conjointe

L'analyse conjointe demande au préalable la résolution d'un certain nombre de questions. En effet, avant même de collecter les données, le chargé d'étude doit s'interroger sur les attributs qu'il cherche à évaluer et leurs niveaux. Cette première sélection des attributs ou facteurs les plus importants et de leurs niveaux ou modalités déterminera le plan d'expérience et le mode de collecte des données.

Le schéma de la figure 7.1 présente les différentes phases de la méthode de l'analyse conjointe et met l'accent sur la première étape de la sélection des attributs et des niveaux.

Figure 7.1
Les étapes de l'analyse conjointe.



Source : adapté de Liqueur et Benavent.

Index

A

- Abaques 11
- Access panels 8
 - on-line 8
- Accord, échelle de mesure 15
- Ad hoc, étendue 4
- Agrégation
 - chaîne 91
 - des données 51
 - méthode 85
- Ajustement
 - courge 187
 - test 42
- Alpha
 - de Cronbach 53
 - seuil d'acceptabilité 53
- Analyse
 - bivariée 20, 36
 - bivariée, variable dépendante 36
 - bivariée, variable indépendante 36
 - conjointe, étapes 158
 - de fiabilité 65
 - de variance 108
 - factorielle 54, 56
 - factorielle, méthode 61
 - multivariée 20
 - multivariée de la variance 115
 - rapport d' 178
 - typologique 81, 90
 - univariée 20
- ANCOVA 116
- ANOVA à un facteur 108
- Aplatissement 32
 - coefficient d' (Kurtosis) 32
- Arbre
 - de décision 90
 - hiérarchique 87

- Association prédictive, coefficient 39
- Asymétrie 32

B

- Baromètre 4
- Barycentres 85
- Biais de l'expérimentation 17, 121
 - effet de l'instrument 121
 - effet de test 121
 - histoire 121
 - maturation 121
 - mortalité 121
- Boîtes à moustaches 31
- Bonferroni, test de 109
- Boule de neige, échantillonnage 11
- Brief de l'étude 178

C

- Carré latin 123
- Catégorisation 80
- Centiles 32
- Centres
 - de groupes 91
 - mobiles 88
- Chaîne des agrégations 91
- Classification 80, 85
 - ascendante 84
 - descendante 84
 - hiérarchique 84
 - hiérarchique ascendante 85
 - hiérarchique descendante 86
 - non hiérarchique 84, 88
- Clustering 80
- Coefficient
 - d'aplatissement (Kurtosis) 32
 - d'association prédictive 39

- de contingence 39
 - de corrélation multiple 136
 - de détermination 137
 - de Pearson 134
 - de symétrie (Skewness) 32
 - de variation 32
 - phi 39
 - standardisé 137
 - structurels 60
 - Collecte de données 3
 - Communalités 60
 - Comparaisons multiples 109
 - Composantes 60
 - principales 61
 - Compréhension, prétest 18
 - Concomitance 137
 - Conditions d'application de la régression 136
 - Confiance
 - intervalle 12
 - seuil 40
 - Construits 16
 - multidimensionnels 59
 - Contamination, effet 18
 - Contingence, coefficient 39
 - Corrélation 137
 - de Pearson, mesure 84
 - linéaire 134
 - matrice 57
 - multiple, coefficient 136
 - Corrélations anti-image, matrice 57
 - Courbe d'ajustement 187
 - Covariable 116
 - Covariance, matrice 57
 - Cramer, V de 39
 - Cronbach, Alpha de 53
- D**
- Data mining 7
 - Data warehouses 3
 - DDL (degrés de liberté) 38
 - Décrire les données 20
 - Degrés de liberté (DDL) 38
 - Démarche d'étude 2
 - Dendogramme 86, 87, 91
 - Descriptive, méthode 20
 - Détermination, coefficient 137
 - Diagramme
 - de dispersion 187
 - en bâtons 31
 - en secteurs 31
 - générateur de 185
 - Différentiel sémantique, échelle de mesure 15
 - Dispersion 31, 32
 - diagramme 187
 - Distance 83
 - de Minkowski, mesure 84
 - de Tchebycheff, mesure 84
 - du diamètre, méthode 85
 - euclidienne 83
 - mesure 83
 - moyenne, méthode 85
 - Distribution 31
 - normale 33
 - Données
 - collecter 3
 - écrire 20
 - expliquer 20
 - non structurées 8
 - normalité 109
 - primaires 8, 9
 - saisir 130
 - secondaires 3, 7, 8
 - secondaires externes 7
 - secondaires internes 7
 - structurées 8
 - textuelles 89
 - Duncan, test de 109
- E**
- Écart type 32
 - Échantillonnage
 - boule de neige 11
 - de convenance 11
 - méthode 10
 - stratifié 10
 - Échantillons 9
 - aléatoires 10
 - appariés 35, 41
 - indépendants 35, 41
 - non probabilistes 10
 - probabilistes 10
 - taille 11
 - uniques 41
 - Échelle 17
 - d'intention 16
 - d'Osgood 15
 - de Likert 16
 - de Stapel 16
 - neutralité 18
 - Échelle de mesure 16, 17
 - accord 15
 - différentiel sémantique 15
 - intensité 15
 - intention 15

Likert 15
 métrique 15
 nominale 15
 ordinaire 15
 Stapel 15
 traduction 17
 Éditeur de diagramme 186
 Effectifs 30
 Effets
 d'interaction 112
 de contamination 18
 de halo 18
 de lassitude 18
 principaux 112
 Égalité des moyennes, hypothèse 108
 Eigenvalue 57
 Emboîtement 184
 Empiler 182
 Entretien 3
 individuel 5
 Épuration des données 59
 Equamax 58
 Erreur 11
 aléatoire 18, 19
 de type 1 40
 de type 2 40
 indépendance des termes 136
 marge d' 12
 systématique 18, 19
 termes d' 18
 types 40
 Étapes de l'analyse conjointe 158
 Étendue 32
 Étude
 ad hoc 4
 brief de l' 178
 d'audience 12
 de cas unique 121
 démarche 2
 descriptive 6
 explicative 6
 exploratoire 18
 omnibus 4
 prédictive 6
 projet 178
 qualitative 5
 quantitative 5, 6
 Euclidienne, mesure de distance 83
 Expérimentation 120
 Explicative, méthode 21
 Expliquer les données 20

F

Facettes 16
 Facteurs 54
 d'inflation de la variance 142
 Factorielle, analyse 54, 56
 Factorisation 63
 Fiabilité 19, 53
 analyse 65
 Formes alternatives, technique 53
 Fractiles 32
 Fréquences 30

G

Générateur de diagramme 185
 Graphiques 31, 181
 Gréco-latin 123
 Groupe statique 121

H

Halo, effet 18
 Hasard, tirage au 10
 Hiérarchique, méthode 84
 Histogrammes 31
 Homogénéité 109
 Homoscédasticité 136
 Hypothèse
 alternative 40
 d'égalité des moyennes 108
 nulle 40
 statistique 40

I

Indépendance des termes d'erreur 136
 Indice
 de Rogers et Tanimoto, mesure de 84
 de Sokal et Michener, mesure de 84
 de Sokal et Sneath, mesure de 84
 Inférence, principe 39
 Inflation de la variance, facteur 142
 Intensité, échelle de mesure 15
 Intention
 échelle 16
 échelle de mesure 15
 Interaction 112, 122
 effet 112
 Intervalle 32
 de confiance 12
 Items 16
 Itinéraires, méthode 11

K

Kaiser-Guttman, règle de 57
 Kaiser-Meyer-Olkin (KMO) 57
 Kolmogorov-Smirnov, test de 42
 Kurtosis (coefficient d'aplatissement) 32

L

Lambda 39
 Lassitude, effet 18
 Liberté, degrés de (DDL) 38
 Likert, échelle de mesure 15, 16
 Linéarité 22, 136
 Loi normale 33

M

MANCOVA 116
 MANOVA 115
 Marché-test 5
 Marge d'erreur 12
 Marketing

- de masse 80
- individualisé 80
- segmenté 80

 Matrice

- de corrélation 57
- de covariance 57
- des corrélations anti-image 57

 McNemar 43
 Measure of Sampling Adequacy (MSA) 57
 Médiane 31
 Mesures

- appariées 41
- de la dispersion 32
- de proximité 83
- indépendantes 41
- outils 17

 Mesures de distance 83

- corrélations de Pearson 84
- distance de Minkowski 84
- distance de Tchebycheff 84
- distance euclidienne 83
- indice de Rogers et Tanimoto 84
- indice de Sokal et Michener 84
- indice de Sokal et Sneath 84

 Méthode

- d'agrégation 85
- d'échantillonnage 10
- de l'analyse factorielle 61
- de sélection des variables de régression 142
- de Ward 86
- des barycentres 85

- des centres mobiles 88
- des itinéraires 11
- des nuées dynamiques 88
- des quotas 10
- des scénarios 121, 160
- des vignettes 160
- descriptive 20
- distance du diamètre 85
- distance moyenne 85
- du plan d'expérience 120
- explicative 21
- hiérarchique 84
- non hiérarchique 84
- non probabiliste 10
- probabiliste 9
- saut minimum 85

 Métrique, échelle de mesure 15
 Minkowski, mesure de distance 84
 Mode 31
 Modèle

- additif 156
- compensatoire 156
- de décomposition 156
- de la vraie valeur 18
- statistique 121

 Moyenne 31
 Multicolinéarité 142
 Multivariée

- de la variance, analyse 115
- analyse 20

N

Netnographie 5, 6
 Neutralité d'une échelle 18
 Nominale, échelle de mesure 15
 Non hiérarchique, méthode 84
 Non probabiliste, méthode 10
 Normalité 136

- des données 109

 Nuage de points 81
 Nuées dynamiques 88

- méthode 88

O

Oblimin direct 58
 Observation 9
 Ordinale, échelle de mesure 15
 Orthogonalité 124
 Osgood, échelles de mesure 15
 Outils de mesure 17
 Outliers 31

P

- Panels 4, 8
 - d'audience 9
 - de distributeurs 9
- Pearson, coefficient 134
- phi, coefficient 39
- Plan
 - complet 160
 - factoriel 122
 - factoriel complet 123
 - factoriel fractionné 123
 - fractionné 160
- Plan d'expérience, méthode 120
- Points clés du rapport 180
- Population 9
- Précision
 - des résultats 11
 - statistique d'un test 12
- Prétest de compréhension 18
- Prétest/post-test et groupe de contrôle 121
- Principaux, effets 112
- Principe d'inférence 39
- Probabiliste, méthode 9
- Projet d'étude 178
- Proximité, mesure de 83
- Puissance du test 40

Q

- Quartiles 32
- Quartimax 58
- Questionnaire 3, 16
- Quotas, méthode 10

R

- R^2 137
- Rapport
 - d'analyse 178
 - d'étude, structure 179
 - points clés 180
- Règle
 - de Kaiser-Guttman 57
 - des valeurs propres 57
- Régression
 - conditions d'application 136
 - linéaire 136
 - linéaire multiple 141
 - linéaire simple 136
 - multiple 141
- Résultats, précision 11
- Rétro-traduction 18
- Réunions de consommateurs 3

- Risque d'artefact 82
- Rotation 62
 - des facteurs 58
 - oblique 58
 - orthogonale 58

S

- Saisir les données 130
- Saut minimum, méthode 85
- Scénarios, méthode 121, 160
- Scree Test 57
- Segmentation 80
- Sélection des variables de régression, méthode 142
- Seuil
 - d'acceptabilité de l'alpha 53
 - de confiance 40
 - de signification 40
- Signe 43
- Signification, seuil 40
- Skewness, coefficient de symétrie 32
- Sondage, taux 11
 - unités de 9
- Split half 53
- Standardisation 84
 - coefficient 137
- Stapel, échelle 15, 16
- Statistique d'un test, précision 12
- Structure d'un rapport d'étude 179
- Structurel, coefficient 60
- Symétrie 32
 - coefficient (Skewness) 32
 - d'une échelle 18

T

- t de Student 42
- Tableaux 181
 - croisés 36
 - personnalisés 182
 - pivotants 182
- Taille de l'échantillon 11
- Taux de sondage 11
- Taxinomie 80
- Tchebycheff, mesure de distance 84
- Techniques
 - des formes alternatives 53
 - qualitatives 4
 - quantitatives 4, 6
- Tendance centrale 31
- Termes d'erreur 18
- Test
 - /retest 53
 - d'ajustement 42

- d'hypothèses 35, 41
 - d'inférence 41
 - de Bonferroni 109
 - de comparaison d'échantillons appariés 42
 - de comparaison d'échantillons indépendants 42
 - de Duncan 109
 - de Kolmogorov-Smirov 42
 - de l'éboulis 57
 - de la médiane 43
 - de Levene 109
 - de McNemar 44
 - de Scheffé 109
 - de Sphéricité de Bartlett 57
 - de Tukey 109
 - de Wilcoxon 43
 - du coude 57
 - du khi-deux 37
 - du signe 43
 - non paramétrique 35, 41, 42, 109
 - paramétriques 35, 41
 - post hoc 109
 - précision statistique 12
 - puissance 40
 - statistiques 39
 - t* 42
 - t* pour échantillon unique 13
 - U de Mann-Whitney 43
 - Z 42
- Tirage au hasard 10
- Traduction d'échelles de mesure 17
- Tri
- à plat 31
 - croisés 36, 37
- Type
- d'analyse de variance 108
 - d'erreurs 40
- Typologie 80
- analyse 81, 90

U

- U de Mann-Whitney, test de 43
- Unités de sondage 9
- Univariée, analyse 20

V-W

- V de Cramer 39
- Valeurs
 - extrêmes 31
 - propres, règle des 57
- Validité 19, 90
 - convergente 52
 - de contenu 52
 - discriminante 52
 - faciale 18, 52
 - nomologique 52
 - prédictive 52
- Variable 30
 - de segmentation 81
 - dépendante 14
 - dépendante, analyse bivariée 36
 - explicative 14
 - expliquée 14
 - indépendante 14
 - indépendante, analyse bivariée 36
 - médiatrice 14
 - modératrice 14
 - qualitative 14, 30
 - quantitative 15, 31
- Variance 32
 - analyse de 108
 - coefficient 32
 - facteur d'inflation 142
 - intragroupe 109
- Varimax 58
- Vignettes 121
 - méthode 160
- Vraie valeur 19
- Ward, méthode 86