

*IBM SPSS Direct Marketing 22*

**IBM**

**Important**

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations générales figurant à la section «Remarques», à la page 27.

**Informations sur le produit**

La présente édition s'applique à la version 22.0.0 d'IBM SPSS Statistics et à toutes les éditions et modifications ultérieures sauf mention contraire dans les nouvelles éditions.

---

## Table des matières

<b>Avis aux lecteurs canadiens . . . . .</b>	<b>v</b>	Création d'un champ de réponse catégorielle . . . . .	15
<b>Chapitre 1. Marketing direct . . . . .</b>	<b>1</b>	<b>Chapitre 5. Taux de réponse des codes postaux . . . . .</b>	<b>17</b>
<b>Chapitre 2. Analyse RFM . . . . .</b>	<b>3</b>	Paramètres . . . . .	18
Scores RFM de Données de transactions . . . . .	3	Création d'un champ de réponse catégorielle . . . . .	19
Scores RFM de Données de clients . . . . .	4	<b>Chapitre 6. Propension à acheter . . . . .</b>	<b>21</b>
Regroupement par casiers RFM . . . . .	4	Paramètres . . . . .	23
Enregistrement des scores RFM de Données de transactions. . . . .	6	Création d'un champ de réponse catégorielle . . . . .	24
Enregistrement des Scores RFM de Données de clients . . . . .	7	<b>Chapitre 7. Test de pack de contrôle . . . . .</b>	<b>25</b>
Sortie RFM . . . . .	8	<b>Remarques . . . . .</b>	<b>27</b>
<b>Chapitre 3. Analyse de cluster. . . . .</b>	<b>11</b>	Marques . . . . .	29
Paramètres . . . . .	12	<b>Index . . . . .</b>	<b>31</b>
<b>Chapitre 4. Profils des clients potentiels . . . . .</b>	<b>13</b>		
Paramètres . . . . .	14		



---

## Avis aux lecteurs canadiens

Le présent document a été traduit en France. Voici les principales différences et particularités dont vous devez tenir compte.

### Illustrations

Les illustrations sont fournies à titre d'exemple. Certaines peuvent contenir des données propres à la France.

### Terminologie

La terminologie des titres IBM peut différer d'un pays à l'autre. Reportez-vous au tableau ci-dessous, au besoin.

IBM France	IBM Canada
ingénieur commercial	représentant
agence commerciale	succursale
ingénieur technico-commercial	informaticien
inspecteur	technicien du matériel

### Claviers

Les lettres sont disposées différemment : le clavier français est de type AZERTY, et le clavier français-canadien de type QWERTY.

### OS/2 et Windows - Paramètres canadiens

Au Canada, on utilise :

- les pages de codes 850 (multilingue) et 863 (français-canadien),
- le code pays 002,
- le code clavier CF.

### Nomenclature

Les touches présentées dans le tableau d'équivalence suivant sont libellées différemment selon qu'il s'agit du clavier de la France, du clavier du Canada ou du clavier des États-Unis. Reportez-vous à ce tableau pour faire correspondre les touches françaises figurant dans le présent document aux touches de votre clavier.

<b>France</b>	<b>Canada</b>	<b>Etats-Unis</b>
 (Pos1)		Home
Fin	Fin	End
 (PgAr)		PgUp
 (PgAv)		PgDn
Inser	Inser	Ins
Suppr	Suppr	Del
Echap	Echap	Esc
Attn	Intrp	Break
Impr écran	ImpEc	PrtSc
Verr num	Num	Num Lock
Arrêt défil	Défil	Scroll Lock
 (Verr maj)	FixMaj	Caps Lock
AltGr	AltCar	Alt (à droite)

## Brevets

Il est possible qu'IBM détienne des brevets ou qu'elle ait déposé des demandes de brevets portant sur certains sujets abordés dans ce document. Le fait qu'IBM vous fournisse le présent document ne signifie pas qu'elle vous accorde un permis d'utilisation de ces brevets. Vous pouvez envoyer, par écrit, vos demandes de renseignements relatives aux permis d'utilisation au directeur général des relations commerciales d'IBM, 3600 Steeles Avenue East, Markham, Ontario, L3R 9Z7.

## Assistance téléphonique

Si vous avez besoin d'assistance ou si vous voulez commander du matériel, des logiciels et des publications IBM, contactez IBM direct au 1 800 465-1234.

---

## Chapitre 1. Marketing direct

L'option de Marketing direct fournit un ensemble d'outils conçus pour améliorer les résultats des campagnes de marketing direct, en identifiant les caractéristiques démographiques, d'achat et autres qui définissent différents groupes de consommateurs et qui ciblent des groupes spécifiques afin d'optimiser les taux de réponse positive.

**Analyse RFM** : Cette technique permet d'identifier les clients existants qui ont plus de chances de répondre à une nouvelle offre.

**Analyse de cluster** : C'est un outil d'exploration conçu pour révéler des groupements naturels (ou clusters) au sein de vos données. Par exemple, elle permet d'identifier différents groupes de clients en fonction de diverses caractéristiques démographiques et d'achat.

**Profils des clients potentiels** : Cette technique utilise les résultats d'une campagne précédente ou d'une campagne test pour créer des profils descriptifs. Vous pouvez utiliser ces profils pour cibler des groupes de contacts spécifiques lors des futures campagnes. Pour plus d'informations, voir Chapitre 4, «Profils des clients potentiels», à la page 13.

**Taux de réponse des codes postaux** : Cette technique utilise les résultats de la campagne précédente pour calculer les taux de réponse par code postal. Ces taux peuvent être utilisés pour cibler des codes postaux spécifiques lors des futures campagnes. Pour plus d'informations, voir Chapitre 5, «Taux de réponse des codes postaux», à la page 17.

**Propension à acheter** : Cette technique utilise les résultats d'un envoi d'essai ou d'une campagne précédente pour générer des scores de propension. Les scores indiquent quels contacts sont les plus susceptibles de répondre. Pour plus d'informations, voir Chapitre 6, «Propension à acheter», à la page 21.

**Test de pack de contrôle** : Cette technique compare des campagnes de marketing pour vérifier s'il existe une différence significative de l'efficacité entre différents packs et offres. Pour plus d'informations, voir Chapitre 7, «Test de pack de contrôle», à la page 25.





---

## Chapitre 2. Analyse RFM

L'analyse RFM est une technique permettant d'identifier des clients existants qui ont plus de chances de répondre à une nouvelle offre. Cette technique est couramment utilisée dans le marketing direct.

L'analyse RFM repose sur la théorie simple suivante :

- Le facteur le plus important dans l'identification des clients qui ont le plus de chances de répondre à une nouvelle offre est la **récence**. Les clients qui ont acheté plus récemment ont plus de chances d'acheter à nouveau que ceux qui ont acheté dans un passé plus lointain.
- Le deuxième facteur le plus important est la **fréquence**. Les clients qui ont acheté davantage dans le passé ont plus de chances de répondre que ceux qui ont moins acheté.
- Le troisième facteur le plus important est le montant total dépensé, appelé **monétaire**. Les clients qui ont dépensé davantage (au total pour tous les achats) dans le passé ont plus de chances de répondre que ceux qui ont dépensé moins.

Fonctionnement de l'analyse RFM

- Les clients obtiennent un score de récence selon la date de l'achat le plus récent ou l'intervalle de temps depuis l'achat le plus récent. Ce score repose sur un simple classement des valeurs de récence en un petit nombre de catégories. Par exemple, si vous utilisez cinq catégories, les clients avec les dates d'achat les plus récentes reçoivent un classement de récence de 5, tandis que ceux avec des dates plus éloignées dans le passé reçoivent un classement de récence de 1.
- De manière semblable, les clients reçoivent ensuite un classement de fréquence, avec des valeurs supérieures représentant une fréquence d'achat supérieure. Par exemple, dans un schéma de classement à cinq catégories, les clients qui achètent le plus souvent reçoivent un classement de fréquence de 5.
- Finalement, les clients sont classés par valeur monétaire, avec les valeurs monétaires les plus élevées ayant le classement le plus haut. Toujours avec le même exemple, les clients qui ont dépensé le plus reçoivent un classement monétaire de 5.

Le résultat correspond à quatre scores pour chaque client : score de récence, de fréquence, monétaire et RFM combiné, qui représente les trois scores individuels concaténés en une seule valeur. Les "meilleurs" clients (ceux qui ont le plus de chances de répondre à une offre) sont ceux ayant les scores RFM combinés les plus élevés. Par exemple, dans un classement à cinq catégories, il existe un total de 125 scores RFM combinés possibles, et celui le plus élevé est 555.

Analyse des données

- Si les lignes de données représentent des transactions (chaque ligne représente une seule transaction, et il peut y avoir plusieurs transactions pour chaque client), utilisez RFM de Transactions. Pour plus d'informations, voir «Scores RFM de Données de transactions».
- Si les lignes de données représentent des clients avec des informations récapitulatives pour toutes les transactions (avec des colonnes qui contiennent des valeurs pour le montant total dépensé, le nombre total de transactions, et la date de transaction la plus récente), utilisez RFM de Données clients. Pour plus d'informations, voir «Scores RFM de Données de clients», à la page 4.

---

### Scores RFM de Données de transactions

Analyse des données

Le jeu de données contient les variables qui comportent les informations suivantes :

- Une variable ou une combinaison de variables qui identifient chaque observation (client).
- Une variable avec la date de chaque transaction.
- Une variable avec la valeur monétaire de chaque transaction.

Création de scores RFM de Données de transactions

1. A partir des menus, sélectionnez :

**Marketing direct > Choisir une technique**

2. Sélectionnez **Aider à identifier mes meilleurs contacts (Analyse RFM)** puis cliquez sur **Poursuivre**.
3. Sélectionnez **Données de transactions**, puis cliquez sur **Poursuivre**.
4. Sélectionnez la variable qui contient les dates de transactions.
5. Sélectionnez la variable qui contient le montant monétaire de chaque transaction.
6. Sélectionnez la méthode de récapitulatif des montants de transaction pour chaque client : Total (somme de toutes les transactions), moyenne, médiane ou maximum (montant de transaction le plus élevé).
7. Sélectionnez la variable ou combinaison de variables qui identifient de façon unique chaque client. Par exemple, les observations peuvent être identifiées par un code d'identification unique ou une combinaison du nom et du prénom.

---

## Scores RFM de Données de clients

Analyse des données

Le jeu de données contient les variables qui comportent les informations suivantes :

- Date d'achat la plus récente ou intervalle de temps depuis la date d'achat la plus récente. Vous l'utilisez pour calculer les scores de récence.
- Nombre total d'achats. Vous l'utilisez pour calculer les scores de fréquence.
- Valeur monétaire récapitulative pour tous les achats. Vous l'utilisez pour calculer les scores monétaires. En général, il s'agit de la somme (du total) de tous les achats, mais peut être la moyenne, le maximum (montant le plus grand) ou autre mesure récapitulative.

Si vous souhaitez écrire des scores RFM dans un nouveau jeu de données, le jeu de données actif doit aussi contenir une variable ou une combinaison de variables qui identifient chaque observation (client).

Création de scores RFM de Données de clients

1. A partir des menus, sélectionnez :

**Marketing direct > Choisir une technique**

2. Sélectionnez **Aider à identifier mes meilleurs contacts (Analyse RFM)** puis cliquez sur **Poursuivre**.
3. Sélectionnez **Données clients**, puis cliquez sur **Poursuivre**.
4. Sélectionnez la variable qui contient la date de transaction la plus récente ou un nombre qui représente un intervalle de temps depuis la transaction la plus récente.
5. Sélectionnez la variable qui contient le nombre total de transactions pour chaque client.
6. Sélectionnez la variable qui contient le montant monétaire récapitulatif de chaque client.
7. Si vous souhaitez écrire des scores RFM dans un nouveau jeu de données, sélectionnez la variable ou la combinaison de variables qui identifient de façon unique chaque client. Par exemple, les observations peuvent être identifiées par un code d'identification unique ou une combinaison du nom et du prénom.

---

## Regroupement par casiers RFM

Le processus de regroupement d'un grand nombre de valeurs numériques en un petit nombre de catégories est parfois appelé **regroupement par casiers**. Dans l'analyse RFM, les casiers sont les catégories classées. L'onglet Regroupement par casiers permet de modifier la méthode utilisée pour affecter les valeurs de récence, de fréquence et monétaires à ces casiers.

## Méthode de regroupement

**Imbriquée** : Dans le regroupement par casier imbriqué, un rang simple est affecté à des valeurs de récence. Dans chaque rang de récence, les clients reçoivent ensuite un rang de fréquence, et dans chacun, les clients reçoivent un rang monétaire. Ceci tend à fournir une distribution plus équitable des scores RFM combinés, mais a pour inconvénient de compliquer l'interprétation des scores de rang de fréquence et monétaire. Par exemple, un rang de fréquence de 5 pour un client avec un rang de récence de 5 risque d'avoir une signification autre qu'un rang de fréquence de 5 pour un client avec un rang de récence de 4, puisque le rang de fréquence dépend de celui de récence.

**Indépendante** : Les rangs simples sont affectés aux valeurs de récence, de fréquence et monétaires. Ces trois rangs sont affectés de manière indépendante. L'interprétation de chacun des trois composants RFM est donc claire. Un score de fréquence de 5 pour un client a la même signification qu'un score de fréquence de 5 pour un autre client, quels que soient les scores de récence. Pour des échantillons plus petits, ceci a l'inconvénient d'entraîner une distribution moins équitable des scores RFM combinés.

## Nombre de casiers

Le nombre de catégories (casiers) à utiliser pour chaque composant pour créer des scores RFM. Le nombre total de scores RFM combinés possibles est le produit des trois scores. Par exemple, 5 casiers de récence, 4 casiers de fréquence et 3 casiers monétaires créent un total de 60 scores RFM combinés possibles, allant de 111 à 543.

- La valeur par défaut est 5 pour chaque composant ; ce qui crée 125 scores RFM combinés possibles, allant de 111 à 555.
- Le nombre maximum de casiers autorisés pour chaque composant de score est neuf.

## Ex aequo

Un "ex aequo" est simplement deux ou plusieurs valeurs de récence, de fréquence ou monétaires égales. Dans l'idéal, vous souhaitez disposer d'environ le même nombre de clients dans chaque casier, mais un grand nombre de valeurs ex aequo peut affecter la distribution des casiers. Il existe deux moyens de gérer les ex aequo :

- **Affecter les ex aequo au même casier** : Cette méthode permet de toujours assigner des valeurs ex aequo au même casier, que ceci affecte ou non la distribution des casiers. Ceci offre une méthode de regroupement par casiers cohérente : Si la même valeur de récence est attribuée à deux clients, ils obtiendront toujours le même score de récence. Dans un cas extrême, toutefois, il se peut que 500 clients sur un total de 1000 fassent leurs achats les plus récents à la même date. Dans un classement à 5 casiers, 50 % des clients reçoivent ainsi un score de récence de 5, au lieu des 20 % attendus.

Notez que la "cohérence" de la méthode imbriquée de regroupement par casiers est un peu plus complexe pour les scores monétaires et de fréquence, puisque ces derniers sont affectés dans les casiers de score de récence, et les scores monétaires sont affectés dans les casiers de score de fréquence. Par conséquent, deux clients ayant la même valeur de fréquence risquent de ne pas avoir le score de fréquence identique s'ils ne possèdent pas aussi le même score de récence, quel que soit le traitement des valeurs ex aequo.

- **Affectation aléatoire des ex aequo** : Ceci garantit une distribution équitable des casiers en affectant un facteur de variance aléatoire très faible avant tout classement. Ainsi, il n'existe aucune valeur ex aequo afin d'affecter les valeurs aux casiers classés. Ce processus n'a aucune incidence sur les valeurs d'origine. Il n'est utilisé que pour désambiguïser les ex aequo. Tandis que ce processus génère une distribution équitable des casiers (environ le même nombre de clients dans chaque casier), il peut entraîner des scores complètement différents pour les clients avec des valeurs de récence, de fréquence, et/ou monétaires semblables voire identiques, surtout si le nombre total de clients est relativement faible et/ou le nombre d'ex aequo élevé.

Tableau 1. Affectation des ex aequo au même casier et Affectation aléatoire des ex aequo.

ID	Achat le plus récent (récence)	Affecter des ex aequo au même casier	Affectation aléatoire des ex aequo
1	29/10/2006	5	5
2	28/10/2006	4	4
3	28/10/2006	4	4
4	28/10/2006	4	5
5	28/10/2006	4	3
6	21/9/2006	3	3
7	21/9/2006	3	2
8	13/8/2006	2	2
9	13/8/2006	2	1
10	20/6/2006	1	1

- Dans cet exemple, l'affectation d'ex aequo au même casier entraîne une distribution inéquitable des casiers : : 5 (10 %), 4 (40 %), 3 (20 %), 2 (20 %), 1 (10 %).
- L'affectation aléatoire d'ex aequo entraîne 20 % dans chaque casier, mais pour obtenir ce résultat, les quatre observations avec une valeur de date de 28/10/2006 sont affectées à 3 casiers différents, et les 2 observations avec une valeur de date de 13/8/2006 sont aussi affectées à des casiers différents.

Notez que la manière dont les ex aequo sont affectés à des casiers différents est entièrement aléatoire (dans les contraintes du résultat final étant un nombre égal d'observations dans chaque casier). Si vous avez calculé un deuxième ensemble de scores à l'aide de la même méthode, le classement d'une observation particulière avec une valeur ex aequo peut changer. Par exemple, les classements de récence de 5 et 3 pour les observations 4 et 5 respectivement, peuvent être changés la seconde fois.

---

## Enregistrement des scores RFM de Données de transactions

RFM de Données de transactions permet de toujours créer un jeu de données agrégé avec une seule ligne pour chaque client. L'onglet Enregistrer permet d'indiquer les scores et autres variables à enregistrer et l'emplacement de leur enregistrement.

### Variables

Les variables d'identification, qui identifient de façon unique chaque client, sont enregistrées automatiquement dans le nouveau jeu de données. Les autres variables suivantes peuvent être enregistrées dans le nouveau jeu de données :

- **Date de la transaction la plus récente pour chaque client :**
- **Nombre de transactions :** Le nombre total de lignes de transaction pour chaque client.
- **Montant :** Le montant récapitulatif pour chaque client selon la méthode récapitulative sélectionnée dans l'onglet Variables.
- **Score de récence :** Le score assigné à chaque client selon la date de transaction la plus récente. Des scores plus élevés indiquent des dates de transaction plus récentes.
- **Score de fréquence :** Le score assigné à chaque client selon le nombre total de transactions. Des scores plus élevés indiquent davantage de transactions.
- **Score monétaire :** Le score affecté à chaque client selon la mesure récapitulative monétaire sélectionnée. Des scores plus élevés indiquent une valeur plus grande pour la mesure récapitulative monétaire.
- **Score RFM :** Les trois scores individuels combinés en une seule valeur :  $(récence \times 100) + (fréquence \times 10) + monétaire$ .

Par défaut, toutes les variables disponibles sont comprises dans le nouveau jeu de données ; ainsi, désélectionnez celles à exclure. Vous pouvez en option préciser vos propres noms de variables. Les noms de variables doivent être conformes aux règles standard de dénomination de variables.

#### Emplacement

RFM de Données de transactions permet de toujours créer un jeu de données agrégé avec une seule ligne pour chaque client. Vous pouvez créer un nouveau jeu de données dans la session en cours ou enregistrer les données de score RFM dans un fichier de données externe. Les noms des jeux de données doivent être conformes aux règles standard de dénomination de variables. (Cette restriction ne s'applique pas aux noms de fichiers de données externes).

---

## Enregistrement des Scores RFM de Données de clients

Pour les données de clients, vous pouvez ajouter les variables de score RFM à le jeu de données actif ou créer un jeu de données qui contient les variables de scores sélectionnées. L'onglet Enregistrer permet d'indiquer les scores de variables à enregistrer et l'emplacement de leur enregistrement.

#### Noms des variables enregistrées

- **Générer automatiquement des noms uniques** : Lors de l'ajout de variables de scores à le jeu de données actif, ceci garantit que les nouveaux noms de variables sont uniques. Ceci est particulièrement utile si vous souhaitez ajouter plusieurs ensembles de scores RFM différents (selon des critères différents) au jeu de données actif.
- **Noms personnalisés** : Ceci permet d'affecter vos propres noms de variables aux variables de scores. Les noms de variables doivent être conformes aux règles standard de dénomination de variables.

#### Variables

Sélectionnez (cochez) les variables de scores à enregistrer :

- **Score de récence** : Le score affecté à chaque client selon la valeur de la variable Date de transaction ou Intervalle sélectionnée dans l'onglet Variables. Des scores élevés sont affectés à des dates plus récentes ou à des valeurs d'intervalles inférieures.
- **Score de fréquence** : Le score affecté à chaque client selon la valeur de la variable Nombre de transactions sélectionnée dans l'onglet Variables. Les scores les plus élevés sont affectés à des valeurs plus grandes.
- **Score monétaire** : Le score affecté à chaque client selon la valeur de la variable Montant sélectionnée dans l'onglet Variables. Les scores les plus élevés sont affectés à des valeurs plus grandes.
- **Score RFM** : Les trois scores individuels combinés en une seule valeur :  
 $(\text{récence} * 100) + (\text{fréquence} * 10) + \text{monétaire}$ .

#### Emplacement

Pour les données de clients, trois alternatives sont possibles pour l'emplacement de l'enregistrement des nouveaux scores RFM :

- **Jeu de données actif** : Les variables de scores RFM sélectionnées sont ajoutées à le jeu de données actif.
- **Nouveau jeu de données** : Les variables de scores RFM sélectionnées et les variables d'identification qui identifient de façon unique chaque client (observation) sont écrites dans un nouveau jeu de données de la session en cours. Les noms des jeux de données doivent être conformes aux règles standard de dénomination de variables. Cette option n'est disponible que si vous sélectionnez une ou plusieurs variables Identificateur client de l'onglet Variables.

- **Fichier** : Les scores RFM sélectionnés et les variables d'identification qui identifient de façon unique chaque client (observation) sont enregistrés dans un fichier de données externes. Cette option n'est disponible que si vous sélectionnez une ou plusieurs variables Identificateur client de l'onglet Variables.

---

## Sortie RFM

### Données groupées

Les graphiques et les tableaux de données regroupées par casiers sont basés sur le calcul des scores de récence, de fréquence et monétaires.

**Carte thermique de la valeur monétaire moyenne par récence et par fréquence** : La carte thermique de la distribution monétaire moyenne indique la valeur monétaire moyenne pour les catégories définies par des scores de récence et de fréquence. Les zones les plus foncées indiquent une valeur monétaire moyenne plus élevée.

**Graphique du nombre de casiers** : Le graphique du nombre de casiers affiche la distribution des casiers pour la méthode de regroupement sélectionnée. Chaque barre représente le nombre d'observations qui sera assigné à chaque score RFM combiné.

- Bien qu'une distribution régulière, avec des barres ayant tous (ou presque tous) approximativement la même taille, soit généralement souhaitée, il faut néanmoins s'attendre à une certaine variance lors de l'utilisation de la méthode de regroupement par défaut qui assigne les valeurs ex aequo au même casier.
- Des fluctuations extrêmes dans la distribution des casiers et/ou de nombreux casiers vides peuvent indiquer que vous devriez essayer une autre méthode de regroupement (casiers moins nombreux et/ou affectation aléatoire d'ex aequo) ou essayer de remettre en question l'adéquation de l'analyse RFM.

**Tableau du nombre de casier** : Informations semblables à celles du graphique du nombre de casiers, sauf qu'elles sont exprimées sous forme d'un tableau, avec des effectifs de casier dans chaque cellule.

### Données non regroupées

Les graphiques et les tableaux de données non regroupées par casiers sont basés sur les variables d'origine utilisées pour créer les scores de récence, de fréquence et monétaires.

**Histogrammes** : Les histogrammes indiquent la distribution relative des valeurs pour les trois variables utilisées pour le calcul des scores de récence, de fréquence et monétaires. Ces histogrammes présentent plus souvent des distributions relativement asymétriques que des distributions normales ou symétriques.

L'axe horizontal de chaque histogramme est toujours ordonné des valeurs les moins élevées sur la gauche aux valeurs les plus élevées sur la droite. Cependant, avec la récence, l'interprétation du graphique dépend du type de mesure de la récence : intervalle de temps ou de date. Pour les dates, les barres sur la gauche représentent des valeurs plus éloignées dans le temps (une date moins récente a une valeur inférieure à celle d'une date plus récente). Pour les intervalles de temps, les barres sur la gauche représentent des valeurs plus récentes (plus l'intervalle de temps est réduit, plus la transaction est récente).

**Nuages de points de paires de variables** : Ces nuages de points présentent la relation entre les trois variables utilisées pour calculer les scores de récence, de fréquence et monétaires.

Il est habituel d'observer des regroupements linéaires de points sur l'échelle de fréquence, car la fréquence représente souvent une plage de valeurs discrètes relativement réduit. Par exemple, si le nombre total des transactions ne dépasse pas 15, alors il n'existe que 15 valeurs de fréquence possibles

(sauf si les transactions fractionnelles sont comptabilisées), alors qu'il existe des centaines de valeurs de récence et des milliers de valeurs monétaires possibles.

L'interprétation de l'axe de la récence dépend du type de mesure de la récence : intervalle de temps ou de date. Pour les dates, les points les plus proches de l'origine représentent des dates plus éloignées dans le temps. Pour les intervalles de temps, des points plus proches de l'origine représentent des valeurs plus récentes.





## Chapitre 3. Analyse de cluster

L'analyse de cluster est un outil d'exploration conçu pour révéler des groupements naturels (ou clusters) au sein de données. Par exemple, elle permet d'identifier différents groupes de clients en fonction de diverses caractéristiques démographiques et d'achat.

**Exemple :** Les entreprises du domaine des produits de consommation et du commerce de détail utilisent régulièrement des techniques de classification des données qui décrivent les habitudes d'achat, le sexe, l'âge, le niveau de revenu, etc. de leurs clients. Ces sociétés adaptent leurs stratégies de marketing et de développement produit à chaque groupe de consommation afin d'augmenter les ventes et de développer la fidélité à la marque.

Remarques sur les données de l'analyse de cluster

**Données :** Cette procédure fonctionne avec des champs continus et catégoriels. Chaque enregistrement (ligne) représente un client à classer et les champs (variables) représentent les attributs sur lesquels est basée la classification.












**Ordre des enregistrements :** Notez que les résultats peuvent dépendre de l'ordre des enregistrements. Pour réduire les effets de tri, classez les enregistrements de manière aléatoire. Vous pouvez exécuter l'analyse plusieurs fois, avec les enregistrements triés de manière aléatoire, afin de vérifier la stabilité d'une solution donnée.

**Niveau de mesure :** L'affectation d'un niveau de mesure correct est important car il affecte le calcul des résultats.

- *Nominal.* Une variable peut être traitée comme étant nominale si ses valeurs représentent des catégories sans classement intrinsèque (par exemple, le service de la société dans lequel travaille un employé). La région, le code postal ou l'appartenance religieuse sont des exemples de variables nominales.
- *Ordinal.* Une variable peut être traitée comme étant ordinale si ses valeurs représentent des catégories associées à un classement intrinsèque (par exemple, des niveaux de satisfaction allant de Très mécontent à Très satisfait). Exemples de variable ordinale : des scores d'attitude représentant le degré de satisfaction ou de confiance, et des scores de classement des préférences.
- *Continue.* Une variable peut être traitée comme une variable d'échelle (continue) si ses valeurs représentent des catégories ordonnées avec une mesure significative, de sorte que les comparaisons de distance entre les valeurs soient adéquates. L'âge en années et le revenu en milliers de dollars sont des exemples de variable d'échelle.

Une icône en regard de chaque champ indique le niveau de mesure actuel.

Tableau 2. Icônes de niveau de mesure

	Numérique	Chaîne	Date	Heure
Echelle (continue).		n/a		
Ordinal				
Nominal				

Vous pouvez modifier le niveau de mesure dans la vue de variable de l'éditeur de données ou utilisez la boîte de dialogue Définir les propriétés de variable pour suggérer un niveau de mesure approprié pour chaque champ .

Champs avec un niveau de mesure inconnu

L'alerte du niveau de mesure apparaît lorsque le niveau de mesure d'une ou plusieurs variables (champs) du jeu de données est inconnu. Le niveau de mesure ayant une incidence sur le calcul des résultats de cette procédure, toutes les variables doivent avoir un niveau de mesure défini.

**Analyser les données** : Lit les données dans le jeu de données actifs et attribue le niveau de mesure par défaut à tous les champs ayant un niveau de mesure inconnu. Si le jeu de données est important, cette action peut prendre un certain temps.

**Affecter manuellement** : Ouvre une boîte de dialogue qui répertorie tous les champs ayant un niveau de mesure inconnu. Vous pouvez utiliser cette boîte de dialogue pour attribuer un niveau de mesure à ces champs. Vous pouvez également attribuer un niveau de mesure dans la vue de variable de l'éditeur de données.

Le niveau de mesure étant important pour cette procédure, vous ne pouvez pas accéder à la boîte de dialogue d'exécution de cette procédure avant que tous les champs n'aient des niveaux de mesure définis.

Pour obtenir une analyse de cluster

A partir des menus, sélectionnez :

#### **Marketing direct > Choisir une technique**

1. Sélectionnez **Segmenter mes contacts en clusters**.
2. Sélectionnez les champs catégoriels (nominaux, ordinaux) et continus (d'échelle) que vous souhaitez utiliser pour créer des segments.
3. Cliquez sur **Exécuter** pour exécuter la procédure.

---

## **Paramètres**

L'onglet Paramètres vous permet d'afficher ou de masquer les graphiques et tableaux qui décrivent les segments, d'enregistrer un nouveau champ qui identifie le segment (cluster) pour chaque enregistrement du jeu de données et de spécifier le nombre de segments à inclure dans la solution de cluster.

**Afficher les graphiques et les tableaux** : Permet d'afficher les tableaux et les graphiques qui décrivent les segments.

**Appartenance aux segments** : Permet d'enregistrer un nouveau champ (variable) qui identifie le segment auquel appartient l'enregistrement.

- Les noms de champ doivent être conformes aux règles de dénomination IBM® SPSS Statistics.
- Le nom de champ de l'appartenance au segment ne peut pas dupliquer un nom de champ qui existe déjà dans le jeu de données. Si vous exécutez cette procédure plusieurs fois sur le même jeu de données, vous devrez spécifier un nom différent à chaque fois.
- **Nombre de segments** : Contrôle le nombre des segments.
- **Déterminer automatiquement** : Cette procédure déterminera automatiquement le meilleur nombre de segments, jusqu'au maximum autorisé.

**Indiquer une valeur fixe** : La procédure va produire le nombre de segments spécifié.

---

## Chapitre 4. Profils des clients potentiels

Cette technique utilise les résultats d'une campagne précédente ou d'une campagne test pour créer des profils descriptifs. Vous pouvez utiliser ces profils pour cibler des groupes de contacts spécifiques lors des futures campagnes. Le champ Réponse indique qui a répondu à la campagne précédente ou à la campagne test. La liste Profils contient les caractéristiques à utiliser pour créer le profil.

**Exemple :** En fonction des résultats d'un envoi d'essai, la division de marketing direct d'une entreprise souhaite générer des profils des types de clients les plus susceptibles de répondre à une offre, sur la base d'informations démographiques.

Sortie

La sortie contient un tableau qui fournit une description de chaque groupe de profils et affiche les taux de réponse (pourcentage de réponses positives) et les taux de réponse cumulés, ainsi qu'un graphique des taux de réponses cumulés. Si vous incluez un taux de réponse minimum cible, le tableau sera codé par couleurs pour indiquer les profils qui satisfont le taux de réponse cumulé minimum, et le graphique contiendra une ligne de référence au niveau de la valeur du taux de réponse minimum spécifié.

Remarques sur les données de profil des clients potentiels

**Champ de réponse :** Le champ de réponse doit être nominal ou ordinal. Il peut s'agir d'une variable de type chaîne ou numérique. Si ce champ contient une valeur indiquant un nombre d'achats ou leur montant, vous devrez créer un nouveau champ dans lequel une valeur unique représente toutes les réponses positives. Pour plus d'informations, voir «Création d'un champ de réponse catégorielle», à la page 15.

**Valeur de réponse positive :** La valeur de réponse positive identifie les clients qui ont répondu positivement (par exemple, qui ont effectué un achat). Toutes les autres valeurs de réponse non manquantes sont supposées indiquer une réponse négative. S'il existe des libellés de valeurs définis pour le champ de réponse, ces libellés s'affichent dans la liste déroulante.












**Créer des profils avec :** Ces champs peuvent être de type nominal, ordinal ou continu (échelle). Il peut s'agir de variables de type chaîne ou numérique.

**Niveau de mesure :** L'affectation d'un niveau de mesure correct est important car il affecte le calcul des résultats.

- *Nominal.* Une variable peut être traitée comme étant nominale si ses valeurs représentent des catégories sans classement intrinsèque (par exemple, le service de la société dans lequel travaille un employé). La région, le code postal ou l'appartenance religieuse sont des exemples de variables nominales.
- *Ordinal.* Une variable peut être traitée comme étant ordinale si ses valeurs représentent des catégories associées à un classement intrinsèque (par exemple, des niveaux de satisfaction allant de Très mécontent à Très satisfait). Exemples de variable ordinale : des scores d'attitude représentant le degré de satisfaction ou de confiance, et des scores de classement des préférences.
- *Continue.* Une variable peut être traitée comme une variable d'échelle (continue) si ses valeurs représentent des catégories ordonnées avec une mesure significative, de sorte que les comparaisons de distance entre les valeurs soient adéquates. L'âge en années et le revenu en milliers de dollars sont des exemples de variable d'échelle.

Une icône en regard de chaque champ indique le niveau de mesure actuel.

Tableau 3. Icônes de niveau de mesure

	Numérique	Chaîne	Date	Heure
Echelle (continue).		n/a		
Ordinal				
Nominal				

Vous pouvez modifier le niveau de mesure dans la vue de variable de l'éditeur de données ou utilisez la boîte de dialogue Définir les propriétés de variable pour suggérer un niveau de mesure approprié pour chaque champ .

Champs avec un niveau de mesure inconnu

L'alerte du niveau de mesure apparaît lorsque le niveau de mesure d'une ou plusieurs variables (champs) du jeu de données est inconnu. Le niveau de mesure ayant une incidence sur le calcul des résultats de cette procédure, toutes les variables doivent avoir un niveau de mesure défini.

**Analyser les données** : Lit les données dans le jeu de données actifs et attribue le niveau de mesure par défaut à tous les champs ayant un niveau de mesure inconnu. Si le jeu de données est important, cette action peut prendre un certain temps.

**Affecter manuellement** : Ouvre une boîte de dialogue qui répertorie tous les champs ayant un niveau de mesure inconnu. Vous pouvez utiliser cette boîte de dialogue pour attribuer un niveau de mesure à ces champs. Vous pouvez également attribuer un niveau de mesure dans la vue de variable de l'éditeur de données.

Le niveau de mesure étant important pour cette procédure, vous ne pouvez pas accéder à la boîte de dialogue d'exécution de cette procédure avant que tous les champs n'aient des niveaux de mesure définis.

Pour obtenir les profils des clients potentiels

A partir des menus, sélectionnez :

#### Marketing direct > Choisir une technique

1. Sélectionnez **Générer les profils de mes contacts qui ont répondu à une offre**.
2. Sélectionnez le champ qui identifie les contacts ayant répondu à l'offre. Ce champ doit être nominal ou ordinal.
3. Entrez la valeur qui indique une réponse positive. Si des valeurs disposent de libellés de valeur définis, vous pouvez sélectionner le libellé de valeur dans la liste déroulante, et la valeur correspondante sera affichée.
4. Sélectionnez les champs à utiliser pour créer les profils.
5. Cliquez sur **Exécuter** pour exécuter la procédure.

---

## Paramètres

L'onglet Paramètres vous permet de contrôler la taille minimale des groupes de profils et inclut un seuil de taux de réponse minimum dans les sorties.

**Taille minimale du groupe de profils :** Chaque profil représente les caractéristiques partagées d'un groupe de contacts dans le jeu de données (par exemple, femmes de moins de 40 ans vivant dans l'Ouest). Par défaut, la plus petite taille de groupe est de 100. Les petites tailles de groupe permettent de révéler davantage de groupes, mais les tailles de groupe plus grandes peuvent fournir des résultats plus fiables. Il doit s'agir d'un nombre entier positif.

**Inclure des informations de seuil minimal de taux de réponse dans les résultats :** Les résultats contiennent un tableau qui affiche les taux de réponse (pourcentage de réponses positives) et les taux de réponse cumulés, ainsi qu'un graphique des taux de réponses cumulés. Si vous saisissez un taux de réponse minimum cible, le tableau sera codé par couleurs pour indiquer les profils qui satisfont le taux de réponse cumulé minimum, et le graphique contiendra une ligne de référence au niveau de la valeur du taux de réponse minimum spécifié. La valeur doit être supérieure à 0 et inférieure à 100.

---

## Création d'un champ de réponse catégorielle

Le champ de réponse doit être catégoriel, avec une valeur représentant toutes les réponses positives. Toutes les autres valeurs non manquantes sont supposées indiquer une réponse négative. Si le champ de réponse représente une valeur continue (échelle), telle que le nombre d'achats ou leur valeur monétaire, vous devez créer un nouveau champ qui assigne une valeur de réponse positive unique à toutes les réponses différentes de zéro.

- Si les réponses négatives sont enregistrées sous la forme d'un 0 (et non un blanc, ce qui est traité comme une valeur manquante), ceci peut être calculé avec la formule suivante :

$NewName = OldName > 0$

où *NewName* est le nom du nouveau champ et *OldName* est le nom du champ d'origine. Il s'agit d'une expression logique qui assigne une valeur de 1 à toutes les valeurs non manquantes supérieures à 0, et 0 à toutes les valeurs non manquantes inférieures ou égales à 0.

- Si aucune valeur n'est enregistrée pour les réponses négatives, ces valeurs sont traitées comme manquantes, et la formule est légèrement plus compliquée :

$NewName = NOT(MISSING(OldName))$

Dans cette expression logique, toutes les valeurs de réponse non manquantes se voient attribuer une valeur de 1 et toutes les valeurs de réponse manquantes se voient attribuer une valeur de 0.

- Si vous ne pouvez pas faire la distinction entre les valeurs de réponse négatives (0) et les valeurs manquantes, le calcul d'une valeur de réponse précise est impossible. S'il existe relativement peu de valeurs véritablement manquantes, cela n'aura peut-être pas d'effet significatif sur les taux de réponse calculés. Toutefois, si de nombreuses valeurs sont manquantes, par exemple lorsque les informations de réponse sont enregistrées uniquement pour un petit échantillon de test du jeu de données total, les taux de réponse calculés ne seront pas pertinents, car ils seront considérablement inférieurs aux vrais taux de réponse.

Pour créer un champ de réponse catégorielle

1. A partir des menus, sélectionnez :

**Transformer > Calculer la variable**

2. Pour la variable cible, entrez le nom du nouveau champ (variable).
3. Si des réponses négatives sont enregistrées en tant que 0, pour l'expression numérique saisissez  $OldName > 0$ , où *OldName* est le nom du champ d'origine.
4. Si des réponses négatives sont enregistrées en tant que valeur manquante (blanc), pour l'expression numérique saisissez  $NOT(MISSING(OldName))$ , où *OldName* est le nom du champ d'origine.



---

## Chapitre 5. Taux de réponse des codes postaux

Cette technique utilise les résultats de la campagne précédente pour calculer les taux de réponse par code postal. Ces taux peuvent être utilisés pour cibler des codes postaux spécifiques lors des futures campagnes. Le champ Réponse indique qui a répondu à la campagne précédente. Le champ Code postal identifie le champ qui contient les codes postaux.

**Exemple :** En fonction des résultats d'un précédent mailing, la division marketing direct d'une entreprise génère des taux de réponse par code postal. Sur la base de plusieurs critères, tels que le taux de réponse minimum acceptable et/ou le nombre maximum de contacts à inclure dans le mailing, il est ensuite possible de cibler des codes postaux spécifiques.

Sortie

La sortie de cette procédure inclut un nouveau jeu de données qui contient les taux de réponse par code postal, ainsi qu'un tableau et un graphique qui récapitulent les résultats par rang de décile (premiers 10%, 20%, etc.). Le tableau peut également faire l'objet d'un codage couleur, en fonction d'un taux de réponse cumulé minimum spécifié par l'utilisateur ou d'un nombre de contacts maximum.

Le nouveau jeu de données contient les champs suivants :

- **Code postal :** Si les groupes de codes postaux reposent sur une partie seulement de la valeur complète, il s'agit de la valeur de cette partie du code postal. Le libellé de ligne pour cette colonne dans le fichier Excel est le nom du champ de code postal du jeu de données d'origine.
- **TauxDeRéponse :** Pourcentage de réponses positives dans chaque code postal.
- **Réponses :** Nombre de réponses positives dans chaque code postal.
- **Contacts :** Nombre total de contacts dans chaque code postal qui contiennent une valeur non manquante dans le champ de réponse.
- **Index :** Réponse "pondérée" basée sur la formule  $N \times P \times (1-P)$ , où  $N$  représente le nombre de contacts et  $P$  le taux de réponses exprimé sous forme de proportion.
- **Rang :** Rang de décile (premiers 10%, 20% , etc.) des taux de réponse de code postal cumulés en ordre décroissant.

Analyse des données de taux de réponse des codes postaux

**Champ de réponse :** Le champ de réponse peut être une variable numérique ou une chaîne. Si ce champ contient une valeur indiquant un nombre d'achats ou leur valeur monétaire, vous devrez créer un nouveau champ dans lequel une valeur unique représente toutes les réponses positives. Pour plus d'informations, voir «Création d'un champ de réponse catégorielle», à la page 19.

**Valeur de réponse positive :** La valeur de réponse positive identifie les clients qui ont répondu positivement (par exemple, qui ont effectué un achat). Toutes les autres valeurs de réponse non manquantes sont supposées indiquer une réponse négative. S'il existe des libellés de valeurs définis pour le champ de réponse, ces libellés s'affichent dans la liste déroulante.

**Champ Code postal :** Le champ de code postal peut être une variable numérique ou une chaîne.

Pour obtenir les taux de réponse des codes postaux

A partir des menus, sélectionnez :

**Marketing direct > Choisir une technique**

1. Sélectionnez **Identifier les codes postaux avec les meilleurs taux de réponse**.
2. Sélectionnez le champ qui identifie les contacts ayant répondu à l'offre.
3. Entrez la valeur qui indique une réponse positive. Si des valeurs disposent de libellés de valeur définis, vous pouvez sélectionner le libellé de valeur dans la liste déroulante, et la valeur correspondante sera affichée.
4. Sélectionnez le champ qui contient le code postal.
5. Cliquez sur **Exécuter** pour exécuter la procédure.

Eventuellement, vous pouvez aussi :

- Générer des taux de réponse en fonction des  $n$  premiers caractères ou chiffres du code postal au lieu de la valeur complète
- Enregistrer les résultats dans un fichier Excel
- Contrôler les options d'affichage des sorties

---

## Paramètres

Grouper les codes postaux selon

Ceci détermine la manière dont les enregistrements sont groupés pour calculer les taux de réponse. Par défaut, le code postal entier est utilisé, et tous les enregistrements avec le même code postal sont regroupés pour calculer le taux de réponse du groupe. Vous pouvez également grouper les enregistrements selon une partie seulement du code postal complet, composée des  $n$  premiers chiffres ou caractères. Vous pouvez par exemple grouper des enregistrements en fonction uniquement des 5 premiers caractères d'un code postal de 10 caractères, ou des 3 premiers caractères d'un code postal de 5 caractères. Le jeu de données de sortie contiendra un enregistrement pour chaque groupe de codes postaux. Si vous saisissez une valeur, il doit s'agir d'un nombre entier positif.

Format numérique du code postal

Si le champ de code postal est numérique et que vous souhaitez grouper des codes postaux selon les  $n$  premiers chiffres au lieu de la valeur entière, vous devez spécifier le nombre de chiffres dans la valeur d'origine. Le nombre de chiffres est le nombre de chiffres *maximum* possible dans le code postal. Par exemple, si le champ de code postal contient un mélange de codes postaux à 5 chiffres et à 9 chiffres, vous devez indiquer 9 comme nombre de chiffres.

Remarque : Selon le format d'affichage, certains codes postaux à 5 chiffres peuvent apparaître avec seulement 4 chiffres, mais ils contiennent un zéro non significatif implicite à gauche.

Sortie

Outre le nouveau jeu de données qui contient les taux de réponse par code postal, vous pouvez afficher un tableau et un graphique qui récapitulent les résultats par rang de décile (premiers 10%, 20%, etc.). Le tableau affiche les taux de réponse, les taux de réponse cumulés, le nombre d'enregistrements, et le nombre cumulé d'enregistrements dans chaque décile. Le graphique affiche les taux de réponse cumulés et le nombre cumulé d'enregistrements dans chaque décile.

**Taux de réponse acceptable minimum** : Si vous saisissez un taux de réponse minimum cible ou une formule de calcul de point-mort, le tableau sera codé par couleurs pour indiquer les déciles qui satisfont le taux de réponse cumulé minimum, et le graphique contiendra une ligne de référence au niveau de la valeur du taux de réponse minimum spécifié.

- **Taux de réponse cible** : Taux de réponse exprimé en pourcentage (pourcentage de réponses positives dans chaque groupe de codes postaux). La valeur doit être supérieure à 0 et inférieure à 100.
- **Calculer le point mort à partir de la formule** : Calculez le taux de réponse cumulé minimum en fonction de la formule :  $(\text{Coût de l'envoi d'un pack}/\text{Revenu net par réponse}) \times 100$ . Les deux valeurs



doivent être positives. Le résultat doit être une valeur supérieure à 0 et inférieure à 100. Par exemple, si le coût d'envoi d'un pack est de 0,75 \$, et que le revenu net par réponse est de 56 \$, le taux de réponse minimum est :  $(0,75/56) \times 100 = 1,34 \%$ .

**Nombre maximal de contacts** : Si vous spécifiez un nombre maximal de contacts, le tableau sera codé par couleurs pour indiquer les déciles qui ne dépassent pas le nombre de contacts (enregistrements) cumulé maximal et le graphique contiendra une ligne de référence au niveau de cette valeur.

- **Pourcentage de contacts** : Maximum exprimé en pourcentage. Par exemple, vous souhaitez peut-être connaître les déciles avec les taux de réponse les plus élevés qui ne contiennent pas plus de 50 % des contacts. La valeur doit être supérieure à 0 et inférieure à 100.
- **Nombre de contacts** : Maximum exprimé en nombre de contacts. Par exemple, si vous ne prévoyez pas d'envoyer plus de 10 000 packs, vous pouvez définir la valeur sur 10000. La valeur doit être un nombre entier positif (sans symbole de regroupement).

Si vous spécifiez un taux de réponse minimum acceptable et un nombre maximal de contacts, la codification couleur du tableau s'effectuera selon la condition satisfaite en premier.

Exporter dans Excel

Cette procédure crée automatiquement un nouveau jeu de données qui contient les taux de réponse par code postal. Chaque enregistrement (ligne) du jeu de données représente un code postal. Vous pouvez enregistrer automatiquement les mêmes informations dans un fichier Excel. Le fichier est enregistré au format Excel 97-2003.

---

## Création d'un champ de réponse catégorielle

Le champ de réponse doit être catégoriel, avec une valeur représentant toutes les réponses positives. Toutes les autres valeurs non manquantes sont supposées indiquer une réponse négative. Si le champ de réponse représente une valeur continue (échelle), telle que le nombre d'achats ou leur valeur monétaire, vous devez créer un nouveau champ qui assigne une valeur de réponse positive unique à toutes les réponses différentes de zéro.

- Si les réponses négatives sont enregistrées sous la forme d'un 0 (et non un blanc, ce qui est traité comme une valeur manquante), ceci peut être calculé avec la formule suivante :

`NewName=OldName>0`

où *NewName* est le nom du nouveau champ et *OldName* est le nom du champ d'origine. Il s'agit d'une expression logique qui assigne une valeur de 1 à toutes les valeurs non manquantes supérieures à 0, et 0 à toutes les valeurs non manquantes inférieures ou égales à 0.

- Si aucune valeur n'est enregistrée pour les réponses négatives, ces valeurs sont traitées comme manquantes, et la formule est légèrement plus compliquée :

`NewName=NOT(MISSING(OldName))`

Dans cette expression logique, toutes les valeurs de réponse non manquantes se voient attribuer une valeur de 1 et toutes les valeurs de réponse manquantes se voient attribuer une valeur de 0.

- Si vous ne pouvez pas faire la distinction entre les valeurs de réponse négatives (0) et les valeurs manquantes, le calcul d'une valeur de réponse précise est impossible. S'il existe relativement peu de valeurs véritablement manquantes, cela n'aura peut-être pas d'effet significatif sur les taux de réponse calculés. Toutefois, si de nombreuses valeurs sont manquantes, par exemple lorsque les informations de réponse sont enregistrées uniquement pour un petit échantillon de test du jeu de données total, les taux de réponse calculés ne seront pas pertinents, car ils seront considérablement inférieurs aux vrais taux de réponse.

Pour créer un champ de réponse catégorielle

1. A partir des menus, sélectionnez :

**Transformer > Calculer la variable**

2. Pour la variable cible, entrez le nom du nouveau champ (variable).
3. Si des réponses négatives sont enregistrées en tant que 0, pour l'expression numérique saisissez  $OldName > 0$ , où *OldName* est le nom du champ d'origine.
4. Si des réponses négatives sont enregistrées en tant que valeur manquante (blanc), pour l'expression numérique saisissez  $NOT(MISSING(OldName))$ , où *OldName* est le nom du champ d'origine.

---

## Chapitre 6. Propension à acheter

La propension à acheter utilise les résultats d'un envoi d'essai ou d'une campagne précédente pour générer les scores. Les scores indiquent quels contacts sont les plus susceptibles de répondre. Le champ Réponse indique qui a répondu à l'envoi d'essai ou à la campagne précédente. Les champs Propension sont les caractéristiques à utiliser pour prévoir la probabilité que les contacts avec des caractéristiques similaires répondent.

Cette technique utilise la régression logistique binaire pour construire un modèle prédictif. Le processus de construction et d'application d'un modèle prédictif se fait en deux étapes de base :

1. La construction du modèle et l'enregistrement du fichier du modèle. Vous construisez le modèle en utilisant un jeu de données pour lequel le résultat qui vous intéresse (souvent appelé la **cible**) est connu. Par exemple, si vous voulez construire un modèle pour prédire qui est susceptible de répondre à une campagne de publipostage, vous devez commencer par un jeu de données qui indique déjà qui a répondu et qui n'a pas répondu. Par exemple, il pourrait s'agir des résultats d'un envoi d'essai à un petit groupe de clients ou d'informations sur les réponses à une campagne similaire par le passé.
2. L'application de ce modèle à un jeu de données différent (pour lequel le résultat qui vous intéresse n'est pas connu) pour prévoir les résultats.

**Exemple :** La division de marketing direct d'une entreprise utilise les résultats d'un envoi d'essai pour assigner des scores de propension au reste de leur base de données de contacts, à l'aide de diverses caractéristiques démographiques permettant d'identifier les contacts les plus susceptibles de répondre et d'effectuer un achat.

Sortie

Cette procédure crée automatiquement un nouveau champ dans le jeu de données qui contient des scores de propension pour les données de test et un fichier de modèle XML qui peut être utilisé pour évaluer d'autres jeux de données. La sortie du diagnostic facultatif comprend un graphique de qualité générale du modèle et une table de classification qui compare les réponses prédites aux réponses réelles.

Remarques sur les données de propension à acheter

**Champ de réponse :** Le champ de réponse peut être une variable numérique ou une chaîne. Si ce champ contient une valeur indiquant un nombre d'achats ou leur valeur monétaire, vous devez créer un nouveau champ dans lequel une valeur unique représente toutes les réponses positives. Pour plus d'informations, voir «Création d'un champ de réponse catégorielle», à la page 24.

**Valeur de réponse positive :** La valeur de réponse positive identifie les clients qui ont répondu positivement (par exemple, qui ont effectué un achat). Toutes les autres valeurs de réponse non manquantes sont supposées indiquer une réponse négative. S'il existe des libellés de valeurs définis pour le champ de réponse, ces libellés s'affichent dans la liste déroulante.












**Prédire la propension avec :** Les champs utilisés pour prédire la propension peuvent être une variable numérique ou une chaîne, et ils peuvent être de type nominal, ordinal ou continu (échelle). Toutefois, il est important d'affecter le niveau de mesure approprié à tous les champs prédicteurs.

**Niveau de mesure :** L'affectation d'un niveau de mesure correct est important car il affecte le calcul des résultats.

- *Nominal*. Une variable peut être traitée comme étant nominale si ses valeurs représentent des catégories sans classement intrinsèque (par exemple, le service de la société dans lequel travaille un employé). La région, le code postal ou l'appartenance religieuse sont des exemples de variables nominales.
- *Ordinal*. Une variable peut être traitée comme étant ordinale si ses valeurs représentent des catégories associées à un classement intrinsèque (par exemple, des niveaux de satisfaction allant de Très mécontent à Très satisfait). Exemples de variable ordinale : des scores d'attitude représentant le degré de satisfaction ou de confiance, et des scores de classement des préférences.
- *Continue*. Une variable peut être traitée comme une variable d'échelle (continue) si ses valeurs représentent des catégories ordonnées avec une mesure significative, de sorte que les comparaisons de distance entre les valeurs soient adéquates. L'âge en années et le revenu en milliers de dollars sont des exemples de variable d'échelle.

Une icône en regard de chaque champ indique le niveau de mesure actuel.

Tableau 4. Icônes de niveau de mesure

	Numérique	Chaîne	Date	Heure
Echelle (continue).		n/a		
Ordinal				
Nominal				

Vous pouvez modifier le niveau de mesure dans la vue de variable de l'éditeur de données ou utilisez la boîte de dialogue Définir les propriétés de variable pour suggérer un niveau de mesure approprié pour chaque champ .

Champs avec un niveau de mesure inconnu

L'alerte du niveau de mesure apparaît lorsque le niveau de mesure d'une ou plusieurs variables (champs) du jeu de données est inconnu. Le niveau de mesure ayant une incidence sur le calcul des résultats de cette procédure, toutes les variables doivent avoir un niveau de mesure défini.

**Analyser les données** : Lit les données dans le jeu de données actifs et attribue le niveau de mesure par défaut à tous les champs ayant un niveau de mesure inconnu. Si le jeu de données est important, cette action peut prendre un certain temps.

**Affecter manuellement** : Ouvre une boîte de dialogue qui répertorie tous les champs ayant un niveau de mesure inconnu. Vous pouvez utiliser cette boîte de dialogue pour attribuer un niveau de mesure à ces champs. Vous pouvez également attribuer un niveau de mesure dans la vue de variable de l'éditeur de données.

Le niveau de mesure étant important pour cette procédure, vous ne pouvez pas accéder à la boîte de dialogue d'exécution de cette procédure avant que tous les champs n'aient des niveaux de mesure définis.

Pour obtenir les scores de propension à acheter

A partir des menus, sélectionnez :

**Marketing direct > Choisir une technique**

1. Choisissez **Sélectionner les contacts les plus susceptibles d'acheter**.

2. Sélectionnez le champ qui identifie les contacts ayant répondu à l'offre.
3. Entrez la valeur qui indique une réponse positive. Si des valeurs disposent de libellés de valeur définis, vous pouvez sélectionner le libellé de valeur dans la liste déroulante, et la valeur correspondante sera affichée.
4. Sélectionnez les champs à utiliser pour prédire la propension.  
Pour enregistrer un fichier de modèle XML afin d'évaluer d'autres fichiers de données :
5. Sélectionnez (cochez) **Exporter les informations du modèle dans un fichier XML**.
6. Saisissez un chemin de répertoire et un nom de fichier ou cliquez sur **Parcourir** pour accéder à l'emplacement où vous souhaitez enregistrer le fichier de modèle XML.
7. Cliquez sur **Exécuter** pour exécuter la procédure.  
Pour utiliser le fichier de modèle afin d'évaluer d'autres jeux de données :
8. Ouvrez le jeu de données que vous souhaitez évaluer.
9. Utilisez l'Assistant d'évaluation pour appliquer le modèle à le jeu de données. A partir des menus, sélectionnez :  
**Utilitaires > Assistant d'évaluation**

---

## Paramètres

### Validation de modèles

La validation de modèles crée des groupes d'apprentissage et de test à des fins de diagnostic. Si vous sélectionnez la table de classification dans la section Sortie du diagnostic, le tableau sera divisé en deux sections, formation (sélectionnée) et test (non sélectionnée), à des fins de comparaison. Ne sélectionnez pas la validation du modèle à moins de sélectionner également la table de classification. Les scores sont basés sur le modèle généré à partir de l'échantillon d'apprentissage, qui contiendra toujours moins d'enregistrements que le nombre total d'enregistrements disponibles. Par exemple, la taille d'échantillon d'apprentissage par défaut est 50%, et il est possible qu'un modèle créé sur la moitié seulement des enregistrements disponibles ne soit pas aussi fiable qu'un modèle créé sur tous les enregistrements disponibles.

- **Echantillon d'apprentissage taille de la partition (%)** : Spécifiez le pourcentage d'enregistrements à affecter à l'échantillon d'apprentissage. Le reste des enregistrements avec des valeurs non manquantes dans le champ de réponse est affecté à l'échantillon de test. La valeur doit être supérieure à 0 et inférieure à 100.
- **Définir une valeur pour répliquer les résultats** : Les résultats étant affectés au hasard aux échantillons d'apprentissage et de test, chaque fois que vous exécutez la procédure, il est possible que obteniez des résultats différents, à moins de toujours spécifier la même valeur de départ pour la valeur de départ de nombre aléatoire.

### Sortie du diagnostic

**Qualité générale du modèle** : Affiche un graphique à barres de la qualité générale du modèle, exprimée sous la forme d'une valeur comprise entre 0 et 1. Un bon modèle doit avoir une valeur supérieure à 0,5.

**Table de classification** : Affiche un tableau qui compare les réponses positives et négatives prédites aux réponses positives et négatives réelles. Le taux de précision globale peut fournir une certaine indication de la performance du modèle, mais le pourcentage de réponses positives prédites correctes peut s'avérer plus intéressant.

- **Probabilité minimale** : Affecte les enregistrements avec une valeur de score supérieure à la valeur spécifiée à la catégorie de réponses positives prédites dans la table de classification. Les scores générés par la procédure représentent la probabilité que le contact réponde de manière positive (par exemple, effectuer un achat). En règle générale, vous devez indiquer une valeur proche de votre taux de réponse cible minimum, exprimé sous forme de proportion. Par exemple, si vous êtes intéressé par un taux de réponse d'au moins 5%, indiquez 0,05. La valeur doit être supérieure à 0 et inférieure à 1.

Nom et libellé du champ de réponses recodé

Cette procédure recode automatiquement le champ de réponse en un nouveau champ dans lequel 1 représente des réponses positives et 0 représente des réponses négatives, et l'analyse est effectuée sur le champ recodé. Vous pouvez ignorer le nom et le libellé par défaut et fournir les vôtres. Les noms doivent être conformes aux règles de dénomination IBM SPSS Statistics.

Enregistrer les scores

Un nouveau champ contenant les scores de propension est automatiquement enregistré dans le jeu de données d'origine. Les scores représentent la probabilité d'une réponse positive, exprimée sous forme de proportion.

- Les noms de champ doivent être conformes aux règles de dénomination IBM SPSS Statistics.
- Le nom de champ ne peut pas dupliquer un nom de champ qui existe déjà dans le jeu de données. Si vous exécutez cette procédure plusieurs fois sur le même jeu de données, vous devrez spécifier un nom différent à chaque fois.

---

## Création d'un champ de réponse catégorielle

Le champ de réponse doit être catégoriel, avec une valeur représentant toutes les réponses positives. Toutes les autres valeurs non manquantes sont supposées indiquer une réponse négative. Si le champ de réponse représente une valeur continue (échelle), telle que le nombre d'achats ou leur valeur monétaire, vous devez créer un nouveau champ qui assigne une valeur de réponse positive unique à toutes les réponses différentes de zéro.

- Si les réponses négatives sont enregistrées sous la forme d'un 0 (et non un blanc, ce qui est traité comme une valeur manquante), ceci peut être calculé avec la formule suivante :

$newName = OldName > 0$

où *NewName* est le nom du nouveau champ et *OldName* est le nom du champ d'origine. Il s'agit d'une expression logique qui assigne une valeur de 1 à toutes les valeurs non manquantes supérieures à 0, et 0 à toutes les valeurs non manquantes inférieures ou égales à 0.

- Si aucune valeur n'est enregistrée pour les réponses négatives, ces valeurs sont traitées comme manquantes, et la formule est légèrement plus compliquée :

$newName = NOT(MISSING(OldName))$

Dans cette expression logique, toutes les valeurs de réponse non manquantes se voient attribuer une valeur de 1 et toutes les valeurs de réponse manquantes se voient attribuer une valeur de 0.

- Si vous ne pouvez pas faire la distinction entre les valeurs de réponse négatives (0) et les valeurs manquantes, le calcul d'une valeur de réponse précise est impossible. S'il existe relativement peu de valeurs véritablement manquantes, cela n'aura peut-être pas d'effet significatif sur les taux de réponse calculés. Toutefois, si de nombreuses valeurs sont manquantes, par exemple lorsque les informations de réponse sont enregistrées uniquement pour un petit échantillon de test du jeu de données total, les taux de réponse calculés ne seront pas pertinents, car ils seront considérablement inférieurs aux vrais taux de réponse.

Pour créer un champ de réponse catégorielle

1. A partir des menus, sélectionnez :

**Transformer > Calculer la variable**

2. Pour la variable cible, entrez le nom du nouveau champ (variable).
3. Si des réponses négatives sont enregistrées en tant que 0, pour l'expression numérique saisissez  $OldName > 0$ , où *OldName* est le nom du champ d'origine.
4. Si des réponses négatives sont enregistrées en tant que valeur manquante (blanc), pour l'expression numérique saisissez  $NOT(MISSING(OldName))$ , où *OldName* est le nom du champ d'origine.

---

## Chapitre 7. Test de pack de contrôle

Cette technique compare des campagnes de marketing pour vérifier s'il existe une différence significative de l'efficacité entre différents packs et offres. L'efficacité de la campagne est mesurée en fonction des réponses. Le champ de campagne identifie différentes campagnes, par exemple Offre A et Offre B. Le champ de réponses indique si un contact a répondu à la campagne. Sélectionnez Montant d'achat lorsque la réponse est enregistrée comme un montant d'achat, par exemple "99,99". Sélectionnez Répondre lorsque la réponse est simplement enregistrée comme une réponse positive ou négative, par exemple "Oui" ou "Non".

**Exemple :** La division de marketing direct d'une société souhaite vérifier si un nouveau plan du pack peut générer plus de réponses positives que le pack existant. Ils procèdent à un envoi d'essai pour déterminer si le nouveau pack génère un taux de réponse significativement plus élevé. L'envoi d'essai consiste en un groupe de contrôle qui reçoit le pack existant et un groupe de test qui reçoit le nouveau plan du pack. Les résultats des deux groupes sont comparés pour juger de la différence.

Sortie

Les sorties comprennent un tableau qui affiche les effectifs et les pourcentages des réponses positives et négatives de chaque groupe défini par le champ Campagne et un tableau qui identifie les groupes qui diffèrent entre eux de manière significative.

Remarques et hypothèses concernant les données du test de pack de contrôle

**Champ Campagne :** Le champ Campagne doit être catégoriel (nominal ou ordinal).

**Champ de réponse d'efficacité :** Si vous avez sélectionné Montant d'achat pour le champ d'efficacité, le champ doit être numérique, et le niveau de mesure doit être continu (échelle).

Si vous ne pouvez pas faire la distinction entre les valeurs de réponse négatives (pour le montant d'achat, une valeur de 0) et les valeurs manquantes, le calcul d'un taux de réponse précise est impossible. S'il existe relativement peu de valeurs véritablement manquantes, cela n'aura peut-être pas d'effet significatif sur les taux de réponse calculés. Toutefois, si de nombreuses valeurs sont manquantes, par exemple lorsque les informations de réponse sont enregistrées uniquement pour un petit échantillon de test du jeu de données total, les taux de réponse calculés ne seront pas pertinents, car ils seront considérablement inférieurs aux vrais taux de réponse.

**Hypothèses :** Cette procédure considère que les contacts ont été affectés de manière aléatoire à chaque groupe de la campagne. En d'autres termes, aucune affectation selon des caractéristiques démographiques, d'habitude d'achat ou autres n'a été effectuée et la probabilité d'être affecté à l'un des groupes est identique pour tous les contacts.

Pour obtenir un test de pack de contrôle

A partir des menus, sélectionnez :

**Marketing direct > Choisir une technique**

1. Sélectionnez **Comparer l'efficacité des campagnes**.
2. Sélectionnez le champ qui identifie le groupe de campagne auquel chaque contact appartient (par exemple, offre A, offre B, etc.) Ce champ doit être nominal ou ordinal.
3. Sélectionnez le champ qui indique l'efficacité des réponses.  
Si le champ de réponse est un montant d'achat, il doit être numérique.

Si le champ de réponse indique simplement si la réponse du contact est positive ou négative, (par exemple "Oui" ou "Non"), sélectionnez **Répondre** et saisissez la valeur qui représente une réponse positive. Si des valeurs disposent de libellés de valeur définis, vous pouvez sélectionner le libellé de valeur dans la liste déroulante, et la valeur correspondante sera affichée.

Un nouveau champ est créé automatiquement, dans lequel 1 représente des réponses positives et 0 représente des réponses négatives, et l'analyse est effectuée sur le nouveau champ. Vous pouvez ignorer le nom et le libellé par défaut et fournir les vôtres. Les noms doivent être conformes aux règles de dénomination IBM SPSS Statistics.

4. Cliquez sur **Exécuter** pour exécuter la procédure.



---

## Remarques

Le présent document peut contenir des informations ou des références concernant certains produits, logiciels ou services IBM non annoncés dans ce pays. Pour plus de détails, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial IBM. Toute référence à un produit, logiciel ou service IBM n'implique pas que seul ce produit, logiciel ou service puisse être utilisé. Tout autre élément fonctionnellement équivalent peut être utilisé, s'il n'enfreint aucun droit d'IBM. Il est de la responsabilité de l'utilisateur d'évaluer et de vérifier lui-même les installations et applications réalisées avec des produits, logiciels ou services non expressément référencés par IBM.

IBM peut détenir des brevets ou des demandes de brevet couvrant les produits mentionnés dans le présent document. La remise de ce document ne vous donne aucun droit de licence sur ces brevets ou demandes de brevet. Si vous désirez recevoir des informations concernant l'acquisition de licences, veuillez en faire la demande par écrit à l'adresse suivante :

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
U.S.A.

Pour le Canada, veuillez adresser votre courrier à :

IBM Director of Commercial Relations  
IBM Canada Ltd.  
3600 Steeles Avenue East  
Markham, Ontario  
L3R 9Z7 Canada

Les informations sur les licences concernant les produits utilisant un jeu de caractères double octet peuvent être obtenues par écrit à l'adresse suivante :

Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.  
1623-14, Shimotsuruma, Yamato-shi  
Kanagawa 242-8502 Japan

Le paragraphe suivant ne s'applique ni au Royaume-Uni, ni dans aucun pays dans lequel il serait contraire aux lois locales : LE PRÉSENT DOCUMENT EST LIVRE "EN L'ÉTAT" SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFAÇON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE. Certaines juridictions n'autorisent pas l'exclusion des garanties implicites, auquel cas l'exclusion ci-dessus ne vous sera pas applicable.

Le présent document peut contenir des inexactitudes ou des coquilles. Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut, à tout moment et sans préavis, modifier les produits et logiciels décrits dans ce document.

Les références à des sites Web non IBM sont fournies à titre d'information uniquement et n'impliquent en aucun cas une adhésion aux données qu'ils contiennent. Les éléments figurant sur ces sites Web ne font pas partie des éléments du présent produit IBM et l'utilisation de ces sites relève de votre seule responsabilité.

IBM pourra utiliser ou diffuser, de toute manière qu'elle jugera appropriée et sans aucune obligation de sa part, tout ou partie des informations qui lui seront fournies.

Les licenciés souhaitant obtenir des informations permettant : (i) l'échange des données entre des logiciels créés de façon indépendante et d'autres logiciels (dont celui-ci), et (ii) l'utilisation mutuelle des données ainsi échangées, doivent adresser leur demande à :

IBM Software Group  
ATTN: Licensing  
200 W. Madison St.  
Chicago, IL; 60606  
U.S.A.

Ces informations peuvent être soumises à des conditions particulières, prévoyant notamment le paiement d'une redevance.

Le logiciel sous licence décrit dans ce document et tous les éléments sous licence disponibles s'y rapportant sont fournis par IBM conformément aux dispositions du Livret Contractuel IBM, des Conditions Internationales d'Utilisation de Logiciels IBM, des Conditions d'Utilisation du Code Machine ou de tout autre contrat équivalent.

Les données de performance indiquées dans ce document ont été déterminées dans un environnement contrôlé. Par conséquent, les résultats peuvent varier de manière significative selon l'environnement d'exploitation utilisé. Certaines mesures évaluées sur des systèmes en cours de développement ne sont pas garanties sur tous les systèmes disponibles. En outre, elles peuvent résulter d'extrapolations. Les résultats peuvent donc varier. Il incombe aux utilisateurs de ce document de vérifier si ces données sont applicables à leur environnement d'exploitation.

Les informations concernant des produits non IBM ont été obtenues auprès des fournisseurs de ces produits, par l'intermédiaire d'annonces publiques ou via d'autres sources disponibles. IBM n'a pas testé ces produits et ne peut confirmer l'exactitude de leurs performances ni leur compatibilité. Elle ne peut recevoir aucune réclamation concernant des produits non IBM. Les questions sur les capacités de produits autres qu'IBM doivent être adressées aux fabricants de ces produits.

Toute instruction relative aux intentions d'IBM pour ses opérations à venir est susceptible d'être modifiée ou annulée sans préavis, et doit être considérée uniquement comme un objectif.

Le présent document peut contenir des exemples de données et de rapports utilisés couramment dans l'environnement professionnel. Ces exemples mentionnent des noms fictifs de personnes, de sociétés, de marques ou de produits à des fins illustratives ou explicatives uniquement. Toute ressemblance avec des noms de personnes, de sociétés ou des données réelles serait purement fortuite.

Toute copie totale ou partielle de ces programmes exemples et des oeuvres qui en sont dérivées doit comprendre une notice de copyright, libellée comme suit :

Le présent document peut contenir des exemples de données et de rapports utilisés couramment dans l'environnement professionnel. Ces exemples mentionnent des noms fictifs de personnes, de sociétés, de marques ou de produits à des fins illustratives ou explicatives uniquement. Toute ressemblance avec des noms de personnes, de sociétés ou des données réelles serait purement fortuite.

Toute copie totale ou partielle de ces programmes exemples et des oeuvres qui en sont dérivées doit comprendre une notice de copyright, libellée comme suit :

© (nom de votre société) (année). Des segments de code sont dérivés des Programmes exemples d'IBM Corp.

© Copyright IBM Corp. \_entrez l'année ou les années\_. Tous droits réservés.

---

## Marques

IBM, le logo IBM et [ibm.com](http://ibm.com) sont des marques d'International Business Machines Corp. dans de nombreux pays. Les autres noms de produits et de services peuvent être des marques d'IBM ou d'autres sociétés. La liste actualisée de toutes les marques d'IBM est disponible sur la page Web «Copyright and trademark information» à l'adresse [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Adobe, le logo Adobe, PostScript et le logo PostScript sont des marques d'Adobe Systems Incorporated aux Etats-Unis et/ou dans certains autres pays.

Intel, le logo Intel, Intel Inside, le logo Intel Inside, Intel Centrino, le logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium et Pentium sont des marques d'Intel Corporation ou de ses filiales aux Etats-Unis et/ou dans certains autres pays.

Linux est une marque de Linus Torvalds aux Etats-Unis et/ou dans certains autres pays.

Microsoft, Windows, Windows NT et le logo Windows sont des marques de Microsoft Corporation aux Etats-Unis et/ou dans certains autres pays.

UNIX est une marque enregistrée de The Open Group aux Etats-Unis et/ou dans certains autres pays.

Java ainsi que toutes les marques et tous les logos incluant Java sont des marques d'Oracle et/ou de ses sociétés affiliées.



---

# Index

## A

analyse de cluster 11  
analyse de cluster (option Marketing direct) 11

## C

cluster 11

## P

profils des clients potentiels (option Marketing direct) 13  
propension à acheter 21

## R

régression logistique (Marketing direct) 21  
RFM 3, 6, 7, 8  
  données de clients 4  
  données de transactions 3  
  regroupement par casiers 4

## T

taux de réponse des codes postaux 17  
test de pack de contrôle 25





