

Fiche synthétique sur les Cartographies de Préférences

Auteur : Jean-François PETIOT

Comment expliquer les préférences des consommateurs par les caractéristiques sensorielles des produits ?

MOTS CLEFS : carte perceptuelle, modèle linéaire, point idéal, modèle vecteur, plan d'expériences, estimation statistique

CONTEXTE

Vous souhaitez optimiser votre offre produit (ou service) en fonction des préférences de vos clients, ou plus généralement vous souhaitez connaître comment évoluent les préférences en fonction des caractéristiques sensorielles ou sémantiques des produits.

Le comportement du client vis à vis d'un produit résulte d'arbitrages complexes mettant en jeu les perceptions et préférences. Pour réussir la conception d'un produit, il est donc important de comprendre ces arbitrages, en analysant les appréciations des clients et leurs préférences par rapport aux perceptions.

Cette fiche présente une méthode, les **cartographies des préférences** (preference mapping en anglais), qui apporte une réponse à cette question. Elle expose les principes de construction de ces cartographies ainsi que l'interprétation des résultats.

L'exploitation de ces cartographies vous permettra :

- D'élaborer une offre produit pertinente,
- De segmenter l'offre produit en fonction de la préférence de différentes cibles,
- De comprendre ce qui structure la préférence client (pourquoi les clients aiment ou n'aiment pas un produit),
- De caractériser sensoriellement les produits préférés.

EN PRATIQUE

ETAPE 1

Constitution de l'espace produits

L'objectif de cette étape est de constituer un ensemble de produits représentatifs sur lesquelles vont porter les analyses perceptuelles et de préférence.

Le choix des produits de **l'espace produit** dépend tout d'abord de l'objectif et du thème de l'étude. Les questions suivantes vont vous permettre de cerner ces éléments :

- quel est le produit que j'étudie (un son d'alarme, un système d'éclairage, un boîtier de téléphone portable, un flacon de parfum,) et quelles sont les modalités sensorielles mises en jeu (audition, vision, olfaction, goût, toucher),
- pour le produit que j'étudie, et parmi l'ensemble des variables de conception qui le définissent, quelles sont celles qui sont fixées (qui constituent des données du problème), et quelles sont celles qui sont variables (qui seront l'objet de décision de conception),
- Quel est l'objectif général de l'étude ?
 - o Améliorer la qualité perçue,
 - o Comprendre l'influence de différentes variables de conception du produit sur la perception,
 - o Mieux maîtriser les aspects connotatifs du produit, en comprenant les liens entre les variables du produit et la sémantique du produit,

Ensuite, le choix de l'**espace produit** dépend également des contraintes matérielles existantes, des disponibilités des moyens expérimentaux, et des sujets :

- Quels produits différents ai-je à disposition pour effectuer l'étude perceptive ?
 - o Des produits réels : produits internes, de la concurrence, prototypes, ...
 - o Des produits virtuels (sons de synthèse, maquette virtuelle CAO). Quelles sont les possibilités de paramétrages des produits virtuels ? Quel mode de présentation des produits peut être utilisé ? (réalité virtuelle, écran, casque, ...)
- Quels moyens expérimentaux ai-je à disposition pour effectuer les épreuves perceptives ? (Salle d'évaluation en réalité virtuelle, salle d'écoute, ...)
- Quel panel de sujets puis-je mobiliser pour cette étude ?

Des réponses à ces différentes questions va découler une proposition d'espace produit, sachant que l'espace produit constitué pour réaliser une épreuve perceptive est toujours un compromis entre un espace de taille faible (pour limiter la fatigue des sujets) et un espace de taille suffisante (pour obtenir une information suffisamment riche). Vous devrez donc choisir des produits qui couvrent correctement l'espace de conception, étant donné le nombre de variables de conception considéré dans l'étude. Il faut éviter de choisir un groupe de produits très similaires, et un seul produit très différent des autres. Les résultats risqueraient d'être caricaturaux, binaires. En pratique, on s'oriente généralement vers un choix de 10 à 30 produits, représentatifs de la diversité sensorielle existante. La théorie des plans d'expérience peut être mobilisée pour définir un ensemble de produits pertinents (plans fractionnaires, tables de Taguchi, ...).

A noter.

Réaliser une épreuve perceptive est une tâche assez lourde à réaliser car elle implique des tests avec des sujets sur des produits. Il faut donc être sûr que c'est bien d'une épreuve perceptive dont on a besoin pour fournir une réponse à l'objectif de l'étude, et que l'analyse d'un expert du domaine, ou qu'une ou plusieurs mesures physiques, ne sont pas suffisantes.

ETAPE 2

Mesurer et caractériser les perceptions

L'objectif de cette étape est de disposer de mesures représentatives des différences perceptives entre les produits de l'espace produit, déclarées par des sujets. L'épreuve la plus courante dans ce cadre est la cotation sur échelle sémantique. Pour cela, vous devez définir un ensemble de termes pertinents pour représenter les différences perçues entre les produits. Une séance de verbalisation libre sur différents produits, impliquant plusieurs sujets, individuelle ou en groupe, va aider à générer ces termes, en suggérant les questions suivantes :

- Décrire chaque produit par des termes caractéristiques du point de vue de votre perception,
- Choisir les 2 produits de l'espace produit que vous considérez comme très différents. Quel(s) terme(s) sont caractéristiques de leurs différences d'un point de vue perceptif ? même question avec des produits très similaires.

Il s'agit ensuite de regrouper les verbalisations générées en familles homogènes (synonymes), en choisissant des termes précis, consensuels, non ambigus, afin de créer la liste de descripteurs. Une séance de groupe impliquant différents acteurs du projet est particulièrement adaptée pour générer cette liste, qui comprend généralement entre 5 et 15 termes. Ensuite, il s'agit de faire coter par un panel de sujets les différents produits de l'espace produit sur les échelles sémantiques définies. On peut utiliser par exemple une échelle catégorielle à 5 niveaux, dont un niveau neutre (figure 1), ou une échelle non structurée (figure 2).

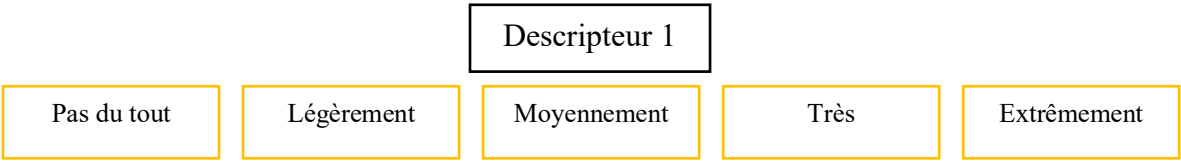


Figure 1 : échelle sémantique structurée

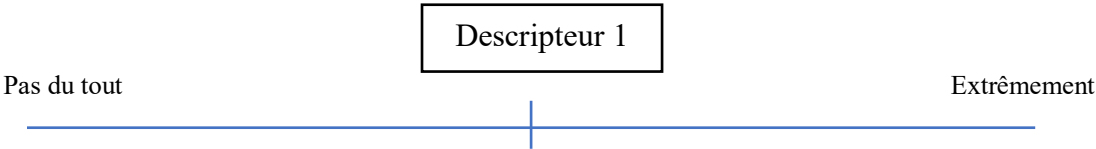


Figure 2 : échelle sémantique non structurée

Différentes épreuves pour les tests perceptifs.

En plus du différentiel sémantique, de très nombreuses méthodes, issues des recherches en psychophysique, sont disponibles. Leur complexité de mise en œuvre et les résultats qu’elles peuvent fournir peuvent être très différents. On peut citer en particulier : estimation de grandeur ; évaluation absolue ; comparaisons par paires ; évaluations comparées ; analyse multidimensionnelle ; catégorisation (tri) ; descriptions verbales.

Exemple de descripteurs pour caractériser le toucher : le Sensotact®.

La caractérisation du toucher d’une surface avec la main a fait l’objet de la création des 10 descripteurs suivants. Ils peuvent par exemple être utilisés pour caractériser la qualité des matériaux d’une planche de bord automobile (Renault Bureau d’études).

Type	Descripteur	Définition/mode d’évaluation
Toucher orthogonal	dureté	Résistance à la pénétration de l’ongle
	nervosité	Résistance à la vitesse de pénétration du doigt
	effet mémoire	Maintien d’une forme après déformation
	collant	Force d’adhésion sur le doigt après pression
Toucher tangentiel	fibreux	Différence de relief entre deux directions orthogonales
	relief	différentes hauteurs perçues à la surface du produit (ongle)
	râpeux	Caractère abrasif de la surface
	bloquant	Force nécessaire pour amorcer le déplacement de l’index à la surface du doigt
	glissant	Force nécessaire pour maintenir une force de déplacement constante de l’index
Toucher thermique	thermique	Flux de chaleur perçu par la main

La mesure perceptive est ainsi réalisée par un panel de sujets, qui cote chaque produit de l’espace produit selon les différentes échelles sémantiques. Pour aller plus loin dans le domaine des mesures perceptives, **l’analyse sensorielle** (cf. encadré) est une méthode très intéressante qui propose différents protocoles expérimentaux suivant les différents cas d’étude.

La sortie de cette étape 2 est un tableau de données qui propose une évaluation de chaque produit de l’espace produit selon chaque descripteur sensoriel DS (figure 3).

		Descripteurs sensoriels			
		DS 1	DS 2	...	DS J
produits	Prod. 1				
	Prod. 2				
	Prod. 3				
	...				
	Prod. I				

Figure 3 : forme du tableau de données présentant les données sensorielles

Pourquoi a t-on besoin d'un panel de sujets ?

L ‘évaluation perceptive est fonction des aptitudes sensorielles (et donc génétiques) du sujet. Pour s’affranchir de cette dépendance, et obtenir une réponse moyenne généralisable, on implique différents sujets dans l’évaluation. Le nombre de sujets à considérer dépend de la fiabilité attendue des résultats et repose sur des considérations liées au risque des statistiques décisionnelles. En pratique, un minimum de 5 à 10 sujets est nécessaire pour une épreuve descriptive.

Comment exploiter les résultats de l'étude perceptive ?

La première chose à faire pour exploiter les résultats est de visualiser les données perceptives. Avec des données quantitatives, vous pouvez tracer la « boîte à moustache » (boxplot) représentant la distribution des notes d’évaluation pour chaque descripteur. Ce graphe permet d’estimer en un seul coup d’œil les différences interproduits et les différences interindividuelles. La figure 4 montre par exemple que les évaluations des produits 1 et 2 sont très consensuelles (faibles différences interindividuelles), contrairement aux évaluations du produit 3. Elle montre également qu’il existe une différence marquée entre les produits 1 et 2.

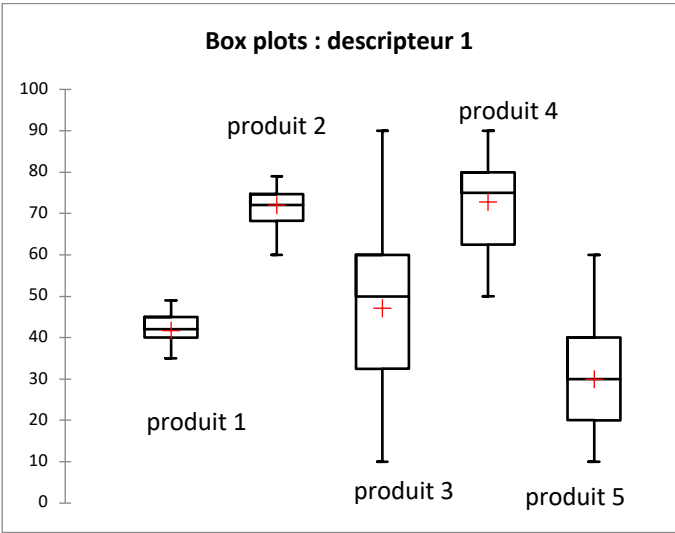


Figure 4 : visualisation de données de l’étude perceptive (boîte à moustache)

La seconde chose à faire est de produire des graphes synthétiques qui représentent les principales différences perceptives entre les produits. Le nombre de descripteurs sensoriels étant généralement important (entre 5 et 15), il n’est pas possible de tracer des cartes avec ce nombre de dimensions. Il s’agit de résumer l’information en produisant des cartes perceptuelles

à 2 ou 3 dimensions, qui présentent un positionnement relatif des produits et des descripteurs afin d'interpréter facilement les différences. A ce niveau, l'analyse en composante principale est une méthode très utilisée (cf. fiche outil2 Analyse en composantes principales).

L'analyse sensorielle.

A l'origine développée dans l'industrie alimentaire, l'analyse sensorielle est désormais pratiquée dans de nombreux secteurs de l'industrie. Le principe est de mesurer et caractériser les propriétés sensorielles d'un produit en utilisant les organes des sens de l'être humain. Elle produit une mesure de l'interaction entre un produit (appelé stimulus) et un sujet, par une évaluation des perceptions. La méthode, qui fait l'objet de nombreuses normes (voir bibliographie) est basée sur l'entraînement d'un **panel d'experts** sur une liste de **descripteurs sensoriels**, et utilise les outils statistiques de test et d'analyse de données pour caractériser et fiabiliser la mesure sensorielle. Au départ limitée aux aspects « contrôle qualité » de produits, elle est devenue maintenant une méthode compétitive pour expliquer les préférences des consommateurs et pour le développement industriel de produits nouveaux.

ETAPE 3

Recueillir les préférences des consommateurs

Pour pouvoir prédire l'acceptabilité d'un produit par les consommateurs, et effectuer une synthèse de produit sur une formulation préférée, il est nécessaire de recueillir un deuxième type de données, les données **hédoniques**, ou de **préférences**. Celles-ci sont recueillies en effectuant des tests de préférences à grande échelle sur un ensemble de consommateurs (plusieurs centaines), qui fournissent généralement des réponses traduisant une grande diversité de points de vue qu'il est important d'exploiter pour la conception.

On distingue trois principaux types d'épreuves pour le recueil de données de préférence :

1. **Évaluation hédonique.** Le consommateur est invité à coter chaque produit sur une échelle prédéfinie (figure 5). Cette épreuve, assez peu instinctive, est employée si le nombre de produits est très important,
2. **Épreuve de classement.** Le sujet doit classer les produits du plus désagréable au plus agréable. Cette épreuve est « simple » pour l'expérimentateur mais « complexe » pour le sujet, la stratégie de classement devenant difficile si le nombre de produits est important (>15),
3. **Épreuve de comparaison par paires.** Le sujet doit indiquer sa préférence entre 2 produits, par exemple sur une échelle de catégories (>>, >, =, <, <<). Cette tâche est très naturelle et facile à réaliser, mais le calcul de « scores » de préférence peut s'avérer complexe car le sujet peut présenter de l'inconsistance ($A >> B >> C$ et $A > C$ voire de l'intransitivité ($A > B > C > A$)).

	1. Très désagréable		5. assez agréable
	2. désagréable		6. agréable
	3. assez désagréable		7. Très agréable
	4. ni désagréable ni agréable		

Figure 5 : exemple de questionnaire pour l'évaluation hédonique

La sortie de cette étape 3 est un tableau de données qui indique une évaluation de préférence de chaque produit par les consommateurs (figure 6).

C préférences des consommateurs

	Conso 1	Conso 2	...	Conso C
produits	Prod. 1			
	Prod. 2			
	Prod. 3			
	...			
	Prod. I			

Figure 6 : forme du tableau de données présentant les données de préférence

Suite au recueil des données de préférence, on pourrait procéder à une classification des consommateurs en groupes homogènes concernant les préférences : les segments. Diverses méthodes peuvent être utilisées, par exemple la classification ascendante hiérarchique (CAH).

La classification ascendante hiérarchique.

Il existe de nombreuses techniques statistiques visant à partitionner une population en différentes classes ou sous-groupes. La classification ascendante hiérarchique (CAH) est l'une d'entre elles. On cherche à ce que les individus regroupés au sein d'une même classe (homogénéité intra-classe) soient le plus semblables possibles tandis que les classes soient le plus dissemblables (hétérogénéité inter-classe). Le principe de la CAH est de rassembler des individus selon un critère de ressemblance défini au préalable qui s'exprimera sous la forme d'une matrice de distances, exprimant la distance entre chaque individu pris deux à deux. Deux observations identiques auront une distance nulle. Plus les deux observations seront dissemblables, plus la distance sera importante. La CAH va ensuite rassembler les individus de manière itérative afin de produire un dendrogramme ou arbre de classification. La classification est *ascendante* car elle part des observations individuelles ; elle est *hiérarchique* car elle produit des classes ou groupes de plus en plus vastes, incluant des sous-groupes en leur sein. En découpant cet arbre à une certaine hauteur choisie (figure 7), on produira la partition désirée (les segments).

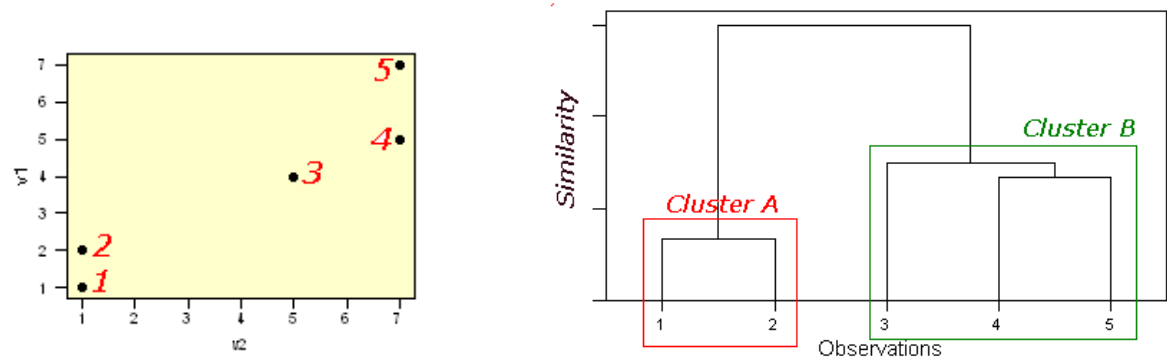


Figure 7 : exemple 5 individus (1, 2, 3, 4, 5) décrits par 2 variables v1 et v2 sont regroupés en 2 segments (cluster A, cluster B)

ETAPE 4

Construire la cartographie et exploiter les résultats

Cette dernière étape consiste à mettre en relation les données sensorielles (ETAPE 2) et les données de préférences (ETAPE 3). La méthodologie générale est présentée figure 8.

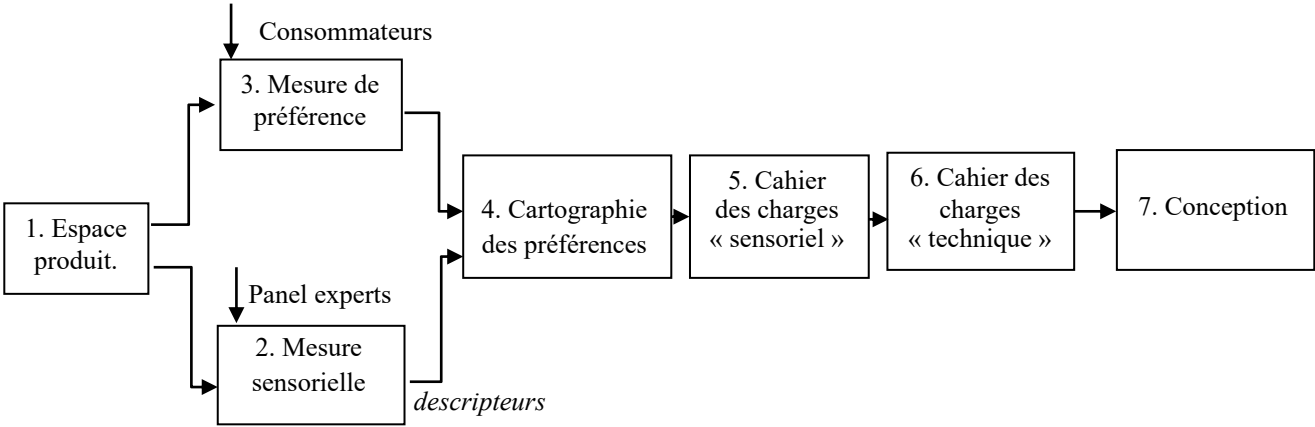


Figure 8 : méthodologie de construction des cartographies de préférences

A l’issu des ETAPE 2 et 3, nous disposons de 2 types de données caractérisant les produits de l’espace produit : les données sensorielles, représentées par l’évaluation quantitative d’un ensemble de descripteurs, et les données de préférences, représentées par les cotations de préférence d’un ensemble de consommateurs. Les données se présentent sous forme de 2 tableaux que l’on peut juxtaposer (figure 9).

C préférences des consommateurs					J descripteurs sensoriels et/ou physico-chimiques			
					DS 1	DS 2	...	DS J
produits	Prod. 1							
	Prod. 2							
	Prod. 3							
	...							
	Prod. I							

Figure 9 : description des données de préférence (en rouge) et des données sensorielles (en bleu) sur l’espace produit

L’idée clé des cartographie externes¹ de préférences est de mettre en relation les évaluations de préférence de chaque consommateur avec les données sensorielles. Le principe est tout d’abord de choisir les variables sensorielles candidates qui sont chargées d’expliquer les préférences. Généralement, afin de pouvoir faire des représentations dans un espace de faible dimension (2 ou 3), et décorréler les dimensions, on choisit comme espace de représentation des produits les premiers facteurs (F1, F2, ...) d’une analyse en composante principale (voir fiche outil1) des données sensorielles.

Ensuite, il s’agit d’identifier les régions de l’espace de représentation où les préférences des consommateurs sont maximales, afin d’orienter la conception vers des formulations préférables

¹ Par opposition aux cartographies internes des préférences, qui étudient directement la structure de préférence indépendamment des données sensorielles.

au niveau sensoriel. Pour cela, on ajuste un modèle de préférence sur les données de chaque consommateur, par une régression linéaire multiple (voir fiche outil2).

L'hypothèse forte de la cartographie externe de préférence consiste à dire que des produits voisins sensoriellement (proches dans l'espace sensoriel) doivent avoir des préférences voisines.

Les méthodes de cartographie externe des préférences consistent à effectuer, pour chaque consommateur, une régression polynomiale en utilisant comme variables explicatives les coordonnées sensoriellles (ou les coordonnées factorielles) notées (F1, F2) de chaque produit, et comme variable à expliquer les notes de préférences P données par le consommateur (par exemple la colonne CONSO 1 du tableau figure 9).

Différents modèles peuvent être proposés, du plus simple au plus complexe.

a) Modèle de degré 1 (modèle vecteur). La surface modélisée est alors un plan.

$$\hat{P} = a + bF1 + cF2$$

Mais plutôt que de représenter ce plan en 3d (ce qui donnerait des graphes peu lisibles), on va représenter simplement la **ligne de plus grande pente** du plan sur le plan sensoriel (F1, F2).

Techniquement, on peut montrer que la ligne de plus grande pente du plan est orientée selon le vecteur V de composantes (b, c) (d'un point de vue mathématique, ce vecteur correspond au vecteur gradient de P).

Pour avoir une interprétation graphique de la régression, on trace sur le plan des descripteurs (figure 10) un vecteur :

- dont l'origine est arbitraire, par ex. au centre de la figure,
- dont la norme est proportionnelle à R^2 ,
- dont la direction est portée par le vecteur (b, c).

Ce vecteur, appelé **modèle vecteur**, représente la direction d'augmentation de préférence du consommateur (figure 10). Les iso-préférence (courbe de niveaux) sont perpendiculaires au modèle vecteur. On peut conclure par exemple en analysant la figure 10 que la préférence pour le produit C5 est supérieure à la préférence pour le produit C2 (le modèle vecteur est orienté selon la direction de l'augmentation de préférence).

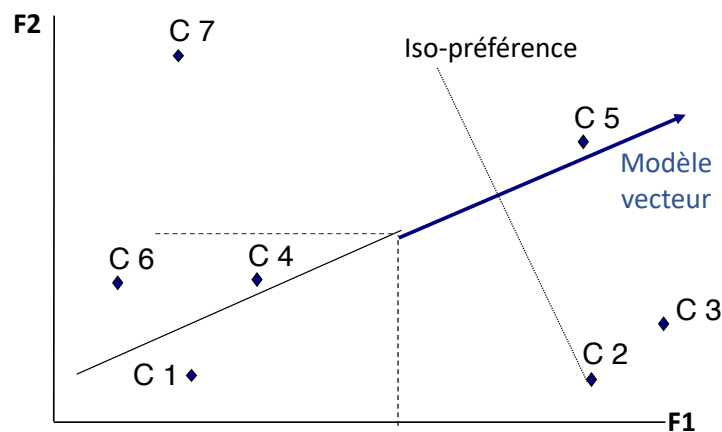


Figure 10 : représentation du modèle vecteur dans le plan sensoriel

Ce modèle vecteur est intéressant pour modéliser les directions d'augmentation de la préférence, **mais il est caricatural** car il prédit que la préférence augmente constamment dans la direction du vecteur, ce qui n'est pas en accord avec l'observation des préférences humaines (on peut aimer un yaourt sucré sans pour autant vouloir du sucre pur !).

Il y a en fait deux limitations dans l'utilisation de ce modèle :

- le modèle n'est utilisable que dans la zone de l'espace sensoriel où il y a des données
- le modèle ne permet pas de modéliser une préférence avec une valeur idéale du descripteur sensoriel.

Pour cette dernière remarque, on constate en effet que les préférences des individus ne sont généralement pas linéaires de degré 1 (figure 11), mais avec une valeur idéale (on veut par exemple que le yaourt soit sucré mais pas trop).

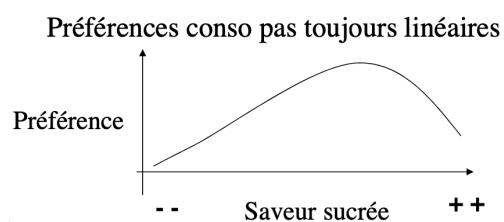


Figure 11 : évolution de la préférence en fonction de la saveur sucrée, avec une valeur idéale de taux de sucre

Pour tenir compte de ce comportement, un second modèle, de degré 2, est proposé.

b) Modèle de degré 2 (modèle circulaire). La surface modélisée est alors un paraboloïde de révolution.

$$\hat{P} = a + bF_1 + cF_2 + d(F_1^2 + F_2^2)$$

La forme de la surface de réponse peut alors représenter un point maximum (point idéal, lorsque $d < 0$ – qui traduit un comportement appelé « préféreur ») ou alors un point minimum (point anti-idéal – $d > 0$ – qui traduit un comportement de « rejeteur ») (figure 12).



Figure 12 : représentation d'une surface de degré 2 (paraboloïde de révolution) avec un comportement de « préféreur » ($d < 0$: plus on s'éloigne de l'idéal, plus la préférence décroît) ou de « rejeteur » ($d > 0$: plus on s'éloigne de l'anti-idéal, plus la préférence croît).

Pour avoir une interprétation graphique de la régression, on trace sur le plan des descripteurs (figure 13) le point extremum et les iso préférence (cercles) :

- un point idéal est représenté par un « + »,
- un point anti-idéal est représenté par un « - »

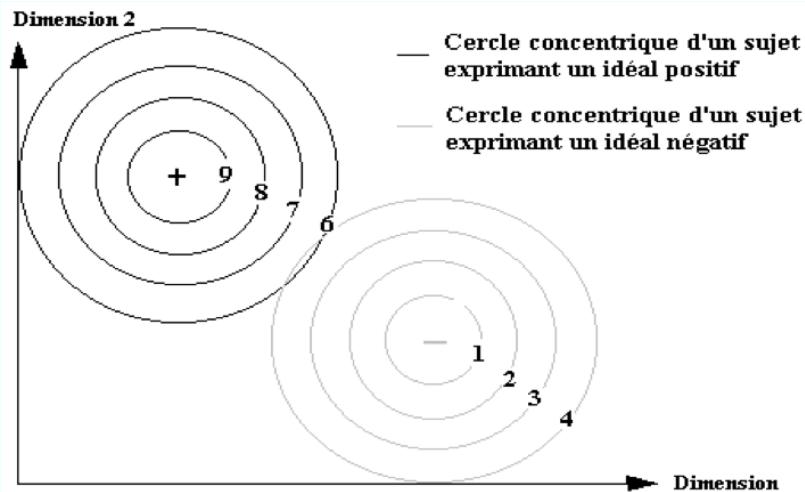


Figure 13 : représentation d'un point idéal positif (+) et point idéal négatif (-).

c) Modèle de degré 2 (modèle elliptique). La surface modélisée est alors un parabolôïde à base elliptique, ou un modèle en selle de cheval (lorsque d et e sont de signe opposé).

$$\hat{P} = a + bF_1 + cF_2 + dF_1^2 + eF_2^2$$

En plus des comportements « préféreur » ou « rejeteur » précédents, la forme de la surface de réponse peut représenter un comportement éclectique lorsque d et e sont de signe opposés (figure 14).

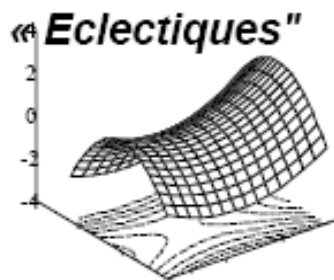


Figure 14 : représentation d'une surface de degré 2 en « selle de cheval » lorsque d et e sont de signe opposé.

Pour avoir une interprétation graphique de la régression, on trace sur le plan des descripteurs le point extremum (idéal ou anti-idéal). On ne trace rien si le modèle est « éclectique ».

d) Modèle de degré 2 (modèle complet). La surface modélisée est soit un parabolôïde, soit un modèle en « selle de cheval ».

$$\hat{P} = a + bF_1 + cF_2 + dF_1^2 + eF_2^2 + fF_1F_2$$

En plus des comportements « préféreur » ou « rejeteur » ou « éclectique » précédents, ce dernier modèle permet une rotation des axes et une interaction entre les dimensions F1 F2.

NB : ce modèle possède beaucoup de paramètres (a, b, c, d, e, f) ce qui nécessite beaucoup d'observations pour en avoir une estimation fiable.

EXEMPLE de cartographie.

La figure 15 représente les points idéaux (+) et anti-idéaux (-) d'un ensemble de consommateurs, pour les produits (A, B, C, D, E, F, G, H). On peut noter le caractère **segmentant** (aimés par les uns, rejetés par les autres) des produits G et E.

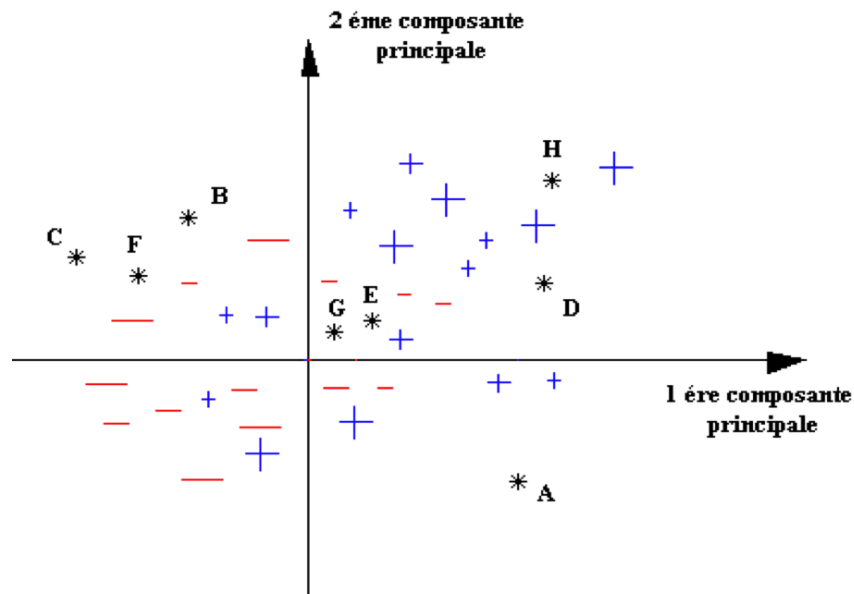


Figure 15 : Exemple de cartographie des préférences avec modèle circulaire pour un ensemble de consommateurs et les produits A, B, C, D, E, F, G, H.

Un test de significativité de la régression (test de Fisher²) permet de décider quel modèle il faut conserver (sachant que l'on commence toujours par le modèle le plus simple). Ce test contrôlera la significativité globale de la régression linéaire multiple pour un certain seuil de confiance. En outre, une régression non significative pour un modèle donné pourra l'être pour un modèle plus complexe. Le coefficient de détermination de la régression, R^2 , est également étudié : il indique le pourcentage d'information pris en compte par la régression.

Attention

L'interprétation des résultats d'un modèle en cartographie des préférences s'effectue sous réserve d'un **ajustement correct du modèle sur les données, et d'une prédiction correcte**³. Un mauvais ajustement peut être dû à une inadéquation du modèle, ou à des réponses du sujet incohérentes ou fantaisistes. Si l'ajustement est mauvais, on peut être amené à retirer les consommateurs de l'étude. En pratique, l'ajustement dans le cas de la méthode des moindres carrés est estimé à l'aide du coefficient de détermination R^2 , qui indique le pourcentage de variance capté par le modèle. On éliminera les sujets pour lesquels le R^2 est inférieur à un certain seuil.

² *Test de Fisher* (test de la significativité de la régression dans son ensemble). Il permet de vérifier la pertinence d'un lien linéaire entre la variable à expliquer Y et les variables explicatives X_k , pour un seuil de risque p appelé seuil de significativité. Pour $p = 5\%$, si $F > F_{\text{seuil}}$, alors au moins une des composantes X_k apporte une contribution significative à l'explication du comportement de la variable Y . Autrement dit, le risque qu'il n'y ait pas de lien linéaire entre Y et une des X_k est plus petit que 5% .

³ La prédiction peut être estimée à l'aide de la statistique PRESS

Préférence cumulée

La dernière étape de la cartographie des préférences va consister à agréger les résultats pour l'ensemble du panel de consommateurs (sous réserve d'avoir vérifié l'homogénéité des préférences, c'est-à-dire que les consommateurs n'ont pas des préférences radicalement différentes).

L'idée est, pour chaque consommateur, de séparer le plan sensoriel de représentation en 2 zones : la zone de préférence (ou la préférence est supérieure à la valeur moyenne de préférence sur le plan sensoriel d'intérêt P_{moy}) et la zone de rejet (ou la préférence est inférieure à la valeur moyenne de préférence P_{moy}). On définit ainsi une indicatrice $C(x, y)$ sur les 2 dimensions F1 et F2 (voir figure 16).

$$C(x, y) = 1 \text{ si } P(x, y) \geq P_{moy}$$

$$C(x, y) = 0 \text{ si } P(x, y) < P_{moy}$$

Ensuite, il suffit de sommer les indicatrices de tous les consommateurs du panel pour obtenir la carte de préférence P .

$$P(x, y) = 100 \cdot \frac{\sum_{i=1}^n C_i(x, y)}{n}$$

Cette fonction représente le % de consommateurs pour lesquels la localisation du point produit est située dans leur zone de préférence. On trace les courbes de niveau de cette fonction, qui peut permettre de mettre en évidence un ou plusieurs points maxima (figure 17).

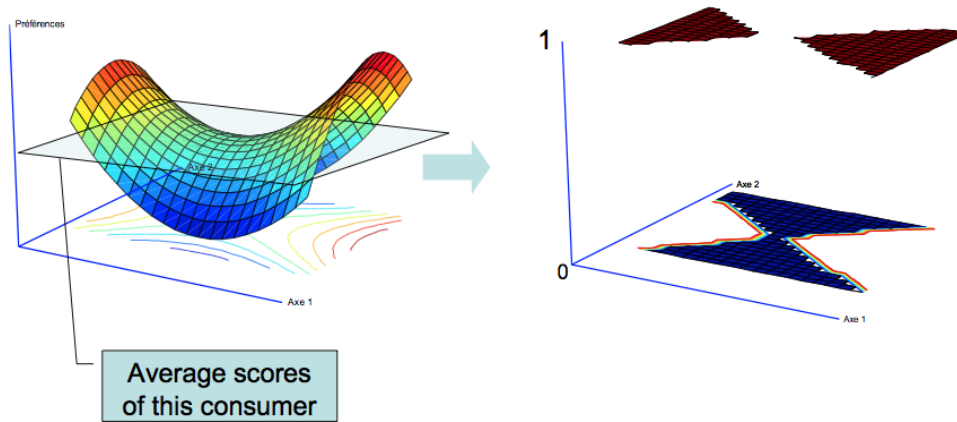


Figure 16 : définition dans le plan sensoriel des zones de préférence et zone de rejet pour un consommateur donné.

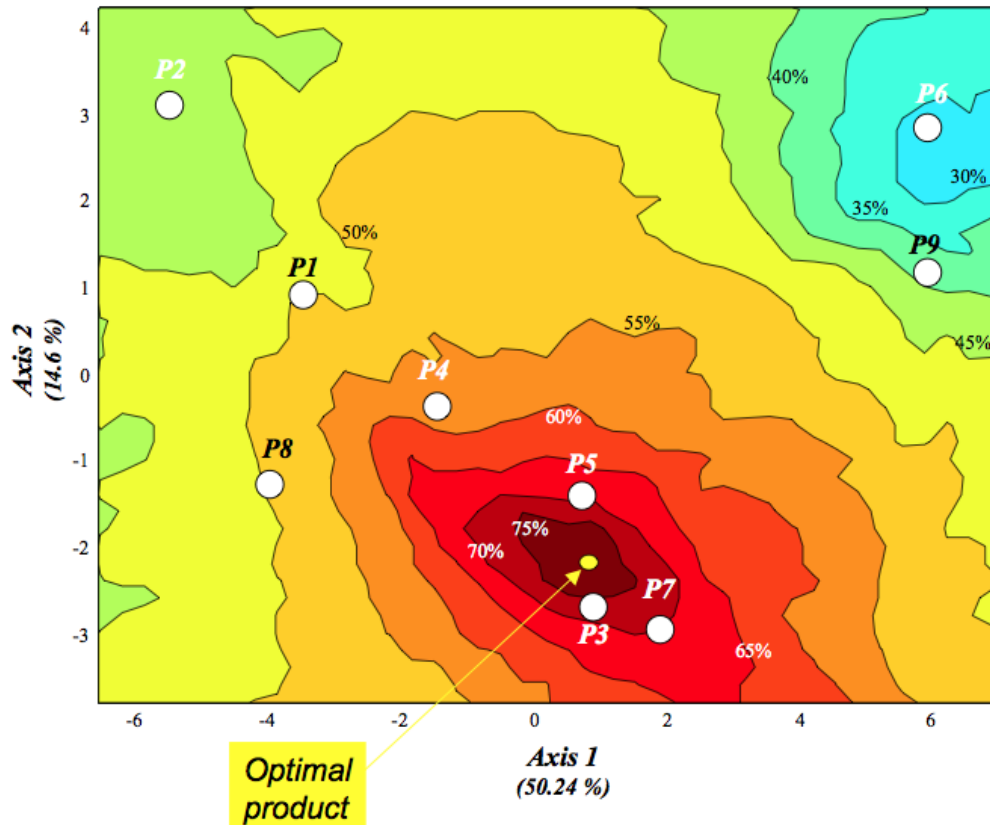


Figure 17 : courbes de niveau des % de consommateurs dans le plan sensoriel, mettant en évidence un point optimal pour lequel 75% des consommateurs apprécient le produit. Le point de rejet (30%) en haut à droite peut également présenter un intérêt pour la conception.

Au final, cette démarche permet de relier les caractéristiques sensorielles d'un produit aux caractéristiques des préférences de consommateurs. C'est pour cette raison un outil privilégié du marketing. Pour pouvoir être utile en conception, il faut ensuite étudier les relations entre les descripteurs sensoriels et des mesures physiques, chimiques ou biochimiques du produit, ainsi que des caractéristiques de fabrication. Ce travail s'effectue en utilisant l'arsenal des méthodes explicatives en statistique (analyse conjointe, analyse discriminante, analyse de la variance, régression, classification,...). Au final, on tient compte directement des préférences des consommateurs pour la conception du produit, ce qui est l'objectif ultime.

EXEMPLE

Pour illustrer la méthode, considérons un exemple très simplifié, issu de l'industrie alimentaire : la conception de la formulation d'une crème à la vanille. On suppose qu'une entreprise souhaite diversifier sa gamme et se lancer sur ce type de produit. Elle doit pour cela analyser les produits concurrents et proposer des tests d'acceptabilité afin de définir la formulation préférable au niveau sensoriel.

On suppose que 7 produits (C1 à C7), de formulations variables, ont été sélectionnés pour les tests (espace produit – ETAPE 1). On considère pour simplifier que les produits sont évalués par le jury d'experts selon deux descripteurs uniquement⁴ : un attribut de saveur : *saveur de vanille* (i.e. intensité de saveur de vanille) et un attribut de texture : *onctuosité* (quantité de

⁴ Dans le cas réel, il y a bien sûr plus de 2 descripteurs et on se ramène à deux ou trois dimensions principales à l'aide d'une technique mathématique dite « d'analyse en composantes principales (ACP) » qui permet de désigner les facteurs d'influence principaux.

particules détectables, de grumeleux à parfaitement lisse ; évaluée en plaçant l'échantillon avec la langue contre le palais). Les données sensorielles concernant les 7 produits, évaluées sur une échelle de 0 à 10, sont données dans la table 1 (ETAPE 2). La représentation correspondante est donnée en figure 18.

Table 1. Données sensorielles des 7 crèmes à la vanille

	Saveur de vanille : d1	Onctuosité : d2
C 1	2	1
C 2	8	1
C 3	9	2
C 4	3	3
C 5	8	6
C 6	1	3
C 7	2	8

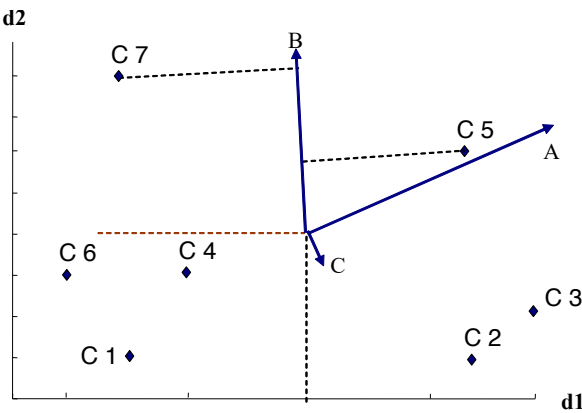


Figure 18 : exemple de cartographie externe des préférences, modèle vecteur, concernant 7 produits et 3 consommateurs A, B, C

On considère trois consommateurs⁵, notés A et B et C, dont les jugements de préférence concernant les 7 produits sont indiqués en table 2 (ETAPE 3). La note P correspond à l'appréciation portée par les consommateurs sur les produits étudiés⁶. Elle n'a de sens que comparée aux notes de la même colonne, aucune comparaison directe entre consommateurs n'a de sens.

Table 2. Données de préférence des 3 sujets A, B, C

P	A	B	C
C 1	2	6	5
C 2	7	6	4
C 3	8	6,5	7
C 4	4,5	7	8
C 5	9	7,5	2
C 6	1,5	7	1
C 7	5	8	1

⁵ Les méthodes de classification, qui visent à regrouper les consommateurs en segments suffisamment homogènes sortent du cadre de cette présentation et ne sont pas abordées ici.

⁶ Cette notation peut s'effectuer soit directement, soit être le résultat d'un calcul à partir de comparaisons relatives par paires.

La méthode de cartographie externe des préférences consiste, pour chaque consommateur, à tenter d'expliquer les notations de préférence des sujets par les données sensorielles définies par les experts (ETAPE 4).

On cherche tout d'abord le modèle le plus simple, un modèle de degré 1 entre la préférence et les données sensorielles (d_1 , d_2).

$$\hat{P} = a + bd_1 + cd_2$$

La surface modélisée est un plan. Les résultats de la régression pour chaque consommateur (estimateurs a , b , c , ainsi que coefficients de détermination R^2 et statistique F), obtenus avec un tableur courant, sont fournis dans la table 3.

Table 3. Résultat des régressions linéaires multiples pour les 3 consommateurs

	b	c	a	R^2	F	Significatif (p=0.05)
A	0,82	0,42	-0,02	0,98	97,84	Oui (p-value<0,05)
B	-0,02	0,27	6,02	0,94	31,17	Oui (p-value<0,05)
C	0,15	-0,59	5,33	0,37	1,17	Non (p-value>0,05)

Pour chaque consommateur, on peut tracer le modèle vecteur V de composantes (b , c) sur le plan sensoriel. Il représente la direction d'augmentation de préférence du consommateur et est tel que :

- l'origine est arbitraire, par ex. au centre de la figure,
- la norme est proportionnelle à R^2 ,
- la direction est portée par le vecteur (b , c).

La figure 18 présente les modèles vecteurs des 3 consommateurs A, B, C. On peut tirer les conclusions suivantes :

- le sujet A, dont les préférences sont bien expliquées par les données sensorielles (régression globalement significative, voir table 3 dernière colonne, R^2 grand). Il aime bien à la fois une saveur vanille marquée et les crèmes lisses,
- le sujet B, dont les préférences sont aussi bien expliquées par les données sensorielles. Il apprécie une crème très lisse et est relativement indifférent au goût de vanille,
- le sujet C, dont la notation de préférence n'est pas expliquée par les descripteurs (la régression n'est pas globalement significative, et le R^2 est faible (37%)).

A partir du modèle vecteur, on peut directement lire les valeurs prédites par la régression en menant la perpendiculaire au vecteur passant par le point considéré : les **iso-préférences sont ainsi des droites perpendiculaires au vecteur**, le vecteur est dirigé vers l'augmentation de la préférence. On voit ainsi sur la figure 8 que pour le consommateur B, la valeur prédite pour le produit C 7 est plus grande que la valeur prédite pour le produit C 5. Pour les sujets A et B, la modélisation est satisfaisante, on peut s'arrêter là.

Pour le sujet C, trois explications sont envisageables pour expliquer les résultats :

- ce sujet n'exprime pas de préférence, soit parce qu'il en est incapable, soit parce qu'il a répondu aux tests de façon fantaisiste,
- ce sujet est influencé par d'autres descripteurs que ceux proposés pour exprimer sa préférence,

- ce sujet a une structure de préférence plus complexe que les précédents, le modèle vectoriel n'est pas adapté pour la décrire. Une étude avec un modèle plus élaboré est nécessaire (par exemple un modèle circulaire).

Pour étudier cette dernière hypothèse, il faut maintenant construire un modèle circulaire pour le sujet C (modèle à point idéal). Le modèle possède la forme suivante (modèle de degré 2).

$$\hat{P} = a + b.d1 + c.d2 + d(d_1^2 + d_2^2)$$

Les résultats de la régression pour le sujet C (estimateurs a, b, c, d ainsi que coefficients de détermination R² et statistique F), obtenus avec un tableur courant, sont fournis dans la table 4. Le R² est de 52%, ce que l'on peut considérer comme suffisant même si la régression n'est pas significative (p-value>5%) (la confiance dans les résultats est limitée – cela est dû au fait qu'il y a peu d'observations (7), et 4 paramètres à estimer (a, b, c, d) – il faudrait plus d'observations pour augmenter la confiance dans les résultats). Néanmoins, on propose quand même d'interpréter les résultats.

Table 4. Résultat de la régression avec modèle circulaire pour le sujet C

	b	c	a	d	R ²	F	Significatif (p=0.05)
C	2	0,98	0,44	-0,18	0,52	1,10	Non (p-value>0,05)

Ce modèle possède un extremum au point :

$$C_{ideal} \begin{cases} -\frac{b}{2d} \\ -\frac{c}{2d} \end{cases} \qquad \text{c'est à dire} \qquad C_{ideal} \begin{cases} 5,5 \\ 2,7 \end{cases}$$

De plus, d étant négatif (-0,18), cet extremum est un maximum. On peut tracer ce point idéal sur le plan sensoriel (figure 19).

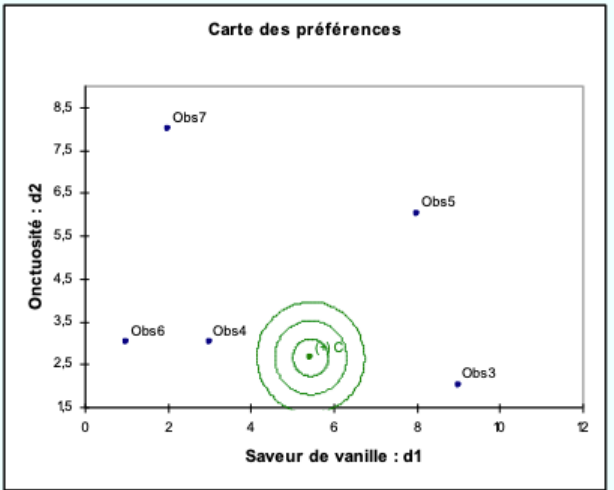


Figure 19 : représentation du point idéal dans le plan sensoriel pour le consommateur C.

Les iso-préférences sont des cercles centrés en C, plus on s'éloigne de C, plus la préférence décroît.

Synthèse des préférences pour les 3 consommateurs A, B, C

En utilisant les modèles de préférence des 3 consommateurs (A : modèle vectoriel, B : modèle vectoriel, C : modèle circulaire), on peut définir la carte qui cumule les zones de préférence (préférence supérieure à la préférence moyenne) pour les 3 sujets (figure 20).

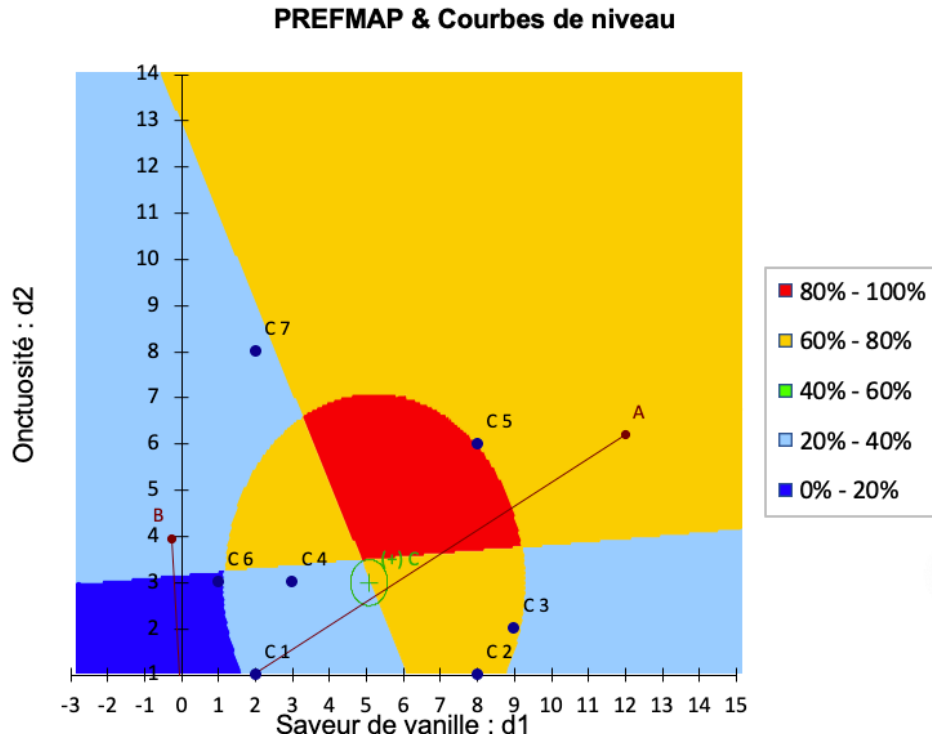


Figure 20 : représentation du cumul des zones de préférence avec les courbes de niveau pour les 3 sujets A, B, C.

Sur cette carte (figure 20), la zone rouge indique la position des produits pour lesquels plus de 80% des consommateurs ont une préférence supérieure à leur préférence moyenne.

En conclusion, il serait intéressant de développer un produit dans cette zone de l'espace sensoriel (cahier des charges sensoriel).

Bibliographie

- Evaluation Sensorielle, manuel méthodologique, 2^{ème} édition. Techniques et Documentation Lavoisier, collection Sciences et Techniques agro-alimentaires, 1998.
- Bonnet C. (1986), *Manuel pratique de psychophysique*, Armand Collin.
- Traité d'Évaluation sensorielle - Aspect cognitifs et métrologiques des perceptions, DUNOD, Paris, 2001.
- Petiot J-F., Yannou B. « Measuring consumer perceptions for a better comprehension, specification and assessment of product semantics». *International Journal of Industrial Ergonomics*. Vol 33/6 pp 507-525, 2004
- **Petiot J-F.** «Analyse sensorielle en conception de produits » Chapitre d'ouvrage : in *La conception industrielle de produits*, sous la direction de B Yannou, H. Christophol, D. Jolly, N. Troussier, Volume 3 (Ingénierie de l'évaluation et de la décision), Hermès, Lavoisier, 2008, pp 79-94.
- AFNOR NF ISO 8586-1, 1983, Analyse sensorielle - Guide général pour la sélection, l'entraînement et le contrôle des sujets - Partie 1 : Sujets qualifiés
- AFNOR NF ISO 8586-2, 1984, Analyse sensorielle - Guide général pour la sélection, l'entraînement et le contrôle des sujets - Partie 1 : Experts.

- AFNOR NF ISO 11035, 1995, Recherche et sélection de descripteurs pour l'élaboration d'un profil sensoriel par approche multidimensionnelle, 271-300.
- Dagher A. « Contribution à l'intégration des préférences client en conception de produits : application au design de formes ». Thèse de Doctorat de l'Ecole Centrale de Nantes, novembre 2008.
- Petiot J-F « Capter la part subjective du besoin : outils et exemples d'application » in *Intelligence en conception des produits et des services*, sous la direction de B. Yannou, Ph. Deshayé, Edition L'Harmattan, 2006.

Fiche Outil 1 : L’analyse en composantes principales

L’analyse en composantes principales est une technique classique d’analyse de données qui vise à **résumer au mieux** l’information contenue dans un tableau de données croisant n individus (par exemple des produits) et p variables (par exemple des descripteurs sensoriels caractérisant les produits), sachant que p peut être grand (plusieurs dizaines). Le principe est de définir un nombre faible q de variables (les composantes principales) qui résument au mieux les différences entre les produits. Lorsque $q = 2$, on pourra par exemple réaliser des représentations graphiques des individus dans un plan, ce qui permet de « visualiser » les proximités (ou au contraire les éloignements) entre les individus.

On suppose que l’on souhaite étudier les ressemblances ou différences entre les produits et les corrélations entre les variables (Y a-t-il des variables pour lesquelles les scores varient de la même façon pour tous les individus ? Ou des produits qui ont les mêmes évaluations pour toutes les variables). Pour cela, il est intéressant de faire une **représentation graphique des données**. Mais pour pouvoir le faire même lorsque le nombre p de variable est supérieur à 3, il impérativement résumer l’information et réduire la dimensionnalité de l’espace de représentation (selon les variables).

L’ACP permet de répondre à cette question. Elle propose de déterminer, à partir du tableau des « données brutes » (produit*variables) des variables annexes, appelées **composantes principales**, en nombre suffisamment faible pour permettre une représentation graphique du nuage de points le moins déformé que possible. Par exemple, à partir de la définition en 3D d’un poisson, une ACP montrerait que les 2 premiers axes principaux (qui conservent le mieux possible la forme du poisson) sont ceux présentés en rouge figure 1.

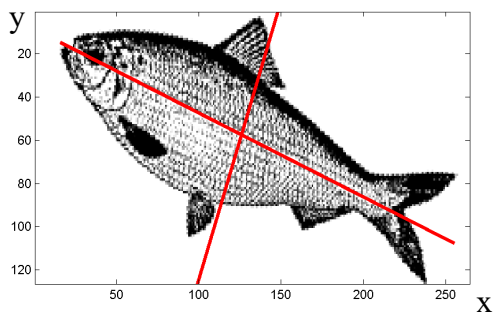


Figure 1 : les 2 directions principales de la forme d’un poisson

Considérons maintenant l’exemple suivant avec 5 voitures V1 à V5 représentées par leur score (score de 1 à 10) selon 4 variables perceptives (*sportivité*, *confort*, *habitabilité*, *sécurité*). La voiture V9 est par exemple très sportive (score = 9) mais peu habitable (score = 5).

Table 1 : évaluation perceptive de 5 véhicules (V1 à V5) selon 4 variables

	<i>sportivité</i>	<i>confort</i>	<i>habitabilité</i>	<i>sécurité</i>
V1	8	5	5	3
V2	4	8	8	7
V3	2	9	6	5
V4	6	6	5	5
V5	9	6	5	8

Chaque véhicule V_i pourrait être repéré dans un espace de dimension 4 (l’espace des 4 variables) par un point. Mais cette représentation, outre le fait qu’elle est impossible à réaliser (notre espace ne comporte que 3 dimensions ...), n’est sans doute pas utile car on constate qu’il existe de fortes corrélations entre les variables (un véhicule jugé confortable est également habitable). D’un point de vue géométrique, cela se traduit par le fait que le nuage de point qui décrit l’ensemble des véhicules ne remplit pas complètement l’espace à 4 dimensions : il

présente des directions principales et est « aplati » selon certaines directions (comme le poisson). Comme pour le cas d'un solide, ces directions sont les directions d'inertie maximale, elles représentent les directions du nuage de point selon lesquelles le nuage est le plus élargi. L'ACP nous permet de trouver ces « direction principales », maximisant l'inertie, donc l'information, du nuage de points. D'un point de vue pratique, on réalise simplement une projection orthogonale des points produits sur les 2 premières directions principales, et on représente les produits selon ces directions (plan principal).

Lorsqu'on effectue une ACP du tableau de données table 1, on obtient les résultats figure 2 et 3. On constate tout d'abord que le pourcentage d'inertie (% d'information) retenu par les 2 premières composantes est de 95,12% et est très élevé (on a perdu seulement $100-95,12 = 4,88\%$ de variance des données avec cette représentation en dimension 2) : donc effectivement, le nuage de point était très « aplati » en dimension 4, et une projection sur les deux dimensions principales permet de représenter plus de 95% de l'information, ce qui est jugé suffisant. Le nuage de point est donc quasiment bidimensionnel.

La figure 2 représente le nuage des individus (projection des voitures sur les facteurs F1 et F2 (les composantes principales)). Le centre du repère (0, 0) représente l'individu « moyen » (la voiture moyenne, moyenne des colonnes de la table 1). On constate que le véhicule V5, très éloigné des autres, est atypique et différent du reste des véhicules. Les individus extrêmes sur l'axe F1 sont V1 et V5, opposés au véhicule V3. Sur l'axe 2, V5 s'oppose à V1. Le véhicule V4 est proche du véhicule moyen (le centre du repère).

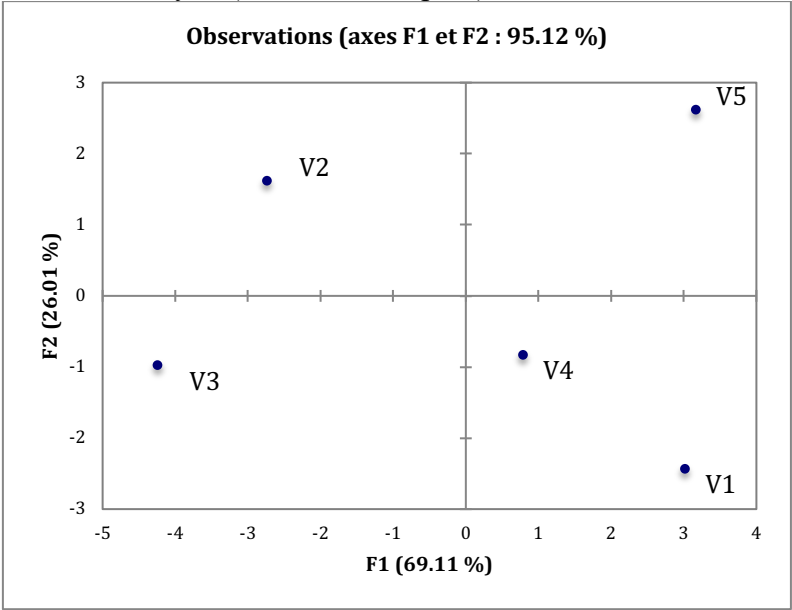


Figure 2 : plan principal (espace des produits)

Maintenant, pour compléter l'interprétation, il va être intéressant de comprendre comment sont créées les composantes principales et ce qu'elles représentent. Les composantes principales représentent des directions dans l'espace des variables initiales, elles sont donc formées comme des **combinaisons linéaires des variables initiales** (les lignes rouges du poisson figure 1 sont par exemple des combinaisons de l'axe x et l'axe y).

La composition des facteurs F1 et F2 en fonction des variables initiales est donnée dans la table 2.

Table 2 : composition des facteurs F1 et F2 (vecteurs propres).

	F1	F2
sportivité	0,839	0,214

<i>confort</i>	-0,468	0,161
<i>habitabilité</i>	-0,275	0,228
<i>sécurité</i>	-0,044	0,936

On a par exemple $F1 = 0,839sportivité - 0,468confort - 0,275habitabilité - 0,044sécurité$.

On constate qu'aucun facteur n'est « pur », c'est-à-dire formé que par une seule variable. Toutes les variables concourent à la création des facteurs, mais à des degrés divers. On peut néanmoins conclure que :

- l'axe F1 est formé principalement par la variable *sportivité* (coeff 0,839), opposé à *confort* (et un petit peu à *habitabilité*)
- l'axe F2 est créé presque exclusivement par la variable *sécurité*

Cette analyse nous permet de tracer une seconde figure qui va nous être très utile pour expliquer les différences entre les produits. De la même façon que précédemment on avait représenté les individus dans le plan des 2 facteurs F1 F2, on va maintenant représenter la projection des variables dans le plan des 2 premiers facteurs.

La figure 3 montre la représentation des variables dans le plan des 2 premières composantes principales. On voit ainsi que la première composante est créée principalement par les variables *sportivité*, en opposition à *confort* : l'angle entre le vecteur *sportivité* et la variable F1 est très petit (d'un point de vue pratique, on peut montrer que le coefficient de corrélation entre 2 variables n'est autre que le cosinus de l'angle entre les variables centrées – l'ACP est donc un excellent moyen de « visualiser » les corrélations entre variables). La deuxième composante est créée par la variable *sécurité* (le vecteur sécurité est quasiment confondu avec F2).

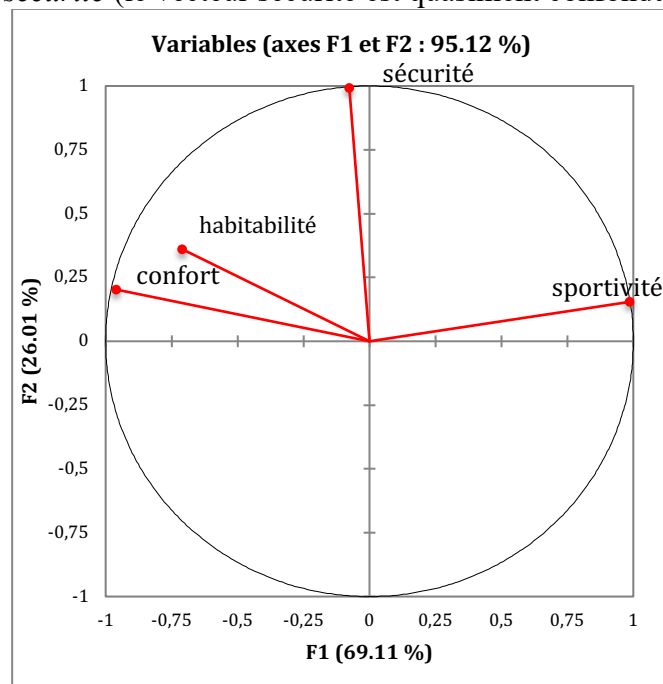


Figure 2 : plan principal (espace des variables)

D'un point de vue perceptif, les différences principales entre les voitures s'expliquent donc par 2 facteurs :

- Facteur F1 : opposition entre *sportivité* et *confort* et *habitabilité*
- Facteur F2 : variable *sécurité*.

On peut donc conclure ici que les voitures V1 et V5, très sportives, mais peu confortables (elles sont situées à l'est sur le plan) s'opposent à la voiture V3, très confortable mais pas sportive. La voiture V5 (au nord), très sécuritaire, s'oppose à V1 (au sud), qui l'est peu.

Pour terminer, il est important de noter que l'ACP, qui a pour objectif la simplification et l'interprétation des données, nécessite toujours de faire un compromis selon deux objectifs contradictoires

- Avoir un espace de faible dimension pour comprendre et visualiser les principales différences entre produits
- Conserver un % d'information suffisant pour ne pas être caricatural, ce qui masquerait les principales différences entre produits.

Fiche Outil 2 : La régression linéaire multiple

La régression linéaire est une méthode de modélisation permettant d’établir une relation linéaire entre une variable continue y dite "variable expliquée" ou dépendante et un ensemble d’autres variables continues x dites "variables explicatives" ou indépendantes. Plus spécifiquement, elle propose un modèle explicatif qui permet de prédire la variable dépendante en fonction des variables indépendantes.

Cette fiche explique comment estimer les coefficients d’un modèle linéaire à partir d’un ensemble d’observations (y_i, x_i) .

La méthode classique utilise la méthode des moindres carrés. Elle fournit également des indicateurs pour estimer la qualité de l’ajustement du modèle sur les données. Nous invitons le lecteur à consulter les ouvrages spécialisés pour plus d’information sur le cadre théorique du modèle linéaire. La présentation sera effectuée sur deux exemples simples commentés.

1 EXEMPLE DE REGRESSION LINEAIRE SIMPLE

On considère le cas de la consommation d’électricité pour le chauffage d’une habitation (en kWh par jour) en fonction de la température extérieure T° (en degré Celsius). Les observations sur 12 jours de l’année sont données table 1.

Table 1 : valeur de la consommation d’électricité en fonction de la température extérieure.

Y Conso (kWh)	T T (°C)
20	18
15	19
9	23
10	22
12	22
45	2
12	20
5	25
32	14
25	13
28	11
22	12

On constate tout d’abord que plus il fait froid, plus la consommation d’électricité est importante. Cette observation est confortée par la figure 1 suivante, qui représente la consommation en fonction de la température pour les 12 jours considérés.

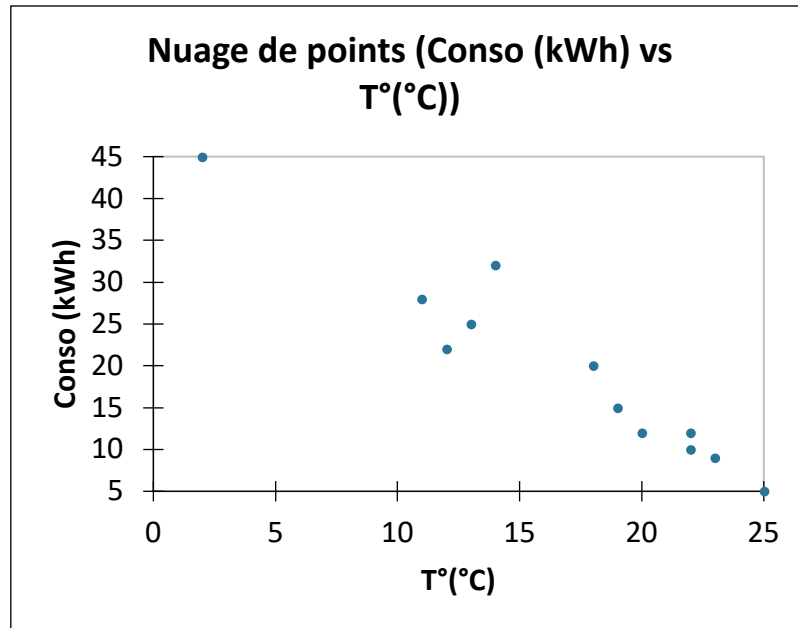


Figure 1 : représentation des observations

On s'intéresse à créer un modèle entre la consommation Y et la température T . On remarque intuitivement ici que plus la température augmente, plus la consommation diminue. On considère donc un modèle linéaire de la forme suivante⁷ :

$$\hat{y} = a + bT$$

Avec : $\hat{y}$: valeur de la consommation, prédite par le modèle

a : ordonnée à l'origine

b : coefficient de la température

Pour une observation i (une ligne de la table 1), la valeur de la consommation donnée par le modèle $\hat{y}_i$ diffère de celle observée y_i avec une erreur e_i

On a :
$$y_i = \hat{y}_i + e_i$$

Ceci est illustré par la figure 2.

⁷ En régression linéaire, il faut toujours **prendre garde à la causalité de la relation** entre la variable à expliquer y et la variable explicative. Dans notre cas, c'est parce qu'il fait froid que la consommation augmente : la température est la cause de l'augmentation de consommation : c'est bien un modèle qui explique y en fonction de T qu'il faut faire (et pas l'inverse ! ce serait stupide d'expliquer la température extérieure par l'observation de la consommation).

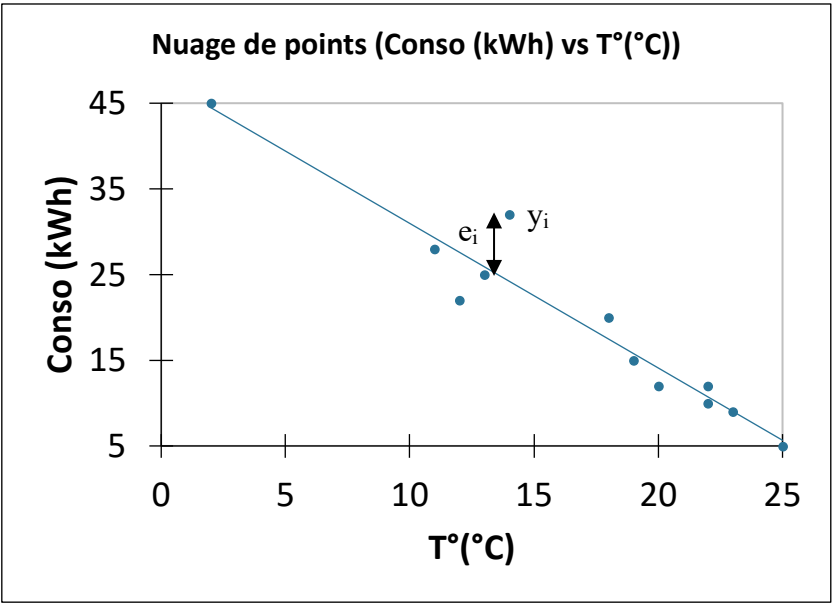


Figure 2 : représentation des observations, de la droite de régression, et de l'erreur e_i pour l'observation i

La méthode des moindres carrés pour l'estimation des paramètres (a,b) du modèle consiste à trouver la valeur des coefficients qui minimise l'erreur quadratique totale (somme sur toutes les observations) :

$$erreur = \sum_i e_i^2$$

De manière pratique, les solutions de ce problème des moindres carrés sont explicites et sont données par :

$$b = \frac{cov(x,y)}{s_x^2} \qquad a = \bar{y} - b\bar{x}$$

Le coefficient de détermination, qui indique l'ajustement du modèle sur les données (ou le pourcentage de variance capté par le modèle), est donné par :

$$R^2 = 1 - \frac{erreur}{\sum_i (y_i - \bar{y})^2}$$

Plus les erreurs de modélisation sont faibles, plus le R^2 est proche de 1.
Lorsque l'on cherche les coefficients de moindres carrés pour les données de la table 1 avec un logiciel de statistique, on obtient les résultats donnés table 2.

Table 2 : résultats des coefficients des moindres carrés pour les données de la table 1.

R² = 0,926						
Source	Valeur	Erreur standard	t	Pr > t	Borne inférieure (95%)	Borne supérieure (95%)
Constante (a)	47,84	2,707	17,675	<0,0001	41,816	53,879
(b) : T°(°C)	-1,68	0,151	-11,152	<0,0001	-2,025	-1,350

Interprétation :

- Le R^2 est très proche de 1 : le modèle s'ajuste très bien sur les données. On peut donc interpréter les résultats
- Le coefficient de la température T (-1,68) est négatif : diminuer la température cause une augmentation de la consommation d'électricité
- Le test statistique sur le coefficient b est significatif ($p\text{-value} < 0,01$: il est significativement différent de zéro). Le degré de confiance sur les résultats est donc grand.

L'équation du modèle est donnée par :

$$\text{Conso (kWh)} = 47,84 - 1,68 * T(^{\circ}\text{C})$$

Elle peut nous servir à prédire la consommation pour une température extérieure donnée.

2 EXEMPLE DE REGRESSION LINEAIRE MULTIPLE

On s'intéresse toujours à la consommation d'électricité pour le chauffage d'une habitation (en kWh par jour) en fonction de la température extérieure T° . Mais afin d'améliorer le modèle et les prédictions, on propose de tenir compte d'une seconde variable qui peut potentiellement jouer un rôle dans la consommation : la vitesse du vent à l'extérieur de la maison. Les données sont présentées en table 3.

Table 3 : valeur de la consommation d'électricité en fonction de la température extérieure et de la vitesse du vent

Conso (kWh)	$T^{\circ} (^{\circ}\text{C})$	Vitesse vent (km/h)
20	18	10
15	19	0
9	23	2
10	22	3
12	22	10
45	2	15
12	20	2
5	25	2
32	14	14
25	13	10
28	11	11
22	12	8

Le modèle linéaire considéré a maintenant la forme suivante :

$$\hat{y} = a + bT + cV$$

Il s'agit maintenant de l'équation d'un plan dans un espace de dimension 3 (conso, T , V)
Lorsque l'on cherche les coefficients de moindres carrés pour les données de la table 3 avec un logiciel de statistique, on obtient les résultats donnés table 4.

Table 4 : résultats des coefficients des moindres carrés pour les données de la table 3.

R² = 0,95						
Source	Valeur	Erreur standard	t	Pr > t	Borne inférieure (95%)	Borne supérieure (95%)
(a) Constante	37,245	4,460	8,351	<0,0001	27,156	47,334
(b) T°(°C)	-1,32	0,180	-7,335	<0,0001	-1,728	-0,914
(c) vitesse vent (km/h)	0,61	0,228	2,702	0,024	0,100	1,131

Interprétation :

- Le R² est très proche de 1 (0,95) : le modèle s'ajuste très bien sur les données. On peut donc interpréter les résultats
- Le coefficient de la température T (-1,32) est négatif : diminuer la température cause une augmentation de la consommation d'électricité, ce qui est logique. A noter que ce coefficient est différent de celui qu'on obtenait précédemment avec une seule variable explicative (-1,68)
- Le coefficient de la vitesse V (0,61) est positif : augmenter la vitesse cause une augmentation de la consommation d'électricité (ce qui est logique : cela augmente les échanges thermiques par convection)
- Le test statistique sur les coefficients b et c est significatif (p-value < 0,05). Le degré de confiance sur les résultats est donc grand.

L'équation du modèle est donnée par :

$$\text{Conso (kWh)} = 37,24 - 1,32 \cdot T^{\circ}(\text{°C}) + 0,61 \cdot \text{vitesse vent (km/h)}$$

Elle peut nous servir à prédire la consommation pour une température extérieure donnée.

3 BILAN SUR LE MODELE

L'examen du modèle linéaire nécessite de plus de s'assurer que :

- les résidus e_i restent petits et ne sont pas dépendants de la variable y ,
- le modèle n'est pas sur-ajusté, c'est-à-dire que sa capacité de généralisation est grande. Pour cela, il faut s'assurer que le nombre d'observations est suffisant étant donné le nombre de paramètres à estimer (il faut généralement avoir entre 1,5 à 3 fois plus d'observations que de paramètres à estimer).